InstaBizIntel

Complete Analysis Guide

All 301 analyses — worked end-to-end on real research & business datasets

What each method does · sample data input · steps · live output · detailed interpretation

Statistics · Econometrics · SEM · Machine Learning · Survival & Reliability · Bayesian · Causal Inference · Pharmacometrics

Generated 07 September 2026

Contents

1. Foundations & Comparative Statistics (101 analyses)

2. Regression & Predictive Modelling (46 analyses)

3. Multivariate, Scaling & Dimension Reduction (39 analyses)

4. Time Series & Forecasting (33 analyses)

5. Panel & Longitudinal Data (5 analyses)

6. Causal Inference & Policy Evaluation (11 analyses)

7. Bayesian Methods (7 analyses)

8. Survival & Reliability Engineering (13 analyses)

9. Clinical, Epidemiology & Pharmacometrics (14 analyses)

10. Design of Experiments & Power (16 analyses)

11. Quality & Statistical Process Control (11 analyses)

12. Consumer & Market Research (5 analyses)

Live output captured for 278 of 301 analyses; the remainder document the exact data shape each method requires.

1. Foundations & Comparative Statistics

1.1 Bartlett's test

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Equal-variance test. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

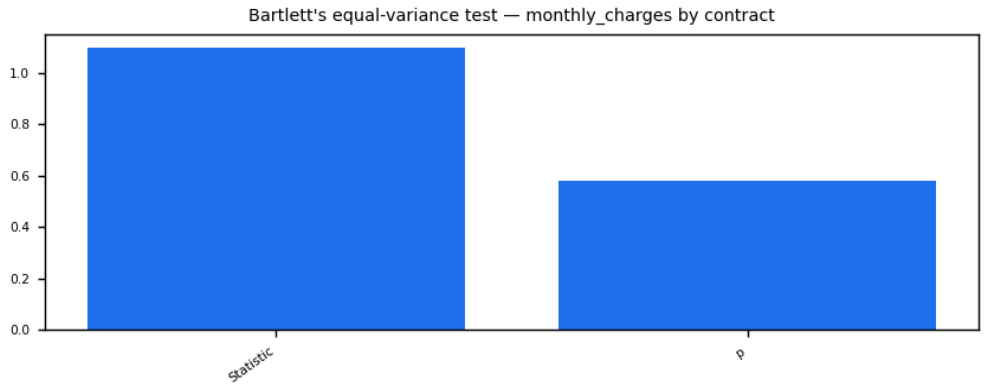
STEPS

1. Statistics → **Compare means** → **Bartlett's test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Bartlett's equal-variance test — monthly_charges by contract

Statistic	p
1.095	0.5784



DETAILED INTERPRETATION

Sensitive test of homogeneity of variance (assumes normality). The lead result here reads Statistic = 1.095; p = 0.5784. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.2 Bayesian t-test (Bayes factor)

LIVE OUTPUT

Group: Compare means · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

JZS Bayes factor for a mean difference (evidence for H1 vs H0). It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, pu1. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

peou1	pu1
3	4
3	3
4	5
5	3
2	3

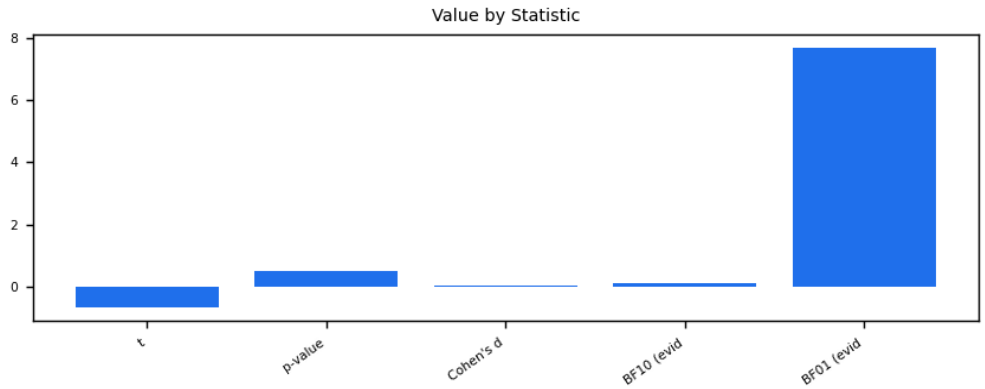
STEPS

1. Statistics → **Compare means** → **Bayesian t-test (Bayes factor)**. 2. Select the demo dataset above. 3. Set parameters — `group1=peou1, group2=pu1, paired=no`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Bayesian t-test — peou1 vs pu1

Statistic	Value
t	-0.657
p-value	0.5118
Cohen's d	0.063
BF10 (evidence for H1)	0.13
BF01 (evidence for H0)	7.692



DETAILED INTERPRETATION

BF10 = 0.13 → evidence for H0 evidence for a difference (JZS/Cauchy prior). BF10>3 supports H1, <1/3 supports H0. The lead result here reads Statistic = t; Value = -0.657. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst decide whether a segment truly behaves differently.

1.3 F-test for equal variances

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Ratio-of-variances test. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

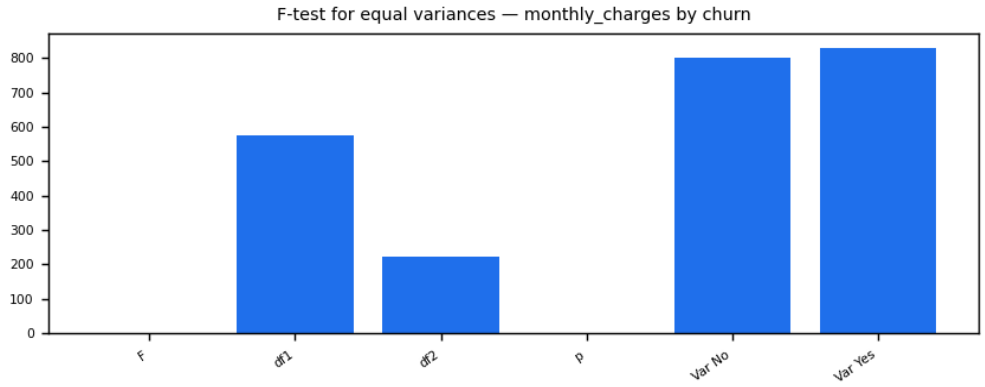
STEPS

1. Statistics → **Compare means** → **F-test for equal variances**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

F-test for equal variances — monthly_charges by churn

F	df1	df2	p	Var No	Var Yes
0.9639	575	223	0.7289	800.8	830.8



DETAILED INTERPRETATION

Ratio of two variances; p < .05 → variances differ. The lead result here reads F = 0.9639; df1 = 575; df2 = 223; p = 0.7289. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.4 Games-Howell post-hoc

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Post-hoc for unequal variances. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

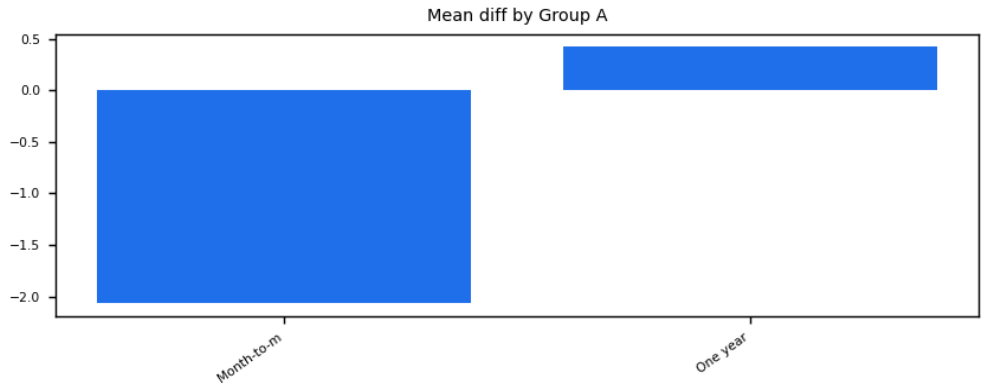
STEPS

1. Statistics → **Compare means** → **Games-Howell post-hoc**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Games-Howell post-hoc — monthly_charges by contract

Group A	Group B	Mean diff	SE	df (Welch)	p
Month-to-month	One year	-2.066	2.425	369.8	
Month-to-month	Two year	-1.644	2.712	278.1	
One year	Two year	0.4223	3.086	333.1	0.9



DETAILED INTERPRETATION

Post-hoc for groups with unequal variances/sizes (does not assume homogeneity). The lead result here reads Group A = Month-to-month; Group B = One year; Mean diff = -2.066; SE = 2.425. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.5 Homogeneity of variance

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Levene, Bartlett and Fligner-Killeen equal-variance tests. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

STEPS

1. Statistics → **Compare means** → **Homogeneity of variance**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

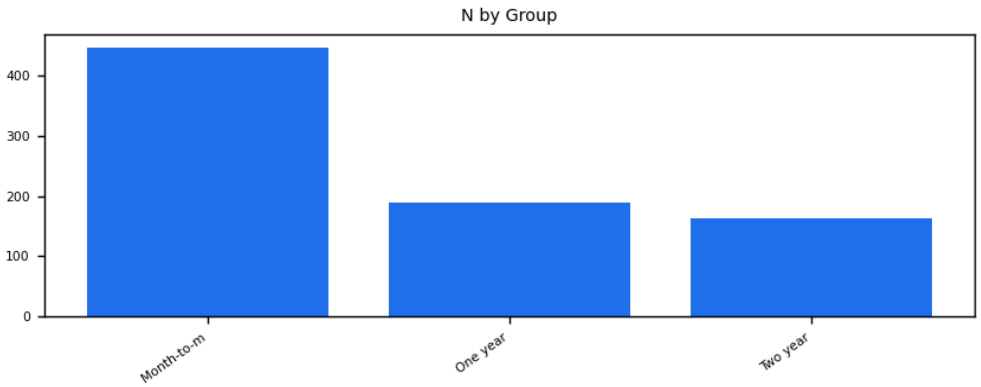
OUTPUT

Group variances

Group	N	Variance	SD
Month-to-month	447	828.5	28.78
One year	190	765.1	27.66
Two year	163	896.3	29.94

Homogeneity of variance tests

Test	Statistic	p
Levene (median-centered)	1.319	0.2679
Bartlett	1.095	0.5784
Fligner-Killeen	2.44	0.2952



DETAILED INTERPRETATION

The lead result here reads Group = Month-to-month; N = 447; Variance = 828.5; SD = 28.78. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.6 Independent-samples t-test

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Compare two group means (Student + Welch, Levene's test, Cohen's d). It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

STEPS

1. Statistics → **Compare means** → **Independent-samples t-test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Group statistics

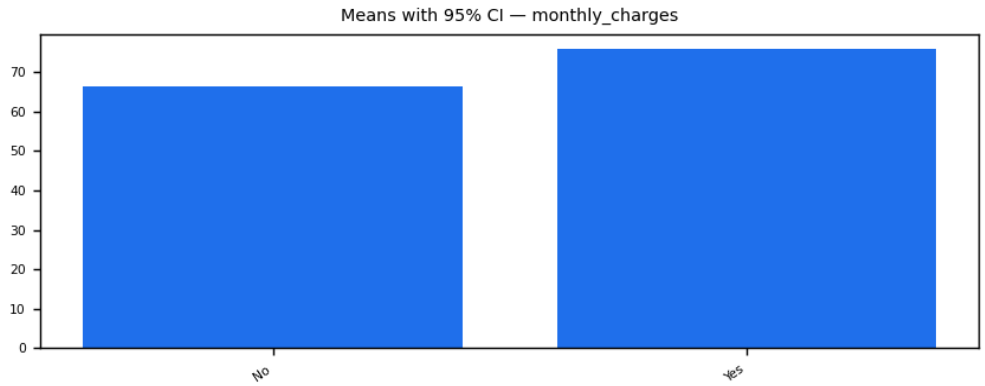
Group	N	Mean	SD	SE
No	576	66.45	28.3	1.179
Yes	224	75.88	28.82	1.926

Levene's test for equality of variances

F	p
0.1873	0.6653

Independent-samples t-test

Assumption	t	df	p (2-tailed)	Mean difference	Cohen's d
Equal variances	-4.209	798	< .001	-9.428	-0.3314
Welch (unequal)	-4.175		< .001	-9.428	-0.3314



DETAILED INTERPRETATION

The lead result here reads Group = No; N = 576; Mean = 66.45; SD = 28.3. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.7 One-sample t-test

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Compare a mean against a fixed test value. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

STEPS

1. Statistics → **Compare means** → **One-sample t-test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, test_value=5`. 4. Click **Run** and read the tables/chart below.

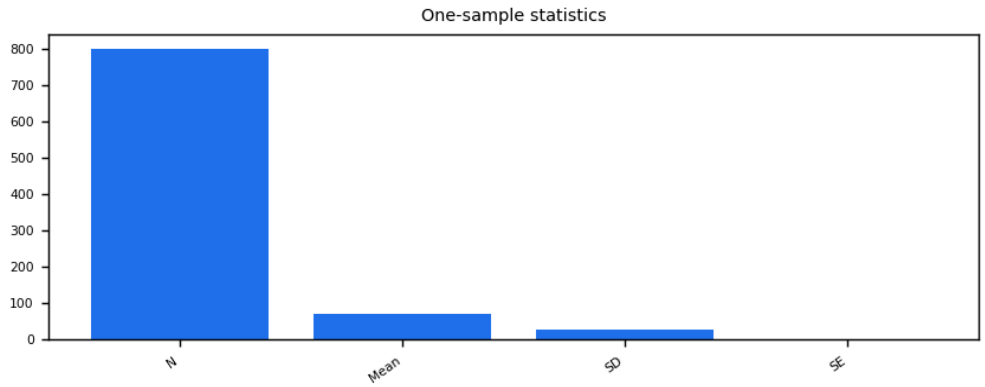
OUTPUT

One-sample statistics

Variable	N	Mean	SD	SE
monthly_charges	800	69.09	28.74	1.016

One-sample t-test (test value = 5.0)

t	df	p (2-tailed)	Mean difference	95% CI low	95% CI high	Cohen's d
63.07	799	< .001	64.09	62.1	66.09	2.23



DETAILED INTERPRETATION

The lead result here reads Variable = monthly_charges; N = 800; Mean = 69.09; SD = 28.74. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.8 One-sample z-test

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Large-sample mean test. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

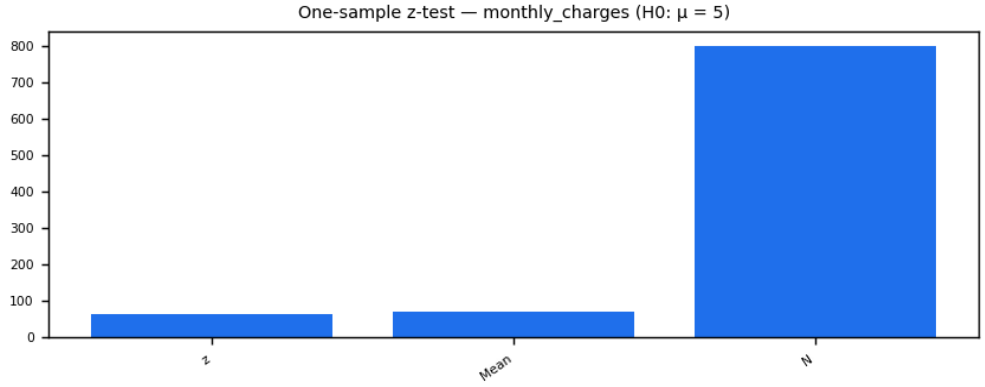
STEPS

1. Statistics → **Compare means** → **One-sample z-test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, value=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

One-sample z-test — monthly_charges (H0: μ = 5)

z	p	Mean	N
63.07	< .001	69.09	800



DETAILED INTERPRETATION

Large-sample test of a mean against a hypothesized value. The lead result here reads z = 63.07; p = < .001; Mean = 69.09; N = 800. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.9 One-way ANOVA

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Compare 3+ group means with Levene, eta² and Tukey HSD post-hoc. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

STEPS

1. Statistics → **Compare means** → **One-way ANOVA**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Group statistics

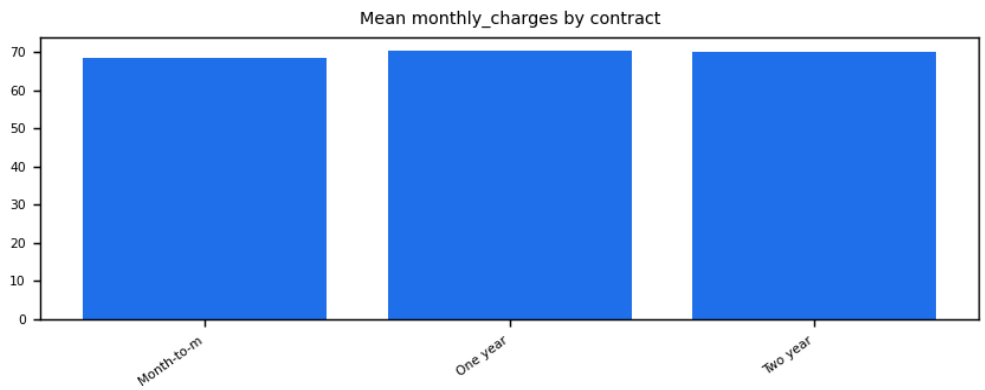
Group	N	Mean	SD	SE
Month-to-month	447	68.26	28.78	1.361
One year	190	70.33	27.66	2.007
Two year	163	69.91	29.94	2.345

Levene's test of homogeneity

F	p
1.319	0.2679

One-way ANOVA

Source	SS	df	MS	F	p	eta²
Between groups	706.3	2	353.1	0.4269	0.6527	0.0011
Within groups	6.593e+05	797	827.3			
Total	6.6e+05	799				



DETAILED INTERPRETATION

The lead result here reads Group = Month-to-month; N = 447; Mean = 68.26; SD = 28.78. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.10 Paired-samples t-test

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Compare two related measurements. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

STEPS

1. Statistics → **Compare means** → **Paired-samples t-test**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

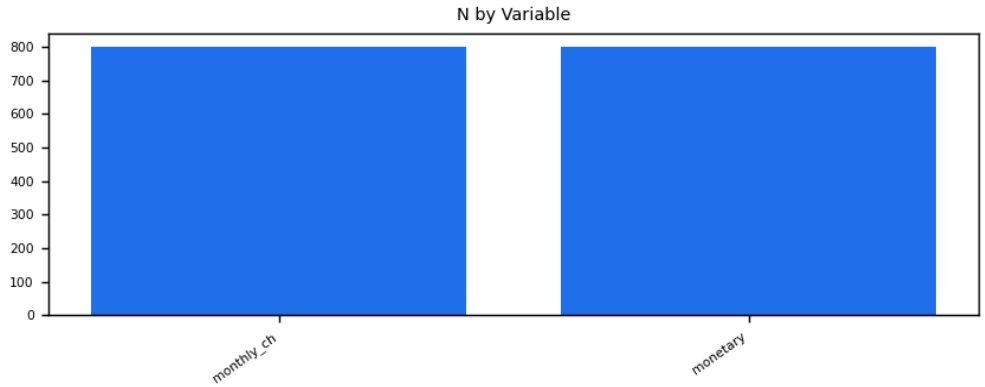
OUTPUT

Paired statistics

Variable	N	Mean	SD
monthly_charges	800	69.09	28.74
monetary	800	2634	1396

Paired-samples t-test

Pair correlation	Mean diff	SD diff	t	df	p (2-tailed)	Cohen's dz
-0.0103	-2565	1396	-51.96	799	< .001	-1.837



DETAILED INTERPRETATION

The lead result here reads Variable = monthly_charges; N = 800; Mean = 69.09; SD = 28.74. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.11 Planned contrasts

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

A priori linear contrast of means. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

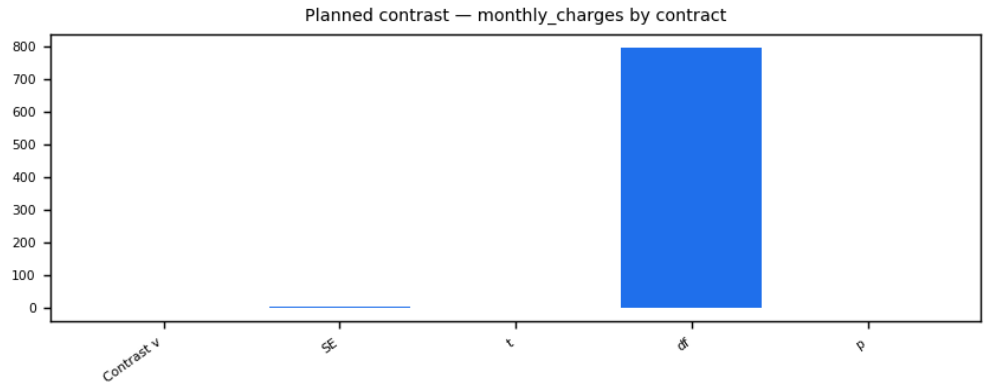
STEPS

1. Statistics → **Compare means** → **Planned contrasts**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract, weights=`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Planned contrast — monthly_charges by contract

Contrast value L	SE	t	df	p	Weights
-1.855	2.051	-0.9043	797	0.3661	1, -0.5, -0.5



DETAILED INTERPRETATION

A priori linear comparison of group means (weights sum to zero). The lead result here reads Contrast value L = -1.855; SE = 2.051; t = -0.9043; df = 797. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.12 Tukey HSD post-hoc

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Family-wise pairwise comparisons. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month

monthly_charges	contract
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

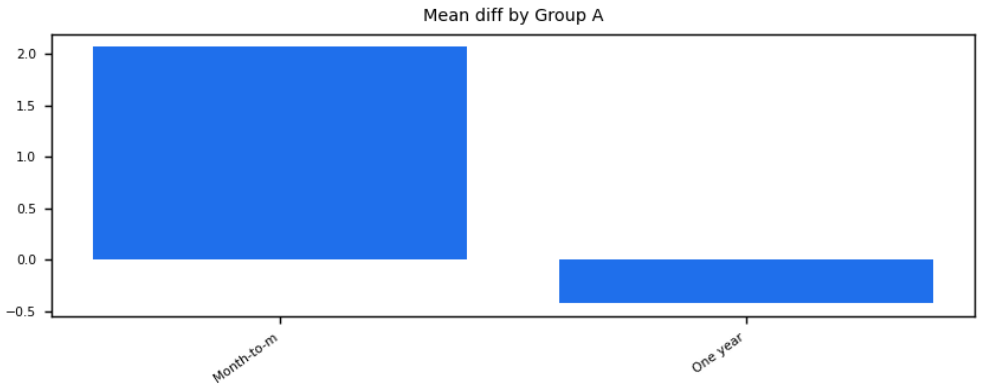
STEPS

1. Statistics → **Compare means** → **Tukey HSD post-hoc**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Tukey HSD post-hoc — monthly_charges by contract

Group A	Group B	Mean diff	p	CI low	CI high	Significant
Month-to-month	One year	2.066	0.6848	-3.783	7.915	False
Month-to-month	Two year	1.644	0.8066	-4.536	7.823	False
One year	Two year	-0.4223	0.9896	-7.633	6.788	False



DETAILED INTERPRETATION

Family-wise controlled pairwise comparisons after a significant ANOVA. The lead result here reads Group A = Month-to-month; Group B = One year; Mean diff = 2.066; p = 0.6848. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.13 Two-proportion z-test

LIVE OUTPUT

Group: Compare means · Dataset: a real-world patient cohort — treatment, recovery and confounders

WHAT THIS DOES

Compare success rates across groups. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a real-world patient cohort — treatment, recovery and confounders, drawing on recovered, treated. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

recovered	treated
0	1
0	0
1	1
0	0
0	0

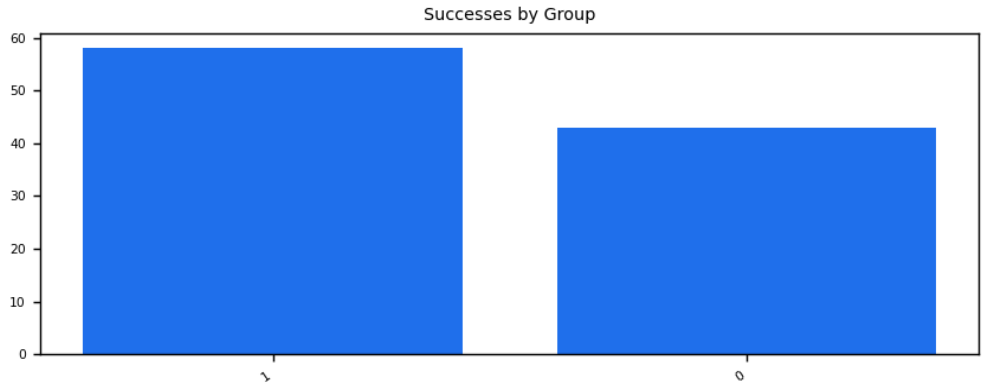
STEPS

1. Statistics → **Compare means** → **Two-proportion z-test**. 2. Select the demo dataset above. 3. Set parameters — `column=recovered, group=treated`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Two-proportion z-test — recovered by treated

Group	Successes	N	Proportion
1	58	170	0.3412
0	43	130	0.3308



DETAILED INTERPRETATION

z = 0.189, p = 0.8501. Compares the success rate of a 0/1 variable across two groups. The lead result here reads Group = 1; Successes = 58; N = 170; Proportion = 0.3412. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a real-world patient cohort — treatment, recovery and confounders, this lets the analyst decide whether a segment truly behaves differently.

1.14 Two-way ANOVA

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Factorial ANOVA with interaction (Type II SS). It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

STEPS

1. Statistics → **Compare means** → **Two-way ANOVA**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, factor1=contract, factor2=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

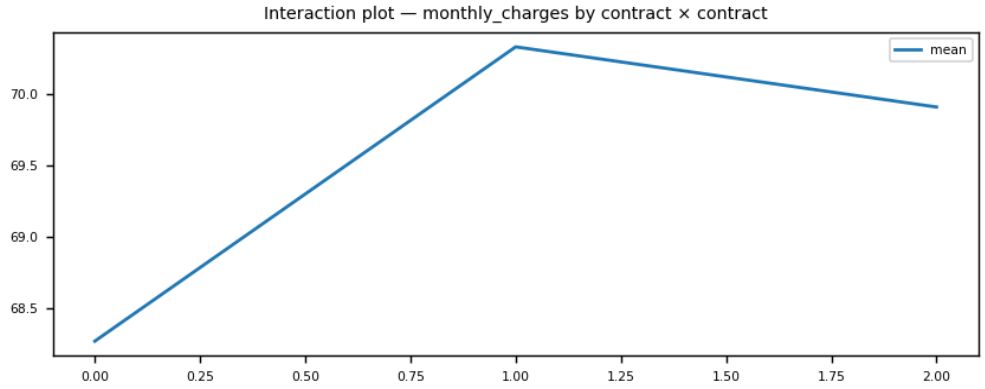
Two-way ANOVA — monthly_charges by contract, contract (Type II SS)

Source	SS	df	MS	F	p
contract		2			
contract	0	2	0	0	1
contract × contract	3875	4	968.8	1.169	0.2799
Residual	6.603e+05	797	828.4		

Cell means

contract	contract	N	Mean	SD
Month-to-month	Month-to-month	447	68.26	28.78
One year	One year	190	70.33	27.66

contract	contract	N	Mean	SD
Two year	Two year	163	69.91	29.94



DETAILED INTERPRETATION

Model R^2 = -0.0004 · adjusted R^2 = -0.0029 · n = 800. The lead result here reads Source = contract; SS = ; df = 2; MS = . A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.15 Welch's ANOVA

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

One-way ANOVA for unequal variances. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

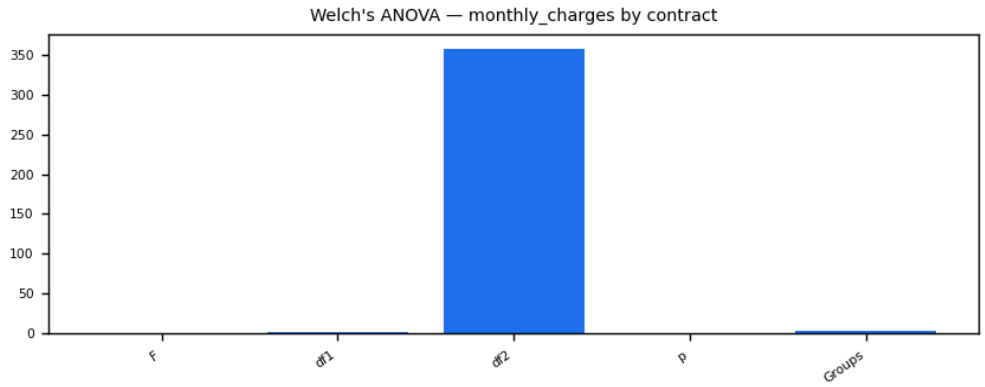
STEPS

1. Statistics → **Compare means** → **Welch's ANOVA**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Welch's ANOVA — monthly_charges by contract

F	df1	df2	p	Groups
0.435	2	358.2	0.6476	3



DETAILED INTERPRETATION

One-way ANOVA that does not assume equal variances (use with Games-Howell post-hoc). The lead result here reads F = 0.435; df1 = 2; df2 = 358.2; p = 0.6476. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.16 Yuen's trimmed t-test

LIVE OUTPUT

Group: Compare means · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Robust two-group mean comparison. It belongs to the **Compare means** family, used to test whether group averages differ beyond chance. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to decide whether a segment truly behaves differently.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

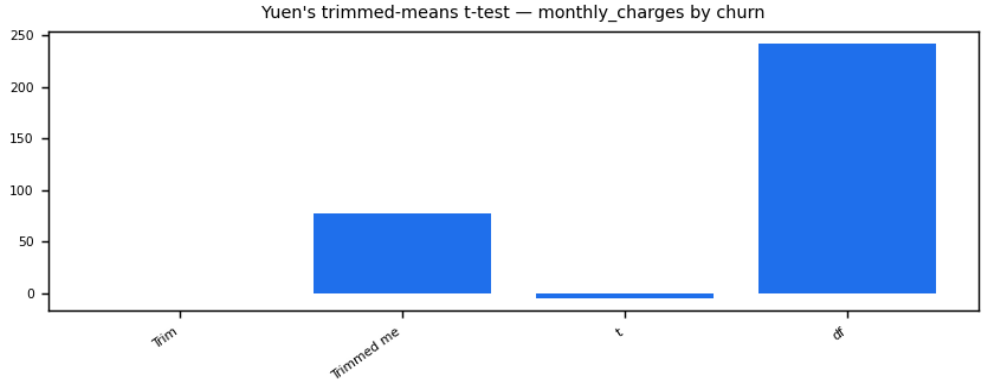
STEPS

1. Statistics → **Compare means** → **Yuen's trimmed t-test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn, trim=20`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Yuen's trimmed-means t-test — monthly_charges by churn

Trim	Trimmed mean 1	Trimmed mean 2	t	df	p
0.2	65.31	77.95	-4.253	241.5	< .001



DETAILED INTERPRETATION

Robust t-test comparing trimmed means (resistant to outliers/skew). The lead result here reads Trim = 0.2; Trimmed mean 1 = 65.31; Trimmed mean 2 = 77.95; t = -4.253. A p-value below .05 means the observed gap between groups is unlikely if they truly did not differ; pair it with the effect size to judge whether the gap matters in practice. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst decide whether a segment truly behaves differently.

1.17 Biweight midcorrelation

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Robust (outlier-resistant) correlation. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

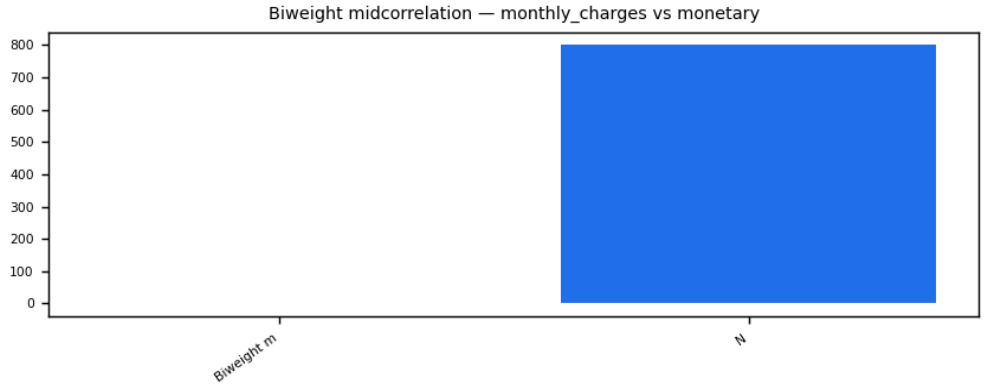
STEPS

1. Statistics → **Correlate** → **Biweight midcorrelation**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Biweight midcorrelation — monthly_charges vs monetary

Biweight midcorrelation	N
-0.0149	800



DETAILED INTERPRETATION

Robust correlation that down-weights outliers (median/MAD based). The lead result here reads Biweight midcorrelation = -0.0149; N = 800. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.18 Correlation matrix

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Pearson, Spearman or Kendall correlations with significance. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

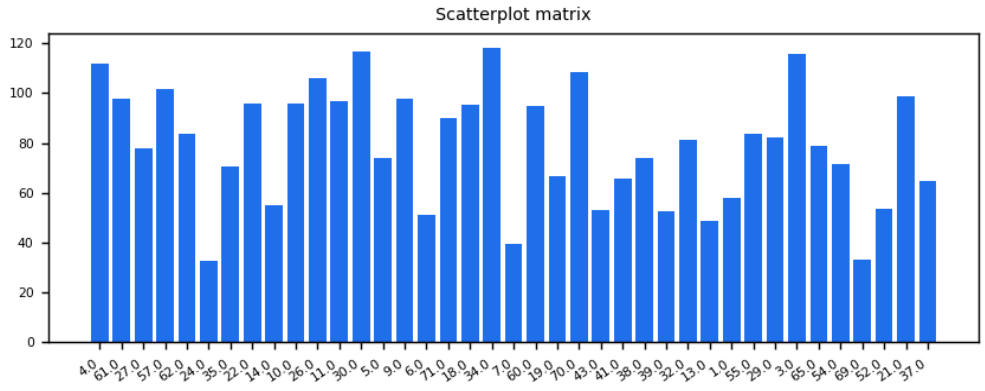
STEPS

1. Statistics → **Correlate** → **Correlation matrix**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, method=pearson`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Pearson correlation matrix

	tenure_months	monthly_charges	support_calls
tenure_months	1	-0.034	0.071*
monthly_charges	-0.034	1	0.019
support_calls	0.071*	0.019	1



DETAILED INTERPRETATION

* p < .05 · ** p < .01 (pairwise deletion). The lead result here reads = tenure_months; tenure_months = 1; monthly_charges = -0.034; support_calls = 0.071*. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.19 Correlation matrix + heatmap

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Correlation matrix with heatmap. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Correlate** → **Correlation matrix + heatmap**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, method=pearson`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Pearson correlation matrix

Variable	tenure_months	monthly_charges	support_calls
tenure_months	1	-0.034	0.0712
monthly_charges	-0.034	1	0.0194
support_calls	0.0712	0.0194	1

DETAILED INTERPRETATION

Pairwise correlations (–1 to 1). The lead result here reads Variable = tenure_months; tenure_months = 1; monthly_charges = -0.034; support_calls = 0.0712. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.20 Correlation with CI

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Pearson r + Fisher-z 95% CI. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

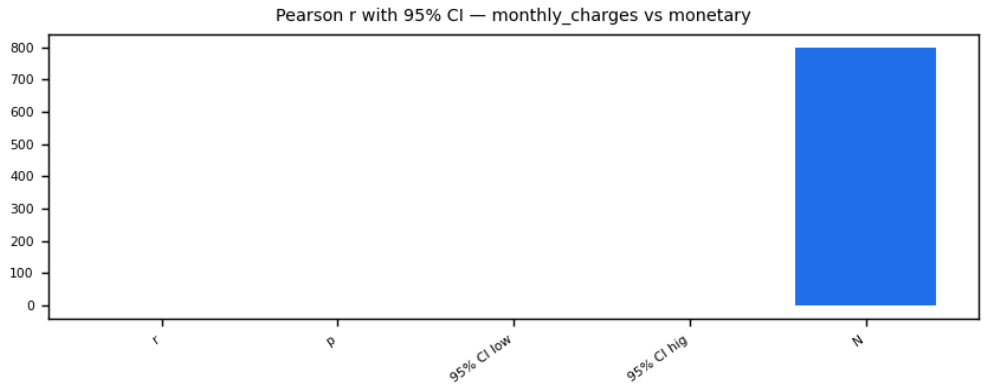
STEPS

1. Statistics → **Correlate** → **Correlation with CI**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Pearson r with 95% CI — monthly_charges vs monetary

r	p	95% CI low	95% CI high	N
-0.0103	0.7707	-0.0796	0.059	800



DETAILED INTERPRETATION

Confidence interval via the Fisher z-transformation. The lead result here reads r = -0.0103; p = 0.7707; 95% CI low = -0.0796; 95% CI high = 0.059. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.21 Covariance matrix

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Variance-covariance table. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

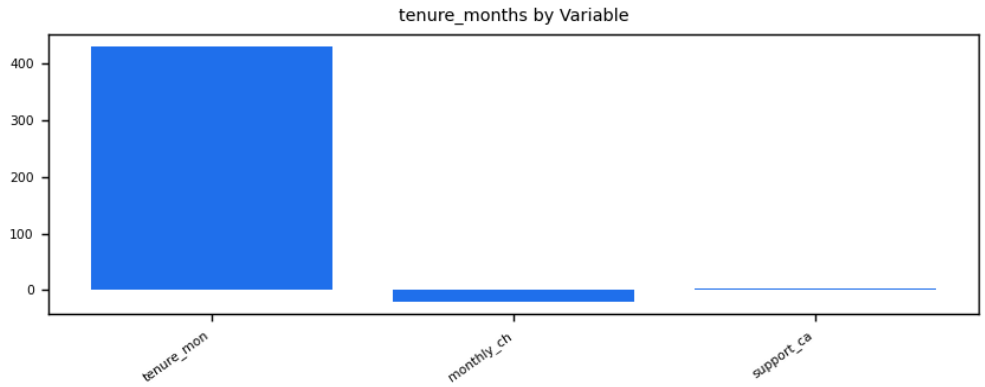
STEPS

1. Statistics → **Correlate** → **Covariance matrix**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Covariance matrix

Variable	tenure_months	monthly_charges	support_calls
tenure_months	429.2	-20.24	2.066
monthly_charges	-20.24	826.1	0.7806
support_calls	2.066	0.7806	1.962



DETAILED INTERPRETATION

Unstandardized co-variation; the diagonal holds each variable's variance. The lead result here reads Variable = tenure_months; tenure_months = 429.2; monthly_charges = -20.24; support_calls = 2.066. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.22 Distance correlation

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Detects any (nonlinear) dependence. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

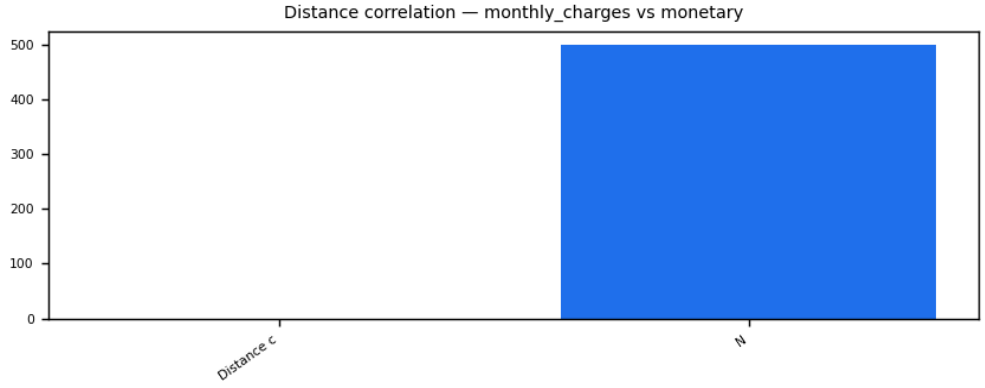
STEPS

1. Statistics → **Correlate** → **Distance correlation**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Distance correlation — monthly_charges vs monetary

Distance correlation	N
0.0811	500



DETAILED INTERPRETATION

Detects ANY dependence (linear or nonlinear); 0 = independent, 1 = perfect. The lead result here reads Distance correlation = 0.0811; N = 500. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.23 Kendall's τ-b correlation

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Rank correlation robust to ties. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

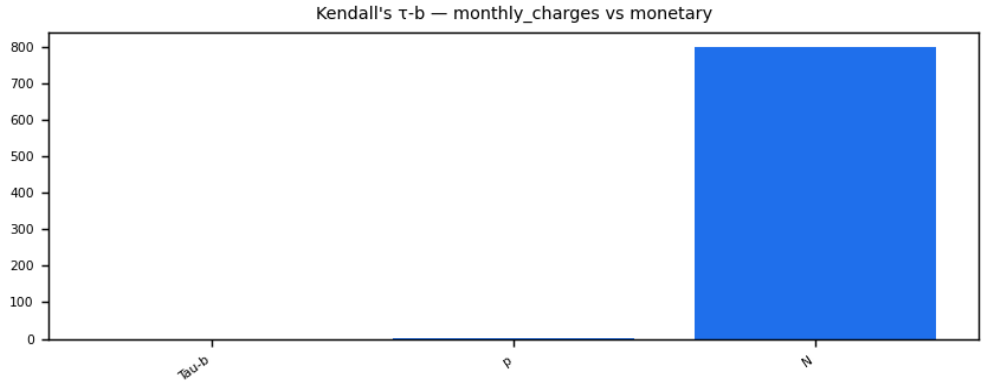
STEPS

1. Statistics → **Correlate** → **Kendall's τ-b correlation**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Kendall's τ-b — monthly_charges vs monetary

Tau-b	p	N
-0.0062	0.7932	800



DETAILED INTERPRETATION

Rank correlation robust to ties; suited to ordinal data. The lead result here reads Tau-b = -0.0062; p = 0.7932; N = 800. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.24 Multiple correlation (R)

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Multiple R of outcome on predictors. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

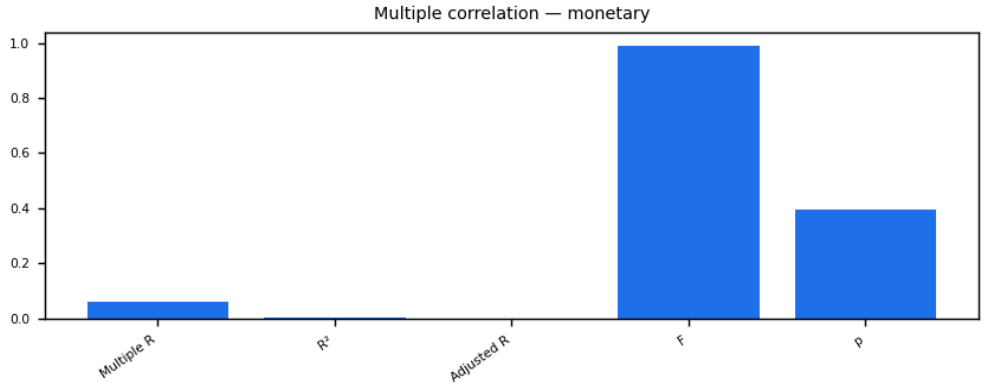
STEPS

1. Statistics → **Correlate** → **Multiple correlation (R)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Multiple correlation — monetary

Multiple R	R²	Adjusted R²	F	p
0.061	0.0037	-0	0.9894	0.3971



DETAILED INTERPRETATION

Correlation between the outcome and its best linear combination of predictors. The lead result here reads Multiple R = 0.061; R² = 0.0037; Adjusted R² = -0; F = 0.9894. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.25 Mutual information

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Nonlinear categorical dependence. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

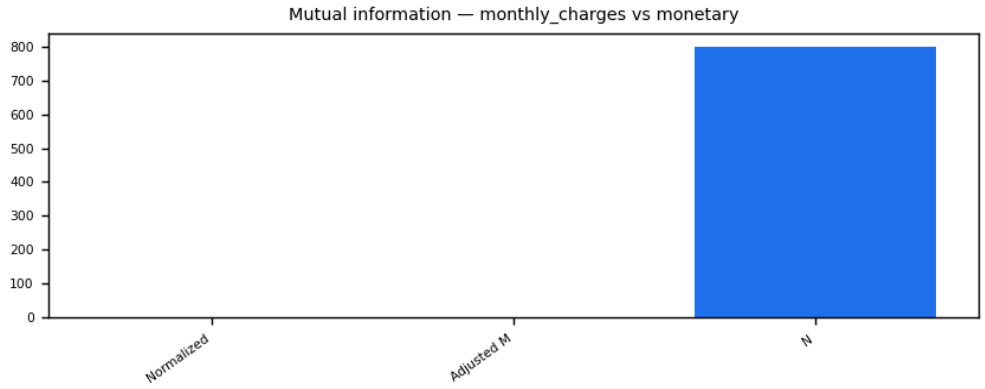
STEPS

1. Statistics → **Correlate** → **Mutual information**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Mutual information — monthly_charges vs monetary

Normalized MI	Adjusted MI	N
0.9968	-0	800



DETAILED INTERPRETATION

Nonlinear dependence between two categorical variables (0 = independent, 1 = identical). The lead result here reads Normalized MI = 0.9968; Adjusted MI = -0; N = 800. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.26 Ordinal association (γ, τ-c)

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Goodman-Kruskal γ, Kendall τ-c. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on contract, monthly_charges. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

contract	monthly_charges
Month-to-month	52.94
One year	113.7
Month-to-month	35.51
Two year	71.45
Two year	29.16

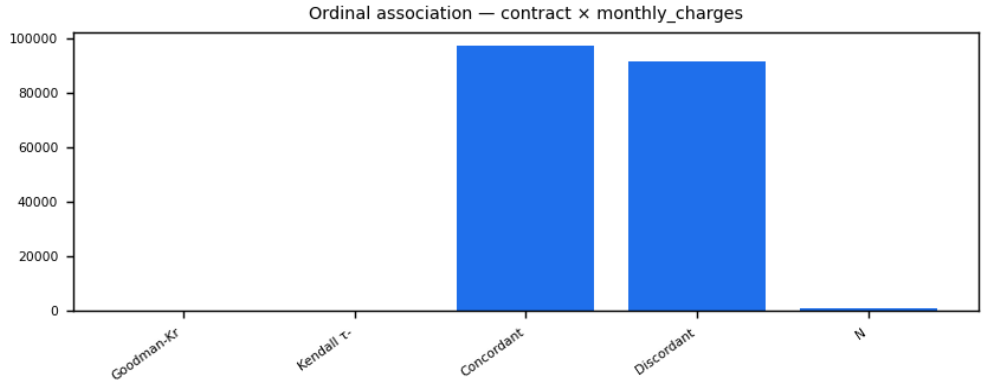
STEPS

1. Statistics → **Correlate** → **Ordinal association (γ, τ-c)**. 2. Select the demo dataset above. 3. Set parameters — `row=contract, column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Ordinal association — contract × monthly_charges

Goodman-Kruskal γ	Kendall τ-c	Concordant	Discordant	N
0.0301	0.0266	9.721e+04	9.154e+04	800



DETAILED INTERPRETATION

γ and τ-c measure the strength/direction of association between two ordinal variables. The lead result here reads Goodman-Kruskal γ = 0.0301; Kendall τ-c = 0.0266; Concordant = 9.721e+04; Discordant = 9.154e+04. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.27 Partial correlation

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Correlation between two variables controlling for others. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary, tenure_months. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

monthly_charges	monetary	tenure_months
52.94	3197	68
113.7	1233	45
35.51	2501	49
71.45	1468	64
29.16	4195	42

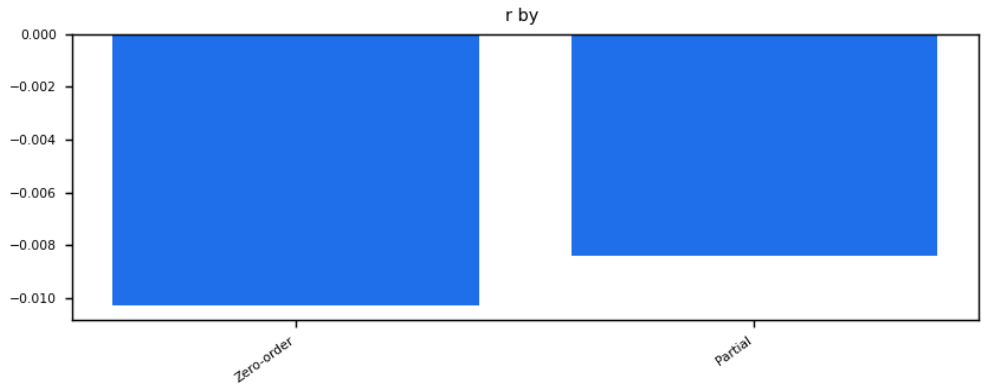
STEPS

1. Statistics → **Correlate** → **Partial correlation**. 2. Select the demo dataset above. 3. Set parameters — `x=monthly_charges, y=monetary, controls=tenure_months`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Partial correlation — monthly_charges & monetary controlling for tenure_months

	r	p	df
Zero-order	-0.0103	0.7707	798
Partial	-0.0084	0.8125	797



DETAILED INTERPRETATION

The lead result here reads = Zero-order; r = -0.0103; p = 0.7707; df = 798. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.28 Spearman ρ with CI

LIVE OUTPUT

Group: Correlate · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Rank correlation + confidence interval. It belongs to the **Correlate** family, used to measure the strength and direction of association between variables. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to screen which variables move together before building a model.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

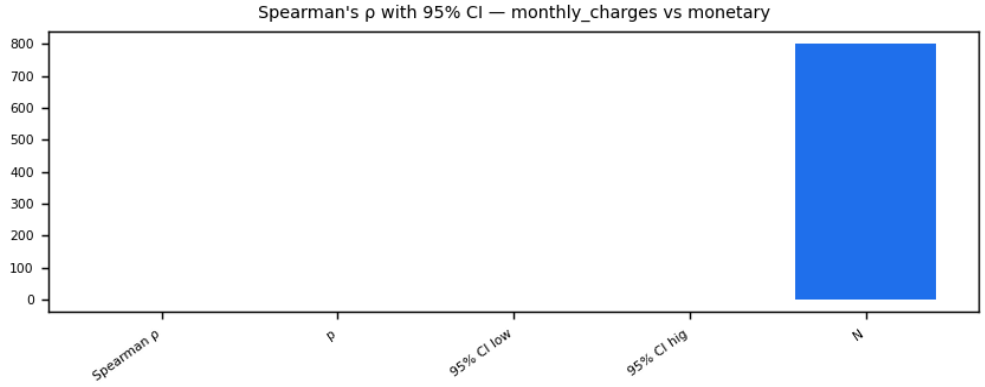
STEPS

1. Statistics → **Correlate** → **Spearman ρ with CI**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Spearman's ρ with 95% CI — monthly_charges vs monetary

Spearman ρ	p	95% CI low	95% CI high	N
-0.0092	0.794	-0.0806	0.0622	800



DETAILED INTERPRETATION

Rank correlation with a confidence interval (Fisher-z approximation). The lead result here reads Spearman ρ = -0.0092; p = 0.794; 95% CI low = -0.0806; 95% CI high = 0.0622. Coefficients run from -1 to +1 - the sign gives direction and the magnitude the strength, while the p-value shows whether it is distinguishable from zero. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst screen which variables move together before building a model.

1.29 Anderson-Darling normality

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Powerful EDF normality test. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

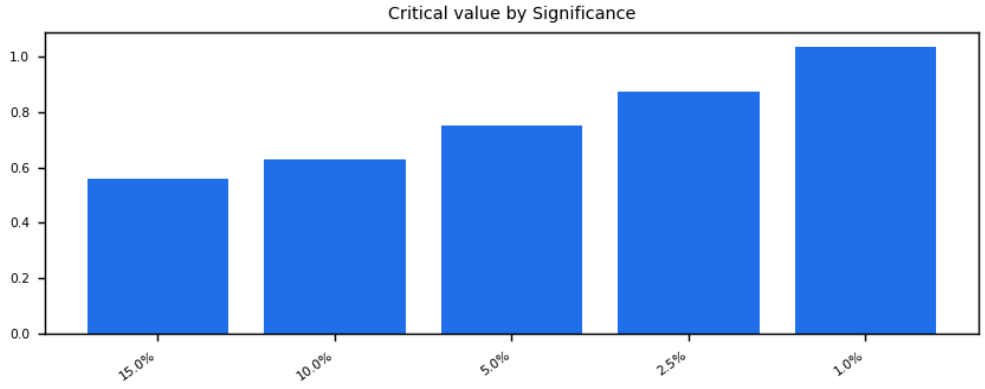
STEPS

1. Statistics → **Descriptive** → **Anderson-Darling normality**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Anderson-Darling normality — monthly_charges

Significance	Critical value	Decision
15.0%	0.56	reject normal
10.0%	0.63	reject normal
5.0%	0.751	reject normal
2.5%	0.872	reject normal
1.0%	1.034	reject normal



DETAILED INTERPRETATION

A^2 = 8.5021. Exceeding the critical value rejects normality at that level. The lead result here reads Significance = 15.0%; Critical value = 0.56; Decision = reject normal. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.30 Bimodality coefficient

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Detect multimodal distributions. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

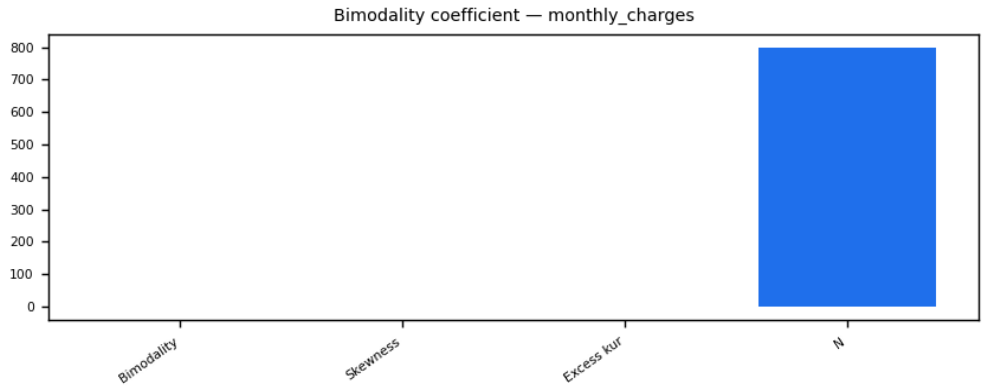
STEPS

1. Statistics → **Descriptive** → **Bimodality coefficient**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Bimodality coefficient — monthly_charges

Bimodality coefficient	Skewness	Excess kurtosis	N
0.5472	0.0465	-1.18	800



DETAILED INTERPRETATION

BC > 0.555 suggests a bimodal/multimodal distribution (uniform ≈ 0.555). The lead result here reads Bimodality coefficient = 0.5472; Skewness = 0.0465; Excess kurtosis = -1.18; N = 800. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.31 Bootstrap CI

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Distribution-free bootstrap confidence interval for a mean/median. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94

monthly_charges
113.7
35.51
71.45
29.16

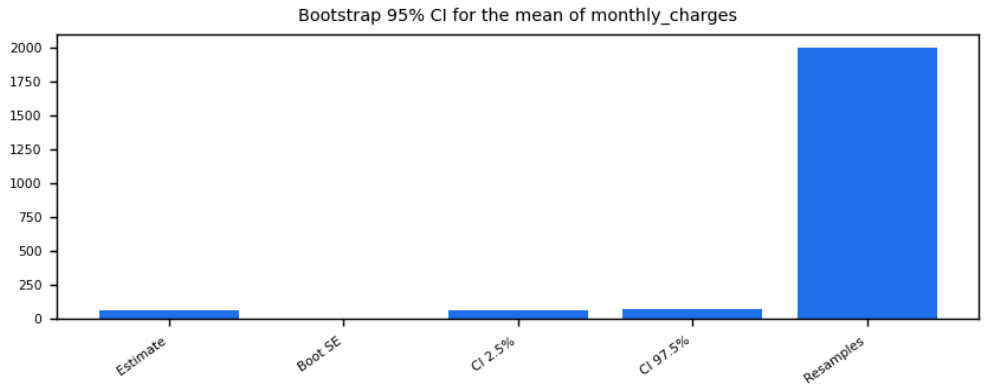
STEPS

1. Statistics → **Descriptive** → **Bootstrap CI**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, statistic=mean`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Bootstrap 95% CI for the mean of monthly_charges

Statistic	Estimate	Boot SE	CI 2.5%	CI 97.5%	Resamples
mean	69.09	1.051	67.03	71.2	2000



DETAILED INTERPRETATION

Percentile bootstrap — distribution-free confidence interval. The lead result here reads Statistic = mean; Estimate = 69.09; Boot SE = 1.051; CI 2.5% = 67.03. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.32 Chi-square goodness-of-fit

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Category counts vs uniform. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

STEPS

1. Statistics → **Descriptive** → **Chi-square goodness-of-fit**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

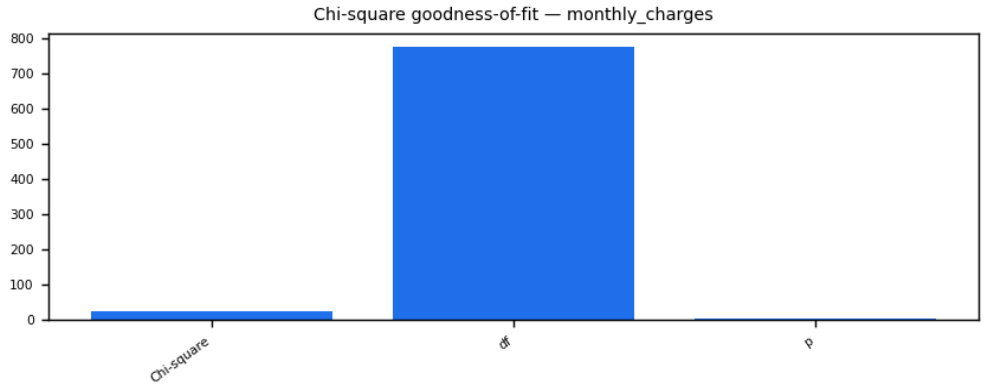
OUTPUT

Chi-square goodness-of-fit — monthly_charges

Chi-square	df	p
24.5	775	1

Observed vs expected

Category	Observed	Expected
20.33	1	1.031
20.43	1	1.031
20.57	1	1.031
20.59	1	1.031
20.82	1	1.031
20.94	1	1.031
21.27	2	1.031
21.33	1	1.031
21.4	1	1.031
21.51	3	1.031
21.54	1	1.031
21.83	1	1.031



DETAILED INTERPRETATION

Tests whether category frequencies match a uniform (equal) expectation. The lead result here reads Chi-square = 24.5; df = 775; p = 1. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.33 Complex samples (survey design)

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Design-based mean/total/proportion with strata, clusters & weights (DEFF, CV, CI). It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, frequency, region. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monetary	frequency	region
3197	46	East
1233	33	North
2501	6	North
1468	6	West
4195	38	West

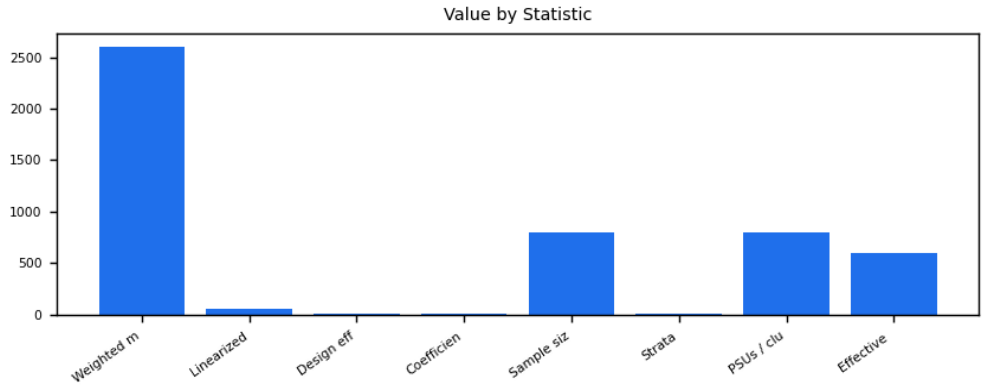
STEPS

1. Statistics → **Descriptive** → **Complex samples (survey design)**. 2. Select the demo dataset above. 3. Set parameters — `variable=monetary, weight=frequency, estimand=mean, strata=region`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Complex-samples estimate — monetary (mean)

Statistic	Value
Weighted mean	2604
Linearized SE	56.84
95% CI	[2492.9151, 2716.0590]
Design effect (DEFF)	1.329
Coefficient of variation (…	2.182
Sample size (n)	800
Strata	4
PSUs / clusters	800
Effective sample size (n/D…	601.8



DETAILED INTERPRETATION

Taylor-linearization (ultimate-cluster) variance honouring strata + weights. DEFF > 1 means the design is less efficient than simple random sampling. The lead result here reads Statistic = Weighted mean; Value = 2604. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.34 Cramér-von Mises normality

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

EDF normality test. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

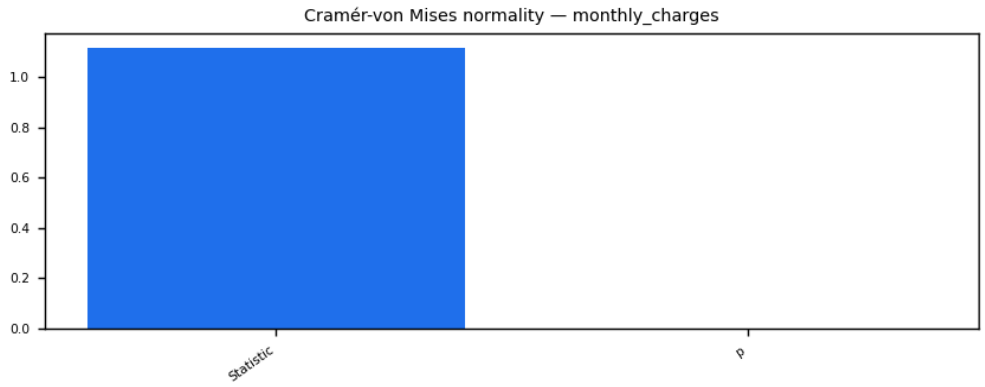
STEPS

1. Statistics → **Descriptive** → **Cramér-von Mises normality**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Cramér-von Mises normality — monthly_charges

Statistic	p
1.115	0.0013



DETAILED INTERPRETATION

EDF-based normality test, sensitive across the whole distribution. The lead result here reads Statistic = 1.115; p = 0.0013. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.35 Crosstab with row %

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Contingency table with percentages. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on contract, monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

contract	monthly_charges
Month-to-month	52.94
One year	113.7
Month-to-month	35.51
Two year	71.45
Two year	29.16

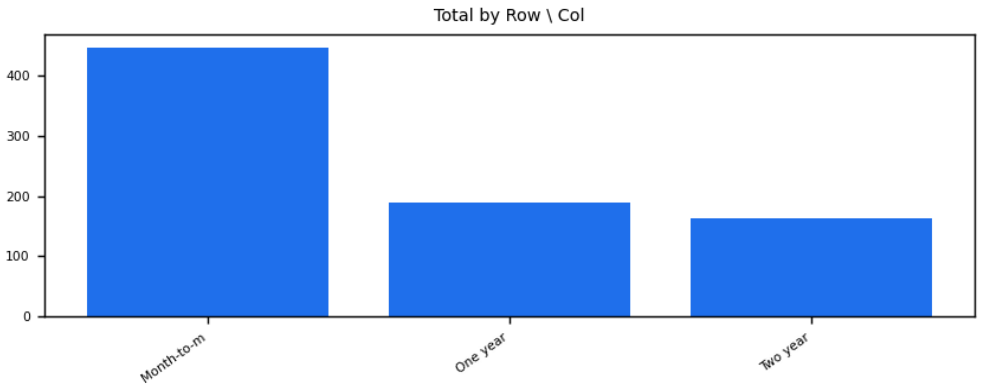
STEPS

1. Statistics → **Descriptive** → **Crosstab with row %**. 2. Select the demo dataset above. 3. Set parameters — `row=contract, column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Crosstab with row % — contract × monthly_charges

Row \ Col	20.33	20.43	20.57	20.59	20.82	20.94	21.27	21.33	...
Month-to-month	0 (0%)	1 (0%)	1 (0%)	0 (0%)	1 (0%)	1 (0%)	0 (0%)	1 (0%)	...
One year	1 (1%)	0 (0%)	0 (0%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	...
Two year	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (1%)	0 (0%)	...



DETAILED INTERPRETATION

$\chi^2(1550) = 1565.93$, $p = 0.3832$. Cells show count (row %). The lead result here reads Row \ Col = Month-to-month; 20.33 = 0 (0%); 20.43 = 1 (0%); 20.57 = 1 (0%). Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.36 Crosstabs & chi-square

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Contingency table with chi-square test and Cramér's V. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on contract, region. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

contract	region
Month-to-month	East
One year	North
Month-to-month	North
Two year	West
Two year	West

STEPS

1. Statistics → **Descriptive** → **Crosstabs & chi-square**. 2. Select the demo dataset above. 3. Set parameters — `row=contract, column=region`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Crosstab — contract × region (observed)

contract	East	North	South	West	Total
Month-to-month	90	117	127	113	447
One year	42	48	56	44	190
Two year	40	39	40	44	163
Total	172	204	223	201	800

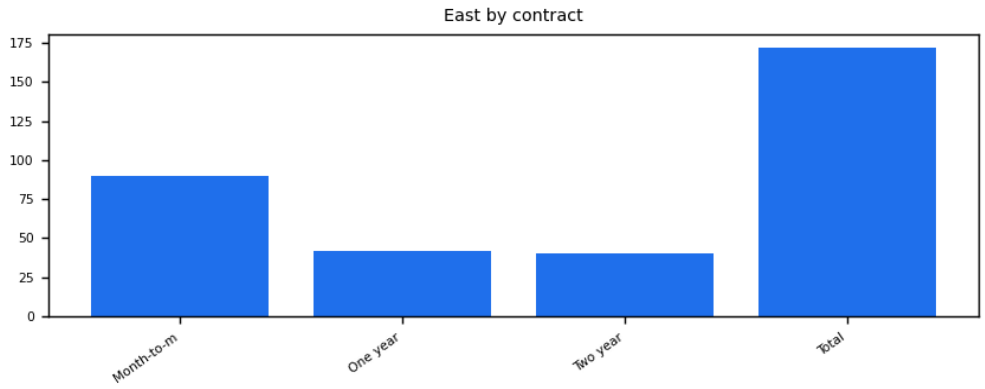
Expected counts

contract	East	North	South	West
Month-to-month	96.11	114	124.6	112.3
One year	40.85	48.45	52.96	47.74
Two year	35.05	41.56	45.44	40.95

Chi-square test

Statistic	Value
Pearson chi-square	2.757
df	6

Statistic	Value
p	0.8386
Cramér's V	0.0415
N	800



DETAILED INTERPRETATION

The lead result here reads contract = Month-to-month; East = 90; North = 117; South = 127. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.37 Custom tables (banner + sig.)

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Presentation crosstab with column % and column-proportion significance letters. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on churn, contract. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

churn	contract
No	Month-to-month
Yes	One year
No	Month-to-month
No	Two year
No	Two year

STEPS

1. Statistics → **Descriptive** → **Custom tables (banner + sig.)**. 2. Select the demo dataset above. 3. Set parameters — `row=churn, column=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Custom table — churn × contract (counts, column %, sig.)

churn \ contract	Month-to-month (A)	One year (B)	Two year (C)
No	281 (63%)	152 (80%) A	143 (88%) A
Yes	166 (37%) BC	38 (20%)	20 (12%)

DETAILED INTERPRETATION

Cell = count (column %). A superscript letter means this column's proportion is significantly greater (p<.05) than the lettered column. Columns: A=Month-to-month, B=One year, C=Two year. Overall $\chi^2(2) = 44.55$, p = < .001. The lead result here reads churn \ contract = No; Month-to-month (A) = 281 (63%); One year (B) = 152 (80%) A; Two year (C) = 143 (88%) A. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.38 D'Agostino K² normality

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Omnibus skew+kurtosis normality test. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

STEPS

1. Statistics → **Descriptive** → **D'Agostino K² normality**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

D'Agostino-Pearson K² normality — monthly_charges

K²	p
485.6	< .001

DETAILED INTERPRETATION

Omnibus normality test combining skewness and kurtosis. The lead result here reads K² = 485.6; p = < .001. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.39 Descriptives

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Mean, SD, CI, quartiles, skewness, kurtosis per variable. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

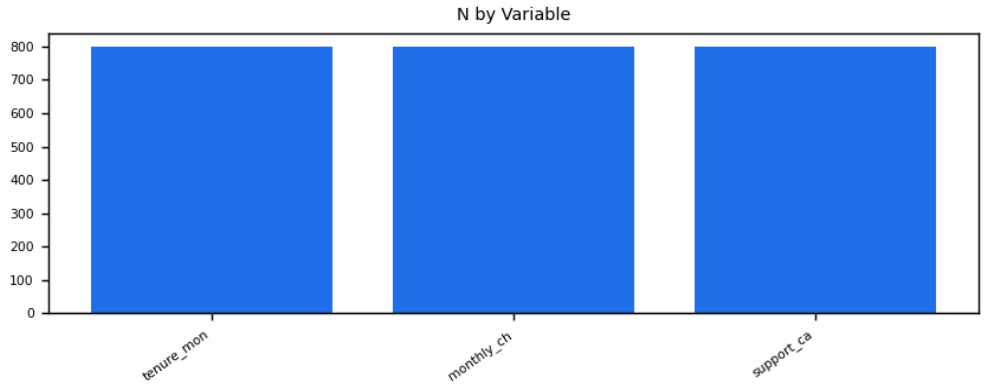
STEPS

1. Statistics → **Descriptive** → **Descriptives**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Descriptive statistics

Variable	N	Mean	SD	SE	95% CI low	95% CI high	Min	Q1	...
tenure_months	800	37.1	20.72	0.7324	35.67	38.54	1	18	...
monthly_charges	800	69.09	28.74	1.016	67.1	71.09	20.33	45.35	...
support_calls	800	1.954	1.401	0.0495	1.857	2.051	0	1	...



DETAILED INTERPRETATION

The lead result here reads Variable = tenure_months; N = 800; Mean = 37.1; SD = 20.72. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.40 Descriptives (all variables)

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Summary of every numeric column. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

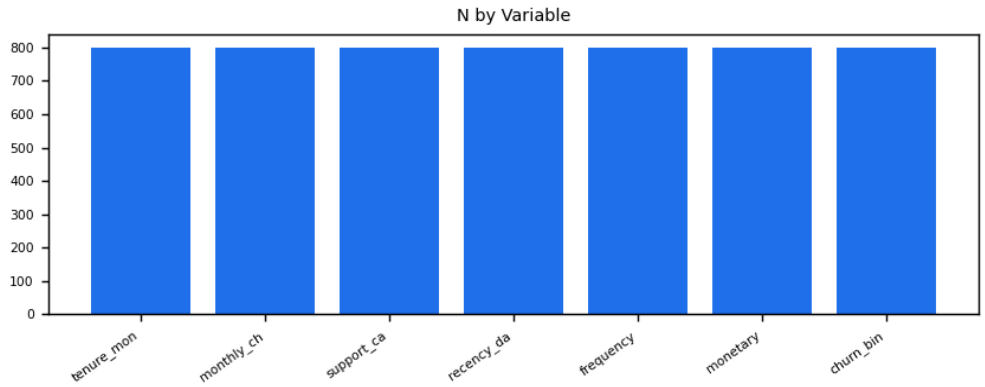
STEPS

1. Statistics → **Descriptive** → **Descriptives (all variables)**. 2. Select the demo dataset above. 3. Set parameters — `(defaults)`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Descriptive statistics — all numeric variables

Variable	N	Mean	SD	Min	Q1	Median	Q3	Max
tenure_months	800	37.1	20.72	1	18	37.5	55.25	71
monthly_charges	800	69.09	28.74	20.33	45.35	67.69	93.76	119.9
support_calls	800	1.954	1.401	0	1	2	3	7
recency_days	800	178.3	103.3	1	88	174.5	263.2	364
frequency	800	24.73	14.04	1	12	26	36	49
monetary	800	2634	1396	65.15	1466	2791	3750	4993
churn_bin	800	0.28	0.4493	0	0	0	1	1



DETAILED INTERPRETATION

One-shot summary of every numeric column. The lead result here reads Variable = tenure_months; N = 800; Mean = 37.1; SD = 20.72. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.41 Descriptives by group

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Group-wise summary statistics. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

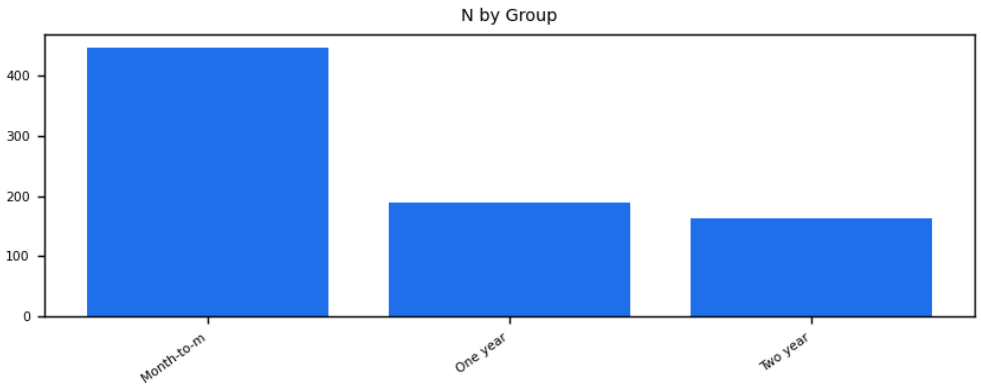
STEPS

1. Statistics → **Descriptive** → **Descriptives by group**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Descriptives by contract — monthly_charges

Group	N	Mean	SD	Min	Median	Max
Month-to-month	447	68.26	28.78	20.43	66.54	119.9
One year	190	70.33	27.66	20.33	70.71	119.7
Two year	163	69.91	29.94	21.27	69.09	119



DETAILED INTERPRETATION

Group-wise descriptive statistics. The lead result here reads Group = Month-to-month; N = 447; Mean = 68.26; SD = 28.78. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.42 Distribution fitting

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Fit & compare Normal/Log-normal/Exp/Gamma/Weibull by AIC. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

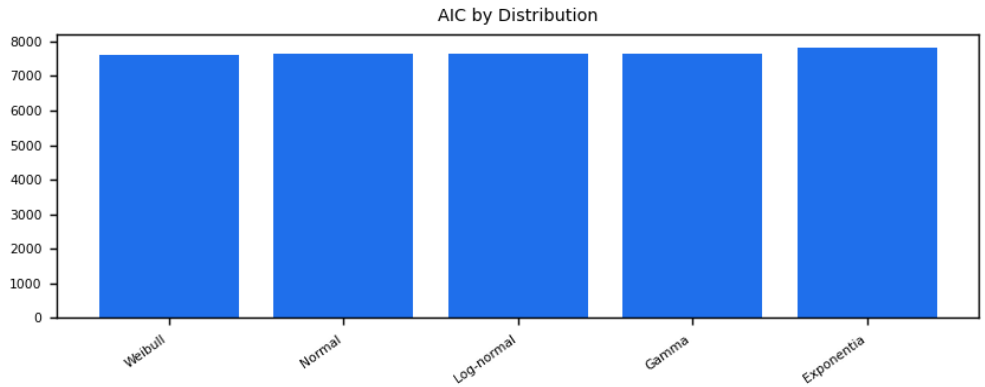
STEPS

1. Statistics → **Descriptive** → **Distribution fitting**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Distribution fitting — monthly_charges

Distribution	AIC	KS statistic	KS p
Weibull	7601	0.0651	0.0022
Normal	7647	0.0639	0.0028
Log-normal	7648	0.0616	0.0044
Gamma	7648	0.0622	0.0039
Exponential	7823	0.1576	< .001



DETAILED INTERPRETATION

Lowest AIC = best fit; KS p > .05 means that distribution is not rejected. The lead result here reads Distribution = Weibull; AIC = 7601; KS statistic = 0.0651; KS p = 0.0022. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.43 Diversity indices

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Shannon/Simpson diversity. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

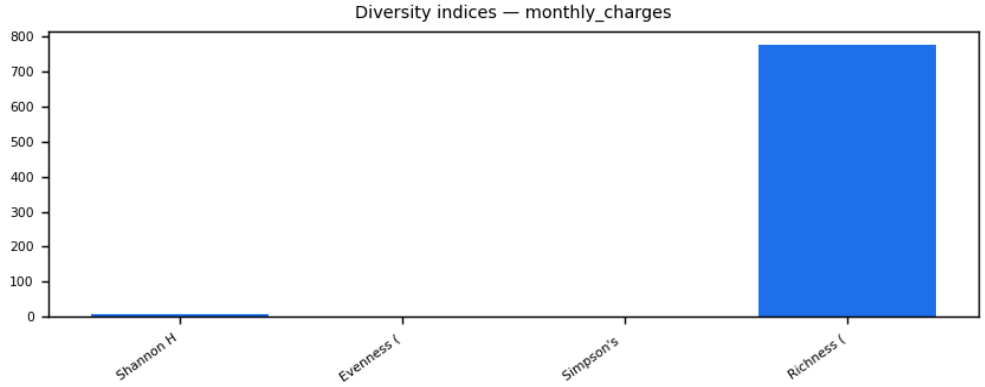
STEPS

1. Statistics → **Descriptive** → **Diversity indices**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Diversity indices — monthly_charges

Shannon H	Evenness (H/Hmax)	Simpson's D	Richness (categories)
6.642	0.9982	0.9987	776



DETAILED INTERPRETATION

Diversity/concentration of a categorical variable (ecology, market share, text). The lead result here reads Shannon H = 6.642; Evenness (H/Hmax) = 0.9982; Simpson's D = 0.9987; Richness (categories) = 776. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.44 Exact binomial test

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Test a success proportion. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

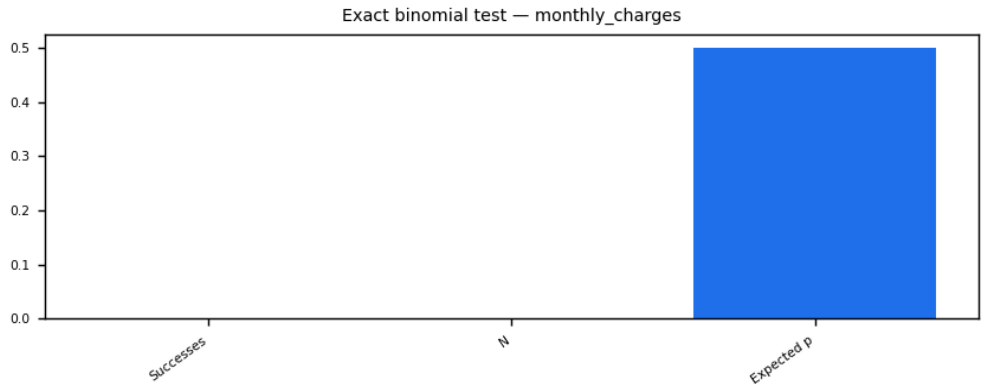
STEPS

1. Statistics → **Descriptive** → **Exact binomial test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, prob=0.5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Exact binomial test — monthly_charges

Successes	N	Observed p	Expected p	p-value
0	0		0.5	



DETAILED INTERPRETATION

Exact test that the success proportion equals the expected value. The lead result here reads Successes = 0; N = 0; Observed p = ; Expected p = 0.5. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.45 Frequencies

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Counts, percentages and cumulative % of a categorical variable. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

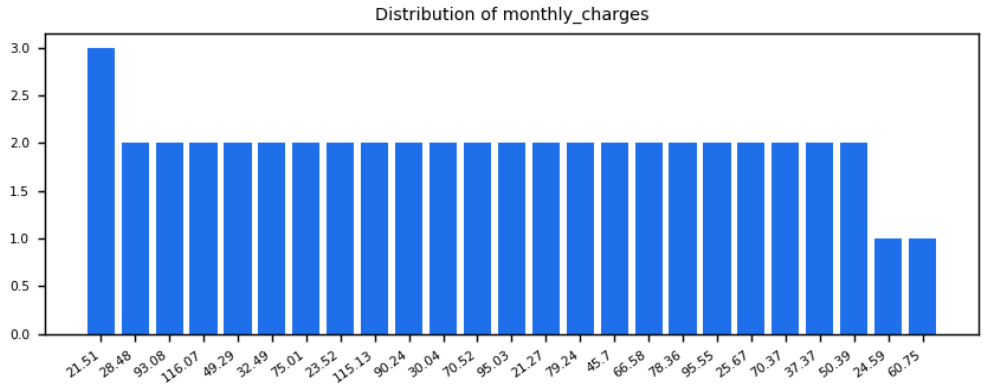
STEPS

1. Statistics → **Descriptive** → **Frequencies**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Frequencies — monthly_charges

Value	Count	%	Cumulative %
21.51	3	0.375	0.375
28.48	2	0.25	0.625
93.08	2	0.25	0.875
116.07	2	0.25	1.125
49.29	2	0.25	1.375
32.49	2	0.25	1.625
75.01	2	0.25	1.875
23.52	2	0.25	2.125
115.13	2	0.25	2.375
90.24	2	0.25	2.625
30.04	2	0.25	2.875
70.52	2	0.25	3.125



DETAILED INTERPRETATION

800 valid observations · showing top 100 of 776 values. The lead result here reads Value = 21.51; Count = 3; % = 0.375; Cumulative % = 0.375. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.46 Frequency distribution (histogram)

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Binned counts + histogram. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

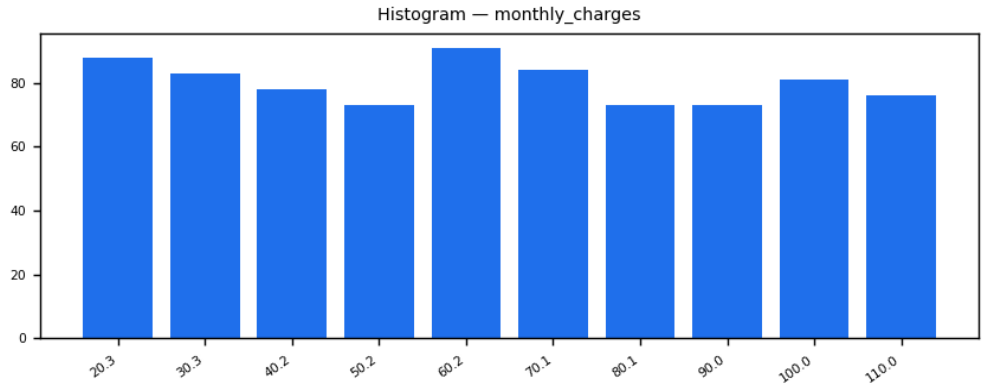
STEPS

1. Statistics → **Descriptive** → **Frequency distribution (histogram)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, bins=10`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Frequency distribution — monthly_charges

Bin	Count	%
20.33–30.29	88	11
30.29–40.25	83	10.38
40.25–50.20	78	9.75
50.20–60.16	73	9.125
60.16–70.12	91	11.38
70.12–80.08	84	10.5
80.08–90.04	73	9.125
90.04–99.99	73	9.125
99.99–109.95	81	10.12
109.95–119.91	76	9.5



DETAILED INTERPRETATION

10 equal-width bins. The lead result here reads Bin = 20.33–30.29; Count = 88; % = 11. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.47 Frequency table

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Counts of a categorical variable. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

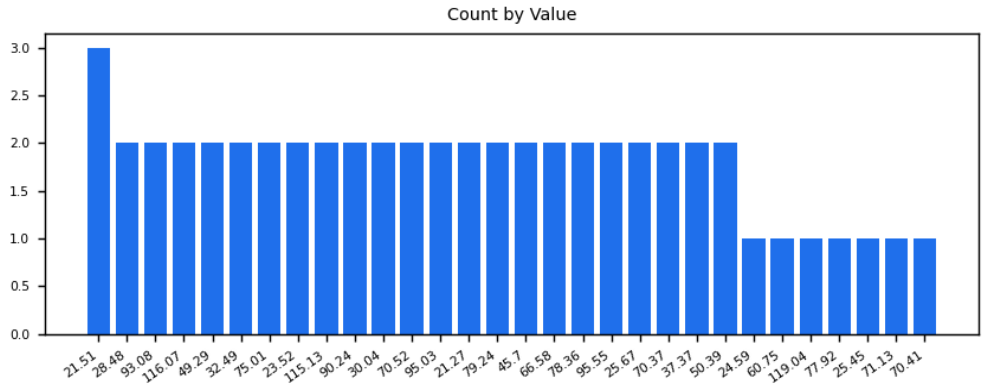
STEPS

1. Statistics → **Descriptive** → **Frequency table**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Frequency table — monthly_charges

Value	Count	%
21.51	3	0.375
28.48	2	0.25
93.08	2	0.25
116.07	2	0.25
49.29	2	0.25
32.49	2	0.25
75.01	2	0.25
23.52	2	0.25
115.13	2	0.25
90.24	2	0.25
30.04	2	0.25
70.52	2	0.25



DETAILED INTERPRETATION

776 distinct values (top 30 shown). The lead result here reads Value = 21.51; Count = 3; % = 0.375. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.48 Geometric & harmonic means

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Geo/harmonic mean, CV, MAD. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

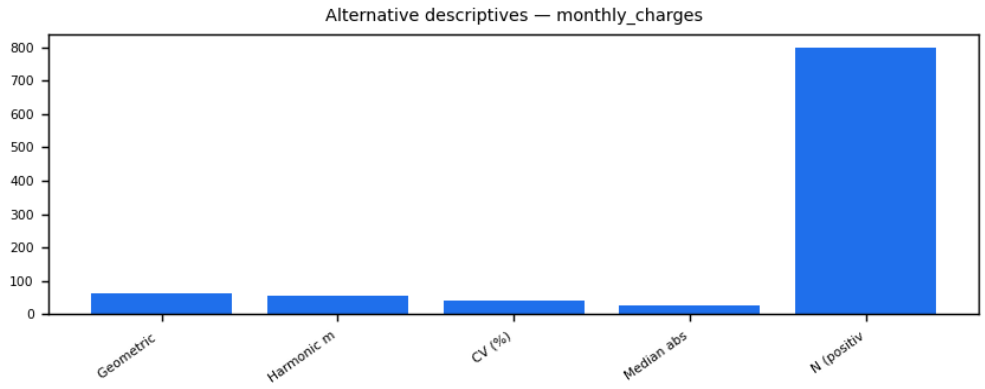
STEPS

1. Statistics → **Descriptive** → **Geometric & harmonic means**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Alternative descriptives — monthly_charges

Geometric mean	Harmonic mean	CV (%)	Median abs. deviation	N (positive)
62.33	55.13	41.6	24.52	800



DETAILED INTERPRETATION

Geometric/harmonic means suit rates & ratios; CV compares dispersion across scales. The lead result here reads Geometric mean = 62.33; Harmonic mean = 55.13; CV (%) = 41.6; Median abs. deviation = 24.52. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.49 Gini coefficient

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Inequality/concentration index. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94

monthly_charges
113.7
35.51
71.45
29.16

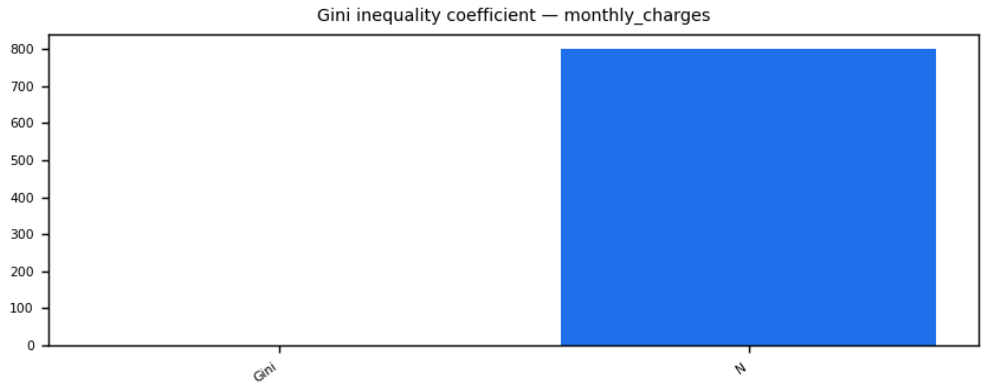
STEPS

1. Statistics → **Descriptive** → **Gini coefficient**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Gini inequality coefficient — monthly_charges

Gini	N
0.2399	800



DETAILED INTERPRETATION

0 = perfect equality, 1 = maximal inequality. Common for income/wealth/market share. The lead result here reads Gini = 0.2399; N = 800. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.50 Grubbs' outlier test

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Detect the single most extreme outlier. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

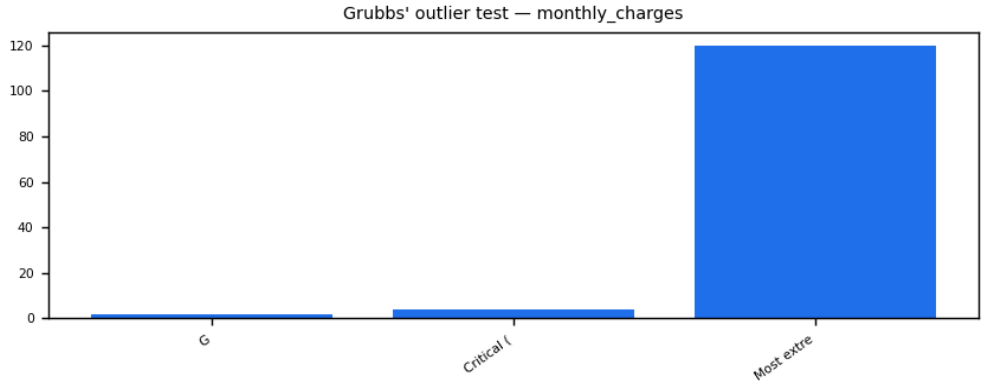
STEPS

1. Statistics → **Descriptive** → **Grubbs' outlier test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Grubbs' outlier test — monthly_charges

G	Critical (α=.05)	Most extreme value	Outlier?
1.768	3.984	119.9	no



DETAILED INTERPRETATION

Detects the single most extreme value as an outlier (assumes near-normal data). The lead result here reads G = 1.768; Critical (α=.05) = 3.984; Most extreme value = 119.9; Outlier? = no. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.51 IQR outlier screen

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Tukey 1.5×IQR outlier detection + boxplot. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

STEPS

1. Statistics → **Descriptive** → **IQR outlier screen**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

IQR outlier screen — monthly_charges

Q1	Q3	IQR	Lower fence	Upper fence	# outliers	% outliers
45.35	93.76	48.41	-27.27	166.4	0	0

DETAILED INTERPRETATION

Tukey's rule: values beyond 1.5×IQR from the quartiles are flagged. The lead result here reads Q1 = 45.35; Q3 = 93.76; IQR = 48.41; Lower fence = -27.27. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.52 Jarque-Bera normality

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Normality via skewness & kurtosis. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

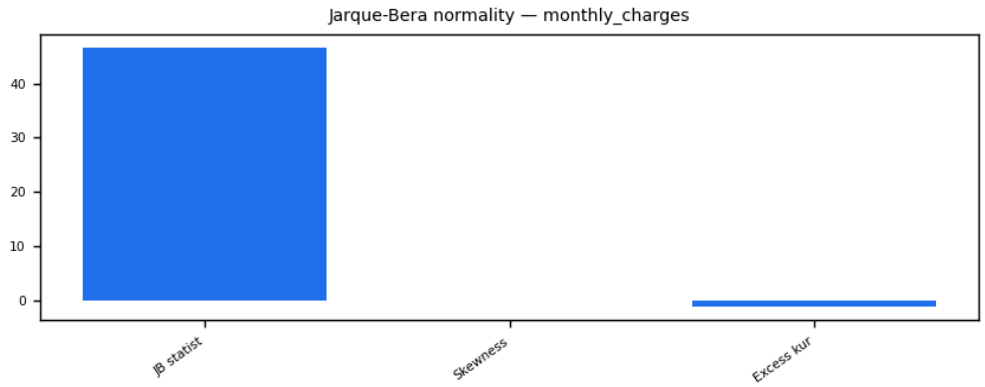
STEPS

1. Statistics → **Descriptive** → **Jarque-Bera normality**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Jarque-Bera normality — monthly_charges

JB statistic	p	Skewness	Excess kurtosis
46.68	< .001	0.0465	-1.18



DETAILED INTERPRETATION

p < .05 rejects normality (jointly tests skewness and kurtosis). The lead result here reads JB statistic = 46.68; p = < .001; Skewness = 0.0465; Excess kurtosis = -1.18. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.53 Little's MCAR test

INPUT REQUIREMENT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Test whether missing data are Missing Completely At Random (EM). It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5

tenure_months	monthly_charges	support_calls
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Descriptive** → **Little's MCAR test**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

No missing values — MCAR test needs missingness

DETAILED INTERPRETATION

This technique has a specific data requirement: *No missing values — MCAR test needs missingness*. Provide inputs of the expected shape. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a descriptive method and supports the decision to establish a clean, well-understood baseline before any modelling.

1.54 Missing-data summary

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Missingness per column. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

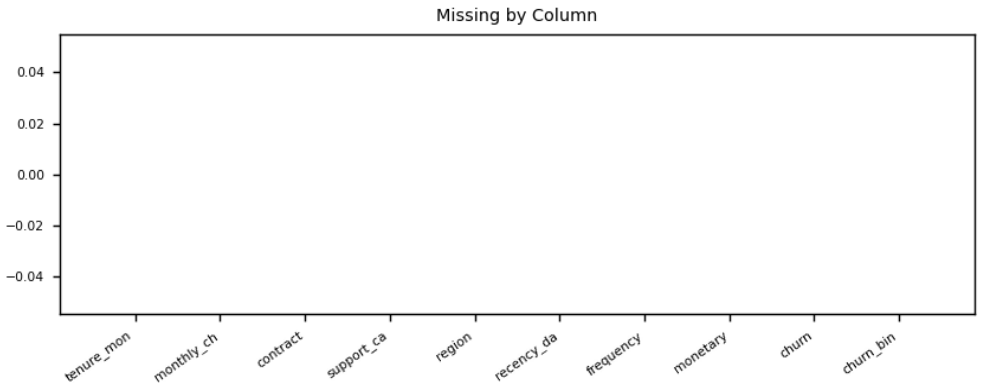
STEPS

1. Statistics → **Descriptive** → **Missing-data summary**. 2. Select the demo dataset above. 3. Set parameters — `(defaults)`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Missing-data summary

Column	Missing	% missing	Type
tenure_months	0	0	int64
monthly_charges	0	0	float64
contract	0	0	object
support_calls	0	0	int64
region	0	0	object
recency_days	0	0	int64
frequency	0	0	int64
monetary	0	0	float64
churn	0	0	object
churn_bin	0	0	int64



DETAILED INTERPRETATION

800 rows. Columns with high missingness may need imputation or exclusion. The lead result here reads Column = tenure_months; Missing = 0; % missing = 0; Type = int64. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.55 Normal tolerance interval

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Population coverage interval. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

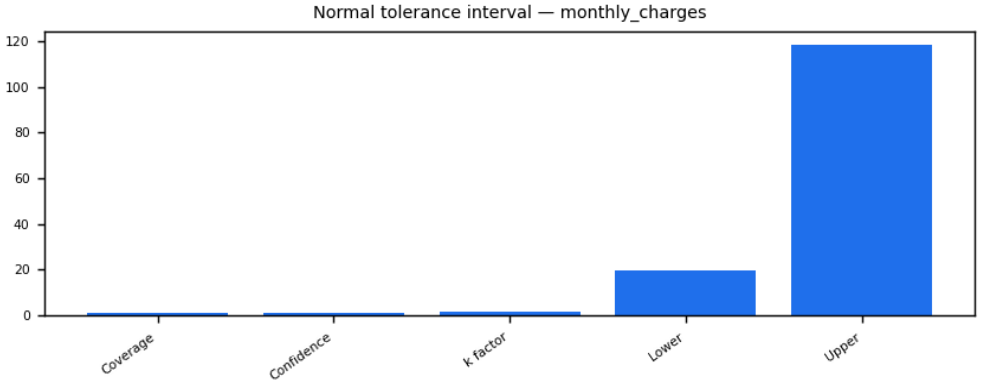
STEPS

1. Statistics → **Descriptive** → **Normal tolerance interval**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, coverage=0.9, confidence=0.95`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Normal tolerance interval — monthly_charges

Coverage	Confidence	k factor	Lower	Upper
0.9	0.95	1.717	19.75	118.4



DETAILED INTERPRETATION

Interval expected to contain 90% of the population with 95% confidence. The lead result here reads Coverage = 0.9; Confidence = 0.95; k factor = 1.717; Lower = 19.75. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.56 Normality test battery

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Shapiro+JB+D'Agostino+KS combined. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

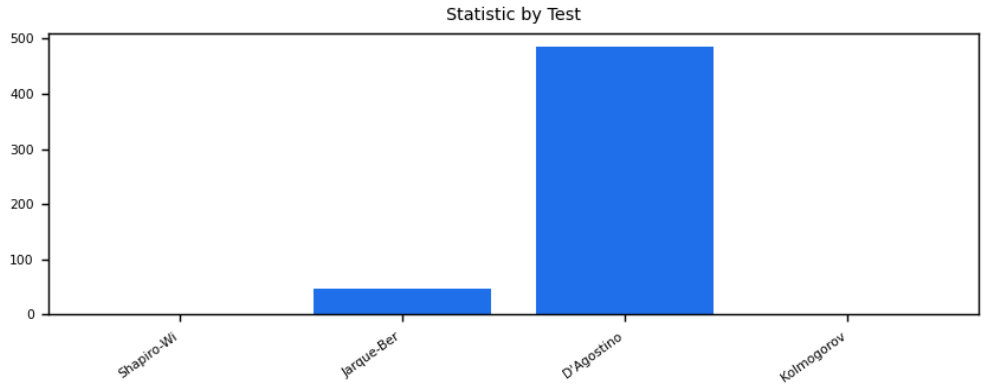
STEPS

1. Statistics → **Descriptive** → **Normality test battery**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Normality test battery — monthly_charges

Test	Statistic	p
Shapiro-Wilk	0.9552	< .001
Jarque-Bera	46.68	< .001
D'Agostino K²	485.6	< .001
Kolmogorov-Smirnov	0.0638	0.0029



DETAILED INTERPRETATION

1/4 tests reject normality at α=.05. Agreement across tests is more reliable than any one. The lead result here reads Test = Shapiro-Wilk; Statistic = 0.9552; p = < .001. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.57 Normality tests

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Shapiro-Wilk and Kolmogorov-Smirnov with histograms. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

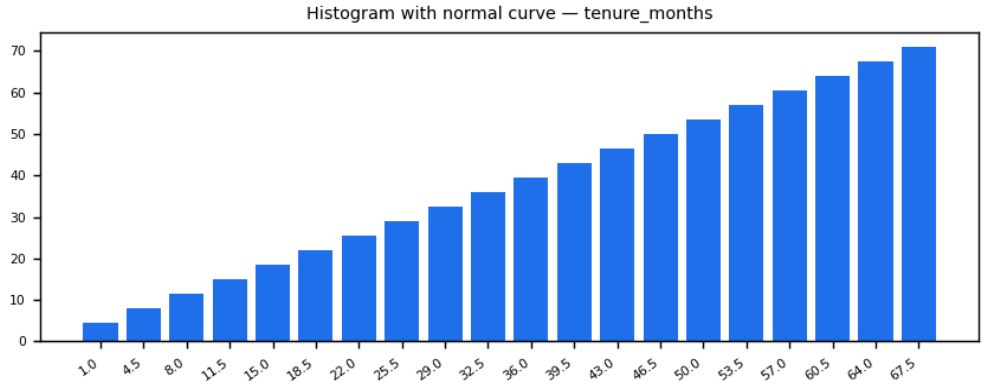
STEPS

1. Statistics → **Descriptive** → **Normality tests**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Tests of normality

Variable	N	Shapiro-Wilk W	SW p	Kolmogorov-Smirnov D	KS p	Skewness	Kurtosis
tenure_months	800	0.9469	< .001	0.0852	< .001	-0.0555	-1.268
monthly_charges	800	0.9552	< .001	0.0638	0.0029	0.0466	-1.18
support_calls	800	0.9167	< .001	0.1881	< .001	0.6963	0.263



DETAILED INTERPRETATION

Significant p (< .05) indicates departure from normality. Shapiro-Wilk computed on a 5,000-row sample for large data. The lead result here reads Variable = tenure_months; N = 800; Shapiro-Wilk W = 0.9469; SW p = < .001. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.58 One-proportion test

INPUT REQUIREMENT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Test a proportion against an expected value (z + exact binomial). It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on churn_bin. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

churn_bin
0
1
0
0
0

STEPS

1. Statistics → **Descriptive** → **One-proportion test**. 2. Select the demo dataset above. 3. Set parameters — `column=churn_bin`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Missing required parameter: expected

DETAILED INTERPRETATION

This technique has a specific data requirement: *Missing required parameter: expected*. Provide inputs of the expected shape. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a descriptive method and supports the decision to establish a clean, well-understood baseline before any modelling.

1.59 Percentile summary

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Deciles & tail percentiles. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

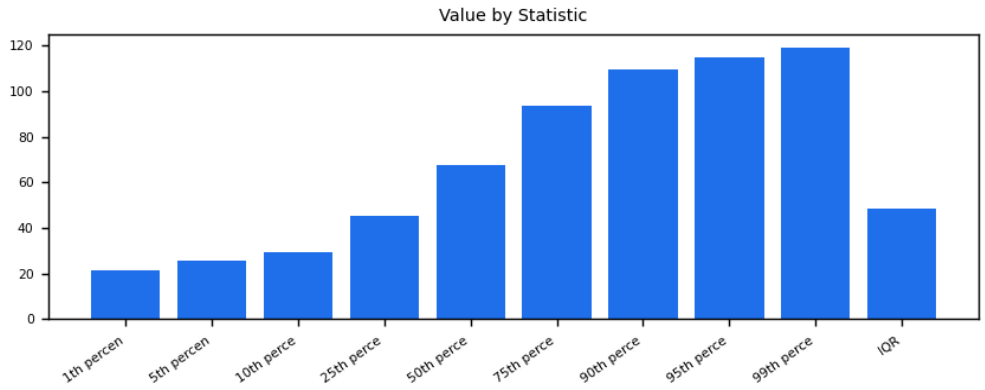
STEPS

1. Statistics → **Descriptive** → **Percentile summary**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Percentile summary — monthly_charges

Statistic	Value
1th percentile	21.33
5th percentile	25.43
10th percentile	29.32
25th percentile	45.35
50th percentile	67.69
75th percentile	93.76
90th percentile	109.4
95th percentile	114.9
99th percentile	119
IQR	48.41



DETAILED INTERPRETATION

Distribution shape at a glance (deciles + tails). The lead result here reads Statistic = 1th percentile; Value = 21.33. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.60 Proportion confidence interval

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Wilson & Clopper-Pearson CIs. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

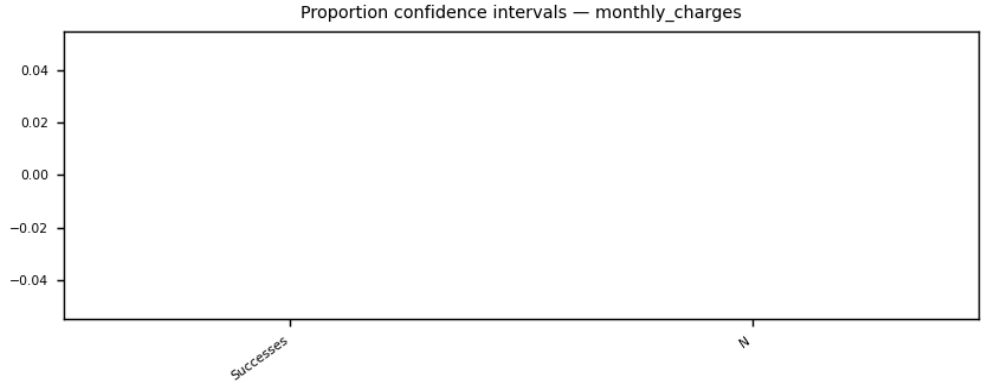
STEPS

1. Statistics → **Descriptive** → **Proportion confidence interval**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Proportion confidence intervals — monthly_charges

Successes	N	Proportion	Wilson 95% CI	Clopper-Pearson 95% CI
0	0		[nan, nan]	[nan, nan]



DETAILED INTERPRETATION

Wilson (score) and Clopper-Pearson (exact) CIs for a binomial proportion. The lead result here reads Successes = 0; N = 0; Proportion = ; Wilson 95% CI = [nan, nan]. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.61 Shapiro-Wilk normality

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Most powerful normality test. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

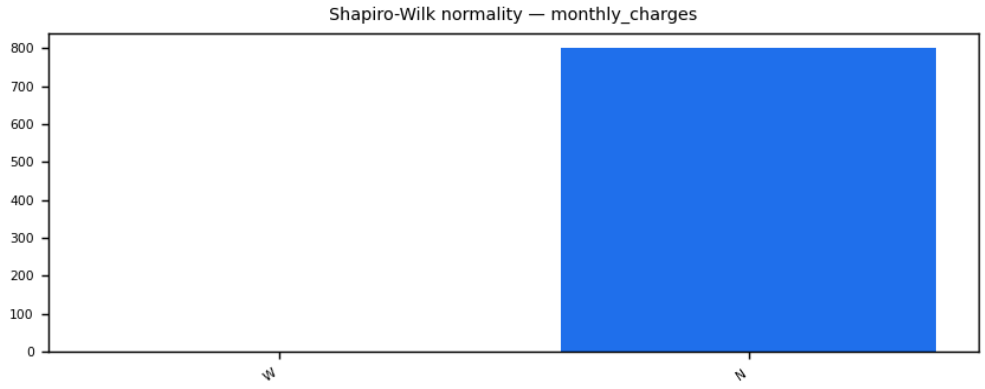
STEPS

1. Statistics → **Descriptive** → **Shapiro-Wilk normality**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Shapiro-Wilk normality — monthly_charges

W	p	N
0.9552	< .001	800



DETAILED INTERPRETATION

The most powerful normality test for small-to-moderate N; p < .05 rejects normality. The lead result here reads W = 0.9552; p = < .001; N = 800. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.62 Skewness & kurtosis tests

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Test shape departures from normal. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

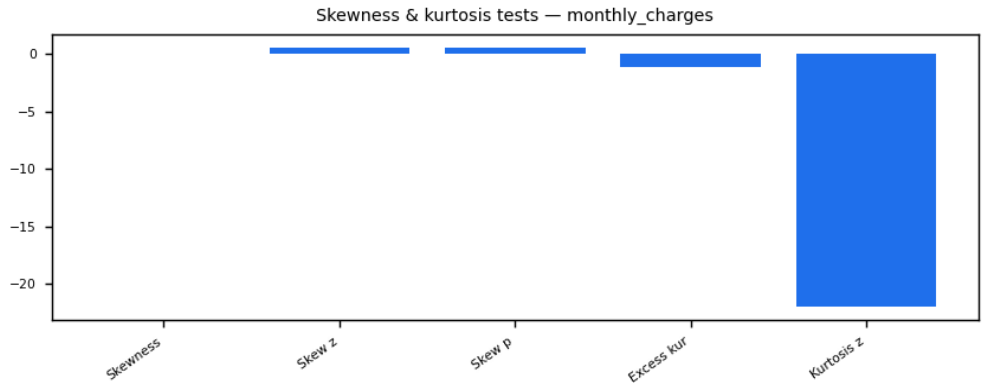
STEPS

1. Statistics → **Descriptive** → **Skewness & kurtosis tests**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Skewness & kurtosis tests — monthly_charges

Skewness	Skew z	Skew p	Excess kurtosis	Kurtosis z	Kurtosis p
0.0465	0.542	0.5878	-1.18	-22.03	< .001



DETAILED INTERPRETATION

Significant skew/kurtosis (p < .05) indicates departure from normality. The lead result here reads Skewness = 0.0465; Skew z = 0.542; Skew p = 0.5878; Excess kurtosis = -1.18. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.63 Stem-and-leaf plot

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Exploratory stem-and-leaf display of one numeric variable. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7

monthly_charges
35.51
71.45
29.16

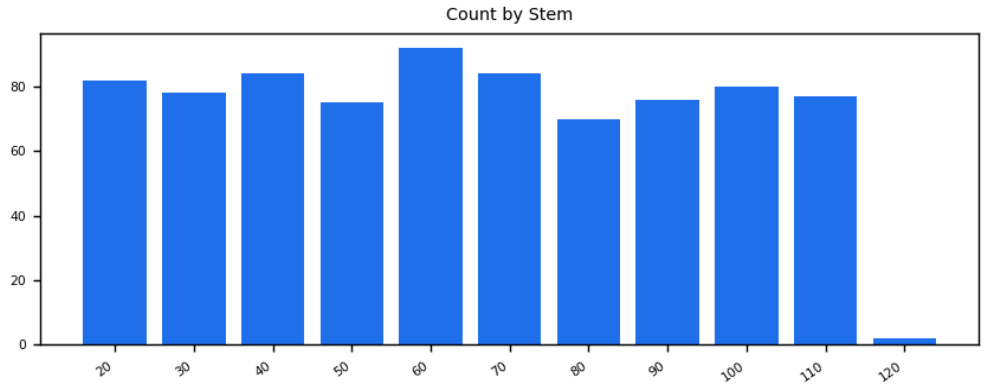
STEPS

1. Statistics → **Descriptive** → **Stem-and-leaf plot**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Stem-and-leaf plot — monthly_charges

Stem	Count	Leaves
20	82	00111111112222223333333333…
30	78	00000001111122222223333333…
40	84	00000000000000001111111122…
50	75	00000000011111111222222222…
60	92	00000000111111111112222222…
70	84	00000001111111111112222222…
80	70	00000001111111122222222333…
90	76	00000000111111111122222222…
100	80	00000000000111111111222222…
110	77	00000111122222222222333333…
120	2	00



DETAILED INTERPRETATION

Leaf unit = 1. Each leaf is one observation; read stem+leaf together (e.g. stem 30, leaf 4 ≈ 34). The lead result here reads Stem = 20; Count = 82; Leaves = 1.111e+79. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.64 Trimmed & winsorized means

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Robust central tendency. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45

monthly_charges
29.16

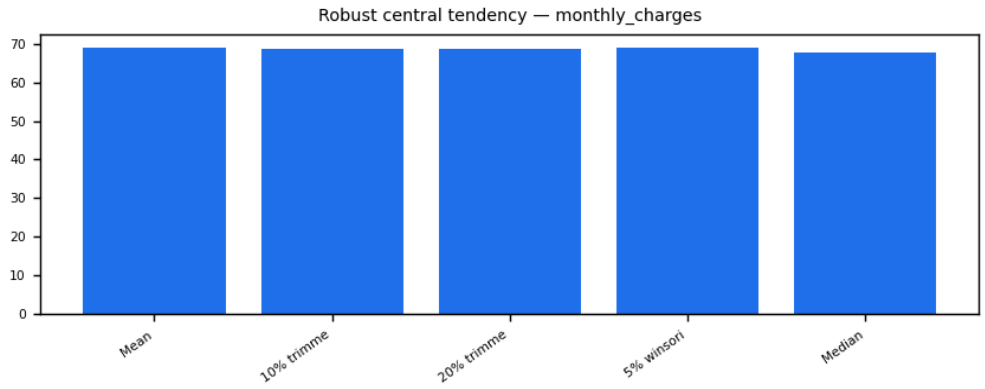
STEPS

1. Statistics → **Descriptive** → **Trimmed & winsorized means**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Robust central tendency — monthly_charges

Mean	10% trimmed	20% trimmed	5% winsorized mean	Median
69.09	68.89	68.75	69.11	67.69



DETAILED INTERPRETATION

Trimmed/winsorized means resist outliers by down-weighting the tails. The lead result here reads Mean = 69.09; 10% trimmed = 68.89; 20% trimmed = 68.75; 5% winsorized mean = 69.11. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.65 Z-score outlier screen

LIVE OUTPUT

Group: Descriptive · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Flag values by |z|. It belongs to the **Descriptive** family, used to summarise and profile data before deeper modelling. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to establish a clean, well-understood baseline before any modelling.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

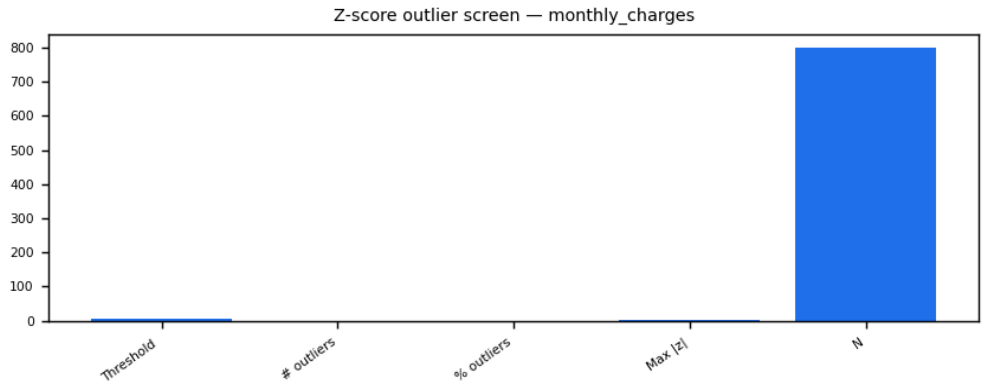
STEPS

1. Statistics → **Descriptive** → **Z-score outlier screen**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, threshold=6.0`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Z-score outlier screen — monthly_charges

Threshold |z|	# outliers	% outliers	Max |z|	N
6	0	0	1.768	800



DETAILED INTERPRETATION

Flags values more than the threshold standard deviations from the mean. The lead result here reads Threshold |z| = 6; # outliers = 0; % outliers = 0; Max |z| = 1.768. Read the tables above alongside the chart to judge direction, magnitude and significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst establish a clean, well-understood baseline before any modelling.

1.66 ANOVA effect sizes (η²/ω²)

LIVE OUTPUT

Group: Effect size · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Eta², omega², epsilon². It belongs to the **Effect size** family, used to quantify the standardised magnitude of a difference or association. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to report how large an effect is, not merely that it exists.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

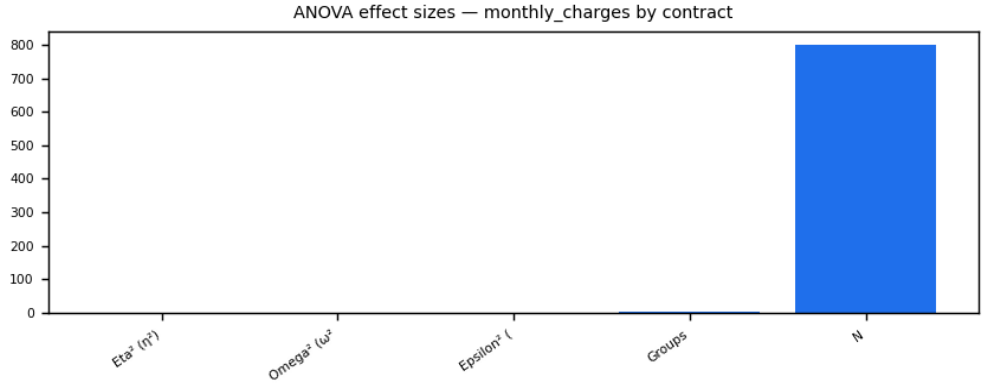
STEPS

1. Statistics → **Effect size** → **ANOVA effect sizes (η²/ω²)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

ANOVA effect sizes — monthly_charges by contract

Eta² (η²)	Omega² (ω²)	Epsilon² (ε²)	Groups	N
0.0011	-0.0014	-0.0014	3	800



DETAILED INTERPRETATION

η² 0.01 small · 0.06 medium · 0.14 large. ω²/ε² are less biased for small samples. The lead result here reads Eta² (η²) = 0.0011; Omega² (ω²) = -0.0014; Epsilon² (ε²) = -0.0014; Groups = 3. The effect size expresses the magnitude on a standardised, sample-size-independent scale, so it can be compared across studies and pooled in meta-analysis. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst report how large an effect is, not merely that it exists.

1.67 Association (Cramér's V)

LIVE OUTPUT

Group: Effect size · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Chi-square, Cramér's V, contingency C, phi. It belongs to the **Effect size** family, used to quantify the standardised magnitude of a difference or association. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on contract, monthly_charges. Analysts reach for it to report how large an effect is, not merely that it exists.

SAMPLE DATA INPUT

contract	monthly_charges
Month-to-month	52.94
One year	113.7
Month-to-month	35.51
Two year	71.45
Two year	29.16

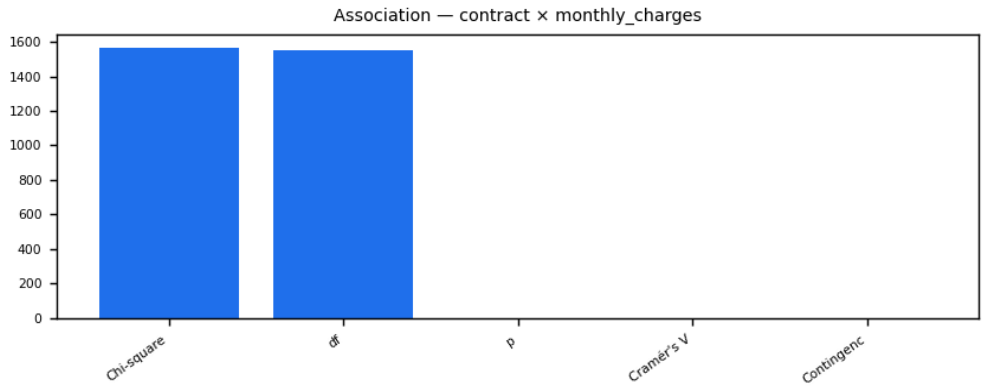
STEPS

1. Statistics → **Effect size** → **Association (Cramér's V)**. 2. Select the demo dataset above. 3. Set parameters — `row=contract, column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Association — contract × monthly_charges

Chi-square	df	p	Cramér's V	Contingency C	Phi (2×2)
1566	1550	0.3832	0.9893	0.8136	



DETAILED INTERPRETATION

Cramér's V ranges 0-1: ~0.1 small · 0.3 medium · 0.5 large association. The lead result here reads Chi-square = 1566; df = 1550; p = 0.3832; Cramér's V = 0.9893. The effect size expresses the magnitude on a standardised, sample-size-independent scale, so it can be compared across studies and pooled in meta-analysis. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst report how large an effect is, not merely that it exists.

1.68 Cliff's delta

LIVE OUTPUT

Group: Effect size · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Nonparametric effect size (2 groups). It belongs to the **Effect size** family, used to quantify the standardised magnitude of a difference or association. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to report how large an effect is, not merely that it exists.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

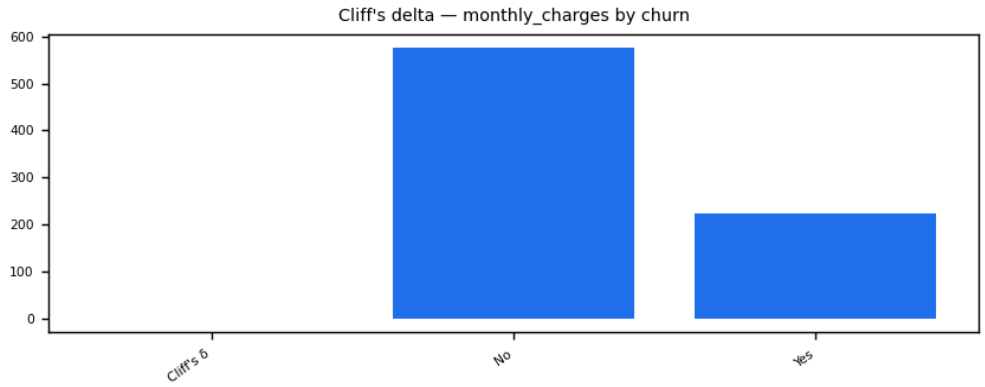
STEPS

1. Statistics → **Effect size** → **Cliff's delta**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Cliff's delta — monthly_charges by churn

Cliff's δ	Magnitude	No	Yes
-0.1885	small	576	224



DETAILED INTERPRETATION

Nonparametric effect size: |δ| 0.147 small · 0.33 medium · 0.474 large. The lead result here reads Cliff's δ = -0.1885; Magnitude = small; No = 576; Yes = 224. The effect size expresses the magnitude on a standardised, sample-size-independent scale, so it can be compared across studies and pooled in meta-analysis. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst report how large an effect is, not merely that it exists.

1.69 Cohen's f² (regression)

LIVE OUTPUT

Group: Effect size · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Effect size for a regression model. It belongs to the **Effect size** family, used to quantify the standardised magnitude of a difference or association. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to report how large an effect is, not merely that it exists.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1

monetary	tenure_months	monthly_charges	support_calls
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

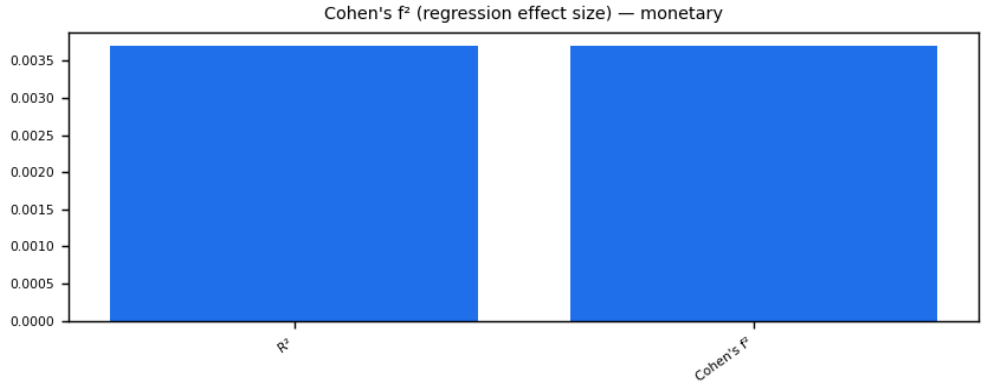
STEPS

1. Statistics → **Effect size** → **Cohen's f² (regression)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Cohen's f² (regression effect size) — monetary

R²	Cohen's f²	Magnitude
0.0037	0.0037	negligible



DETAILED INTERPRETATION

f²: 0.02 small · 0.15 medium · 0.35 large. Effect size for a regression model. The lead result here reads R² = 0.0037; Cohen's f² = 0.0037; Magnitude = negligible. The effect size expresses the magnitude on a standardised, sample-size-independent scale, so it can be compared across studies and pooled in meta-analysis. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst report how large an effect is, not merely that it exists.

1.70 Correlation ratio (η)

LIVE OUTPUT

Group: Effect size · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Categorical → numeric association. It belongs to the **Effect size** family, used to quantify the standardised magnitude of a difference or association. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to report how large an effect is, not merely that it exists.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

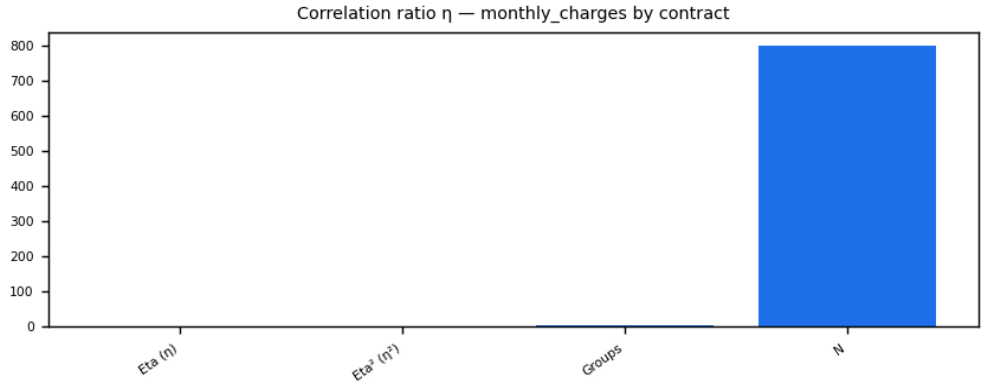
STEPS

1. Statistics → **Effect size** → **Correlation ratio (η)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Correlation ratio η — monthly_charges by contract

Eta (η)	Eta² (η²)	Groups	N
0.0327	0.0011	3	800



DETAILED INTERPRETATION

Strength of association between a categorical predictor and a numeric outcome (0-1). The lead result here reads Eta (η) = 0.0327; Eta² (η²) = 0.0011; Groups = 3; N = 800. The effect size expresses the magnitude on a standardised, sample-size-independent scale, so it can be compared across studies and pooled in meta-analysis. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst report how large an effect is, not merely that it exists.

1.71 Effect size (Cohen's d)

LIVE OUTPUT

Group: Effect size · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

d, Hedges' g, Glass' Δ and CLES for two groups. It belongs to the **Effect size** family, used to quantify the standardised magnitude of a difference or association. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to report how large an effect is, not merely that it exists.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

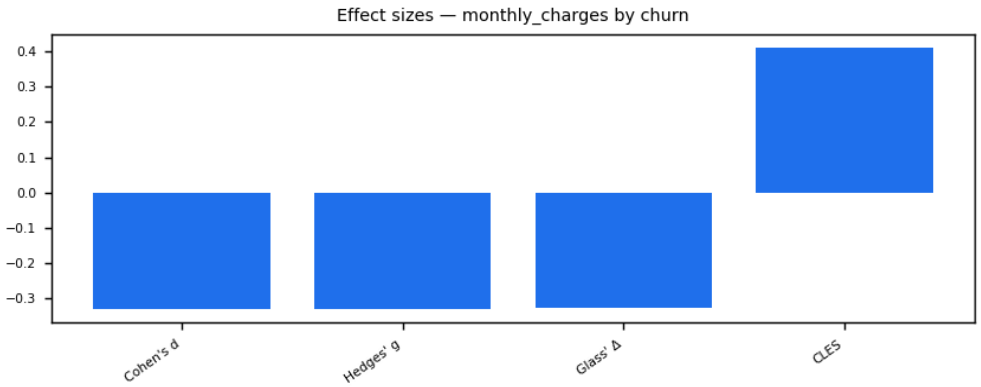
STEPS

1. Statistics → **Effect size** → **Effect size (Cohen's d)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Effect sizes — monthly_charges by churn

Cohen's d	Hedges' g	Glass' Δ	CLES	Magnitude
-0.3314	-0.3311	-0.3271	0.4118	small



DETAILED INTERPRETATION

Groups: No vs Yes. |d|: 0.2 small · 0.5 medium · 0.8 large. CLES = probability a random case from group 1 exceeds one from group 2. The lead result here reads Cohen's d = -0.3314; Hedges' g = -0.3311; Glass' Δ = -0.3271; CLES = 0.4118. The effect size expresses the magnitude on a standardised, sample-size-independent scale, so it can be compared across studies and pooled in meta-analysis. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst report how large an effect is, not merely that it exists.

1.72 Paired effect size (dz)

LIVE OUTPUT

Group: Effect size · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Standardized paired difference. It belongs to the **Effect size** family, used to quantify the standardised magnitude of a difference or association. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to report how large an effect is, not merely that it exists.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

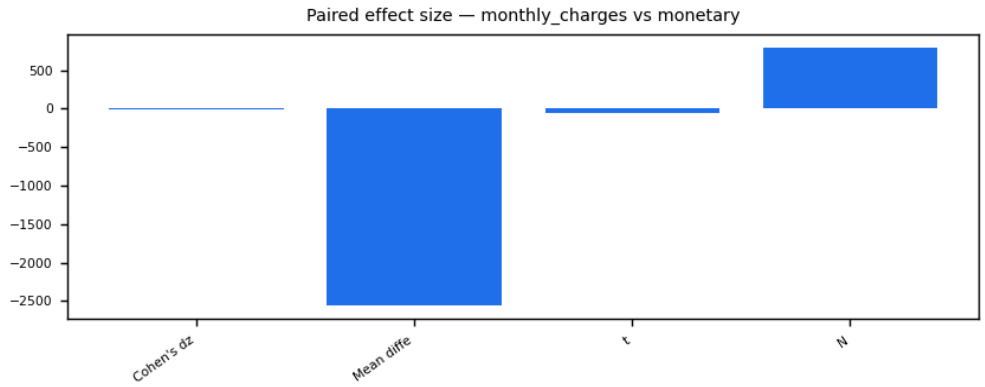
STEPS

1. Statistics → **Effect size** → **Paired effect size (dz)**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Paired effect size — monthly_charges vs monetary

Cohen's dz	Mean difference	t	p	N
-1.837	-2565	-51.96	< .001	800



DETAILED INTERPRETATION

dz is the standardized mean change for paired/repeated measures. The lead result here reads Cohen's dz = -1.837; Mean difference = -2565; t = -51.96; p = < .001. The effect size expresses the magnitude on a standardised, sample-size-independent scale, so it can be compared across studies and pooled in meta-analysis. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst report how large an effect is, not merely that it exists.

1.73 Point-biserial correlation

LIVE OUTPUT

Group: Effect size · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Correlation between continuous and binary variable. It belongs to the **Effect size** family, used to quantify the standardised magnitude of a difference or association. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, tenure_months. Analysts reach for it to report how large an effect is, not merely that it exists.

SAMPLE DATA INPUT

monthly_charges	tenure_months
52.94	68
113.7	45
35.51	49
71.45	64
29.16	42

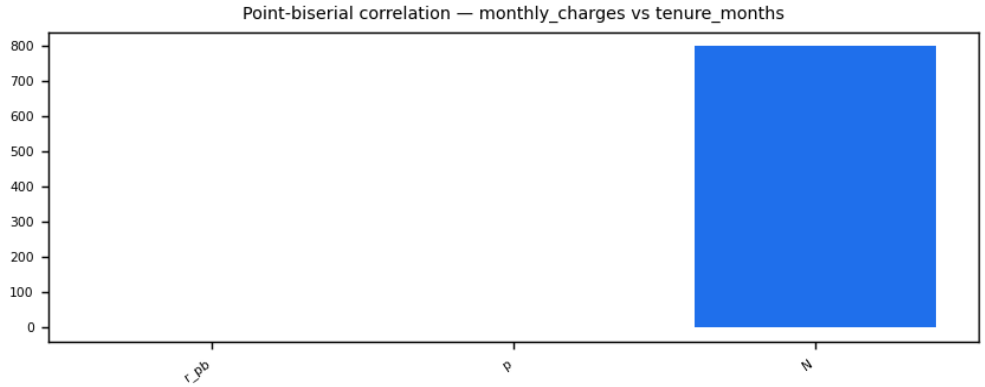
STEPS

1. Statistics → **Effect size** → **Point-biserial correlation**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, binary=tenure_months`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Point-biserial correlation — monthly_charges vs tenure_months

r_pb	p	N
-0.034	0.3371	800



DETAILED INTERPRETATION

Correlation between a continuous variable and a binary (0/1) variable. The lead result here reads r_pb = -0.034; p = 0.3371; N = 800. The effect size expresses the magnitude on a standardised, sample-size-independent scale, so it can be compared across studies and pooled in meta-analysis. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst report how large an effect is, not merely that it exists.

1.74 Rank-biserial (Mann-Whitney)

LIVE OUTPUT

Group: Effect size · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Nonparametric effect size. It belongs to the **Effect size** family, used to quantify the standardised magnitude of a difference or association. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to report how large an effect is, not merely that it exists.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

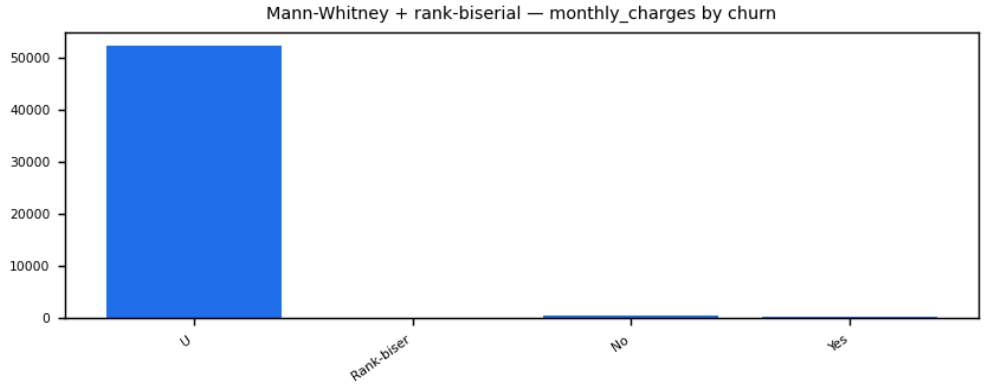
STEPS

1. Statistics → **Effect size** → **Rank-biserial (Mann-Whitney)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Mann-Whitney + rank-biserial — monthly_charges by churn

U	p	Rank-biserial r	No	Yes
5.235e+04	< .001	0.1885	576	224



DETAILED INTERPRETATION

Rank-biserial r is the nonparametric effect size for a Mann-Whitney test. The lead result here reads U = 5.235e+04; p = < .001; Rank-biserial r = 0.1885; No = 576. The effect size expresses the magnitude on a standardised, sample-size-independent scale, so it can be compared across studies and pooled in meta-analysis. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst report how large an effect is, not merely that it exists.

1.75 Ansari-Bradley scale test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Rank test for equal spread. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

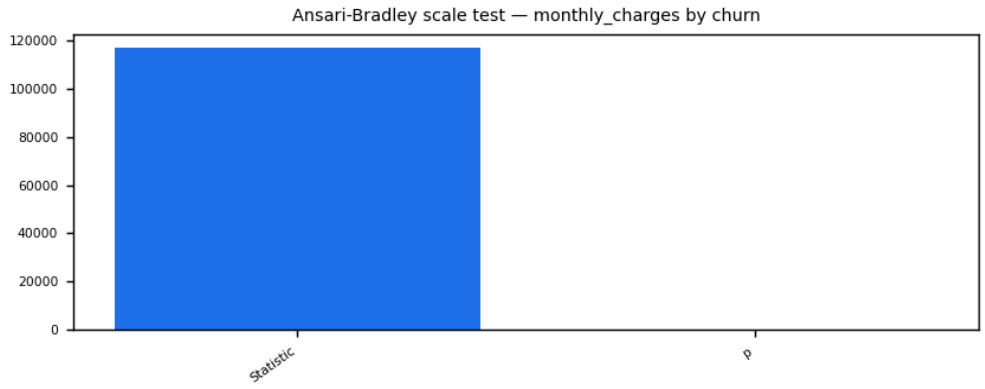
STEPS

1. Statistics → **Nonparametric** → **Ansari-Bradley scale test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Ansari-Bradley scale test — monthly_charges by churn

Statistic	p
1.169e+05	0.3419



DETAILED INTERPRETATION

Rank test for equality of scale/spread between two groups. The lead result here reads Statistic = 1.169e+05; p = 0.3419. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.76 Barnard's exact test

INPUT REQUIREMENT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Unconditional exact 2×2 test. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on contract, monthly_charges. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

contract	monthly_charges
Month-to-month	52.94

contract	monthly_charges
One year	113.7
Month-to-month	35.51
Two year	71.45
Two year	29.16

STEPS

1. Statistics → **Nonparametric** → **Barnard's exact test**. 2. Select the demo dataset above. 3. Set parameters — `row=contract, column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

'contract' × 'monthly_charges' must both be binary (2×2 table)

DETAILED INTERPRETATION

This technique has a specific data requirement: *'contract' × 'monthly_charges' must both be binary (2×2 table)*. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a nonparametric method and supports the decision to reach a robust conclusion even with skewed or ordinal data.

1.77 Brunner-Munzel test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Robust 2-sample rank test. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

STEPS

1. Statistics → **Nonparametric** → **Brunner-Munzel test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Brunner-Munzel test — monthly_charges by churn

W	p
4.185	< .001

DETAILED INTERPRETATION

Nonparametric alternative to Mann-Whitney that does not assume equal variances/shapes. The lead result here reads W = 4.185; p = < .001. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.78 Cochran's Q

INPUT REQUIREMENT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Repeated-measures test for 3+ related binary variables. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Nonparametric** → **Cochran's Q**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

'tenure_months' must be binary 0/1

DETAILED INTERPRETATION

This technique has a specific data requirement: *'tenure_months' must be binary 0/1*. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a nonparametric method and supports the decision to reach a robust conclusion even with skewed or ordinal data.

1.79 Dunn's post-hoc

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Post-hoc after Kruskal-Wallis. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

STEPS

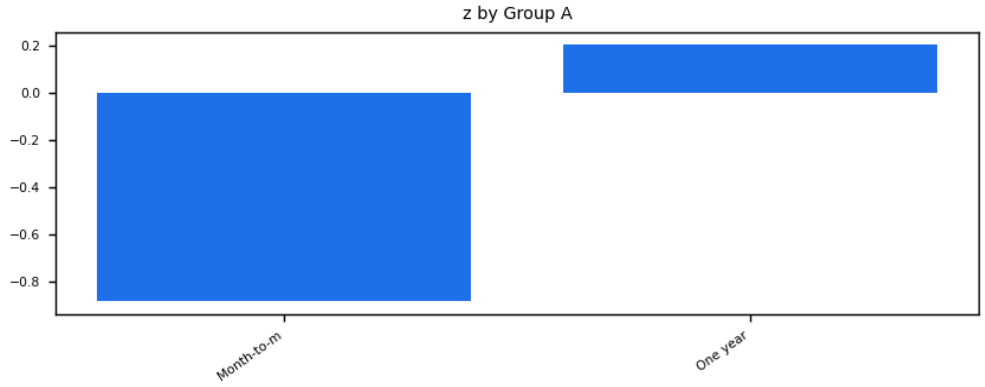
1. Statistics → **Nonparametric** → **Dunn's post-hoc**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Dunn's post-hoc (rank sums) — monthly_charges by contract

Group A	Group B	z	p	Bonferroni p
Month-to-month	One year	-0.8808	0.3784	1
Month-to-month	Two year	-0.5972	0.5504	1

Group A	Group B	z	p	Bonferroni p
One year	Two year	0.2027	0.8394	1



DETAILED INTERPRETATION

Pairwise post-hoc after a significant Kruskal-Wallis test. The lead result here reads Group A = Month-to-month; Group B = One year; z = -0.8808; p = 0.3784. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.80 Epps-Singleton test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Equal-distribution test. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

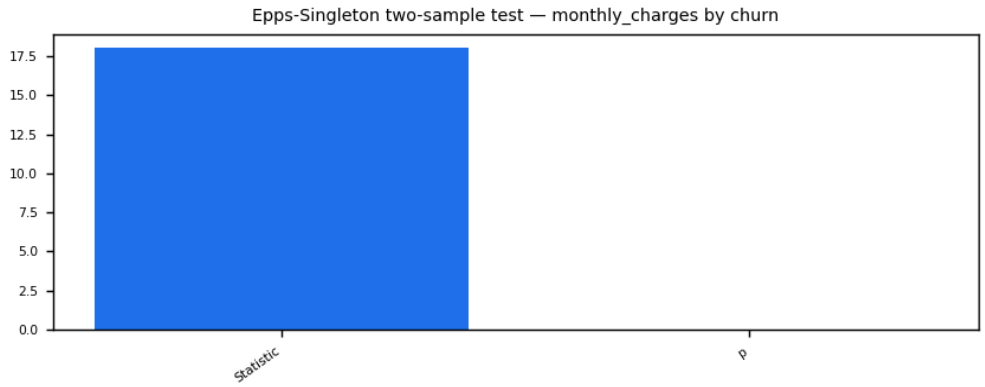
STEPS

1. Statistics → **Nonparametric** → **Epps-Singleton test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Epps-Singleton two-sample test — monthly_charges by churn

Statistic	p
18.02	0.0012



DETAILED INTERPRETATION

Characteristic-function test for equal distributions (works on discrete data too). The lead result here reads Statistic = 18.02; p = 0.0012. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.81 Exact Mann-Whitney U

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Exact-conditional rank-sum test for two small groups. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

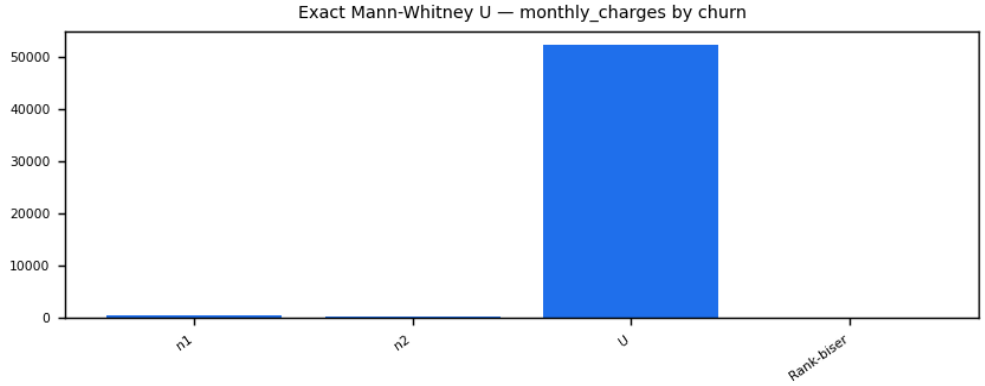
STEPS

1. Statistics → **Nonparametric** → **Exact Mann-Whitney U**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Exact Mann-Whitney U — monthly_charges by churn

Group 1	n1	Group 2	n2	U	p (asymptotic)	Rank-biserial r
No	576	Yes	224	5.235e+04	< .001	0.1885



DETAILED INTERPRETATION

Exact conditional p-value when samples are small and untied; otherwise the normal approximation is used. The lead result here reads Group 1 = No; n1 = 576; Group 2 = Yes; n2 = 224. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.82 Exact Wilcoxon signed-rank

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Exact paired rank test for small samples. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

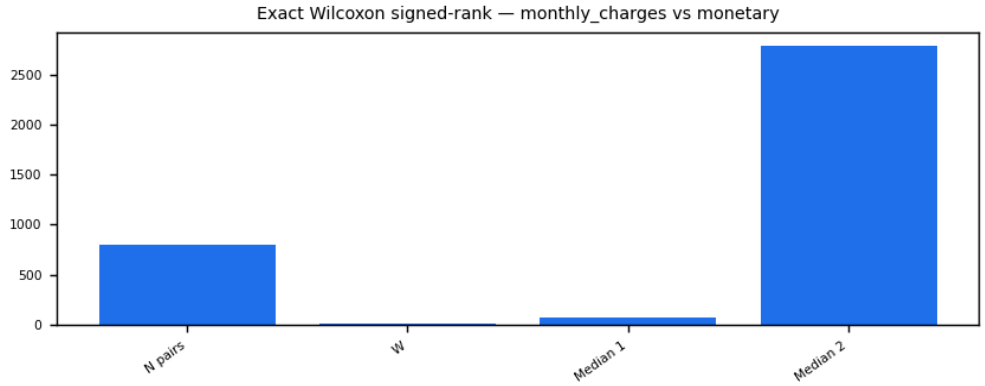
STEPS

1. Statistics → **Nonparametric** → **Exact Wilcoxon signed-rank**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Exact Wilcoxon signed-rank — monthly_charges vs monetary

N pairs	W	p (approx)	Median 1	Median 2
800	8	< .001	67.69	2791



DETAILED INTERPRETATION

Exact p-value for small, untied paired samples. The lead result here reads N pairs = 800; W = 8; p (approx) = < .001; Median 1 = 67.69. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.83 Exact contingency test (r×c)

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Fisher (2×2) or Monte-Carlo exact p-value for small/sparse tables. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on contract, monthly_charges. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

contract	monthly_charges
Month-to-month	52.94
One year	113.7
Month-to-month	35.51
Two year	71.45
Two year	29.16

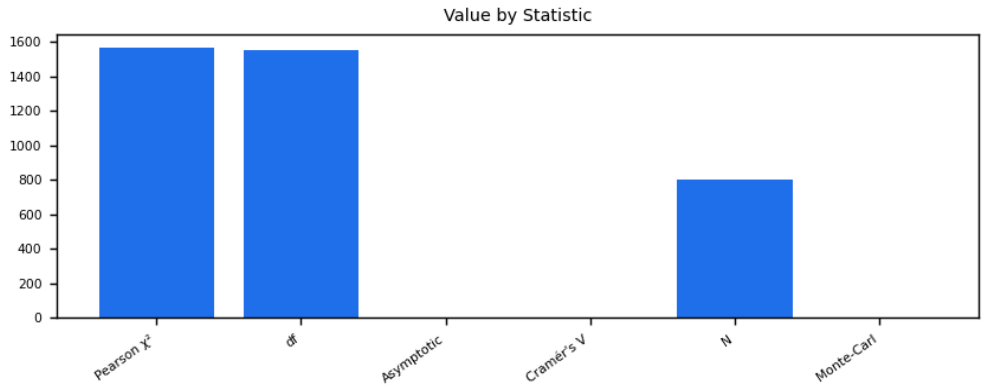
STEPS

1. Statistics → **Nonparametric** → **Exact contingency test (r×c)**. 2. Select the demo dataset above. 3. Set parameters — `row=contract, column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Exact contingency test — contract × monthly_charges

Statistic	Value
Pearson χ^2	1566
df	1550
Asymptotic p	0.3832
Cramér's V	0.9893
N	800
Monte-Carlo exact p (3000 …	0.042



DETAILED INTERPRETATION

Monte-Carlo exact test with fixed margins (3000 permutations). Use when expected cell counts are small. The lead result here reads Statistic = Pearson χ^2; Value = 1566. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.84 Fisher's exact test

LIVE OUTPUT

Group: Nonparametric · Dataset: a real-world patient cohort — treatment, recovery and confounders

WHAT THIS DOES

Exact 2×2 association. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a real-world patient cohort — treatment, recovery and confounders, drawing on treated, recovered. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

treated	recovered
1	0
0	0
1	1
0	0
0	0

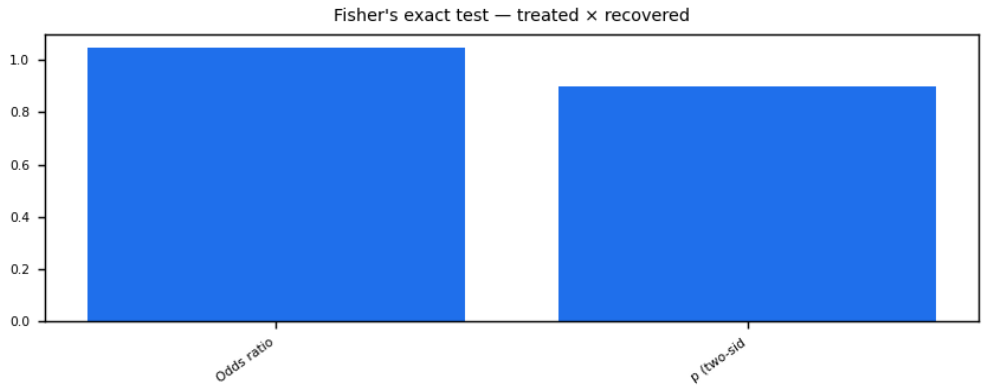
STEPS

1. Statistics → **Nonparametric** → **Fisher's exact test**. 2. Select the demo dataset above. 3. Set parameters — `row=treated, column=recovered`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Fisher's exact test — treated × recovered

Odds ratio	p (two-sided)
1.048	0.9021



DETAILED INTERPRETATION

Exact test of association in a 2×2 table (valid for small samples). The lead result here reads Odds ratio = 1.048; p (two-sided) = 0.9021. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a real-world patient cohort — treatment, recovery and confounders, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.85 Fligner-Killeen test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Robust equal-variance test. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month

monthly_charges	contract
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

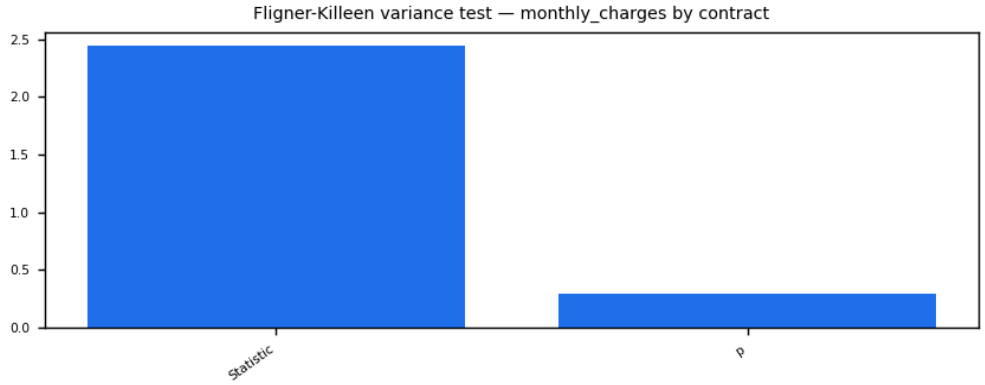
STEPS

1. Statistics → **Nonparametric** → **Fligner-Killeen test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Fligner-Killeen variance test — monthly_charges by contract

Statistic	p
2.44	0.2952



DETAILED INTERPRETATION

Robust nonparametric test of equal variances. The lead result here reads Statistic = 2.44; p = 0.2952. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.86 Friedman test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Nonparametric repeated-measures comparison of 3+ variables. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

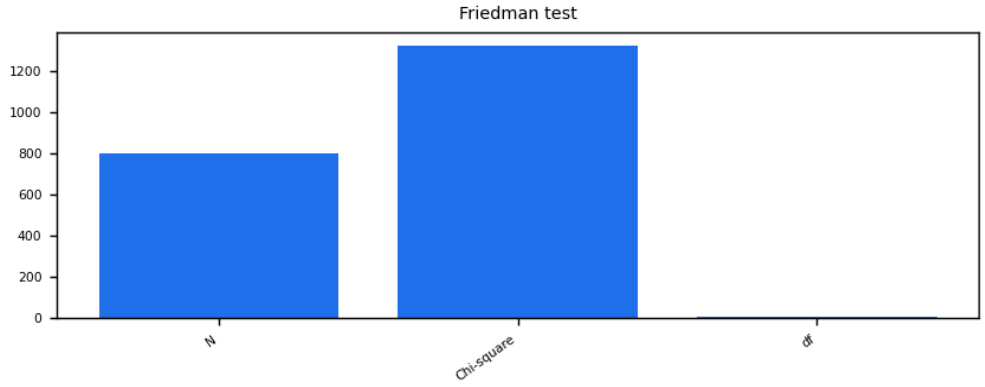
STEPS

1. Statistics → **Nonparametric** → **Friedman test**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Friedman test

N	Chi-square	df	p
800	1323	2	< .001



DETAILED INTERPRETATION

Medians: tenure_months=37.500, monthly_charges=67.690, support_calls=2.000. The lead result here reads N = 800; Chi-square = 1323; df = 2; p = < .001. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.87 Jonckheere-Terpstra trend

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Ordered-alternatives trend test. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

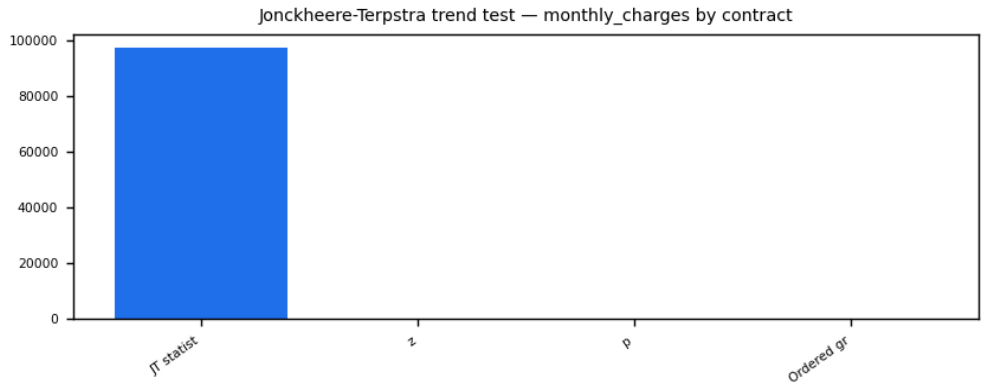
STEPS

1. Statistics → **Nonparametric** → **Jonckheere-Terpstra trend**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Jonckheere-Terpstra trend test — monthly_charges by contract

JT statistic	z	p	Ordered groups
9.722e+04	0.8386	0.4017	3



DETAILED INTERPRETATION

Tests for a monotonic (ordered) trend across the groups. The lead result here reads JT statistic = 9.722e+04; z = 0.8386; p = 0.4017; Ordered groups = 3. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.88 Kruskal-Wallis + effect size

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Rank ANOVA with ε²/η². It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

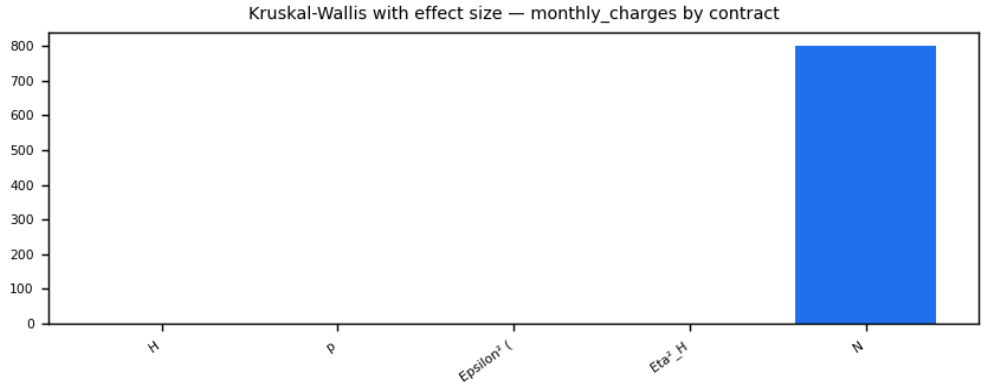
STEPS

1. Statistics → **Nonparametric** → **Kruskal-Wallis + effect size**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Kruskal-Wallis with effect size — monthly_charges by contract

H	p	Epsilon² (ε²)	Eta²_H	N
0.9079	0.6351	0.0011	-0.0014	800



DETAILED INTERPRETATION

Nonparametric ANOVA with rank-based effect sizes (ε²: 0.01 small · 0.08 medium · 0.26 large). The lead result here reads H = 0.9079; p = 0.6351; Epsilon² (ε²) = 0.0011; Eta²_H = -0.0014. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.89 Kruskal-Wallis H

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Nonparametric alternative to one-way ANOVA. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

STEPS

1. Statistics → **Nonparametric** → **Kruskal-Wallis H**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

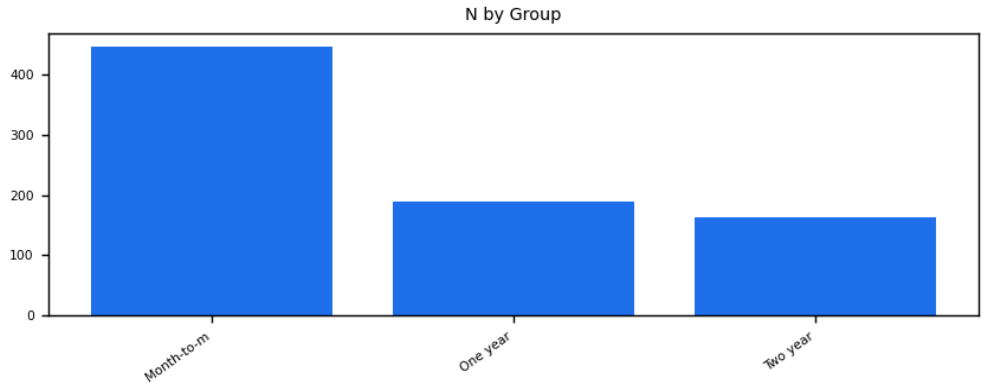
OUTPUT

Group medians

Group	N	Median
Month-to-month	447	66.54
One year	190	70.71
Two year	163	69.09

Kruskal-Wallis H test

H	df	p	epsilon²
0.9079	2	0.6351	-0.0014



DETAILED INTERPRETATION

The lead result here reads Group = Month-to-month; N = 447; Median = 66.54. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.90 Loglinear analysis

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Poisson model of a multiway contingency table; tests independence. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on contract, region. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

contract	region
Month-to-month	East
One year	North
Month-to-month	North
Two year	West
Two year	West

STEPS

1. Statistics → **Nonparametric** → **Loglinear analysis**. 2. Select the demo dataset above. 3. Set parameters — `factors=contract, region`. 4. Click **Run** and read the tables/chart below.

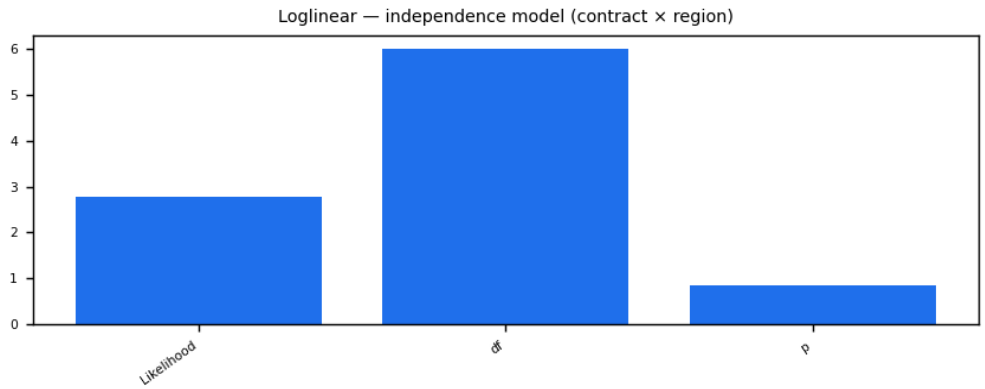
OUTPUT

Loglinear — independence model (contract × region)

Likelihood-ratio G²	df	p
2.764	6	0.8378

Cell departures from independence (Pearson residuals)

Cell	Observed	Expected	Std. residual
Two year × East	40	35	0.84
Two year × South	40	45.4	-0.81
Month-to-month × East	90	96.1	-0.62
One year × West	44	47.7	-0.54
Two year × West	44	41	0.48
One year × South	56	53	0.42
Two year × North	39	41.6	-0.4
Month-to-month × North	117	114	0.28
Month-to-month × South	127	124.6	0.21
One year × East	42	40.9	0.18
Month-to-month × West	113	112.3	0.07
One year × North	48	48.4	-0.06



DETAILED INTERPRETATION

Tests independence of the factors. p < .05 → the factors are associated (a higher-order term is needed). The lead result here reads Likelihood-ratio G² = 2.764; df = 6; p = 0.8378. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by

how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.91 Mann-Whitney U

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Nonparametric alternative to the independent t-test. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

STEPS

1. Statistics → **Nonparametric** → **Mann-Whitney U**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

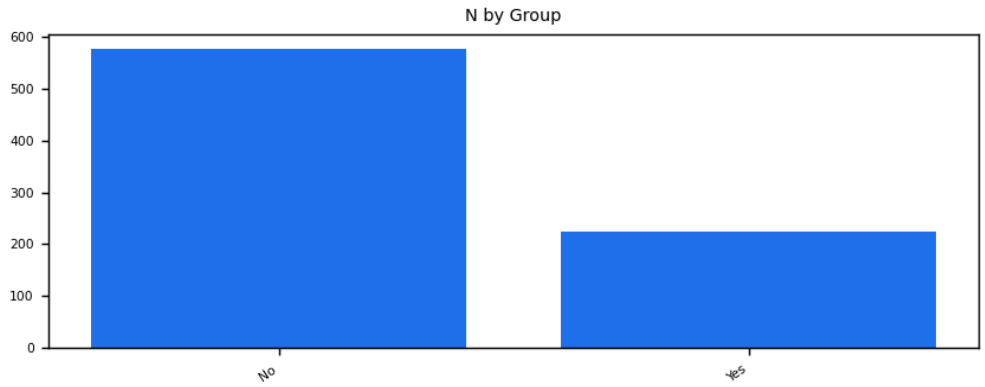
OUTPUT

Ranks

Group	N	Median	Mean rank
No	576	65.36	379.4
Yes	224	78.3	454.8

Mann-Whitney U test

U	p (2-tailed)	Rank-biserial r
5.235e+04	< .001	0.1885



DETAILED INTERPRETATION

The lead result here reads Group = No; N = 576; Median = 65.36; Mean rank = 379.4. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.92 McNemar's test

INPUT REQUIREMENT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Paired binary test (before/after on the same subjects). It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

STEPS

1. Statistics → **Nonparametric** → **McNemar's test**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

'monthly_charges' must be binary 0/1 for McNemar's test

DETAILED INTERPRETATION

This technique has a specific data requirement: *'monthly_charges' must be binary 0/1 for McNemar's test*. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a nonparametric method and supports the decision to reach a robust conclusion even with skewed or ordinal data.

1.93 Mood's median test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Nonparametric equality-of-medians across groups. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

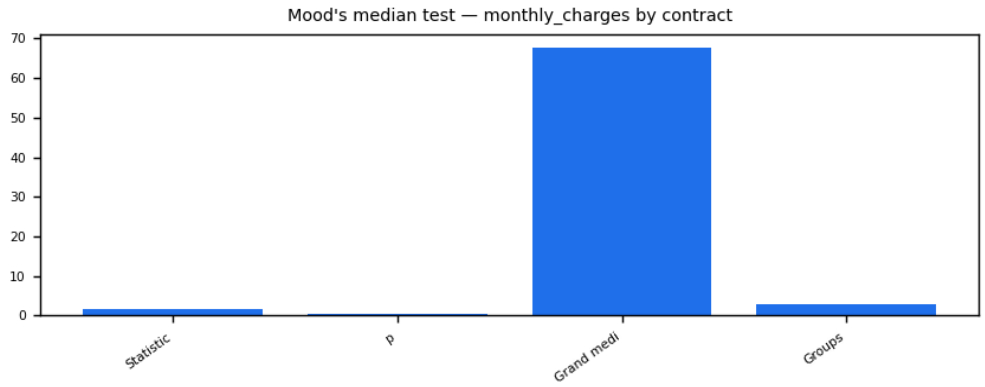
STEPS

1. Statistics → **Nonparametric** → **Mood's median test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Mood's median test — monthly_charges by contract

Statistic	p	Grand median	Groups
1.733	0.4204	67.69	3



DETAILED INTERPRETATION

Nonparametric test that groups share a common median. The lead result here reads Statistic = 1.733; p = 0.4204; Grand median = 67.69; Groups = 3. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.94 Mood's scale test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Equal-dispersion test (2 groups). It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

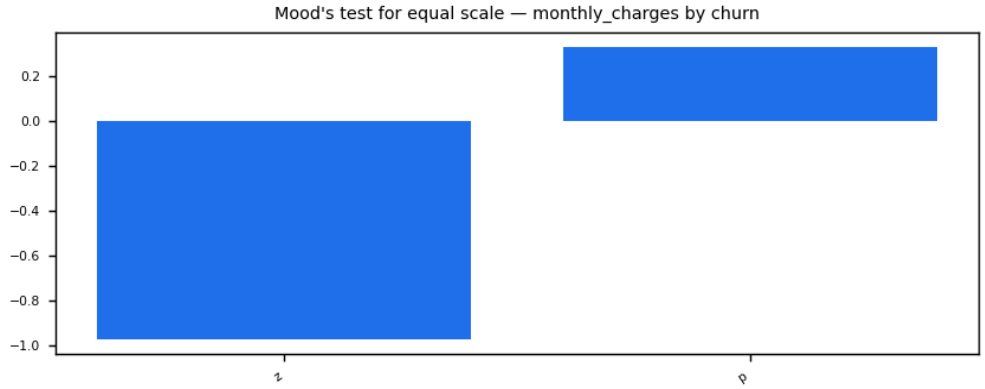
STEPS

1. Statistics → **Nonparametric** → **Mood's scale test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Mood's test for equal scale — monthly_charges by churn

z	p
-0.9718	0.3311



DETAILED INTERPRETATION

Nonparametric test that two samples have the same dispersion/scale. The lead result here reads z = -0.9718; p = 0.3311. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.95 One-sample Wilcoxon

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Signed-rank test of the median. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

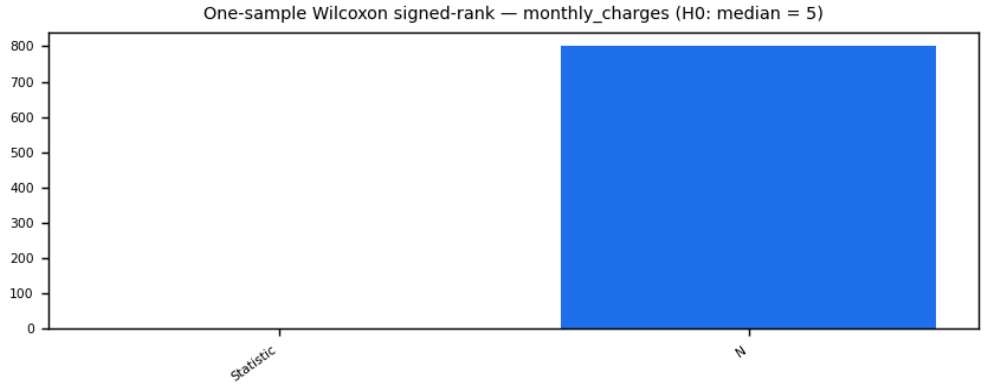
STEPS

1. Statistics → **Nonparametric** → **One-sample Wilcoxon**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, median=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

One-sample Wilcoxon signed-rank — monthly_charges (H0: median = 5)

Statistic	p	N
0	< .001	800



DETAILED INTERPRETATION

Nonparametric one-sample test of the median. The lead result here reads Statistic = 0; p = < .001; N = 800. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.96 Quade test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Repeated-measures (Friedman alternative). It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

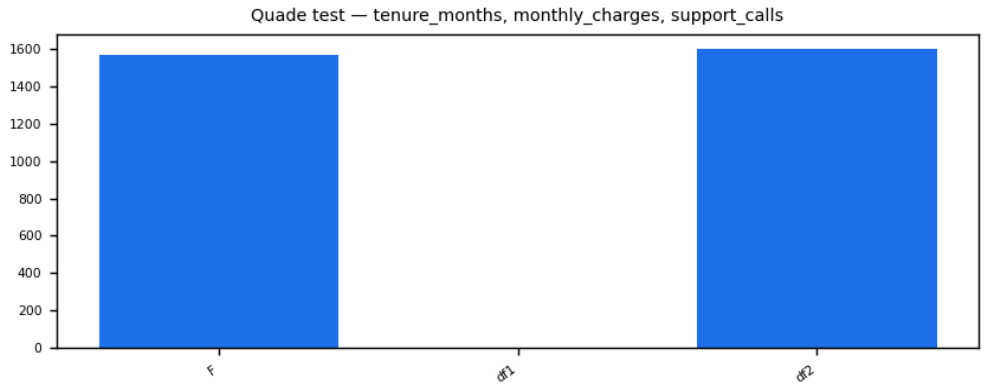
STEPS

1. Statistics → **Nonparametric** → **Quade test**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Quade test — tenure_months, monthly_charges, support_calls

F	df1	df2	p
1570	2	1598	< .001



DETAILED INTERPRETATION

Nonparametric repeated-measures test (a more powerful alternative to Friedman). The lead result here reads F = 1570; df1 = 2; df2 = 1598; p = < .001. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.97 Runs test (randomness)

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Wald-Wolfowitz test for randomness of a sequence. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges
52.94

monthly_charges
113.7
35.51
71.45
29.16

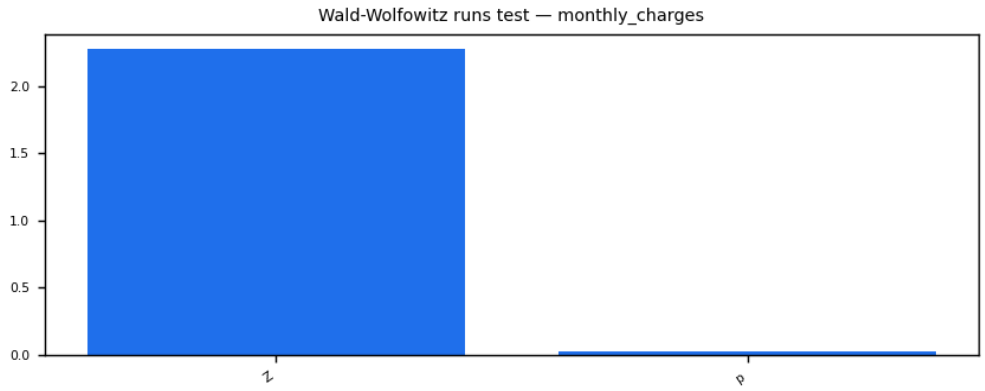
STEPS

1. Statistics → **Nonparametric** → **Runs test (randomness)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Wald-Wolfowitz runs test — monthly_charges

Z	p
2.273	0.023



DETAILED INTERPRETATION

Tests randomness of a sequence (p < .05 → non-random ordering). The lead result here reads Z = 2.273; p = 0.023. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.98 Sign test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Distribution-free test of the median. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

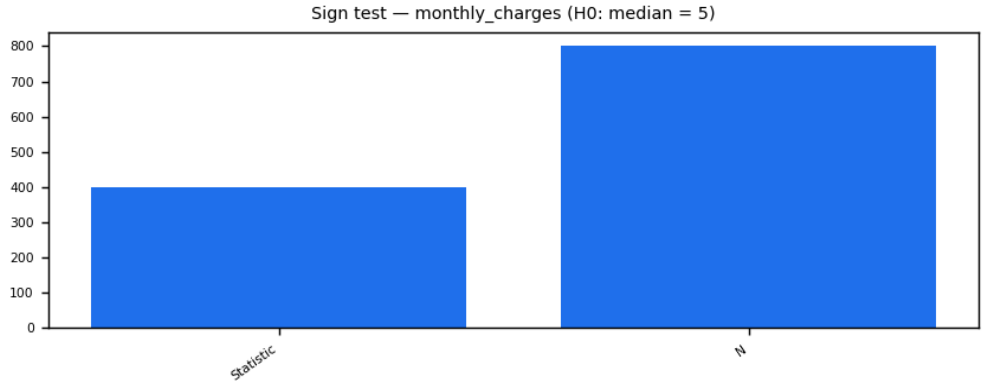
STEPS

1. Statistics → **Nonparametric** → **Sign test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, median=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Sign test — monthly_charges (H0: median = 5)

Statistic	p	N
400	< .001	800



DETAILED INTERPRETATION

Distribution-free test that the median equals the hypothesized value. The lead result here reads Statistic = 400; p = < .001; N = 800. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.99 Two-sample KS test

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Compare two distributions. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

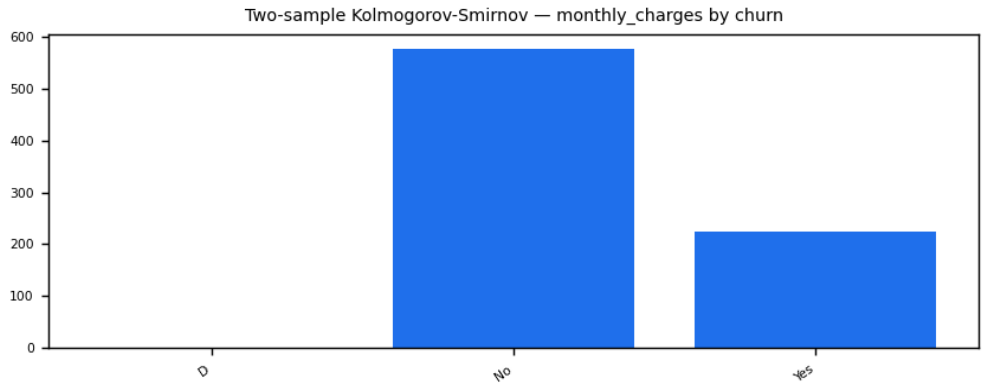
STEPS

1. Statistics → **Nonparametric** → **Two-sample KS test**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Two-sample Kolmogorov-Smirnov — monthly_charges by churn

D	p	No	Yes
0.156	< .001	576	224



DETAILED INTERPRETATION

Tests whether two samples come from the same distribution (any difference). The lead result here reads D = 0.156; p = < .001; No = 576; Yes = 224. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.100 Wilcoxon signed-rank

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Nonparametric alternative to the paired t-test. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

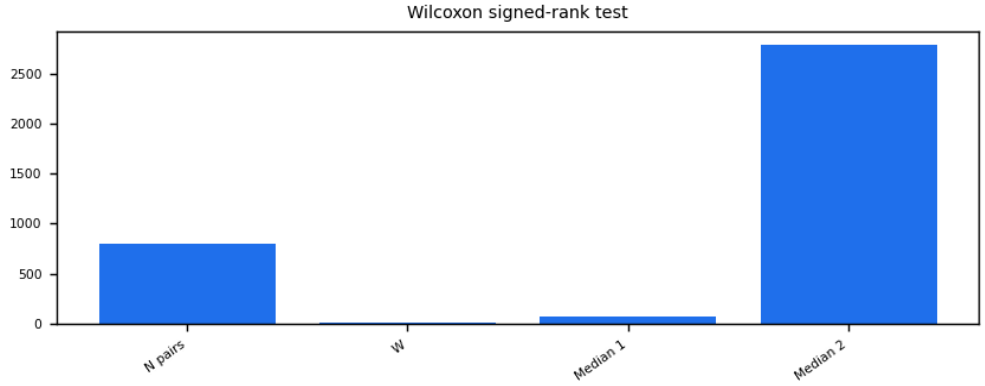
STEPS

1. Statistics → **Nonparametric** → **Wilcoxon signed-rank**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Wilcoxon signed-rank test

N pairs	W	p (2-tailed)	Median 1	Median 2
800	8	< .001	67.69	2791



DETAILED INTERPRETATION

The lead result here reads N pairs = 800; W = 8; p (2-tailed) = < .001; Median 1 = 67.69. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

1.101 k-sample Anderson-Darling

LIVE OUTPUT

Group: Nonparametric · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Same-distribution test across groups. It belongs to the **Nonparametric** family, used to compare groups or distributions without assuming normality. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to reach a robust conclusion even with skewed or ordinal data.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

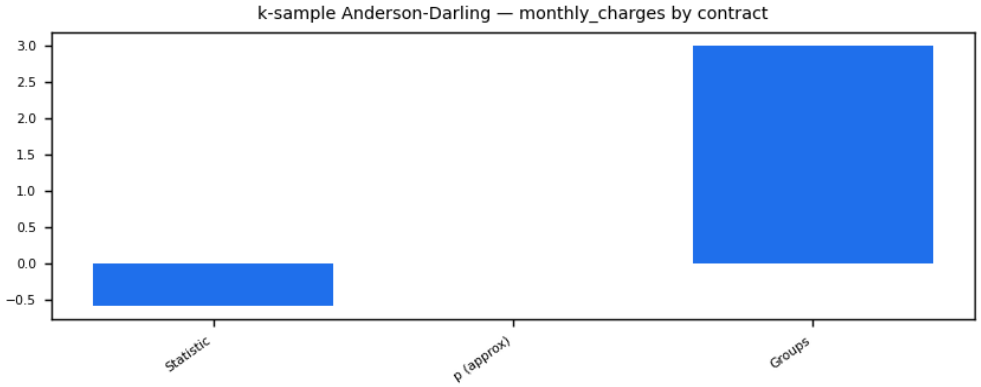
STEPS

1. Statistics → **Nonparametric** → **k-sample Anderson-Darling**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

k-sample Anderson-Darling — monthly_charges by contract

Statistic	p (approx)	Groups
-0.5882	0.0025	3



DETAILED INTERPRETATION

Tests whether several samples share the same distribution. The lead result here reads Statistic = -0.5882; p (approx) = 0.0025; Groups = 3. These rank-based tests need no normality assumption; a significant result signals the groups differ in location or distribution, and the effect size shows by how much. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reach a robust conclusion even with skewed or ordinal data.

2. Regression & Predictive Modelling

2.1 ANCOVA

LIVE OUTPUT

Group: General linear models · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

ANOVA with covariates and adjusted (marginal) means. It belongs to the **General linear models** family, used to partition variance in a response across factors and covariates. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, contract, tenure_months, monthly_charges. Analysts reach for it to attribute variation in the outcome to specific factors.

SAMPLE DATA INPUT

monetary	contract	tenure_months	monthly_charges
3197	Month-to-month	68	52.94
1233	One year	45	113.7
2501	Month-to-month	49	35.51
1468	Two year	64	71.45
4195	Two year	42	29.16

STEPS

1. Statistics → **General linear models** → **ANCOVA**. 2. Select the demo dataset above. 3. Set parameters — `column=monetary, factor=contract, covariates=tenure_months, monthly_charges`. 4. Click **Run** and read the tables/chart below.

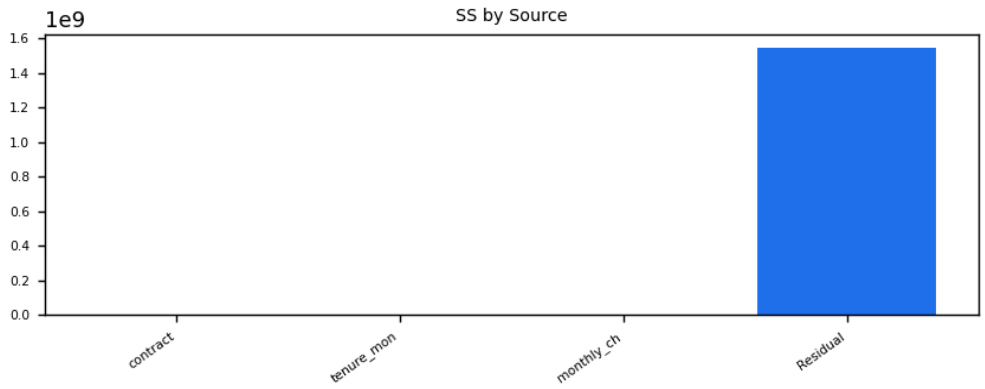
OUTPUT

ANCOVA — monetary by contract controlling for tenure_months, monthly_charges

Source	SS	df	F	p
contract	2.883e+06	2	0.7403	0.4773
tenure_months	5.081e+06	1	2.61	0.1066
monthly_charges	1.4e+05	1	0.0719	0.7886
Residual	1.548e+09	795		

Estimated marginal (adjusted) means

contract	N	Adjusted mean
Month-to-month	447	2597
One year	190	2741
Two year	163	2611



DETAILED INTERPRETATION

R^2 = 0.0052 · n = 800. The lead result here reads Source = contract; SS = 2.883e+06; df = 2; F = 0.7403. The model splits variance across factors and covariates; terms with small p-values contribute meaningfully to explaining the response. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst attribute variation in the outcome to specific factors.

2.2 Box's M test

LIVE OUTPUT

Group: General linear models · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Equality of covariance matrices. It belongs to the **General linear models** family, used to partition variance in a response across factors and covariates. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls, contract. Analysts reach for it to attribute variation in the outcome to specific factors.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls	contract
68	52.94	1	Month-to-month
45	113.7	5	One year
49	35.51	1	Month-to-month
64	71.45	2	Two year
42	29.16	1	Two year

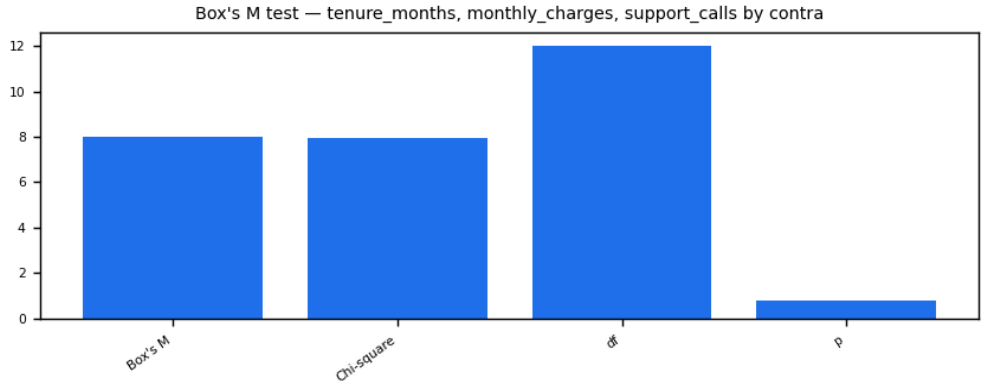
STEPS

1. Statistics → **General linear models** → **Box's M test**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Box's M test — tenure_months, monthly_charges, support_calls by contract

Box's M	Chi-square approx.	df	p
8.008	7.955	12	0.7887



DETAILED INTERPRETATION

Tests equality of covariance matrices across groups (assumption of MANOVA/LDA). The lead result here reads Box's M = 8.008; Chi-square approx. = 7.955; df = 12; p = 0.7887. The model splits variance across factors and covariates; terms with small p-values contribute meaningfully to explaining the response. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst attribute variation in the outcome to specific factors.

2.3 Hotelling's T²

LIVE OUTPUT

Group: General linear models · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Multivariate two-group mean comparison. It belongs to the **General linear models** family, used to partition variance in a response across factors and covariates. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls, churn. Analysts reach for it to attribute variation in the outcome to specific factors.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls	churn
68	52.94	1	No
45	113.7	5	Yes
49	35.51	1	No
64	71.45	2	No
42	29.16	1	No

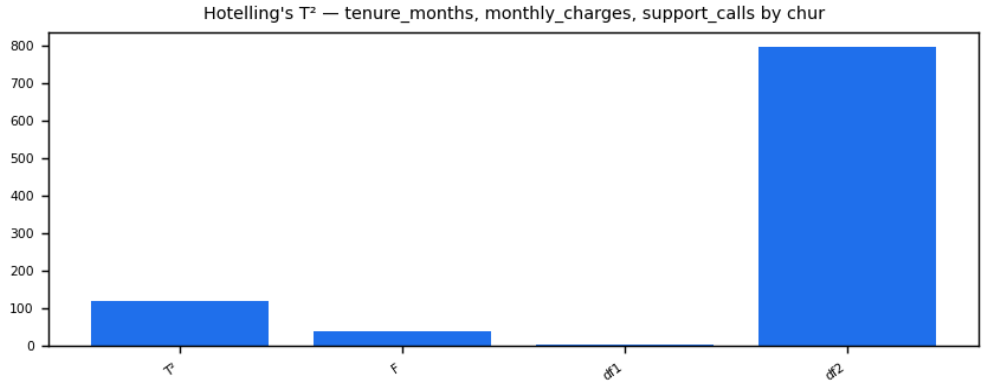
STEPS

1. Statistics → **General linear models** → **Hotelling's T²**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, group=churn`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Hotelling's T² — tenure_months, monthly_charges, support_calls by churn

T²	F	df1	df2	p
121.3	40.32	3	796	< .001



DETAILED INTERPRETATION

Multivariate 2-group mean comparison (No vs Yes). p < .05 → mean vectors differ. The lead result here reads T² = 121.3; F = 40.32; df1 = 3; df2 = 796. The model splits variance across factors and covariates; terms with small p-values contribute meaningfully to explaining the response. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst attribute variation in the outcome to specific factors.

2.4 Linear mixed model

LIVE OUTPUT

Group: General linear models · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Multilevel regression with a random intercept (ICC reported). It belongs to the **General linear models** family, used to partition variance in a response across factors and covariates. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls, contract. Analysts reach for it to attribute variation in the outcome to specific factors.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls	contract
3197	68	52.94	1	Month-to-month
1233	45	113.7	5	One year
2501	49	35.51	1	Month-to-month
1468	64	71.45	2	Two year
4195	42	29.16	1	Two year

STEPS

1. Statistics → **General linear models** → **Linear mixed model**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, group=contract`. 4. Click **Run** and read the tables/chart below.

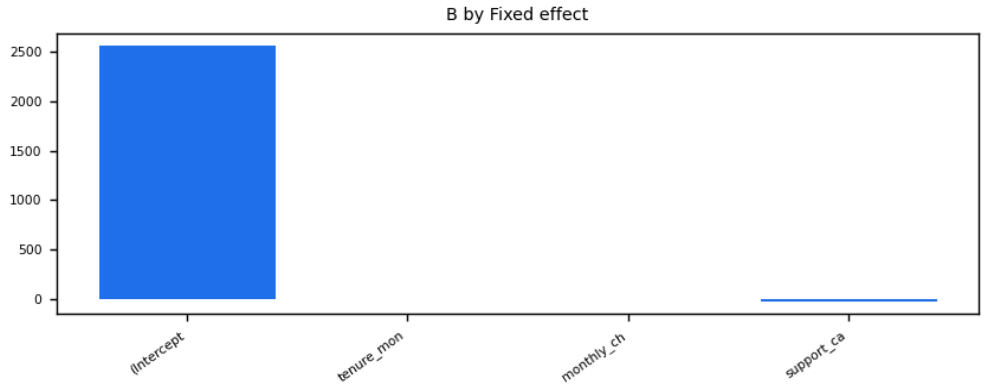
OUTPUT

Linear mixed model — monetary with random intercept by contract

Fixed effect	B	SE	z	p	CI low	CI high
(Intercept)	2555	173.1	14.76	< .001	2216	2894
tenure_months	3.913	2.391	1.637	0.1017	-0.7731	8.599
monthly_charges	-0.3873	1.721	-0.225	0.822	-3.761	2.986
support_calls	-20.13	35.37	-0.5692	0.5692	-89.45	49.19

Variance components

Component	Variance
Between contract (random i…	120
Residual	1.948e+06
ICC	0.0001



DETAILED INTERPRETATION

The lead result here reads Fixed effect = (Intercept); B = 2555; SE = 173.1; z = 14.76. The model splits variance across factors and covariates; terms with small p-values contribute meaningfully to explaining the response. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst attribute variation in the outcome to specific factors.

2.5 MANOVA

LIVE OUTPUT

Group: General linear models · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Multivariate ANOVA — several DVs by one factor (Wilks, Pillai…). It belongs to the **General linear models** family, used to partition variance in a response across factors and covariates. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls, contract. Analysts reach for it to attribute variation in the outcome to specific factors.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls	contract
68	52.94	1	Month-to-month
45	113.7	5	One year
49	35.51	1	Month-to-month
64	71.45	2	Two year
42	29.16	1	Two year

STEPS

1. Statistics → **General linear models** → **MANOVA**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, group=contract`. 4. Click **Run** and read the tables/chart below.

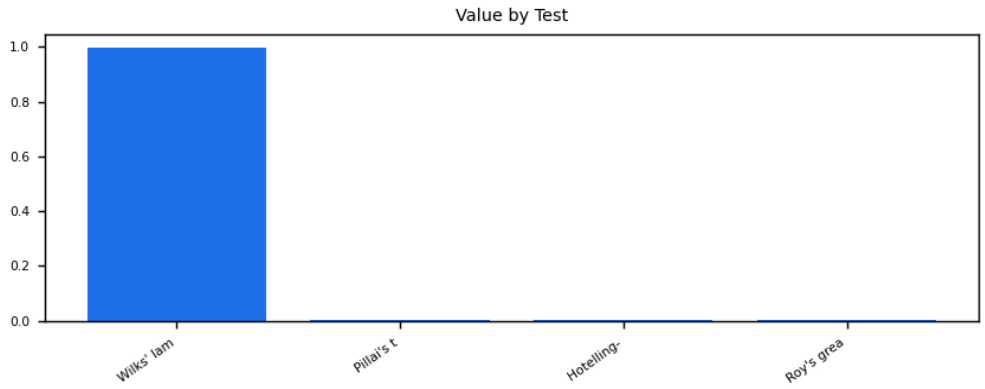
OUTPUT

MANOVA — tenure_months, monthly_charges, support_calls by contract

Test	Value	F	df num	df den	p
Wilks' lambda	0.9968	0.4315	6	1590	0.8582
Pillai's trace	0.0033	0.4321	6	1592	0.8579
Hotelling-Lawley trace	0.0033	0.4313	6	1058	0.8583
Roy's greatest root	0.0021	0.5538	3	796	0.6457

Group means

contract	tenure_months	monthly_charges	support_calls
Month-to-month	36.78	68.26	1.926
One year	36.43	70.33	1.995
Two year	38.77	69.91	1.982



DETAILED INTERPRETATION

Wilks' lambda is the most commonly reported statistic. The lead result here reads Test = Wilks' lambda; Value = 0.9968; F = 0.4315; df num = 6. The model splits variance across factors and covariates; terms with small p-values contribute meaningfully to explaining the response. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst attribute variation in the outcome to specific factors.

2.6 RM-ANOVA (sphericity + power)

LIVE OUTPUT

Group: General linear models · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Repeated-measures ANOVA with Mauchly, Greenhouse-Geisser & observed power. It belongs to the **General linear models** family, used to partition variance in a response across factors and covariates. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to attribute variation in the outcome to specific factors.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **General linear models** → **RM-ANOVA (sphericity + power)**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

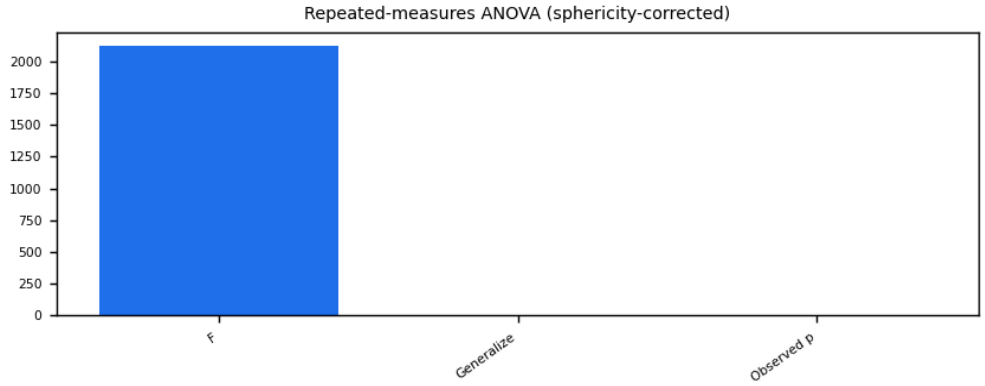
OUTPUT

Repeated-measures ANOVA (sphericity-corrected)

F	p (uncorrected)	p (Greenhouse-Geisser)	Generalized η^2	Observed power
2123	< .001	< .001	0.6424	1

Mauchly's test of sphericity

Mauchly W	p	Greenhouse-Geisser ε	Verdict
0.6509	< .001	0.7412	sphericity violated — use …



DETAILED INTERPRETATION

n = 800 subjects, 3 conditions. The lead result here reads F = 2123; p (uncorrected) = < .001; p (Greenhouse-Geisser) = < .001; Generalized η^2 = 0.6424. The model splits variance across factors and covariates; terms with small p-values contribute meaningfully to explaining the response. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst attribute variation in the outcome to specific factors.

2.7 Repeated-measures ANOVA

LIVE OUTPUT

Group: General linear models · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Within-subject comparison of 2+ repeated measurements. It belongs to the **General linear models** family, used to partition variance in a response across factors and covariates. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to attribute variation in the outcome to specific factors.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **General linear models** → **Repeated-measures ANOVA**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

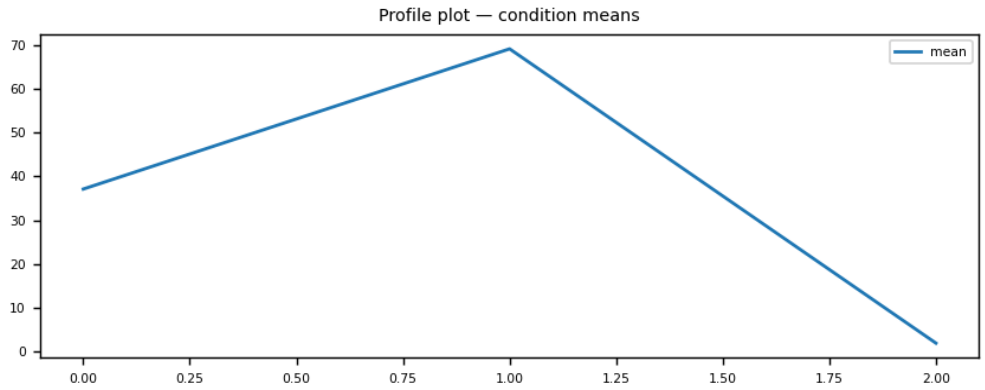
OUTPUT

Repeated-measures ANOVA (within-subject)

Effect	F	df num	df den	p
condition	2123	2	1598	< .001

Condition means

Condition	Mean	SD
tenure_months	37.1	20.72
monthly_charges	69.09	28.74
support_calls	1.954	1.401



DETAILED INTERPRETATION

n = 800 complete cases. Sphericity is assumed (Greenhouse-Geisser correction not applied). The lead result here reads Effect = condition; F = 2123; df num = 2; df den = 1598. The model splits variance across factors and covariates; terms with small p-values contribute meaningfully to explaining the response. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst attribute variation in the outcome to specific factors.

2.8 Breusch-Pagan test

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Heteroskedasticity diagnostic. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

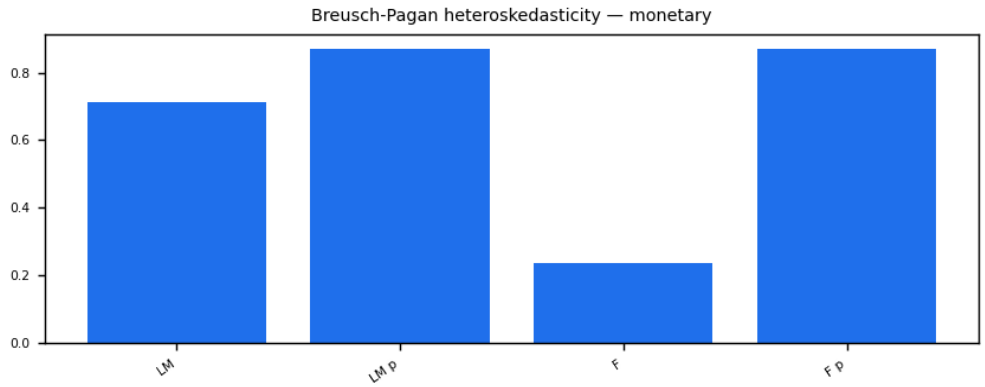
STEPS

1. Statistics → **Regression** → **Breusch-Pagan test**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Breusch-Pagan heteroskedasticity — monetary

LM	LM p	F	F p
0.7121	0.8704	0.2364	0.871



DETAILED INTERPRETATION

p < .05 → error variance is not constant (heteroskedasticity) → use robust SEs. The lead result here reads LM = 0.7121; LM p = 0.8704; F = 0.2364; F p = 0.871. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.9 Categorical regression (CATREG)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Optimal-scaling regression with nominal/ordinal predictors (ALS). It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges
3197	68	52.94
1233	45	113.7
2501	49	35.51
1468	64	71.45
4195	42	29.16

STEPS

1. Statistics → **Regression** → **Categorical regression (CATREG)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges`. 4. Click **Run** and read the tables/chart below.

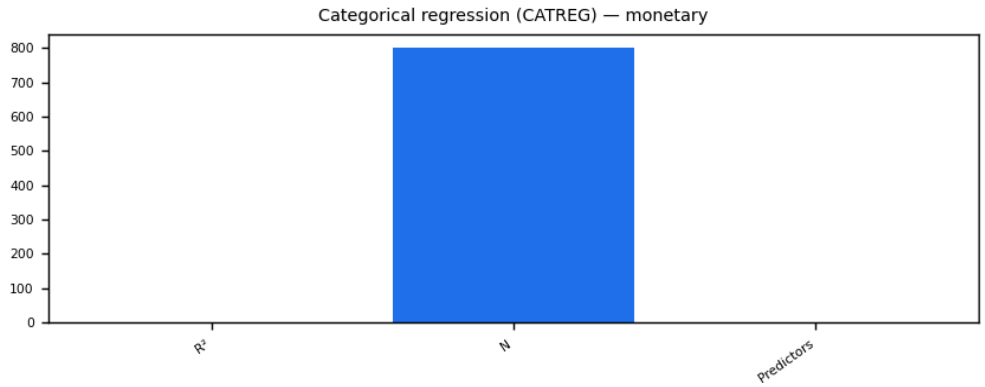
OUTPUT

Categorical regression (CATREG) — monetary

R^2	N	Predictors
0.0033	800	2

Standardized coefficients (importance = |β|)

Predictor	Beta	|Beta|	Scaling
tenure_months	0.0566	0.0566	numeric
monthly_charges	-0.0084	0.0084	numeric



DETAILED INTERPRETATION

The lead result here reads R² = 0.0033; N = 800; Predictors = 2. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.10 Conformal prediction intervals

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Distribution-free prediction intervals with coverage guarantee. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

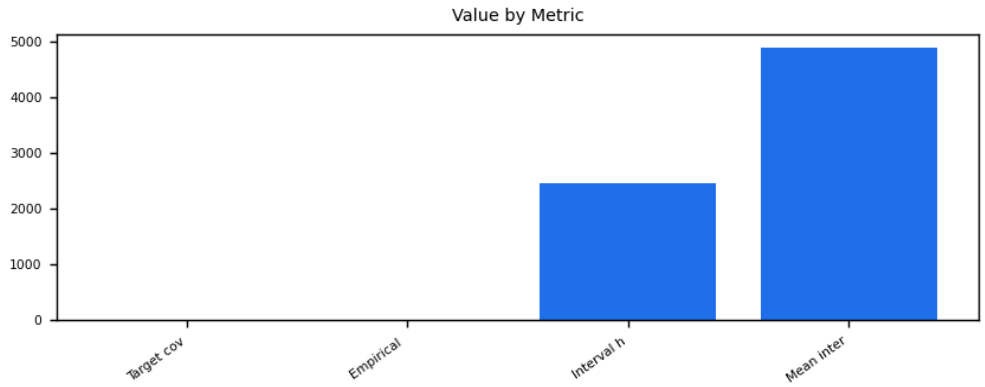
STEPS

1. Statistics → **Regression** → **Conformal prediction intervals**. 2. Select the demo dataset above. 3. Set parameters — `target=monetary, predictors=tenure_months, monthly_charges, support_calls, alpha=0.1`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Split-conformal prediction intervals — monetary

Metric	Value
Target coverage (1–α)	0.9
Empirical coverage (test)	0.915
Interval half-width (q)	2444
Mean interval width	4887
n train / calib / test	400 / 200 / 200



DETAILED INTERPRETATION

Distribution-free intervals with finite-sample coverage guarantee (Vovk; wraps a gradient-boosting model). Empirical coverage should ≈ 1–α. The lead result here reads Metric = Target coverage (1–α); Value = 0.9. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.11 Cook's distance (influence)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Influential-observation diagnostics. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

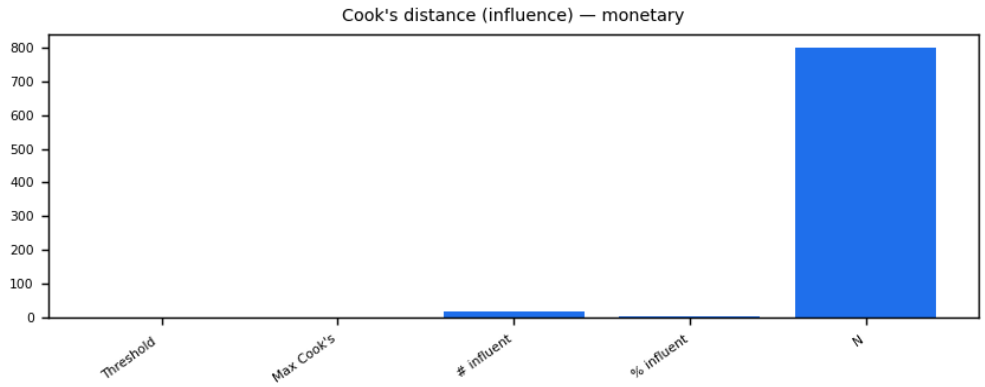
STEPS

1. Statistics → **Regression** → **Cook's distance (influence)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Cook's distance (influence) — monetary

Threshold (4/n)	Max Cook's D	# influential	% influential	N
0.005	0.0106	18	2.25	800



DETAILED INTERPRETATION

Rows above 4/n exert outsized influence on the fitted regression. The lead result here reads Threshold (4/n) = 0.005; Max Cook's D = 0.0106; # influential = 18; % influential = 2.25. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.12 Curve estimation

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Fit & rank linear/quadratic/cubic/log/inverse/power/exponential/S models. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months
3197	68
1233	45
2501	49
1468	64
4195	42

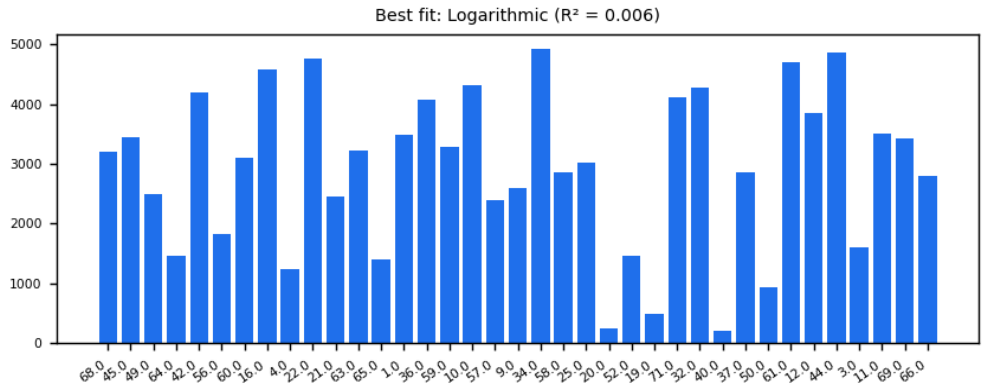
STEPS

1. Statistics → **Regression** → **Curve estimation**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, independent=tenure_months`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Curve estimation — monetary on tenure_months

Model	Params	R^2	Adjusted R^2	AIC	
Logarithmic	2	0.0064	0.0052	1.158e+04	★ best
Inverse	2	0.0059	0.0047	1.158e+04	
Cubic	4	0.0098	0.0061	1.158e+04	
Quadratic	3	0.007	0.0045	1.159e+04	
Linear	2	0.0032	0.002	1.159e+04	
Power	2	-0.1541	-0.1555	1.17e+04	
S-curve	2	-0.1545	-0.1559	1.17e+04	
Exponential	2	-0.1583	-0.1597	1.171e+04	



DETAILED INTERPRETATION

Models ranked by AIC (lower = better). R² is on the original scale. Non-applicable models (e.g. needing positive values) are omitted. The lead result here reads Model = Logarithmic; Params = 2; R² = 0.0064; Adjusted R² = 0.0052. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.13 Durbin-Watson test

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Residual autocorrelation. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

STEPS

1. Statistics → **Regression** → **Durbin-Watson test**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Durbin-Watson test — monetary residuals

DW statistic	Interpretation
2.018	no strong autocorrelation

DETAILED INTERPRETATION

≈2 = no autocorrelation; <1.5 positive; >2.5 negative residual autocorrelation. The lead result here reads DW statistic = 2.018; Interpretation = no strong autocorrelat. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.14 Gaussian-process regression (kriging)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

GP emulator for smooth/expensive functions with uncertainty. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on churn_bin, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

churn_bin	tenure_months	monthly_charges	support_calls
0	68	52.94	1
1	45	113.7	5
0	49	35.51	1
0	64	71.45	2
0	42	29.16	1

STEPS

1. Statistics → **Regression** → **Gaussian-process regression (kriging)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=churn_bin, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

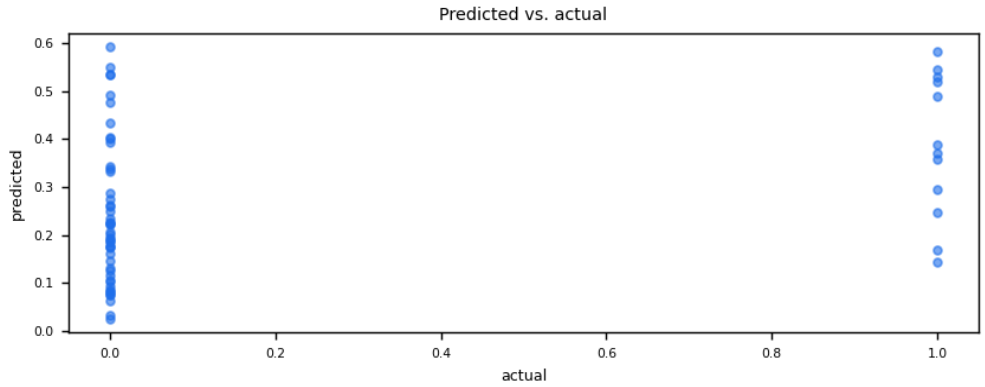
OUTPUT

Gaussian-process regression (kriging) — churn_bin

R² (fit)	Log marginal likelihood	N	Predictors
0.1388	-1088	800	3

Kernel length-scales (larger = smoother / less influential)

Predictor	Length scale
tenure_months	3.211
monthly_charges	9.374
support_calls	9.566



DETAILED INTERPRETATION

GP emulator with an RBF kernel + noise term. Use as a fast surrogate for expensive simulations; predictions carry uncertainty. The lead result here reads R² (fit) = 0.1388; Log marginal likelihood = -1088; N = 800; Predictors = 3. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.15 Generalized estimating equations (GEE)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Population-averaged model for clustered data with robust SEs. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls, contract. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls	contract
3197	68	52.94	1	Month-to-month
1233	45	113.7	5	One year
2501	49	35.51	1	Month-to-month
1468	64	71.45	2	Two year
4195	42	29.16	1	Two year

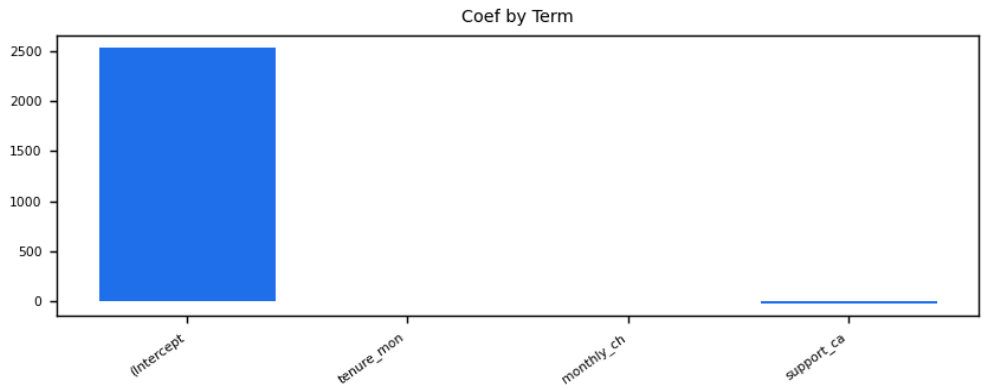
STEPS

1. Statistics → **Regression** → **Generalized estimating equations (GEE)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, group=contract, family=gaussian`. 4. Click **Run** and read the tables/chart below.

OUTPUT

GEE — monetary ~ tenure_months, monthly_charges, support_calls (family: gaussian, exchangeable, robust SE)

Term	Coef	Robust SE	z	p	95% CI low	95% CI high
(Intercept)	2529	227.6	11.11	< .001	2083	2975
tenure_months	3.913	1.074	3.644	< .001	1.809	6.018
monthly_charges	-0.3289	0.9208	-0.3572	0.721	-2.134	1.476
support_calls	-19.33	59.5	-0.3248	0.7453	-135.9	97.29



DETAILED INTERPRETATION

Population-averaged over 3 clusters of 'contract'. The lead result here reads Term = (Intercept); Coef = 2529; Robust SE = 227.6; z = 11.11. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.16 Generalized linear mixed model (GLMM)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Random-intercept logistic/Poisson mixed model. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls, contract. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls	contract
3197	68	52.94	1	Month-to-month
1233	45	113.7	5	One year
2501	49	35.51	1	Month-to-month
1468	64	71.45	2	Two year
4195	42	29.16	1	Two year

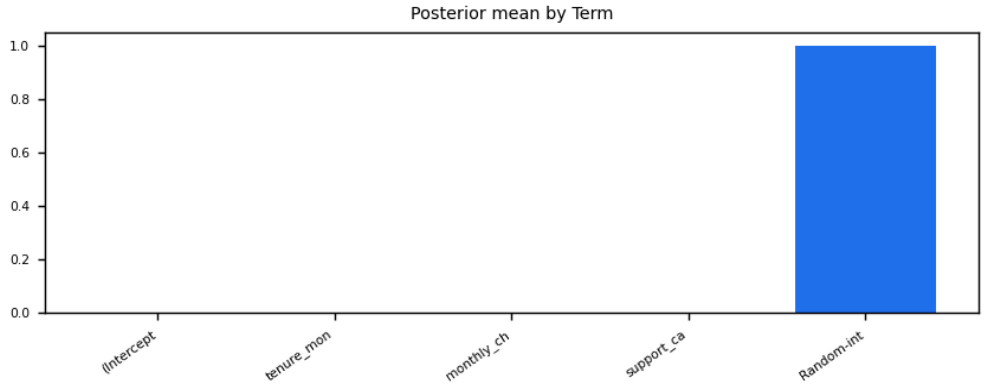
STEPS

1. Statistics → **Regression** → **Generalized linear mixed model (GLMM)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

GLMM — monetary ~ tenure_months, monthly_charges, support_calls + (1|contract) [Poisson (log), variational Bayes]

Term	Posterior mean	Posterior SD
(Intercept)	0	0.5186
tenure_months	0	0.5744
monthly_charges	0	0.6431
support_calls	0	0.575
Random-intercept SD (grp)	1	



DETAILED INTERPRETATION

Fixed effects on the link scale; random-intercept SD summarizes between-'contract' variation. The lead result here reads Term = (Intercept); Posterior mean = 0; Posterior SD = 0.5186. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.17 Generalized linear model (GLM)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

GLM with selectable family/link. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

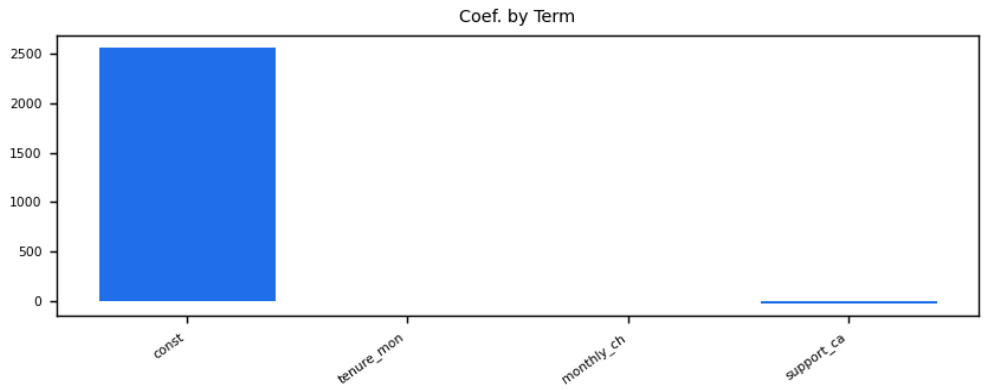
STEPS

1. Statistics → **Regression** → **Generalized linear model (GLM)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, family=gaussian`. 4. Click **Run** and read the tables/chart below.

OUTPUT

GLM (gaussian) — monetary

Term	Coef.	SE	z	p
const	2554	169.4	15.08	< .001
tenure_months	3.913	2.391	1.637	0.1017
monthly_charges	-0.3863	1.719	-0.2247	0.8222
support_calls	-20.12	35.35	-0.5691	0.5693



DETAILED INTERPRETATION

Deviance = 1550317104.9 · AIC = 13860.0. Generalized linear model with a link function. The lead result here reads Term = const; Coef. = 2554; SE = 169.4; z = 15.08. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.18 Heckman selection model

LIVE OUTPUT

Group: Regression · Dataset: a labour survey where wages are observed only for participants (sample selection)

WHAT THIS DOES

Two-step sample-selection correction (inverse Mills ratio). It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a labour survey where wages are observed only for participants (sample selection), drawing on wage, participate, education, experience, num_kids, other_income. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

wage	participate	education	experience	num_kids	other_income
	0	11	11	3	39.2
	0	15	28	1	31.6
	0	12	33	3	27.4
	0	8	2	1	36.3
14.51	1	14	14	1	2.8

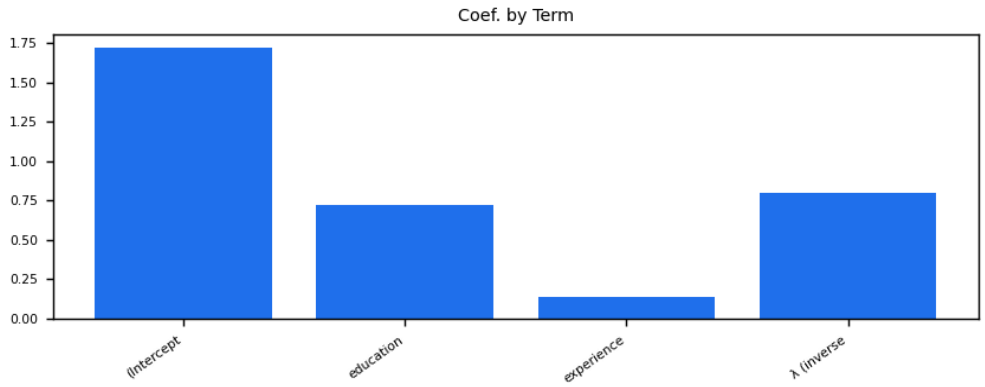
STEPS

1. Statistics → **Regression** → **Heckman selection model**. 2. Select the demo dataset above. 3. Set parameters — `outcome=wage, selection=participate, outcome_predictors=education, experience, selection_predictors=education, experience, num_kids, other_income`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Heckman selection model (two-step) — wage

Term	Coef.	SE	p
(Intercept)	1.722	3.054	0.574
education	0.7209	0.1173	< .001
experience	0.137	0.0263	< .001
λ (inverse Mills)	0.8002	1.217	0.5122



DETAILED INTERPRETATION

Outcome equation corrected for sample selection via the inverse Mills ratio. n selected = 120 of 300. A significant λ (p = 0.5122) indicates selection bias — OLS on the selected sample alone would be biased. The lead result here reads Term = (Intercept); Coef. = 1.722; SE = 3.054; p = 0.574. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a labour survey where wages are observed only for participants (sample selection), this lets the analyst identify and size the levers that move the outcome.

2.19 IV / two-stage least squares

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Instrumental-variables regression for endogenous regressors. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, monthly_charges, support_calls, tenure_months. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	monthly_charges	support_calls	tenure_months
3197	52.94	1	68
1233	113.7	5	45
2501	35.51	1	49
1468	71.45	2	64
4195	29.16	1	42

STEPS

1. Statistics → **Regression** → **IV / two-stage least squares**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, endogenous=monthly_charges, instruments=support_calls, exogenous=tenure_months`. 4. Click **Run** and read the tables/chart below.

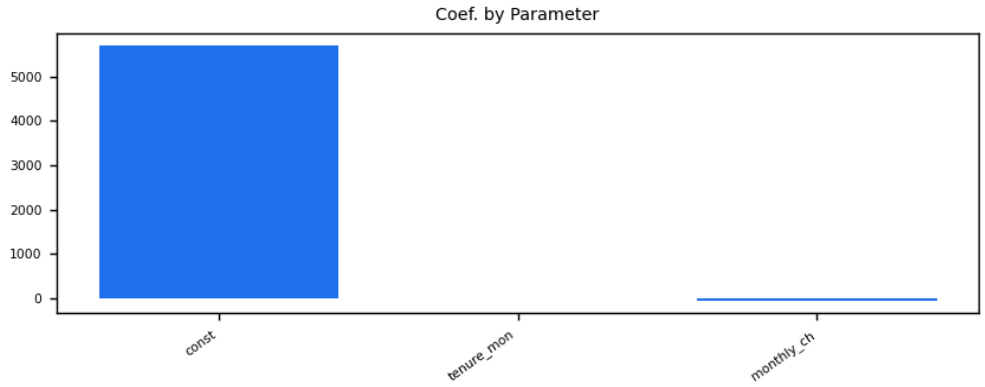
OUTPUT

Two-stage least squares (IV) — monetary

Parameter	Coef.	Robust SE	t	p	CI low	CI high
const	5686	7338	0.7749	0.4384	-8697	2.007e+04
tenure_months	1.708	5.762	0.2964	0.767	-9.585	13
monthly_charges	-45.1	103.6	-0.4355	0.6632	-248.1	157.9

Wu-Hausman endogeneity test

Statistic	p
0.3239	0.5694



DETAILED INTERPRETATION

Endogenous: monthly_charges · instruments: support_calls. R^2 = -0.8430 · n = 800. The lead result here reads Parameter = const; Coef. = 5686; Robust SE = 7338; t = 0.7749. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.20 Linear regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

OLS with standardized betas, CIs, VIF and Durbin-Watson. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

STEPS

1. Statistics → **Regression** → **Linear regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

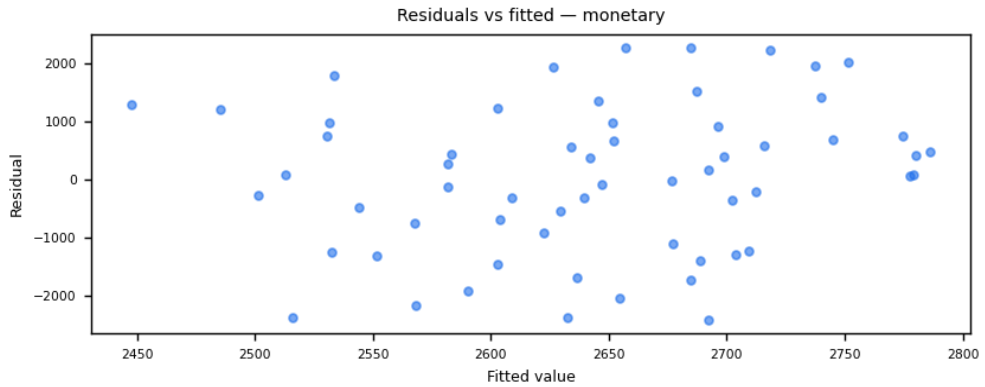
OUTPUT

Model summary

R	R²	Adjusted R²	SE of estimate	F	df	p	Durbin-Watson	N
0.061	0.0037	-0	1396	0.9894	3, 796	0.3971	2.018	800

Coefficients

Predictor	B	SE	Beta (std)	t	p	95% CI low	95% CI high	VIF
(Constant)	2554	169.4		15.08	< .001	2222	2887	
tenure_months	3.913	2.391	0.0581	1.637	0.1021	-0.7803	8.606	1.006
monthly_charges	-0.3863	1.719	-0.008	-0.2247	0.8222	-3.761	2.988	1.002
support_calls	-20.12	35.35	-0.0202	-0.5691	0.5694	-89.51	49.27	1.006



DETAILED INTERPRETATION

The lead result here reads R = 0.061; R² = 0.0037; Adjusted R² = -0; SE of estimate = 1396. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.21 Logistic diagnostics (H-L)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Hosmer-Lemeshow calibration. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on churn_bin, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

churn_bin	tenure_months	monthly_charges	support_calls
0	68	52.94	1
1	45	113.7	5
0	49	35.51	1
0	64	71.45	2
0	42	29.16	1

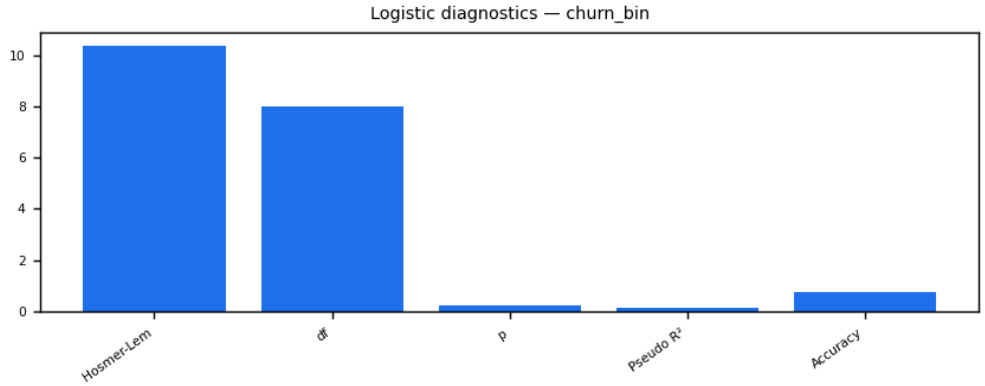
STEPS

1. Statistics → **Regression** → **Logistic diagnostics (H-L)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=churn_bin, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Logistic diagnostics — churn_bin

Hosmer-Lemeshow χ^2	df	p	Pseudo R^2	Accuracy
10.37	8	0.2402	0.1179	0.7438



DETAILED INTERPRETATION

Hosmer-Lemeshow p > .05 indicates adequate calibration (good fit). The lead result here reads Hosmer-Lemeshow χ^2 = 10.37; df = 8; p = 0.2402; Pseudo R^2 = 0.1179. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.22 Logistic regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Binary logit with odds ratios, pseudo-R^2 and classification table. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on churn_bin, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

churn_bin	tenure_months	monthly_charges	support_calls
0	68	52.94	1
1	45	113.7	5
0	49	35.51	1
0	64	71.45	2
0	42	29.16	1

STEPS

1. Statistics → **Regression** → **Logistic regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=churn_bin, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Logistic regression — P(churn_bin = 1)

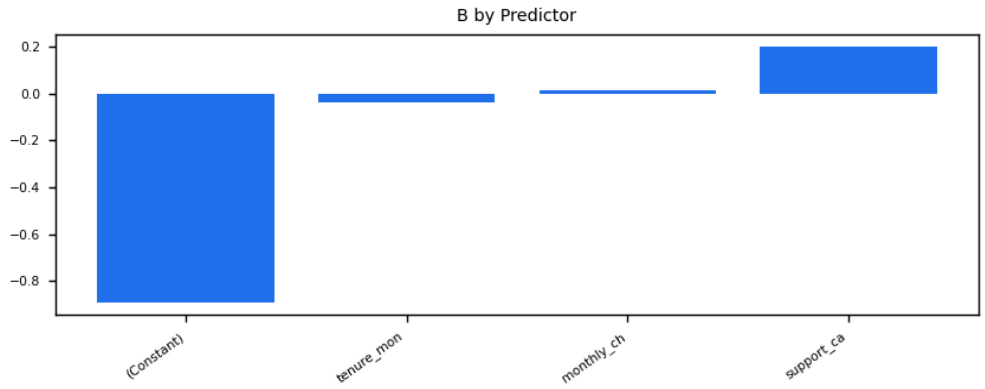
Predictor	B	SE	Wald	p	Odds ratio	OR CI low	OR CI high
(Constant)	-0.8899	0.2849	9.757	0.0018	0.4107	0.235	0.7178
tenure_months	-0.0389	0.0044	77.14	< .001	0.9619	0.9536	0.9703
monthly_charges	0.0121	0.003	16.05	< .001	1.012	1.006	1.018
support_calls	0.1972	0.0598	10.88	< .001	1.218	1.083	1.369

Model fit

-2 Log likelihood	McFadden R²	Nagelkerke R²	Accuracy	N
836.9	0.1179	0.1878	0.7438	800

Classification table (cut = 0.5)

	Predicted 0	Predicted 1	% correct
Observed 0	537	39	93.23
Observed 1	166	58	25.89



DETAILED INTERPRETATION

The lead result here reads Predictor = (Constant); B = -0.8899; SE = 0.2849; Wald = 9.757. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.23 Multinomial logistic

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Categorical outcome with 3+ levels; relative risk ratios. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on contract, tenure_months, monthly_charges, monetary. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

contract	tenure_months	monthly_charges	monetary
Month-to-month	68	52.94	3197
One year	45	113.7	1233
Month-to-month	49	35.51	2501
Two year	64	71.45	1468
Two year	42	29.16	4195

STEPS

1. Statistics → **Regression** → **Multinomial logistic**. 2. Select the demo dataset above. 3. Set parameters — `dependent=contract, predictors=tenure_months, monthly_charges, monetary`. 4. Click **Run** and read the tables/chart below.

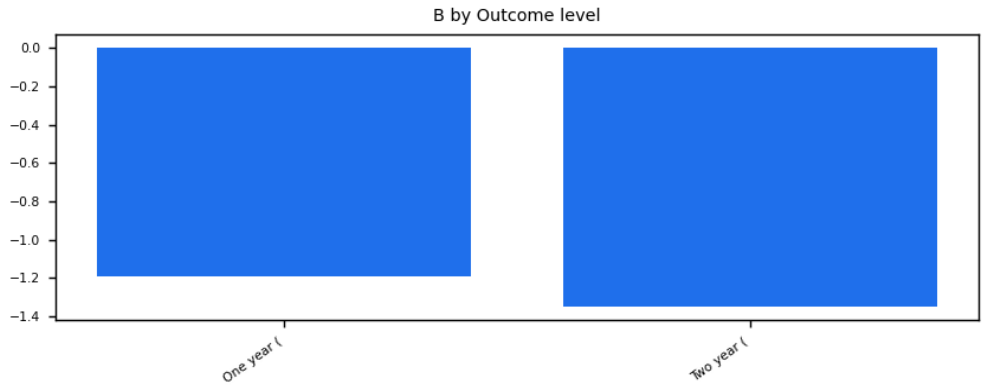
OUTPUT

Multinomial logistic regression — contract (base: Month-to-month)

Outcome level	Predictor	B	SE	p	Relative risk ratio
One year (vs Month-to-mont…	(Constant)	-1.193	0.3219	< .001	0.3032
One year (vs Month-to-mont…	tenure_months	-0.001	0.0042	0.8118	0.999
One year (vs Month-to-mont…	monthly_charges	0.0025	0.003	0.4035	1.002
One year (vs Month-to-mont…	monetary	0.0001	0.0001	0.2316	1
Two year (vs Month-to-mont…	(Constant)	-1.352	0.3402	< .001	0.2586
Two year (vs Month-to-mont…	tenure_months	0.0047	0.0044	0.2875	1.005
Two year (vs Month-to-mont…	monthly_charges	0.0021	0.0032	0.5078	1.002
Two year (vs Month-to-mont…	monetary	0	0.0001	0.9109	1

Model fit

McFadden R^2	Accuracy	N	Categories
0.0024	0.5587	800	3



DETAILED INTERPRETATION

The lead result here reads Outcome level = One year (vs Month-to-; Predictor = (Constant); B = -1.193; SE = 0.3219. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.24 Multiple imputation (MICE)

INPUT REQUIREMENT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Impute missing data m times and pool with Rubin's rules. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Regression** → **Multiple imputation (MICE)**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, m=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

No missing values in the selected columns — nothing to impute

DETAILED INTERPRETATION

This technique has a specific data requirement: *No missing values in the selected columns — nothing to impute*. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a regression method and supports the decision to identify and size the levers that move the outcome.

2.25 Negative binomial regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Overdispersed count-data regression. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

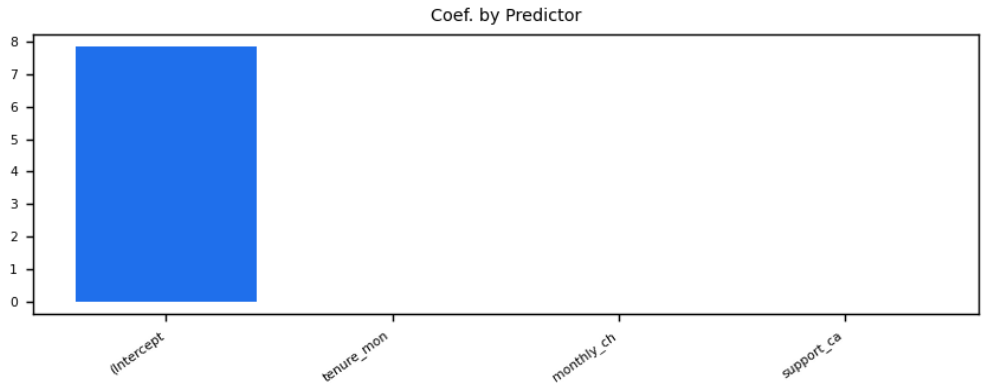
STEPS

1. Statistics → **Regression** → **Negative binomial regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Negative binomial regression (overdispersed counts)

Predictor	Coef.	SE	z	p	IRR
(Intercept)	7.846	0.1214	64.63	< .001	2556
tenure_months	0.0015	0.0017	0.8967	0.3699	1.002
monthly_charges	-0.0002	0.0012	-0.1385	0.8899	0.9998
support_calls	-0.0081	0.0253	-0.3181	0.7504	0.992



DETAILED INTERPRETATION

n = 800 · deviance = 382.85 · AIC = 14209.26. IRR = exp(coef) (incidence-rate ratio). The lead result here reads Predictor = (Intercept); Coef. = 7.846; SE = 0.1214; z = 64.63. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.26 Nonlinear regression (NLS)

INPUT REQUIREMENT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Fit logistic/Gompertz/Michaelis-Menten/exponential-plateau growth models. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary
3197
1233
2501
1468
4195

STEPS

1. Statistics → **Regression** → **Nonlinear regression (NLS)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, independent=monetary, model=logistic`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Model did not converge — try a different model. (Result from function call is not a proper array of floats.)

DETAILED INTERPRETATION

This technique has a specific data requirement: *Model did not converge — try a different model. (Result from function call is not a proper array of floats.)*. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a regression method and supports the decision to identify and size the levers that move the outcome.

2.27 Ordinal regression

LIVE OUTPUT

Group: Regression · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Proportional-odds model for ordered outcomes (e.g. Likert). It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on int1, peou1, pu1. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

int1	peou1	pu1
4	3	4
3	3	3
5	4	5
5	5	3
2	2	3

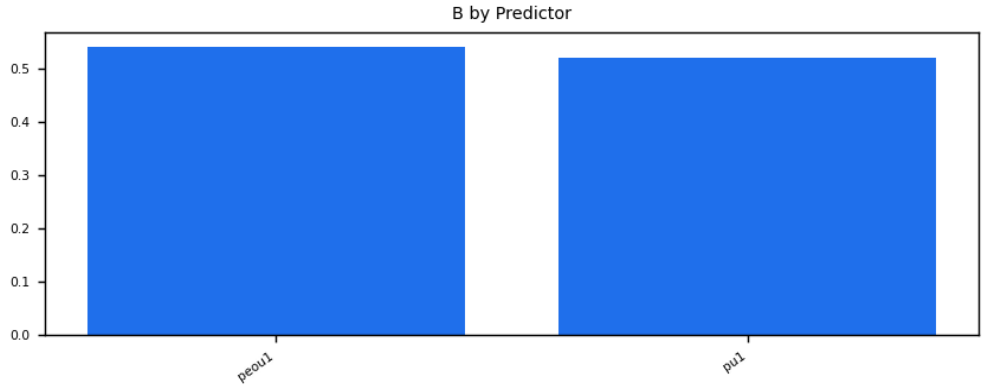
STEPS

1. Statistics → **Regression** → **Ordinal regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=int1, predictors=peou1, pu1`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Ordinal (proportional-odds) regression — int1

Predictor	B	SE	p	Odds ratio
peou1	0.5417	0.1108	< .001	1.719
pu1	0.5213	0.1182	< .001	1.684



DETAILED INTERPRETATION

Logit link · 7 ordered levels · n = 220 · positive B = higher predictor values push toward higher categories. The lead result here reads Predictor = peou1; B = 0.5417; SE = 0.1108; p = < .001. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst identify and size the levers that move the outcome.

2.28 Partial least squares regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

For many/collinear predictors. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

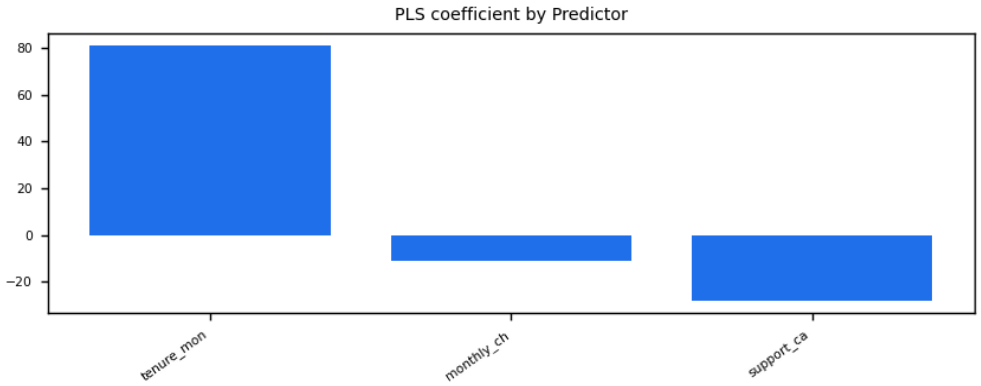
STEPS

1. Statistics → **Regression** → **Partial least squares regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, components=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Partial least squares regression — monetary

Predictor	PLS coefficient
tenure_months	81.01
monthly_charges	-11.1
support_calls	-28.16



DETAILED INTERPRETATION

3 components · in-sample R^2 = 0.004. Handles many/collinear predictors. The lead result here reads Predictor = tenure_months; PLS coefficient = 81.01. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.29 Poisson regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Count-data regression with incidence-rate ratios. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on support_calls, tenure_months, monthly_charges. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

support_calls	tenure_months	monthly_charges
1	68	52.94
5	45	113.7
1	49	35.51
2	64	71.45
1	42	29.16

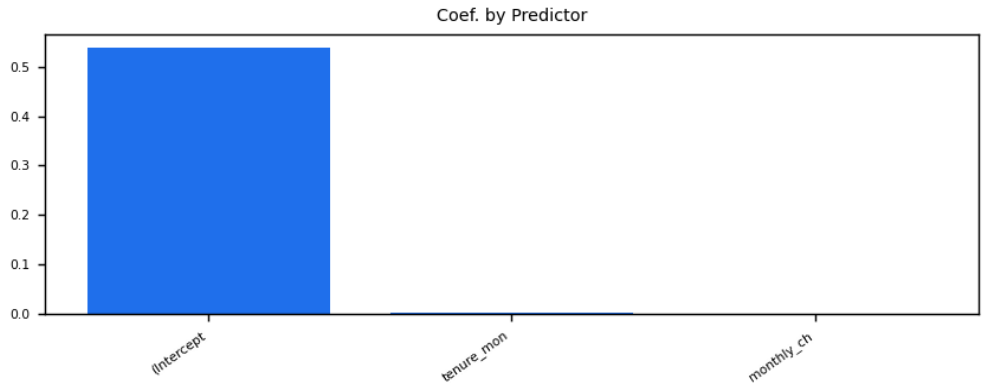
STEPS

1. Statistics → **Regression** → **Poisson regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=support_calls, predictors=tenure_months, monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Poisson regression (count data)

Predictor	Coef.	SE	z	p	IRR
(Intercept)	0.5383	0.0821	6.554	< .001	1.713
tenure_months	0.0025	0.0012	2.036	0.0418	1.002
monthly_charges	0.0005	0.0009	0.6163	0.5377	1



DETAILED INTERPRETATION

n = 800 · deviance = 911.27 · AIC = 2709.73. IRR = exp(coef) (incidence-rate ratio). The lead result here reads Predictor = (Intercept); Coef. = 0.5383; SE = 0.0821; z = 6.554. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.30 Principal-component regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Regress on orthogonal PCs. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

STEPS

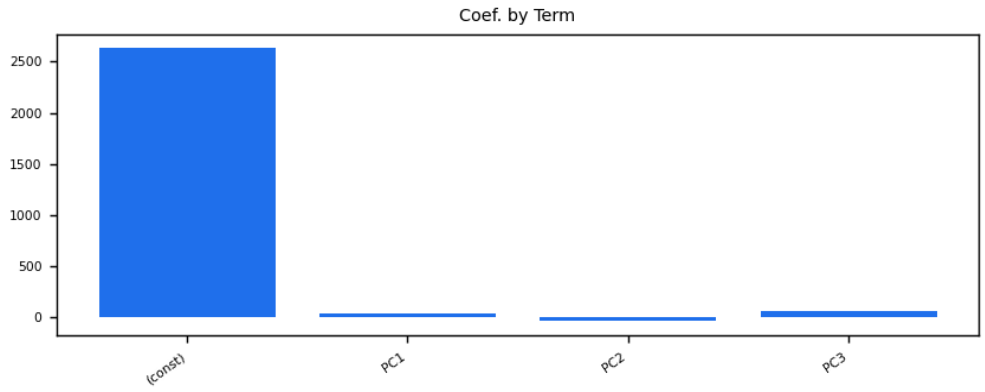
1. Statistics → **Regression** → **Principal-component regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, components=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Principal-component regression — monetary

Term	Coef.	p	Variance explained
(const)	2634	< .001	
PC1	41.9	0.3793	0.3576

Term	Coef.	p	Variance explained
PC2	-34.48	0.4816	0.3382
PC3	67.33	0.1928	0.3041



DETAILED INTERPRETATION

3 components · R^2 = 0.004. Regresses on orthogonal PCs to beat collinearity. The lead result here reads Term = (const); Coef. = 2634; p = < .001; Variance explained = . Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.31 Probit regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Binary model with probit link. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on churn_bin, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

churn_bin	tenure_months	monthly_charges	support_calls
0	68	52.94	1
1	45	113.7	5
0	49	35.51	1
0	64	71.45	2
0	42	29.16	1

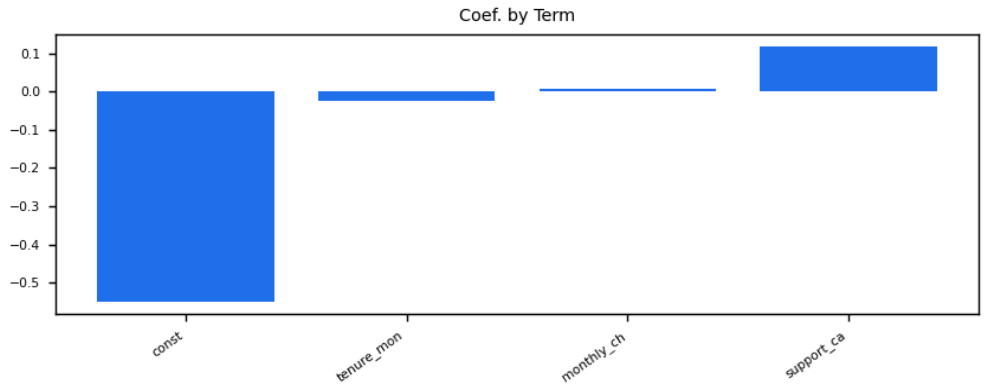
STEPS

1. Statistics → **Regression** → **Probit regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=churn_bin, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Probit regression — P(churn_bin = 1)

Term	Coef.	SE	z	p
const	-0.55	0.1688	-3.259	0.0011
tenure_months	-0.0228	0.0025	-9.057	< .001
monthly_charges	0.0073	0.0018	4.124	< .001
support_calls	0.1172	0.0353	3.323	< .001



DETAILED INTERPRETATION

Pseudo R^2 = 0.118. Binary model with a normal (probit) link. The lead result here reads Term = const; Coef. = -0.55; SE = 0.1688; z = -3.259. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.32 Quantile regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Model a conditional quantile (robust to outliers). It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

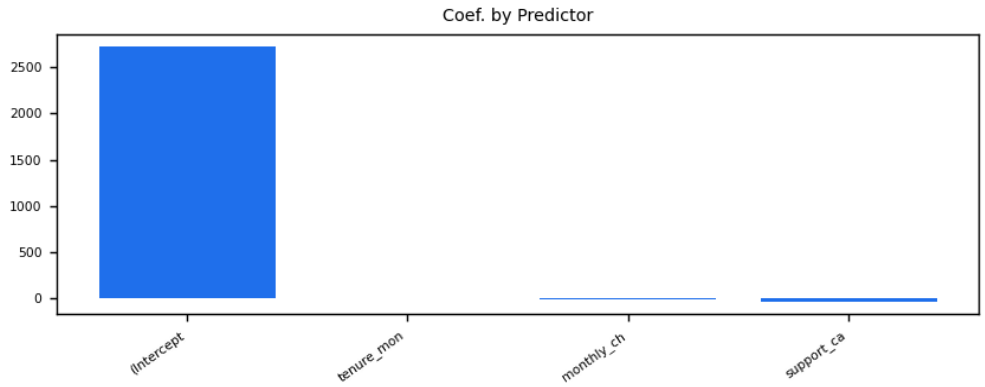
STEPS

1. Statistics → **Regression** → **Quantile regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, quantile=0.5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Quantile regression (τ = 0.5) — monetary

Predictor	Coef.	SE	t	p
(Intercept)	2717	255.3	10.64	< .001
tenure_months	5.821	3.603	1.615	0.1066
monthly_charges	-1.835	2.591	-0.7081	0.4791
support_calls	-32.42	53.28	-0.6085	0.543



DETAILED INTERPRETATION

Models the conditional 50% quantile of monetary (robust to outliers; compare across quantiles for heterogeneous effects). The lead result here reads Predictor = (Intercept); Coef. = 2717; SE = 255.3; t = 10.64. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.33 Ramsey RESET test

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Functional-form/mis-specification test. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

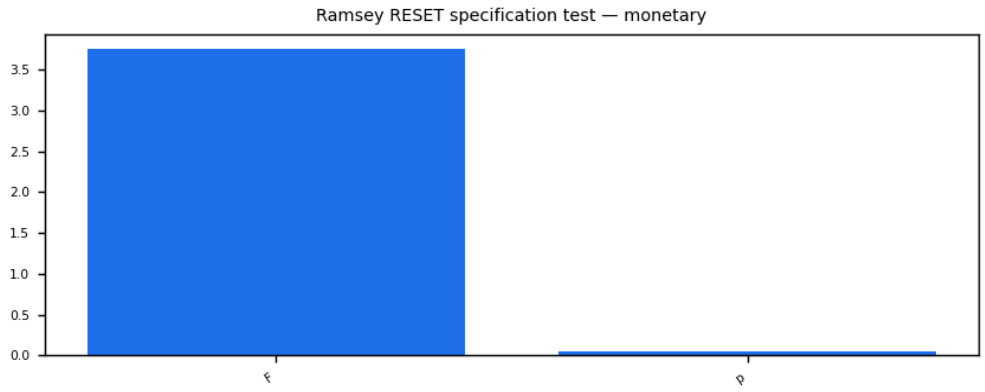
STEPS

1. Statistics → **Regression** → **Ramsey RESET test**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Ramsey RESET specification test — monetary

F	p
3.749	0.0532



DETAILED INTERPRETATION

p < .05 → the linear model is mis-specified (omitted nonlinearity). The lead result here reads F = 3.749; p = 0.0532. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.34 Ridge / Lasso regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Penalized regression (CV-tuned). It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

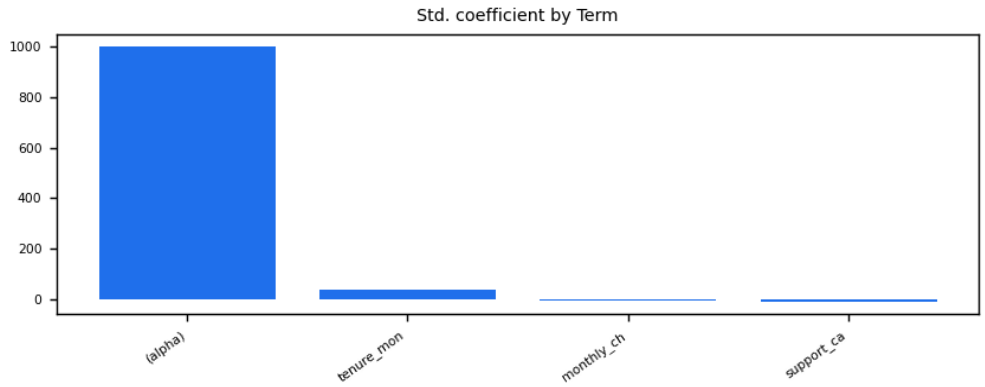
STEPS

1. Statistics → **Regression** → **Ridge / Lasso regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, penalty=ridge`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Ridge regression (standardized) — monetary

Term	Std. coefficient	Note
(alpha)	1000	
tenure_months	35.54	
monthly_charges	-5.766	
support_calls	-11.12	



DETAILED INTERPRETATION

Cross-validated penalty. In-sample R^2 = 0.003. Ridge shrinks correlated predictors for stability. The lead result here reads Term = (alpha); Std. coefficient = 1000; Note = . Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.35 Robust regression (Huber)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Outlier-resistant regression. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

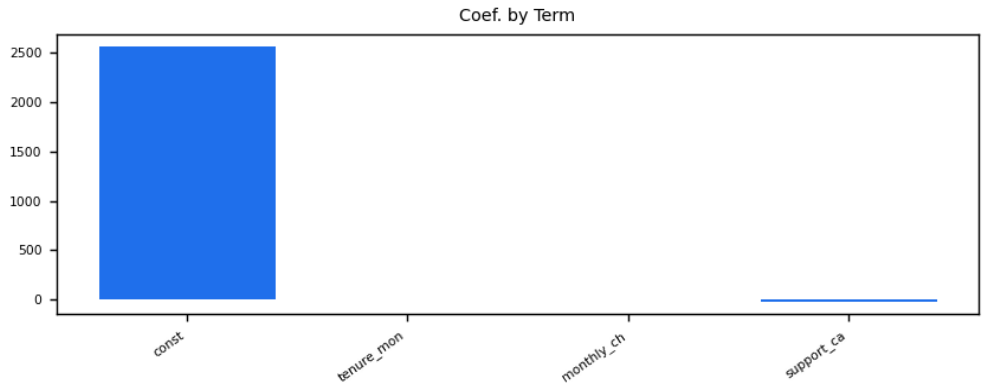
STEPS

1. Statistics → **Regression** → **Robust regression (Huber)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Robust regression (Huber M) — monetary

Term	Coef.	SE	z	p
const	2557	177.2	14.43	< .001
tenure_months	4.067	2.501	1.626	0.1039
monthly_charges	-0.4246	1.798	-0.2361	0.8134
support_calls	-20.13	36.98	-0.5443	0.5862



DETAILED INTERPRETATION

Down-weights outliers via iteratively reweighted least squares (Stata rreg). The lead result here reads Term = const; Coef. = 2557; SE = 177.2; z = 14.43. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.36 Score-code export (formula depot)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Export a fitted linear/logistic model as Python, SQL & JavaScript scoring code. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

STEPS

1. Statistics → **Regression** → **Score-code export (formula depot)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, model_type=linear`. 4. Click **Run** and read the tables/chart below.

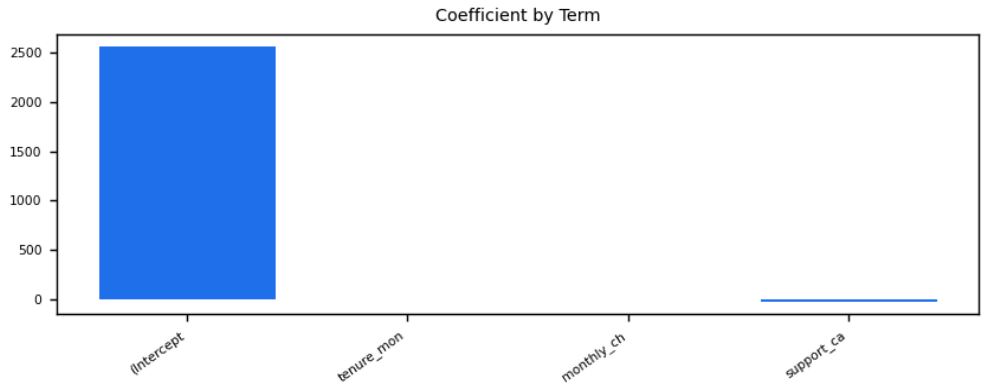
OUTPUT

Model coefficients — linear (monetary)

Term	Coefficient
(Intercept)	2554
tenure_months	3.913
monthly_charges	-0.3863
support_calls	-20.12

Score code (copy into your pipeline)

Language	Code
Python	score = 2554.498747 + 3.91…
SQL	(2554.498747 + 3.912727*te…
JavaScript	const score = 2554.498747 …



DETAILED INTERPRETATION

n = 800. Use the code below to score new rows in any system. The lead result here reads Term = (Intercept); Coefficient = 2554. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.37 Spatial regression (lag / error)

INPUT REQUIREMENT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Spatial-lag or spatial-error model with k-nearest-neighbour weights. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

STEPS

1. Statistics → **Regression** → **Spatial regression (lag / error)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, x_coord=monthly_charges, y_coord=tenure_months, model=lag, k=3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Requires geographic coordinates (latitude/longitude or projected x/y) per observation so spatial weights and neighbourhood structure can be built.

DETAILED INTERPRETATION

This technique has a specific data requirement: *Requires geographic coordinates (latitude/longitude or projected x/y) per observation so spatial weights and neighbourhood structure can be built*. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a regression method and supports the decision to identify and size the levers that move the outcome.

2.38 Specification-curve (multiverse)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Robustness: the focal effect across every covariate specification. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, monthly_charges, tenure_months, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	monthly_charges	tenure_months	support_calls
3197	52.94	68	1
1233	113.7	45	5
2501	35.51	49	1
1468	71.45	64	2
4195	29.16	42	1

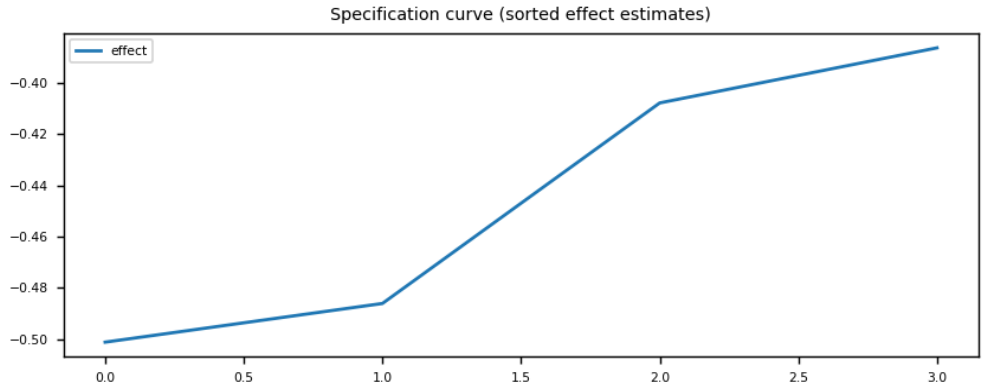
STEPS

1. Statistics → **Regression** → **Specification-curve (multiverse)**. 2. Select the demo dataset above. 3. Set parameters — `outcome=monetary, focal=monthly_charges, covariates=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Specification-curve (multiverse) — effect of monthly_charges on monetary

Metric	Value
Specifications	4
Median effect	-0.4469
Min / Max	-0.5012 / -0.3863
% significant (p<.05)	0%
% median sign	100%



DETAILED INTERPRETATION

4 specifications over 2 candidate covariates. Focal effect of 'monthly_charges' on 'monetary': median -0.4469 (range -0.5012 to -0.3863). 0/4 (0%) are significant at p<.05; 4/4 share the median sign. High agreement ⇒ a robust, non-cherry-picked effect. The lead result here reads Metric = Specifications; Value = 4. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.39 Spline regression

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Flexible nonlinear B-spline fit. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, monthly_charges. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	monthly_charges
3197	52.94
1233	113.7
2501	35.51
1468	71.45
4195	29.16

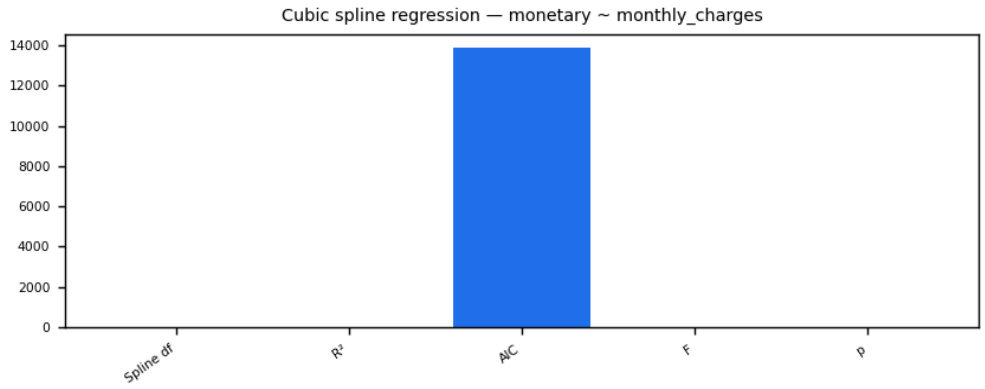
STEPS

1. Statistics → **Regression** → **Spline regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictor=monthly_charges, df=4`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Cubic spline regression — monetary ~ monthly_charges

Spline df	R²	AIC	F	p
4	0.0017	1.386e+04	0.3443	0.8481



DETAILED INTERPRETATION

Flexible nonlinear fit via a B-spline basis (captures curvature without a global polynomial). The lead result here reads Spline df = 4; R² = 0.0017; AIC = 1.386e+04; F = 0.3443. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.40 Stepwise regression (AIC)

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Forward selection by AIC. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1

monetary	tenure_months	monthly_charges	support_calls
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

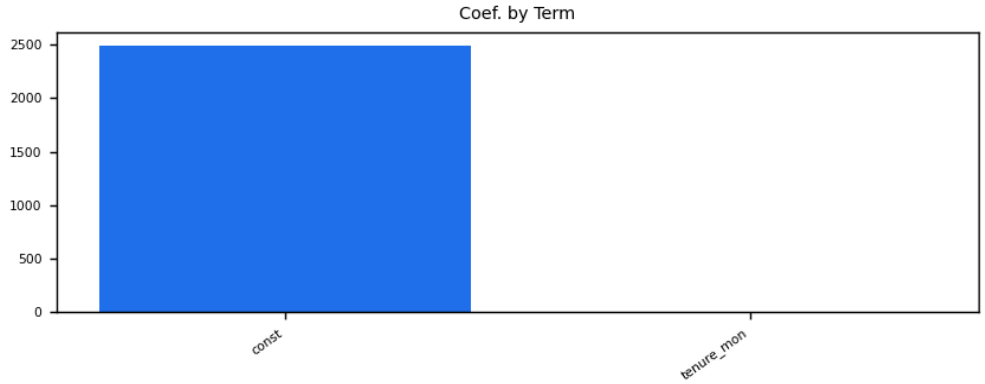
STEPS

1. Statistics → **Regression** → **Stepwise regression (AIC)**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Forward stepwise regression — monetary

Term	Coef.	p
const	2491	< .001
tenure_months	3.834	0.1077



DETAILED INTERPRETATION

Selected 1 predictor(s) by AIC. Final AIC = 13856.4, R^2 = 0.003. The lead result here reads Term = const; Coef. = 2491; p = < .001. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.41 Stochastic frontier analysis (SFA)

LIVE OUTPUT

Group: Regression · Dataset: firms with inputs (capital, labour, materials) and output (efficiency)

WHAT THIS DOES

Efficiency frontier with inefficiency term (normal-halfnormal) + technical efficiency. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to firms with inputs (capital, labour, materials) and output (efficiency), drawing on output, capital, labor, materials. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

output	capital	labor	materials
13.08	3.19	4.78	7.53
19.29	9.02	8.89	3.84
9.28	2.26	6.96	1.28
22.56	3.22	11.75	2.39
24.91	9.46	12.07	7.49

STEPS

1. Statistics → **Regression** → **Stochastic frontier analysis (SFA)**. 2. Select the demo dataset above. 3. Set parameters — `output=output, inputs=capital, labor, materials`. 4. Click **Run** and read the tables/chart below.

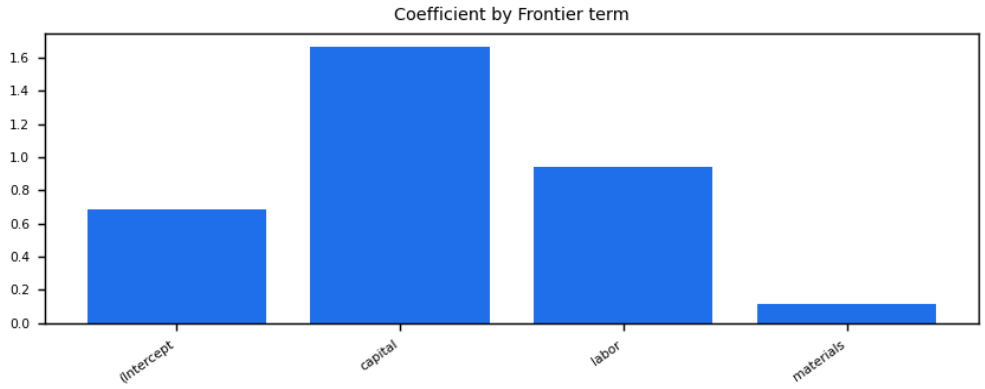
OUTPUT

Stochastic frontier (production) — output

Frontier term	Coefficient
(Intercept)	0.6816
capital	1.663
labor	0.9408
materials	0.1127

Variance decomposition & efficiency

σ_u	σ_v	λ = σ_u/σ_v	γ = σ_u²/σ²	Mean technical efficiency
0.001	4.459	0.0002	0	0.9992



DETAILED INTERPRETATION

Frontier coefficients (β). The composed error is v±u where u ≥ 0 is inefficiency. The lead result here reads Frontier term = (Intercept); Coefficient = 0.6816. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For firms with inputs (capital, labour, materials) and output (efficiency), this lets the analyst identify and size the levers that move the outcome.

2.42 Theil-Sen robust slope

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Median-slope robust regression. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

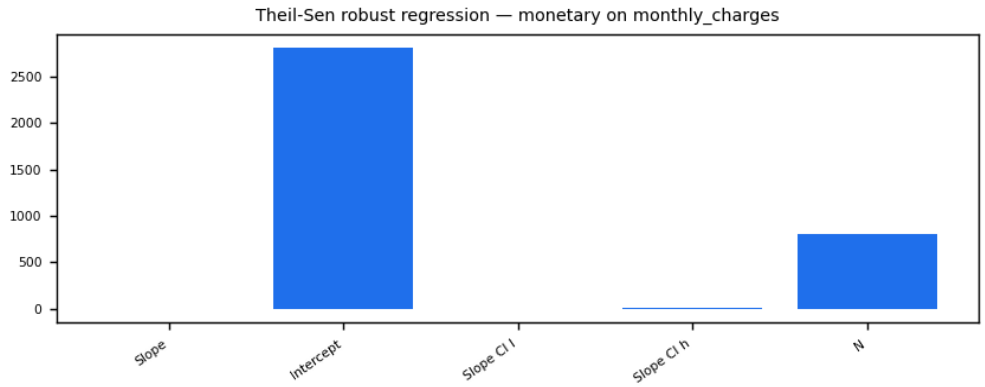
STEPS

1. Statistics → **Regression** → **Theil-Sen robust slope**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Theil-Sen robust regression — monetary on monthly_charges

Slope	Intercept	Slope CI low	Slope CI high	N
-0.4364	2820	-3.949	2.971	800



DETAILED INTERPRETATION

Median-of-slopes estimator; highly robust to outliers. The lead result here reads Slope = -0.4364; Intercept = 2820; Slope CI low = -3.949; Slope CI high = 2.971. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.43 Variance inflation factors

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Collinearity diagnostics. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

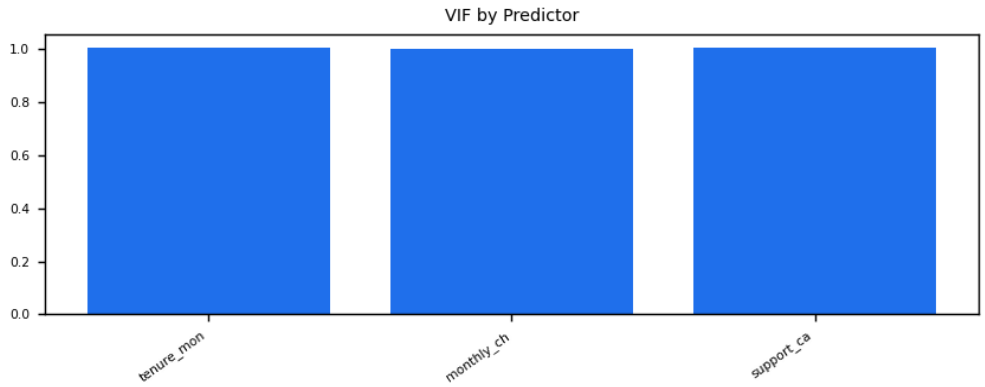
STEPS

1. Statistics → **Regression** → **Variance inflation factors**. 2. Select the demo dataset above. 3. Set parameters — `predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Variance inflation factors (collinearity)

Predictor	VIF	Assessment
tenure_months	1.006	ok
monthly_charges	1.002	ok
support_calls	1.006	ok



DETAILED INTERPRETATION

VIF > 5 signals collinearity; > 10 is serious. 1 = no correlation with other predictors. The lead result here reads Predictor = tenure_months; VIF = 1.006; Assessment = ok. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.44 Weighted least squares

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Regression with observation weights. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls, frequency. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls	frequency
3197	68	52.94	1	46
1233	45	113.7	5	33
2501	49	35.51	1	6
1468	64	71.45	2	6
4195	42	29.16	1	38

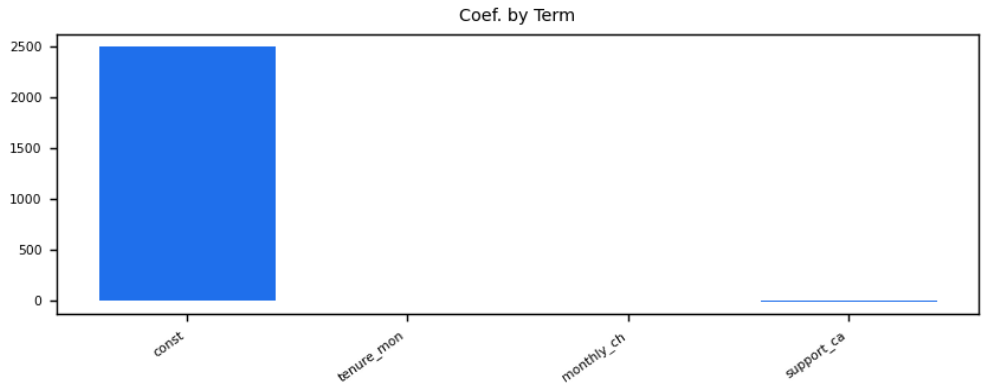
STEPS

1. Statistics → **Regression** → **Weighted least squares**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls, weight=frequency`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Weighted least squares — monetary

Term	Coef.	SE	p
const	2497	171.5	< .001
tenure_months	4.783	2.411	0.0476
monthly_charges	-0.73	1.731	0.6733
support_calls	-9.216	34.73	0.7908



DETAILED INTERPRETATION

R^2 = 0.005. Observations weighted by 'frequency'. The lead result here reads Term = const; Coef. = 2497; SE = 171.5; p = < .001. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.45 White's test

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

General heteroskedasticity test. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

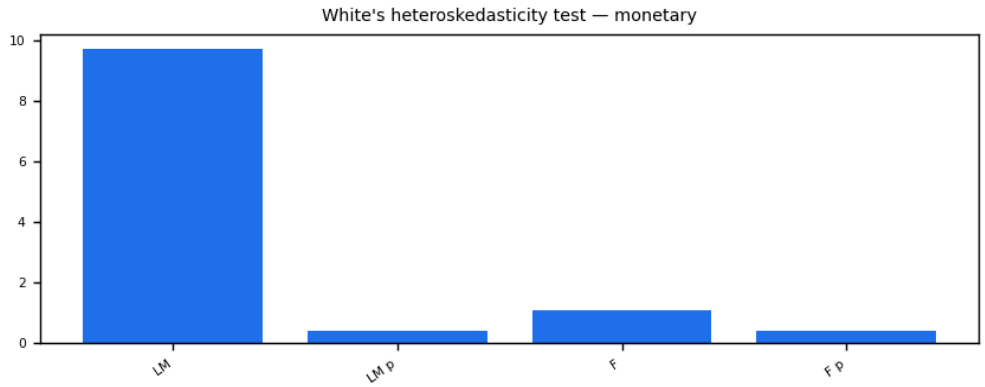
STEPS

1. Statistics → **Regression** → **White's test**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

White's heteroskedasticity test — monetary

LM	LM p	F	F p
9.72	0.3736	1.08	0.3753



DETAILED INTERPRETATION

General test for heteroskedasticity (includes nonlinear forms). The lead result here reads LM = 9.72; LM p = 0.3736; F = 1.08; F p = 0.3753. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

2.46 Zero-inflated Poisson

LIVE OUTPUT

Group: Regression · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Counts with excess zeros. It belongs to the **Regression** family, used to model and predict an outcome from one or more drivers. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on churn_bin, tenure_months, monthly_charges, support_calls. Analysts reach for it to identify and size the levers that move the outcome.

SAMPLE DATA INPUT

churn_bin	tenure_months	monthly_charges	support_calls
0	68	52.94	1
1	45	113.7	5
0	49	35.51	1
0	64	71.45	2
0	42	29.16	1

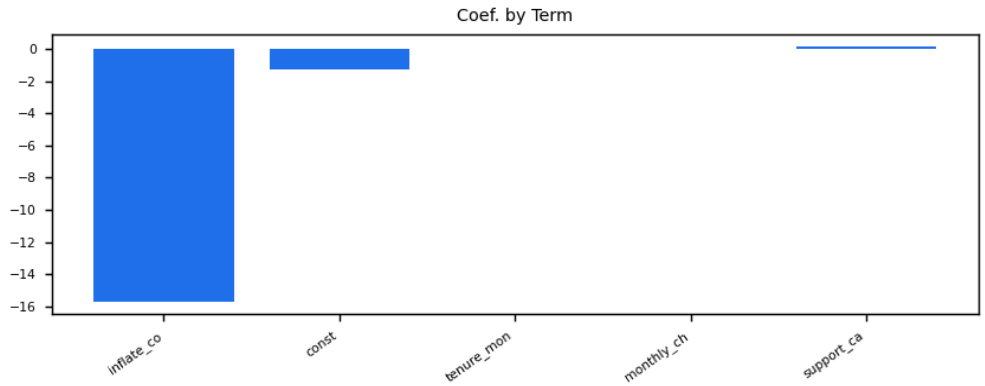
STEPS

1. Statistics → **Regression** → **Zero-inflated Poisson**. 2. Select the demo dataset above. 3. Set parameters — `dependent=churn_bin, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Zero-inflated Poisson — churn_bin

Term	Coef.	p
inflate_const	-15.69	0.9625
const	-1.253	< .001
tenure_months	-0.0262	< .001
monthly_charges	0.0076	0.0013
support_calls	0.1242	0.0063



DETAILED INTERPRETATION

AIC = 950.9. Models count data with excess zeros (two-part model). The lead result here reads Term = inflate_const; Coef. = -15.69; p = 0.9625. Read each coefficient as the expected change in the outcome per one-unit rise in that predictor with the others held fixed; judge significance from the p-value or confidence interval and overall fit from R2/AIC. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst identify and size the levers that move the outcome.

3. Multivariate, Scaling & Dimension Reduction

3.1 CHAID decision tree

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Chi-square automatic interaction detection tree with gains & classification. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on churn, tenure_months, monthly_charges, contract. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

churn	tenure_months	monthly_charges	contract
No	68	52.94	Month-to-month
Yes	45	113.7	One year
No	49	35.51	Month-to-month
No	64	71.45	Two year
No	42	29.16	Two year

STEPS

1. Statistics → **Classify** → **CHAID decision tree**. 2. Select the demo dataset above. 3. Set parameters — `target=churn, predictors=tenure_months, monthly_charges, contract, max_depth=3, min_node=30`. 4. Click **Run** and read the tables/chart below.

OUTPUT

CHAID tree — churn

Node	Rule	N	Majority class	Purity	Type
0	(root)	800	No	72.0%	split on tenure_months
1	· tenure_months = (0.999, …	204	Yes	51.0%	split on contract
2	· tenure_months = (18.0, 2…	301	No	73.4%	split on contract
3	· tenure_months = (47.0, 5…	295	No	86.4%	split on contract
4	· · contract = Month-to-mo…	168	Yes	57.1%	terminal
5	· · contract = Two year	36	No	77.8%	terminal
6	· · contract = Month-to-mo…	167	No	64.1%	split on monthly_charges
7	· · contract = One year / …	134	No	85.1%	terminal
8	· · · monthly_charges = (1…	39	Yes	53.8%	terminal
9	· · · monthly_charges = (2…	128	No	69.5%	terminal
10	· · contract = Month-to-mo…	162	No	79.6%	split on monthly_charges
11	· · contract = One year / …	133	No	94.7%	terminal

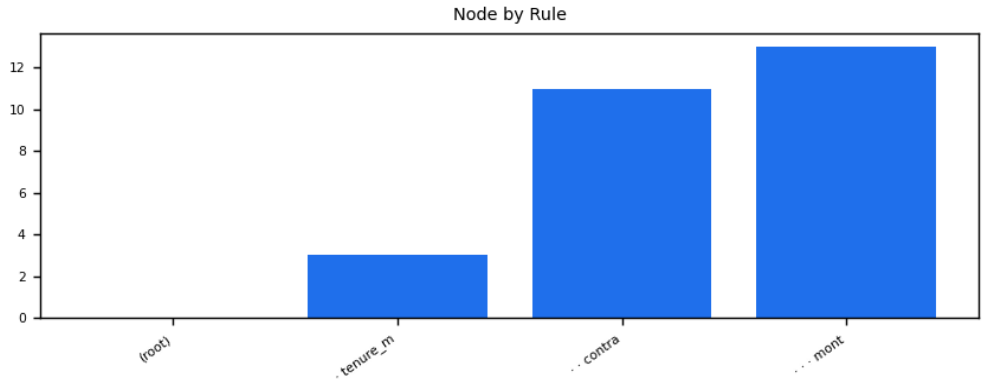
Gains by terminal node (target class = No)

Node	Rule	N	% No	Index
4	contract = Month-to-month …	168	42.9%	60
7	contract = One year / Two …	134	85.1%	118
11	contract = One year / Two …	133	94.7%	132
9	monthly_charges = (20.3289…	128	69.5%	97

Node	Rule	N	% No	Index
12	monthly_charges = (107.154…	98	71.4%	99
13	monthly_charges = (20.3289…	64	92.2%	128
8	monthly_charges = (107.154…	39	46.2%	64
5	contract = Two year	36	77.8%	108

Classification table (accuracy = 75.4%)

Actual \ Predicted	No	Yes
No	486	90
Yes	107	117



DETAILED INTERPRETATION

n = 800. Splits chosen by smallest Bonferroni-adjusted chi-square p-value; predictor categories merged when their split is non-significant. The lead result here reads Node = 0; Rule = (root); N = 800; Majority class = No. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.2 DBSCAN clustering

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Density-based clustering with automatic noise detection. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Classify** → **DBSCAN clustering**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, eps=5, min_samples=5`. 4. Click **Run** and read the tables/chart below.

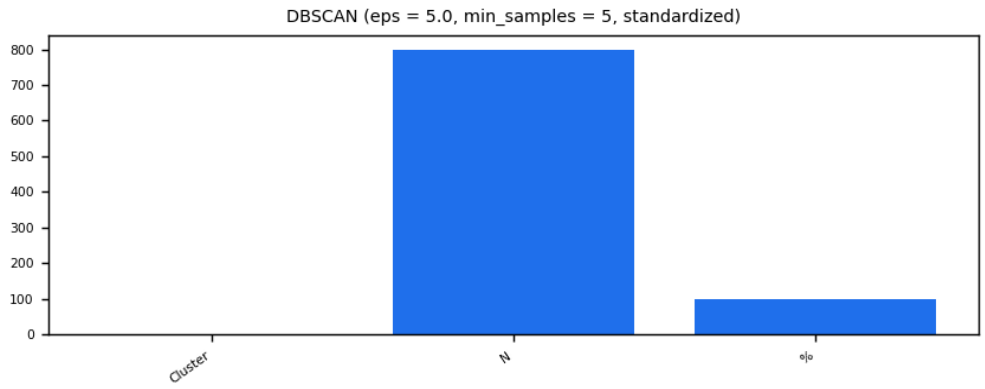
OUTPUT

DBSCAN (eps = 5.0, min_samples = 5, standardized)

Cluster	N	%
0	800	100

Cluster means (original units)

Cluster	tenure_months	monthly_charges	support_calls
0	37.1	69.09	1.954



DETAILED INTERPRETATION

0 noise points; 1 clusters found. The lead result here reads Cluster = 0; N = 800; % = 100. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.3 Discriminant analysis (LDA)

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Linear discriminant functions with cross-validated classification. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls, contract. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls	contract
68	52.94	1	Month-to-month
45	113.7	5	One year
49	35.51	1	Month-to-month
64	71.45	2	Two year
42	29.16	1	Two year

STEPS

1. Statistics → **Classify** → **Discriminant analysis (LDA)**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Discriminant function coefficients

Variable	→ Month-to-month	→ One year	→ Two year
tenure_months	-0.0007	-0.0016	0.0039
monthly_charges	-0.001	0.0014	0.0011
support_calls	-0.0129	0.022	0.0097

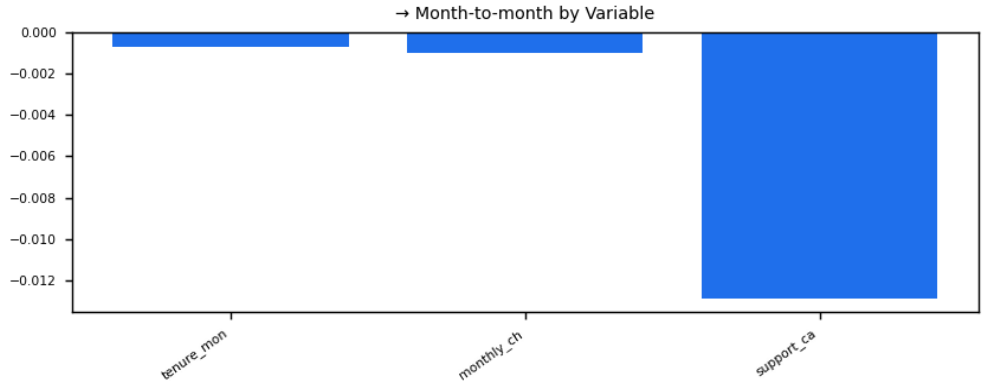
Group centroids

contract	tenure_months	monthly_charges	support_calls
Month-to-month	36.78	68.26	1.926
One year	36.43	70.33	1.995

contract	tenure_months	monthly_charges	support_calls
Two year	38.77	69.91	1.982

Classification (5-fold cross-validated)

Observed \ Predicted	Month-to-month
Month-to-month	447
One year	190
Two year	163



DETAILED INTERPRETATION

The lead result here reads Variable = tenure_months; → Month-to-month = -0.0007; → One year = -0.0016; → Two year = 0.0039. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.4 Gaussian mixture clustering

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Probabilistic (soft) clustering. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

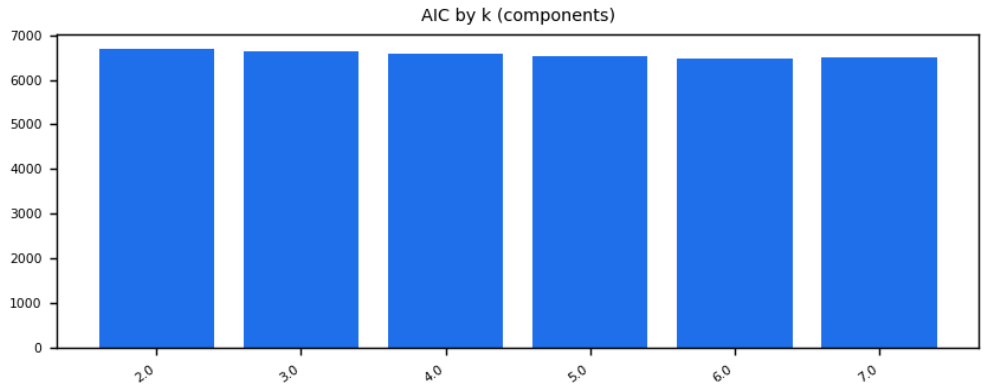
STEPS

1. Statistics → **Classify** → **Gaussian mixture clustering**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Gaussian mixture model (cluster selection)

k (components)	BIC	AIC
2	6774	6685
3	6764	6628
4	6777	6594
5	6764	6535
6	6751	6475
7	6812	6489



DETAILED INTERPRETATION

Best k = 6 by BIC. Soft, probabilistic clustering (elliptical clusters). The lead result here reads k (components) = 2; BIC = 6774; AIC = 6685. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.5 Hierarchical clustering

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Agglomerative clustering with agglomeration schedule. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Classify** → **Hierarchical clustering**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, k=3, method=ward`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Hierarchical clustering (ward linkage, cut at k = 3)

Cluster	N	%
1	386	48.25
2	171	21.38
3	243	30.38

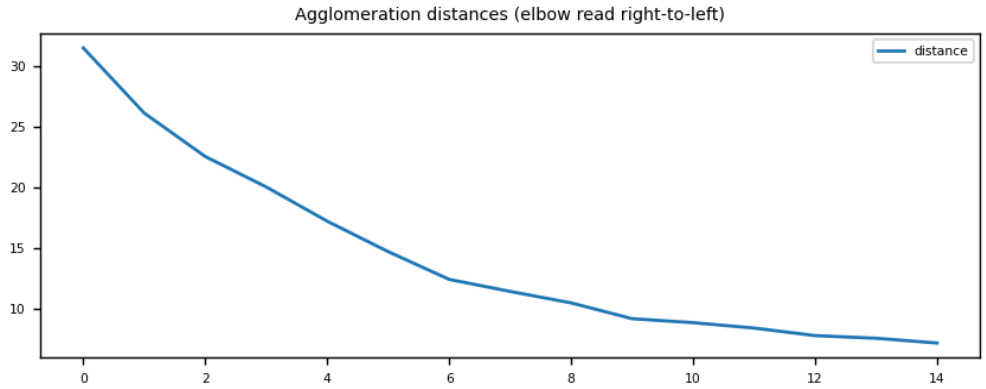
Cluster means (original units)

Cluster	tenure_months	monthly_charges	support_calls
1	50.75	57.53	1.585
2	34.18	91.74	3.532
3	17.49	71.51	1.428

Last merge distances (agglomeration schedule tail)

Merges from end	Distance
1	31.49
2	26.13

Merges from end	Distance
3	22.56
4	20.06
5	17.24
6	14.74
7	12.46
8	11.47
9	10.52
10	9.224
11	8.899
12	8.454



DETAILED INTERPRETATION

Standardized inputs; capped at 5,000 rows (random sample). The lead result here reads Cluster = 1; N = 386; % = 48.25. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.6 K-means clustering

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Cluster observations with center profiles and separation ANOVA. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Classify** → **K-means clustering**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, k=3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

K-means clustering (k = 3, standardized inputs)

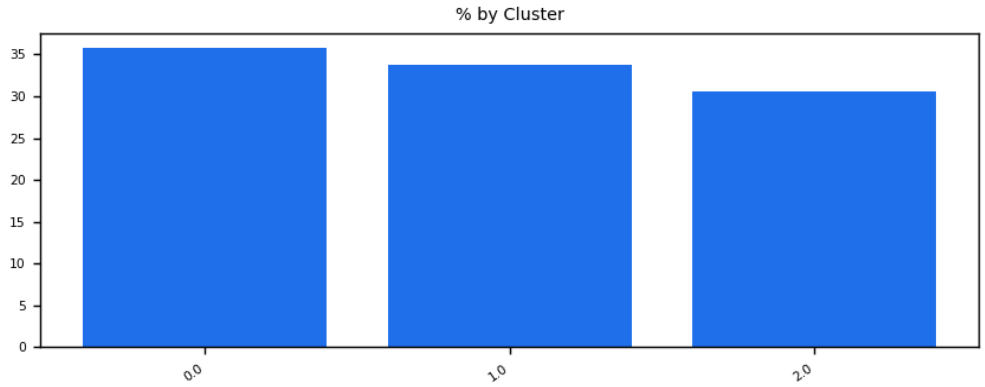
Cluster	N	%
0	286	35.75
1	270	33.75
2	244	30.5

Cluster centers (original units)

Cluster	tenure_months	monthly_charges	support_calls
0	16.02	69.93	1.308
1	50	43.62	2.104
2	47.55	96.3	2.545

ANOVA across clusters

Variable	F	p
tenure_months	548.7	< .001
monthly_charges	466.8	< .001
support_calls	61.92	< .001



DETAILED INTERPRETATION

Inertia (within-cluster SS) = 1395.76. The lead result here reads Cluster = 0; N = 286; % = 35.75. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.7 Latent class analysis (LCA)

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Person-centred typology from categorical/binary indicators (BIC, entropy). It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Classify** → **Latent class analysis (LCA)**. 2. Select the demo dataset above. 3. Set parameters — `indicators=tenure_months, monthly_charges, support_calls, classes=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Latent class analysis — 5 classes

Metric	Value
Log-likelihood	2.905e+06
AIC	-5.811e+06

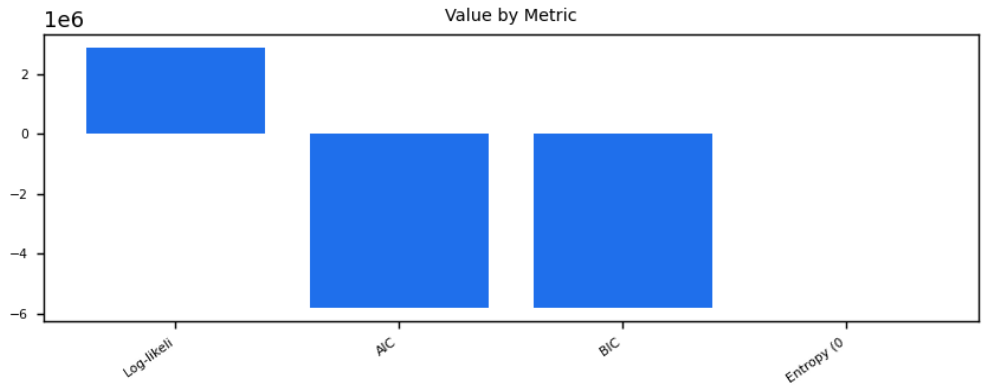
Metric	Value
BIC	-5.811e+06
Entropy (0–1, higher=clean…	0

Class sizes

Class	n	%
Class 1	800	100
Class 2	0	0
Class 3	0	0
Class 4	0	0
Class 5	0	0

Class profiles (means by indicator)

Class	tenure_months	monthly_charges	support_calls
Class 1	37.1	69.09	1.954
Class 2			
Class 3			
Class 4			
Class 5			



DETAILED INTERPRETATION

Compare BIC across class solutions (lower is better); entropy > 0.8 indicates well-separated classes. The lead result here reads Metric = Log-likelihood; Value = 2.905e+06. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.8 Latent profile analysis (LPA)

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Person-centred profiles from continuous indicators (BIC, entropy). It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Classify** → **Latent profile analysis (LPA)**. 2. Select the demo dataset above. 3. Set parameters — `indicators=tenure_months, monthly_charges, support_calls, classes=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Latent profile analysis — 5 classes

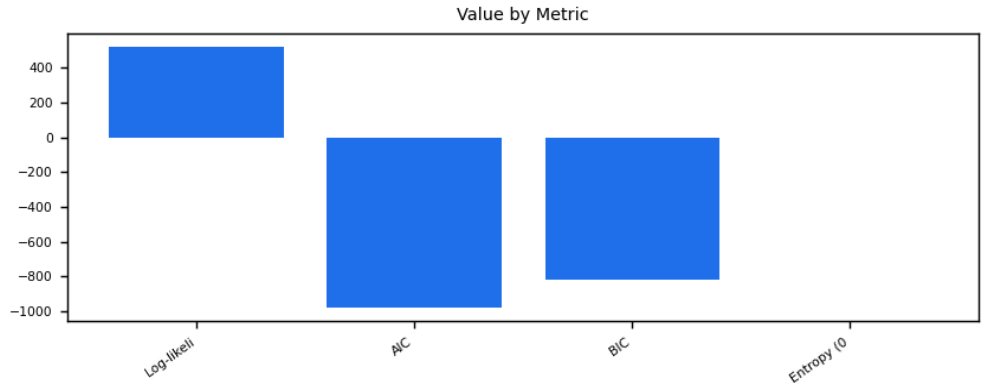
Metric	Value
Log-likelihood	523
AIC	-977.9
BIC	-818.7
Entropy (0–1, higher=clean…	1

Class sizes

Class	n	%
Class 1	114	14.2
Class 2	112	14
Class 3	127	15.9
Class 4	230	28.8
Class 5	217	27.1

Class profiles (means by indicator)

Class	tenure_months	monthly_charges	support_calls
Class 1	33.24	63.8	0
Class 2	39.29	68.08	4.509
Class 3	39.13	72.9	3
Class 4	37.01	67.36	2
Class 5	36.93	72	1



DETAILED INTERPRETATION

Compare BIC across class solutions (lower is better); entropy > 0.8 indicates well-separated classes. The lead result here reads Metric = Log-likelihood; Value = 523. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.9 Mahalanobis outliers

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Multivariate outlier detection. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

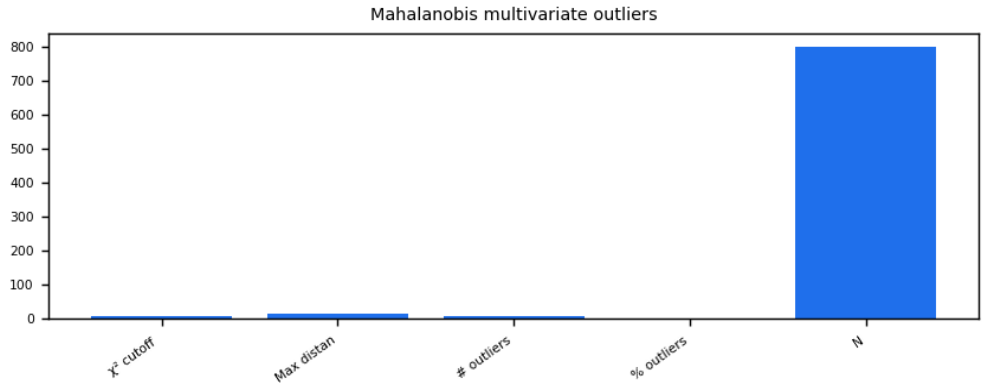
STEPS

1. Statistics → **Classify** → **Mahalanobis outliers**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Mahalanobis multivariate outliers

χ² cutoff (p=.975)	Max distance²	# outliers	% outliers	N
9.348	15.23	10	1.25	800



DETAILED INTERPRETATION

Rows with Mahalanobis distance² beyond the χ² cutoff are multivariate outliers. The lead result here reads χ² cutoff (p=.975) = 9.348; Max distance² = 15.23; # outliers = 10; % outliers = 1.25. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.10 Market-basket (association rules)

INPUT REQUIREMENT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Apriori association rules (support/confidence/lift) from transaction data. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, tenure_months. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

monthly_charges	tenure_months
52.94	68
113.7	45
35.51	49
71.45	64
29.16	42

STEPS

1. Statistics → **Classify** → **Market-basket (association rules)**. 2. Select the demo dataset above. 3. Set parameters — `transaction=monthly_charges, item=tenure_months, min_support=0.1, min_confidence=0.95`. 4. Click **Run** and read the tables/chart below.

OUTPUT

No itemsets at support ≥ 0.1 — lower the minimum support

DETAILED INTERPRETATION

This technique has a specific data requirement: *No itemsets at support ≥ 0.1 — lower the minimum support*. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a classify method and supports the decision to route or target cases based on predicted membership.

3.11 Quadratic discriminant analysis

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Curved-boundary classifier. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls, contract. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls	contract
68	52.94	1	Month-to-month
45	113.7	5	One year
49	35.51	1	Month-to-month
64	71.45	2	Two year
42	29.16	1	Two year

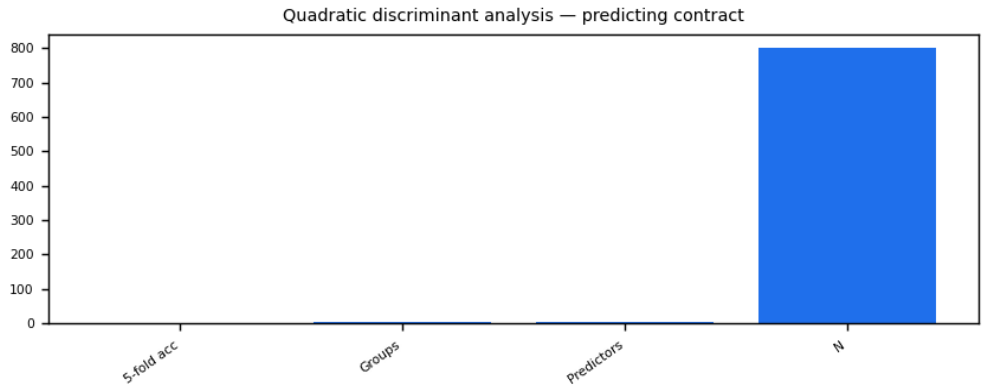
STEPS

1. Statistics → **Classify** → **Quadratic discriminant analysis**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Quadratic discriminant analysis — predicting contract

5-fold accuracy	Groups	Predictors	N
0.5575	3	3	800



DETAILED INTERPRETATION

QDA allows each class its own covariance (curved boundaries), unlike LDA. The lead result here reads 5-fold accuracy = 0.5575; Groups = 3; Predictors = 3; N = 800. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.12 Silhouette (optimal k)

LIVE OUTPUT

Group: Classify · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Choose the number of clusters. It belongs to the **Classify** family, used to group cases or predict class membership. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to route or target cases based on predicted membership.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

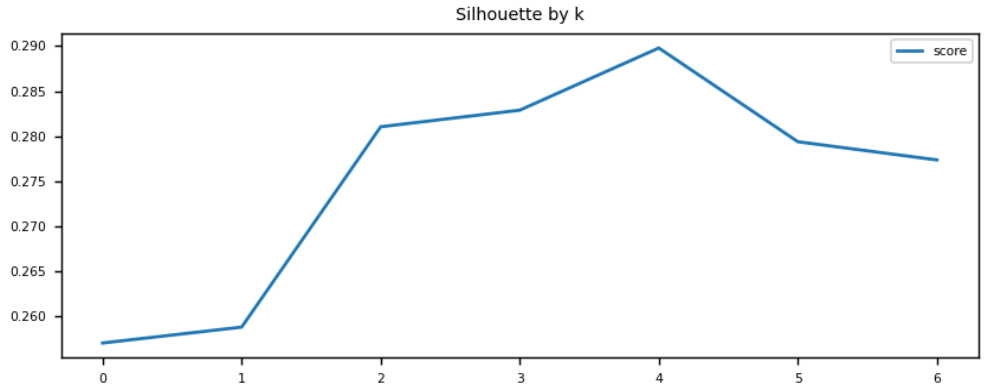
STEPS

1. Statistics → **Classify** → **Silhouette (optimal k)**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Silhouette analysis (optimal k)

k (clusters)	Silhouette score
2	0.257
3	0.2588
4	0.281
5	0.2829
6	0.2898
7	0.2794
8	0.2774



DETAILED INTERPRETATION

Best k = 6 (silhouette 0.290). Higher = better-separated clusters (max 1). The lead result here reads k (clusters) = 2; Silhouette score = 0.257. Judge the model by accuracy or AUC and by how cleanly the groups separate; the rules or clusters reveal the structure behind the classes. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst route or target cases based on predicted membership.

3.13 Canonical correlation

LIVE OUTPUT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Correlate two variable sets. It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

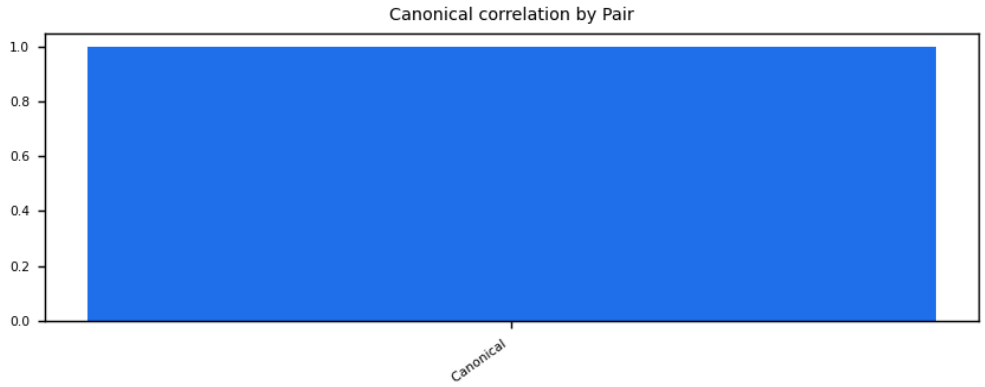
STEPS

1. Statistics → **Dimension reduction** → **Canonical correlation**. 2. Select the demo dataset above. 3. Set parameters — `set1=peou1, peou2, peou3, set2=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Canonical correlation analysis

Pair	Canonical correlation
Canonical pair 1	1
Canonical pair 2	1
Canonical pair 3	1



DETAILED INTERPRETATION

Correlations between linear combinations of the two variable sets (0-1). The lead result here reads Pair = Canonical pair 1; Canonical correlation = 1. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst visualise and compress high-dimensional data.

3.14 Confirmatory factor analysis (CFA)

LIVE OUTPUT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Test a measurement model with fit indices (CFI/TLI/RMSEA), loadings, AVE & CR. It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on the selected variables. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

STEPS

1. Statistics → **Dimension reduction** → **Confirmatory factor analysis (CFA)**. 2. Select the demo dataset above. 3. Set parameters — `model=PEOU =~ peou1 + peou2 + peou3 PU =~ pu1 + pu2 + pu3 INT =~ int1 + int2 + int3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

CFA model fit

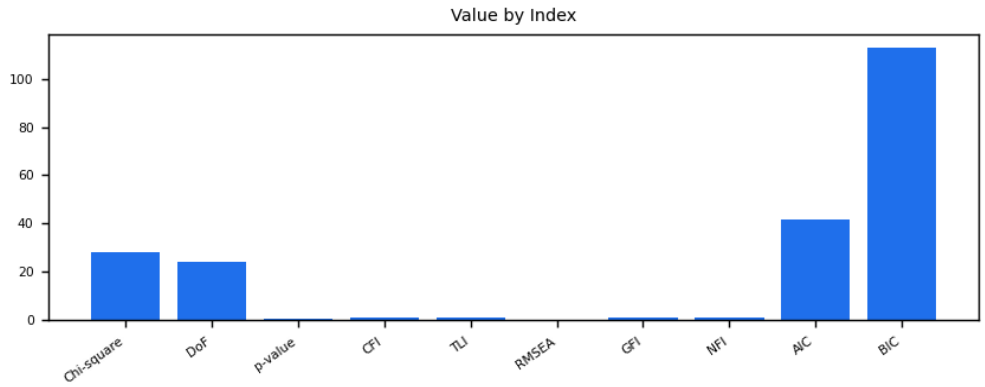
Index	Value
Chi-square	28.01
DoF	24
p-value	0.2596
CFI	0.996
TLI	0.994
RMSEA	0.0276
GFI	0.9728
NFI	0.9728
AIC	41.75
BIC	113

Standardised factor loadings

Factor	Indicator	Loading
PEOU	peou1	1
PEOU	peou2	0.927
PEOU	peou3	0.846
PU	pu1	1
PU	pu2	0.978
PU	pu3	0.844
INT	int1	1
INT	int2	0.93
INT	int3	0.95

Convergent validity (AVE & composite reliability)

Factor	AVE	CR	Verdict
PEOU	0.858	0.948	ok
PU	0.89	0.96	ok
INT	0.923	0.973	ok



DETAILED INTERPRETATION

CFI/TLI ≥ 0.95 and RMSEA ≤ 0.06 indicate good fit (≥0.90 / ≤0.08 acceptable). Good fit. The lead result here reads Index = Chi-square; Value = 28.01. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst visualise and compress high-dimensional data.

3.15 Correspondence analysis

LIVE OUTPUT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Map categorical associations. It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2
3	4
3	3
4	4
5	4
2	1

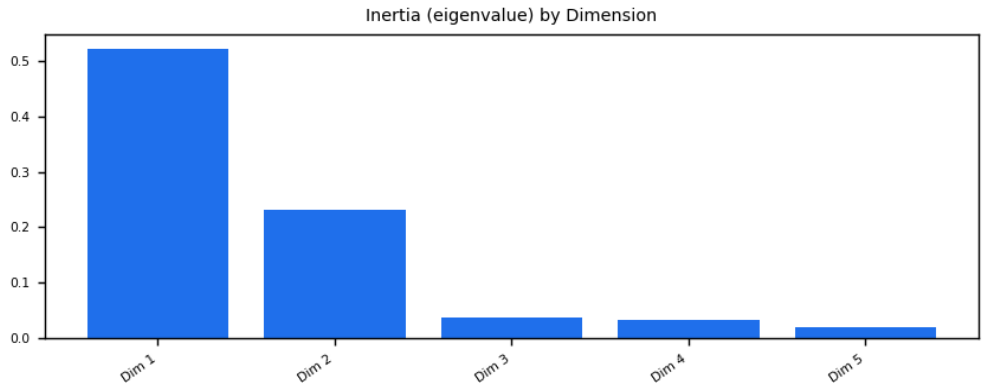
STEPS

1. Statistics → **Dimension reduction** → **Correspondence analysis**. 2. Select the demo dataset above. 3. Set parameters — `row=peou1, column=peou2`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Correspondence analysis — peou1 × peou2

Dimension	Inertia (eigenvalue)	% of inertia
Dim 1	0.5228	62.07
Dim 2	0.2317	27.51
Dim 3	0.0371	4.4
Dim 4	0.0318	3.778
Dim 5	0.018	2.143



DETAILED INTERPRETATION

Total inertia = 0.8423. Reveals associations between row and column categories. The lead result here reads Dimension = Dim 1; Inertia (eigenvalue) = 0.5228; % of inertia = 62.07. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst visualise and compress high-dimensional data.

3.16 Embedding (UMAP / t-SNE / PCA)

INPUT REQUIREMENT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

2-D map of high-dimensional data for visualization & cluster discovery. It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3, pu1, pu2, pu3. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2	peou3	pu1	pu2	pu3	int1	int2
3	4	5	4	4	3	4	4
3	3	3	3	4	3	3	4
4	4	5	5	5	4	5	5
5	4	5	3	4	3	5	5
2	1	1	3	4	2	2	3

STEPS

1. Statistics → **Dimension reduction** → **Embedding (UMAP / t-SNE / PCA)**. 2. Select the demo dataset above. 3. Set parameters — `features=peou1, peou2, peou3, pu1, pu2, pu3, int1, int2, int3, method=umap, label=`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Columns not in dataset: ['']

DETAILED INTERPRETATION

This technique has a specific data requirement: *Columns not in dataset: ['']*. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure.. Supplied with a suitable dataset for a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, it returns the standard output for a dimension reduction method and supports the decision to visualise and compress high-dimensional data.

3.17 Factor analysis / PCA

LIVE OUTPUT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Principal components with varimax rotation, KMO, Bartlett, scree. It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

STEPS

1. Statistics → **Dimension reduction** → **Factor analysis / PCA**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3, n_components=5, rotation=varimax`. 4. Click **Run** and read the tables/chart below.

OUTPUT

KMO and Bartlett's test

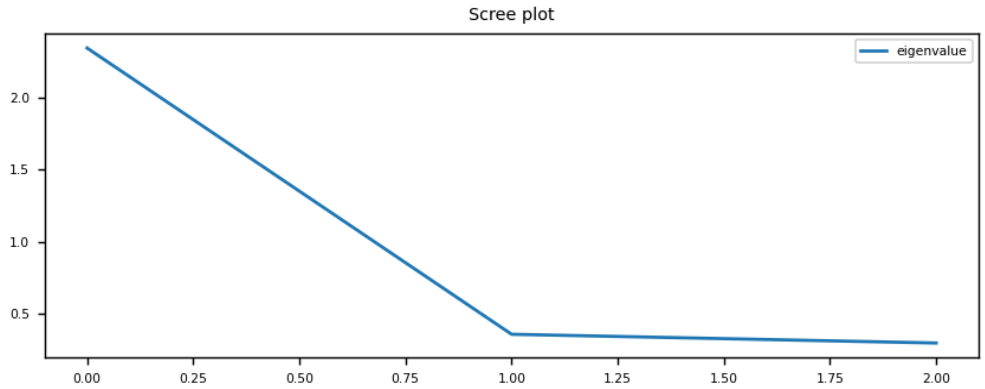
Measure	Value
Kaiser-Meyer-Olkin (KMO)	0.734
Bartlett's chi-square	303.3
df	3
p	< .001

Total variance explained

Component	Eigenvalue	% of variance	Cumulative %
1	2.348	78.26	78.26
2	0.3565	11.88	90.15
3	0.2956	9.854	100

Component loadings (varimax rotation, |λ| < .2 suppressed)

Variable	C1	C2	C3	Communality
peou1	-0.3424	0.3195	-0.8835	1
peou2	-0.8867	0.3135	-0.3399	1
peou3	-0.3024	0.9028	-0.3059	1



DETAILED INTERPRETATION

KMO ≥ 0.6 and significant Bartlett support factorability. The lead result here reads Measure = Kaiser-Meyer-Olkin (KM; Value = 0.734. The reduced dimensions retain most of the information in far fewer variables; use explained

variance and the projection to interpret and visualise the structure. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst visualise and compress high-dimensional data.

3.18 Functional PCA (curve data)

INPUT REQUIREMENT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Smooth curves and extract functional principal components (FDA). It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

STEPS

1. Statistics → **Dimension reduction** → **Functional PCA (curve data)**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Need at least 4 measurement points (columns)

DETAILED INTERPRETATION

This technique has a specific data requirement: *Need at least 4 measurement points (columns)*. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure.. Supplied with a suitable dataset for a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, it returns the standard output for a dimension reduction method and supports the decision to visualise and compress high-dimensional data.

3.19 ML factor analysis

LIVE OUTPUT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Latent common-factor model. It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

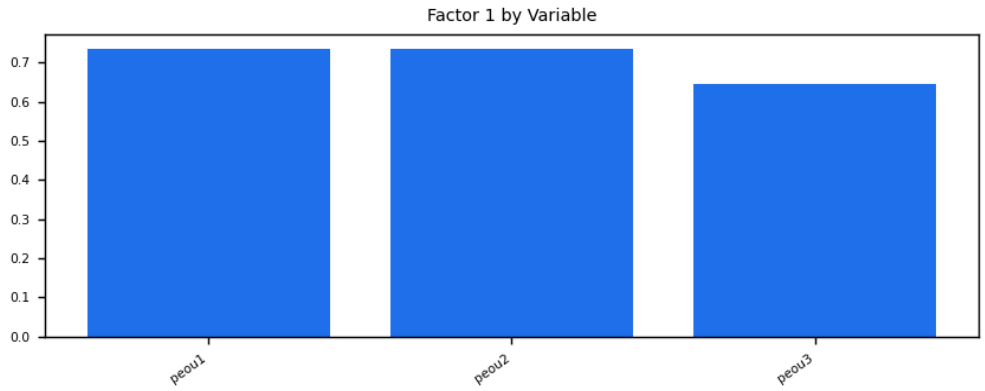
STEPS

1. Statistics → **Dimension reduction** → **ML factor analysis**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3, n_factors=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Maximum-likelihood factor analysis (varimax)

Variable	Factor 1	Factor 2
peou1	0.7338	0.4124
peou2	0.7354	0.3958
peou3	0.6434	0.4603



DETAILED INTERPRETATION

Loadings |λ| < 0.2 suppressed. ML factor analysis models common latent factors. The lead result here reads Variable = peou1; Factor 1 = 0.7338; Factor 2 = 0.4124. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst visualise and compress high-dimensional data.

3.20 Multidimensional scaling

LIVE OUTPUT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

2D distance-preserving map. It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

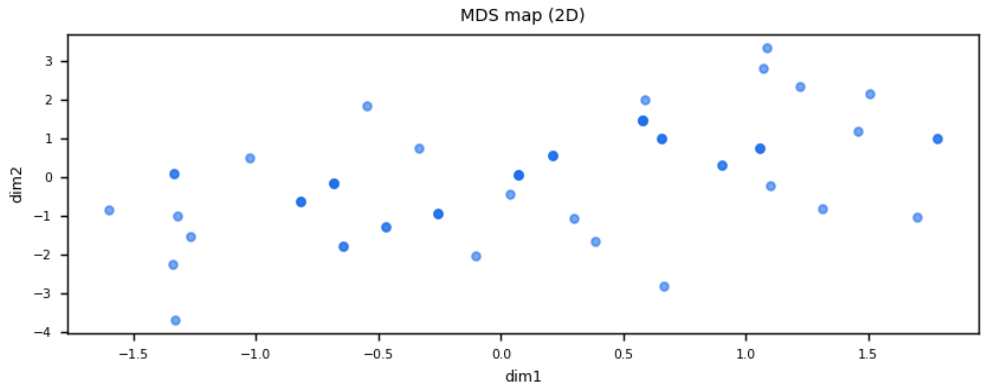
STEPS

1. Statistics → **Dimension reduction** → **Multidimensional scaling**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Multidimensional scaling (MDS)

Stress	Points	Dimensions
1561	220	2



DETAILED INTERPRETATION

Projects cases into 2D preserving distances; lower stress = better representation. The lead result here reads Stress = 1561; Points = 220; Dimensions = 2. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst visualise and compress high-dimensional data.

3.21 OVERALS (multiple-set canonical)

LIVE OUTPUT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Generalized canonical correlation across 2-3 sets of variables (optimal scaling). It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3, pu1, pu2, pu3. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2	peou3	pu1	pu2	pu3	int1	int2
3	4	5	4	4	3	4	4
3	3	3	3	4	3	3	4
4	4	5	5	5	4	5	5
5	4	5	3	4	3	5	5
2	1	1	3	4	2	2	3

STEPS

1. Statistics → **Dimension reduction** → **OVERALS (multiple-set canonical)**. 2. Select the demo dataset above. 3. Set parameters — `set1=peou1, peou2, peou3, set2=pu1, pu2, pu3, set3=int1, int2, int3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

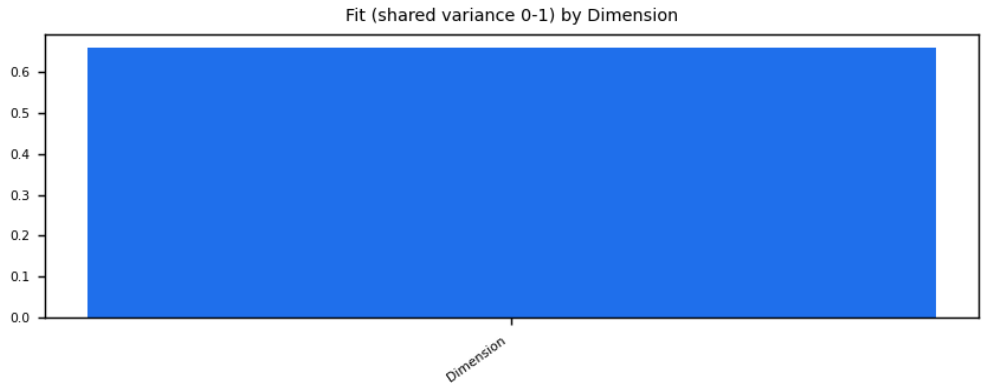
OVERALS — multiple-set canonical correlation (fit per dimension)

Dimension	Fit (shared variance 0-1)
Dimension 1	0.6596
Dimension 2	0.428
Dimension 3	0.377

Component loadings (correlation of each variable with the consensus dimension)

Variable	Loading
Set 1: peou1	0.7005
Set 1: peou2	0.7196
Set 1: peou3	0.6219
Set 2: pu1	0.7183
Set 2: pu2	0.761

Variable	Loading
Set 2: pu3	0.6204
Set 3: int1	0.7504
Set 3: int2	0.7127
Set 3: int3	0.7478



DETAILED INTERPRETATION

3 sets analysed. Fit near 1 means the sets share a strong common dimension (generalized canonical correlation via the MAXVAR criterion). The lead result here reads Dimension = Dimension 1; Fit (shared variance 0-1) = 0.6596. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst visualise and compress high-dimensional data.

3.22 Parallel analysis

LIVE OUTPUT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

How many factors to retain. It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

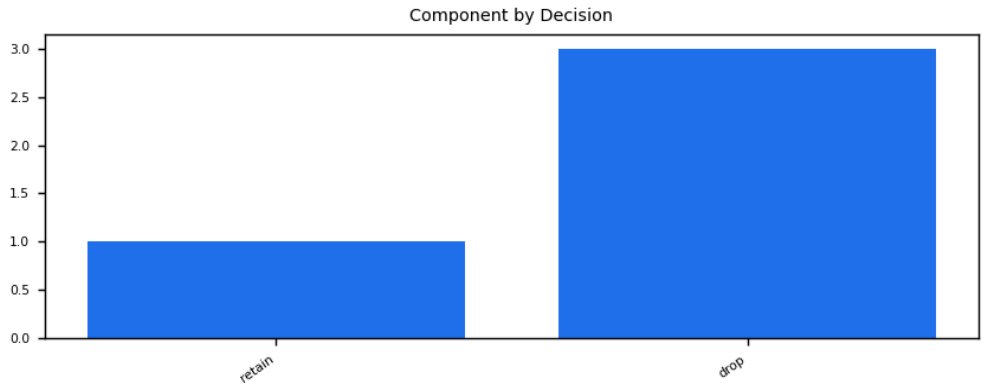
STEPS

1. Statistics → **Dimension reduction** → **Parallel analysis**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Parallel analysis (number of factors)

Component	Data eigenvalue	Random eigenvalue	Decision
1	2.348	1.101	retain
2	0.3565	1	drop
3	0.2956	0.8985	drop



DETAILED INTERPRETATION

Retain 1 factor(s): components whose eigenvalue exceeds the random-data benchmark. The lead result here reads Component = 1; Data eigenvalue = 2.348; Random eigenvalue = 1.101; Decision = retain. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst visualise and compress high-dimensional data.

3.23 Preference scaling / unfolding (PREFSCAL)

LIVE OUTPUT

Group: Dimension reduction · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Perceptual object map from preference/rating data via multidimensional unfolding. It belongs to the **Dimension reduction** family, used to compress many variables into a few interpretable dimensions. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3, pu1, pu2. Analysts reach for it to visualise and compress high-dimensional data.

SAMPLE DATA INPUT

peou1	peou2	peou3	pu1	pu2
3	4	5	4	4
3	3	3	3	4
4	4	5	5	5
5	4	5	3	4
2	1	1	3	4

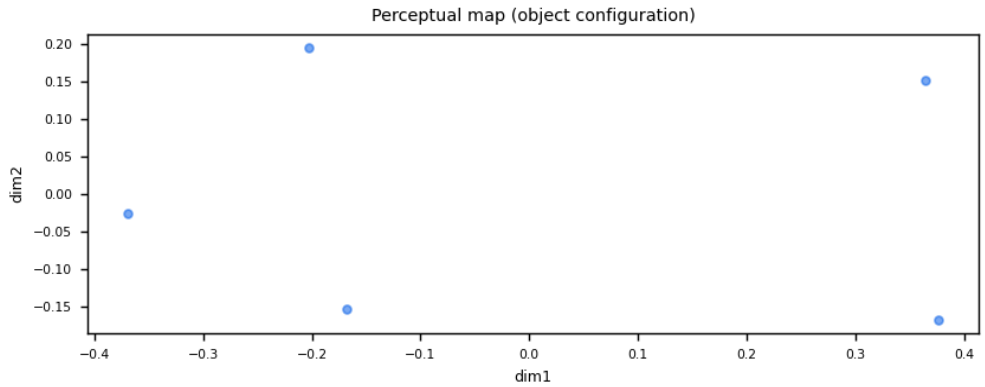
STEPS

1. Statistics → **Dimension reduction** → **Preference scaling / unfolding (PREFSCAL)**. 2. Select the demo dataset above. 3. Set parameters — `objects=peou1, peou2, peou3, pu1, pu2`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Preference scaling / unfolding — object configuration (2-D)

Object	Dimension 1	Dimension 2
peou1	-0.2032	0.195
peou2	-0.1683	-0.1536
peou3	-0.3691	-0.0252
pu1	0.3762	-0.1677
pu2	0.3644	0.1515



DETAILED INTERPRETATION

Objects placed in a common 2-D perceptual space from preference/rating data (stress = 0.04; lower is better). Objects close together are perceived similarly. The lead result here reads Object = peou1; Dimension 1 = -0.2032; Dimension 2 = 0.195. The reduced dimensions retain most of the information in far fewer variables; use explained variance and the projection to interpret and visualise the structure. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst visualise and compress high-dimensional data.

3.24 Categorical PCA (CATPCA)

LIVE OUTPUT

Group: Multivariate · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Nonlinear PCA on optimally-scaled categorical variables. It belongs to the **Multivariate** family, used to uncover latent structure among many variables at once. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to reduce many indicators to a few actionable constructs.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

STEPS

1. Statistics → **Multivariate** → **Categorical PCA (CATPCA)**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

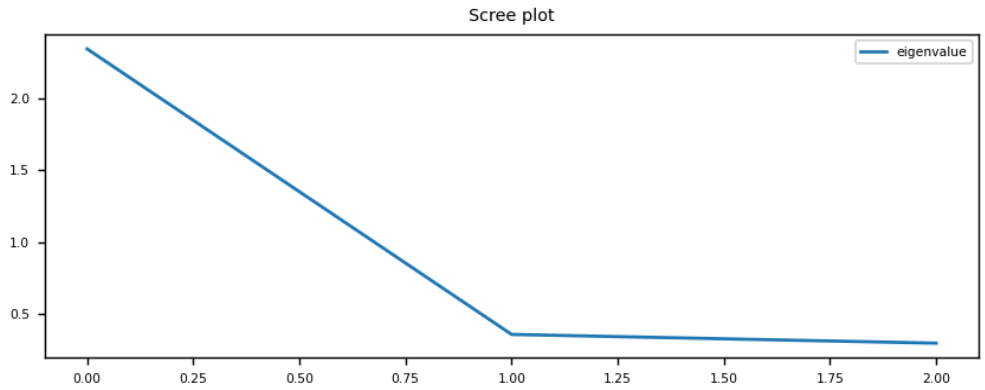
OUTPUT

CATPCA — variance explained

Component	Eigenvalue	% variance	Cumulative %
1	2.348	78.26	78.26
2	0.3565	11.88	90.15
3	0.2956	9.854	100

Component loadings

Variable	Dim 1	Dim 2	Dim 3
peou1	0.8928	-0.2109	0.398
peou2	0.8898	-0.2681	-0.3692
peou3	0.8712	0.49	-0.0307



DETAILED INTERPRETATION

Categorical variables ordinally quantified (optimal scaling). The lead result here reads Component = 1; Eigenvalue = 2.348; % variance = 78.26; Cumulative % = 78.26. Loadings and weights show how observed variables map onto latent dimensions - retain factors that explain substantial variance and remain interpretable. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst reduce many indicators to a few actionable constructs.

3.25 Data Envelopment Analysis (DEA)

LIVE OUTPUT

Group: Multivariate · Dataset: firms with inputs (capital, labour, materials) and output (efficiency)

WHAT THIS DOES

CCR input-oriented efficiency scores for decision-making units. It belongs to the **Multivariate** family, used to uncover latent structure among many variables at once. In this example it is applied to firms with inputs (capital, labour, materials) and output (efficiency), drawing on capital, labor, output. Analysts reach for it to reduce many indicators to a few actionable constructs.

SAMPLE DATA INPUT

capital	labor	output
3.19	4.78	13.08
9.02	8.89	19.29
2.26	6.96	9.28
3.22	11.75	22.56
9.46	12.07	24.91

STEPS

1. Statistics → **Multivariate** → **Data Envelopment Analysis (DEA)**. 2. Select the demo dataset above. 3. Set parameters — `inputs=capital, labor, outputs=output`. 4. Click **Run** and read the tables/chart below.

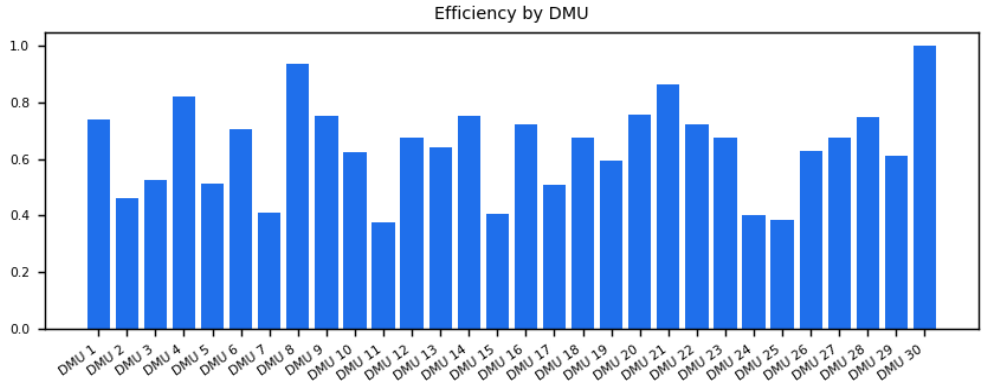
OUTPUT

Data Envelopment Analysis (CCR, input-oriented) — efficiency scores

DMU	Efficiency	
DMU 1	0.7425	
DMU 2	0.462	
DMU 3	0.5249	
DMU 4	0.8217	
DMU 5	0.5128	
DMU 6	0.7049	
DMU 7	0.4121	
DMU 8	0.9357	
DMU 9	0.7539	
DMU 10	0.6255	
DMU 11	0.3775	
DMU 12	0.6747	

Summary

Efficient DMUs	Mean efficiency	Min efficiency
3	0.6209	0.2691



DETAILED INTERPRETATION

n = 200 DMUs. Score = 1 means on the efficient frontier; < 1 means inputs could be reduced to that fraction and still produce the same outputs. (First 40 shown.). The lead result here reads DMU = DMU 1; Efficiency = 0.7425; = . Loadings and weights show how observed variables map onto latent dimensions - retain factors that explain substantial variance and remain interpretable. For firms with inputs (capital, labour, materials) and output (efficiency), this lets the analyst reduce many indicators to a few actionable constructs.

3.26 Key influencers

LIVE OUTPUT

Group: Multivariate · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Rank which factors most increase or decrease a target (Power BI-style). It belongs to the **Multivariate** family, used to uncover latent structure among many variables at once. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to reduce many indicators to a few actionable constructs.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

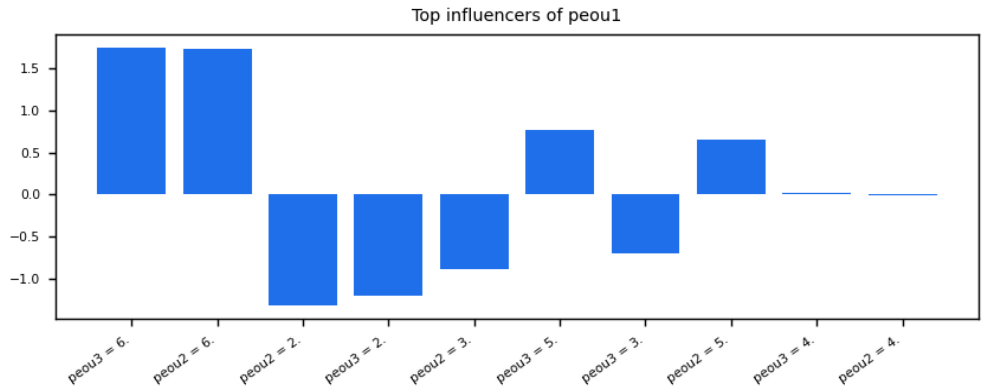
STEPS

1. Statistics → **Multivariate** → **Key influencers**. 2. Select the demo dataset above. 3. Set parameters — `target=peou1, factors=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Key influencers of peou1 (baseline = 3.91)

Rank	When…	Effect on peou1	N	Detail
1	peou3 = 6.0	1.753	18	segment mean 5.67
2	peou2 = 6.0	1.736	20	segment mean 5.65
3	peou2 = 2.0	-1.314	25	segment mean 2.6
4	peou3 = 2.0	-1.199	21	segment mean 2.71
5	peou2 = 3.0	-0.8914	45	segment mean 3.02
6	peou3 = 5.0	0.7682	44	segment mean 4.68
7	peou3 = 3.0	-0.6914	63	segment mean 3.22
8	peou2 = 5.0	0.6604	54	segment mean 4.57
9	peou3 = 4.0	0.0149	70	segment mean 3.93
10	peou2 = 4.0	-0.0095	73	segment mean 3.9



DETAILED INTERPRETATION

Effect = how much this segment shifts the target vs the overall average (numeric factors: change per +1 standard deviation). Ranked by |effect|. The lead result here reads Rank = 1; When… = peou3 = 6.0; Effect on peou1 = 1.753; N = 18. Loadings and weights show how observed variables map onto latent dimensions - retain factors that explain substantial variance and remain interpretable. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst reduce many indicators to a few actionable constructs.

3.27 Network analysis (SNA)

LIVE OUTPUT

Group: Multivariate · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Centrality + community detection on an edge list (co-authorship, SNA, etc.). It belongs to the **Multivariate** family, used to uncover latent structure among many variables at once. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to reduce many indicators to a few actionable constructs.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

STEPS

1. Statistics → **Multivariate** → **Network analysis (SNA)**. 2. Select the demo dataset above. 3. Set parameters — `source=peou1, target=peou2, weight=peou3, directed=no`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Network summary

Metric	Value
Nodes	7
Edges	18
Density	0.8571
Avg degree	5.14
Connected components	1
Communities (Louvain)	3
Modularity	0.2901

Most central nodes

Node	Degree	Betweenness	Eigenvector	PageRank	Community
5.0	1.167	0.3333	0.5009	—	3
3.0	1	0.1333	0.4236	—	1

Node	Degree	Betweenness	Eigenvector	PageRank	Community
4.0	1	0.1	0.4532	—	2
2.0	1	0.1333	0.4236	—	1
6.0	0.8333	0.0333	0.3199	—	3
7.0	0.6667	0	0.223	—	3
1.0	0.3333	0	0.181	—	1

DETAILED INTERPRETATION

Density = actual/possible edges; modularity > 0.3 indicates meaningful community structure. The lead result here reads Metric = Nodes; Value = 7. Loadings and weights show how observed variables map onto latent dimensions - retain factors that explain substantial variance and remain interpretable. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst reduce many indicators to a few actionable constructs.

3.28 Topic model (BERTopic-style)

LIVE OUTPUT

Group: Multivariate · Dataset: free-text customer product reviews

WHAT THIS DOES

Embedding + HDBSCAN + c-TF-IDF topics (auto number of topics, outlier detection). It belongs to the **Multivariate** family, used to uncover latent structure among many variables at once. In this example it is applied to free-text customer product reviews, drawing on review. Analysts reach for it to reduce many indicators to a few actionable constructs.

SAMPLE DATA INPUT

review
The dashboard is fast and …
Loved the AI analyst — it …
Setup was confusing and th…
Great value for money, doe…
The forecasting module sav…

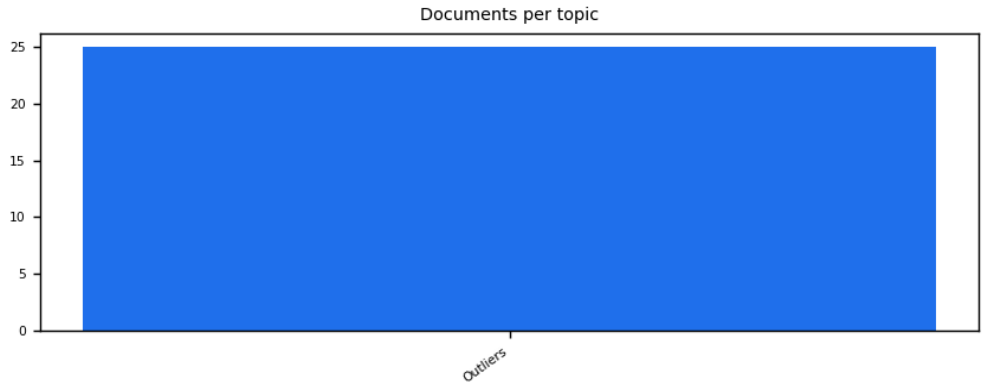
STEPS

1. Statistics → **Multivariate** → **Topic model (BERTopic-style)**. 2. Select the demo dataset above. 3. Set parameters — `text=review, min_cluster_size=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Topic model (embedding + HDBSCAN, BERTopic-style) — 0 topics from 25 docs

Topic	Top terms (c-TF-IDF)	Documents	Share
Outliers	great, wish, customer, acc…	25	100.0%



DETAILED INTERPRETATION

Documents are embedded (TF-IDF → UMAP), clustered by density (HDBSCAN, which auto-selects the number of topics and isolates outliers), and each topic is described by class-based TF-IDF terms. The lead result here reads Topic = Outliers; Top terms (c-TF-IDF) = great, wish, customer,; Documents = 25; Share = 100.0%. Loadings and weights show how observed variables map onto latent dimensions - retain factors that explain substantial variance

and remain interpretable. For free-text customer product reviews, this lets the analyst reduce many indicators to a few actionable constructs.

3.29 Topic modelling (LDA / NMF)

LIVE OUTPUT

Group: Multivariate · Dataset: free-text customer product reviews

WHAT THIS DOES

Discover latent topics and their top terms from a text column. It belongs to the **Multivariate** family, used to uncover latent structure among many variables at once. In this example it is applied to free-text customer product reviews, drawing on review. Analysts reach for it to reduce many indicators to a few actionable constructs.

SAMPLE DATA INPUT

review
The dashboard is fast and …
Loved the AI analyst — it …
Setup was confusing and th…
Great value for money, doe…
The forecasting module sav…

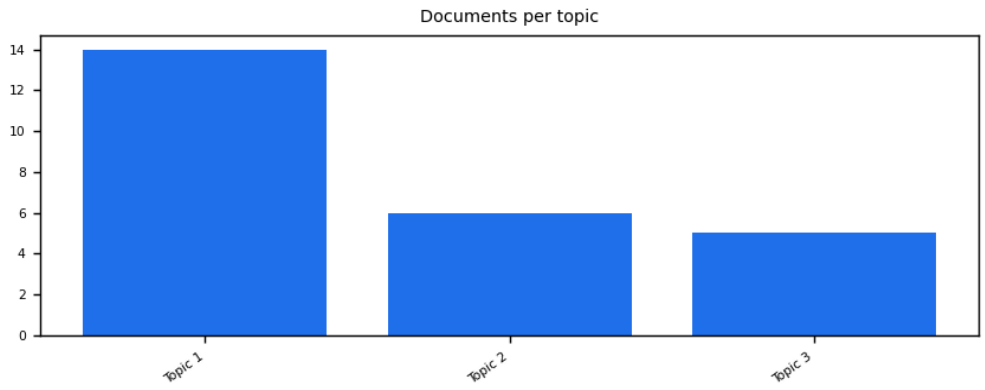
STEPS

1. Statistics → **Multivariate** → **Topic modelling (LDA / NMF)**. 2. Select the demo dataset above. 3. Set parameters — `text=review, num_topics=3, method=nmf`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Topic modelling (NMF) — 3 topics from 25 documents

Topic	Top terms	Documents	Share
Topic 1	great, fast, ai, large, co…	14	56.0%
Topic 2	customer, large, confusing…	6	24.0%
Topic 3	wish, reliable, accurate, …	5	20.0%



DETAILED INTERPRETATION

Each topic's most characteristic terms, with how many documents it dominates. LDA = probabilistic; NMF = parts-based on TF-IDF. The lead result here reads Topic = Topic 1; Top terms = great, fast, ai, large; Documents = 14; Share = 56.0%. Loadings and weights show how observed variables map onto latent dimensions - retain factors that explain substantial variance and remain interpretable. For free-text customer product reviews, this lets the analyst reduce many indicators to a few actionable constructs.

3.30 TwoStep cluster

LIVE OUTPUT

Group: Multivariate · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Auto-selects cluster count by BIC, then profiles the clusters. It belongs to the **Multivariate** family, used to uncover latent structure among many variables at once. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to reduce many indicators to a few actionable constructs.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

STEPS

1. Statistics → **Multivariate** → **TwoStep cluster**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3, max_k=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

TwoStep cluster — model selection

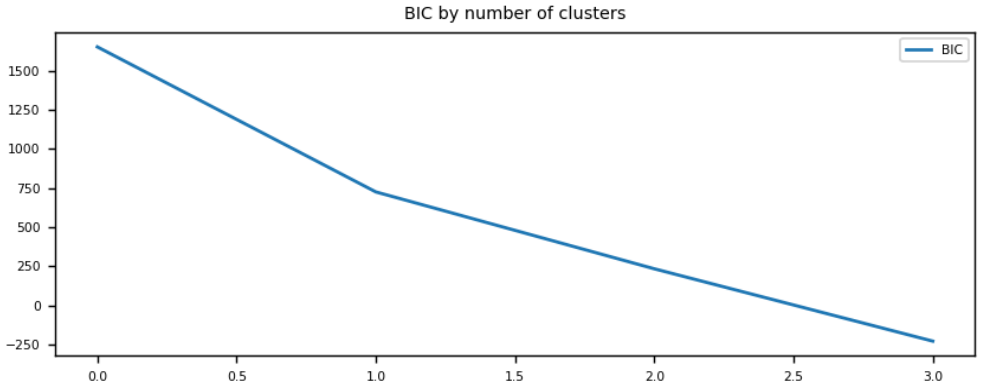
Chosen clusters	BIC (lower = better)
5	-229.2

Cluster sizes

Cluster	N	%
0	27	12.27
1	63	28.64
2	70	31.82
3	14	6.364
4	46	20.91

Cluster profile means (original units)

Cluster	peou1	peou2	peou3
0	4.741	4.222	5.222
1	3.222	3.46	3
2	3.929	4.1	4
3	3	3.214	1.929
4	4.63	4.739	4.739



DETAILED INTERPRETATION

Cluster count chosen automatically by the lowest BIC. The lead result here reads Chosen clusters = 5; BIC (lower = better) = -229.2. Loadings and weights show how observed variables map onto latent dimensions - retain factors

that explain substantial variance and remain interpretable. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst reduce many indicators to a few actionable constructs.

3.31 fuzzy-set QCA (fsQCA)

LIVE OUTPUT

Group: Multivariate · Dataset: firms scored on innovation, capability, funding and market reach vs. high performance

WHAT THIS DOES

Configurational analysis — calibration, necessity & sufficiency (consistency/coverage). It belongs to the **Multivariate** family, used to uncover latent structure among many variables at once. In this example it is applied to firms scored on innovation, capability, funding and market reach vs. high performance, drawing on high_performance, innovation, capability, funding, market_reach. Analysts reach for it to reduce many indicators to a few actionable constructs.

SAMPLE DATA INPUT

high_performance	innovation	capability	funding	market_reach
0.867	0.563	0.672	0.855	0.006
0.796	0.81	0.82	0.848	0.583
0.582	0.084	0.476	0.569	0.245
0.923	0.361	0.832	0.964	0.615
0.63	0.696	0.744	0.519	0.346

STEPS

1. Statistics → **Multivariate** → **fuzzy-set QCA (fsQCA)**. 2. Select the demo dataset above. 3. Set parameters — `outcome=high_performance, conditions=innovation, capability, funding, market_reach, calibrate=yes`. 4. Click **Run** and read the tables/chart below.

OUTPUT

fsQCA — analysis of necessity (high_performance)

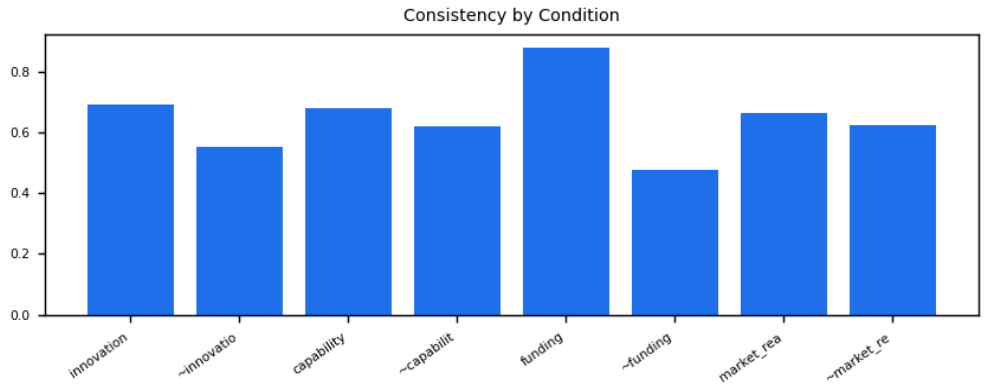
Condition	Consistency	Coverage	Verdict
innovation	0.6897	0.6768	
~innovation	0.5526	0.5633	
capability	0.6795	0.6547	
~capability	0.6181	0.6424	
funding	0.8788	0.8832	
~funding	0.4771	0.4747	
market_reach	0.6617	0.6911	
~market_reach	0.6238	0.5983	

fsQCA — analysis of sufficiency (truth table)

Configuration	Cases	Consistency	Raw coverage	
~innovation·~capability·fu…	8	0.931	0.0997	■ sufficient
innovation·~capability·fun…	8	0.9459	0.1114	■ sufficient
innovation·~capability·~fu…	8	0.2633	0.0287	
~innovation·capability·~fu…	7	0.3499	0.0304	
innovation·capability·~fun…	7	0.7928	0.0795	
innovation·capability·fund…	6	1	0.0806	■ sufficient
~innovation·capability·fun…	6	0.9224	0.0811	■ sufficient
~innovation·~capability·fu…	6	0.9865	0.0818	■ sufficient
~innovation·capability·~fu…	6	0.1335	0.0125	
innovation·capability·~fun…	5	0.826	0.0568	■ sufficient
~innovation·~capability·~f…	5	0.2049	0.0139	
innovation·capability·fund…	4	0.9548	0.0496	■ sufficient

fsQCA — solution

Sufficient configurations	Overall solution coverage
~innovation·~capability·fu…	0.8181



DETAILED INTERPRETATION

A condition is necessary when consistency ≥ 0.90 — it is (near-)always present when the outcome is present. The lead result here reads Condition = innovation; Consistency = 0.6897; Coverage = 0.6768; Verdict = . Loadings and weights show how observed variables map onto latent dimensions - retain factors that explain substantial variance and remain interpretable. For firms scored on innovation, capability, funding and market reach vs. high performance, this lets the analyst reduce many indicators to a few actionable constructs.

3.32 Bland-Altman agreement

LIVE OUTPUT

Group: Scale · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Agreement between two methods + plot. It belongs to the **Scale** family, used to assess whether survey items form a reliable, unidimensional construct. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2. Analysts reach for it to validate a measurement instrument before using its scores.

SAMPLE DATA INPUT

peou1	peou2
3	4
3	3
4	4
5	4
2	1

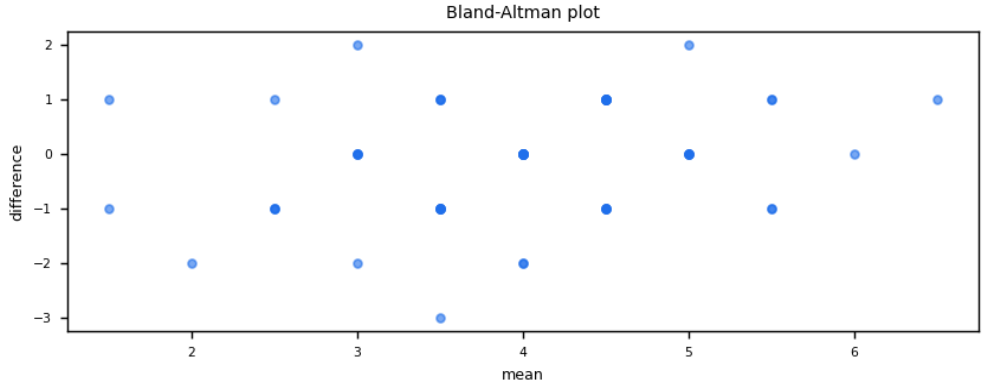
STEPS

1. Statistics → **Scale** → **Bland-Altman agreement**. 2. Select the demo dataset above. 3. Set parameters — `method1=peou1, method2=peou2`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Bland-Altman agreement — peou1 vs peou2

Bias (mean diff)	SD of diff	Lower LoA (−1.96SD)	Upper LoA (+1.96SD)	N
-0.0955	0.9484	-1.954	1.764	220



DETAILED INTERPRETATION

Assesses agreement between two measurement methods; ~95% of differences fall within the limits of agreement. The lead result here reads Bias (mean diff) = -0.0955; SD of diff = 0.9484; Lower LoA (–1.96SD) = -1.954; Upper LoA (+1.96SD) = 1.764. A reliability coefficient such as Cronbach's alpha at or above .70, together with clean loadings, indicates the items measure one coherent construct. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst validate a measurement instrument before using its scores.

3.33 Cohen's kappa (agreement)

LIVE OUTPUT

Group: Scale · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Inter-rater agreement for two raters. It belongs to the **Scale** family, used to assess whether survey items form a reliable, unidimensional construct. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2. Analysts reach for it to validate a measurement instrument before using its scores.

SAMPLE DATA INPUT

peou1	peou2
3	4
3	3
4	4
5	4
2	1

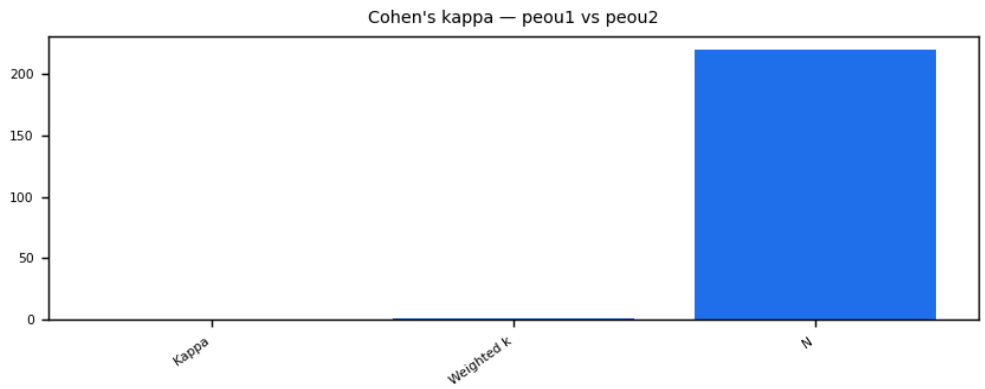
STEPS

1. Statistics → **Scale** → **Cohen's kappa (agreement)**. 2. Select the demo dataset above. 3. Set parameters — `rater1=peou1, rater2=peou2`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Cohen's kappa — peou1 vs peou2

Kappa	Weighted kappa	N	Agreement
0.1976	0.4753	220	poor



DETAILED INTERPRETATION

κ: <0.2 poor · 0.2-0.4 fair · 0.4-0.6 moderate · 0.6-0.8 substantial · >0.8 excellent. The lead result here reads Kappa = 0.1976; Weighted kappa = 0.4753; N = 220; Agreement = poor. A reliability coefficient such as Cronbach's alpha at or above .70, together with clean loadings, indicates the items measure one coherent construct. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst validate a measurement instrument before using its scores.

3.34 Fleiss' kappa

LIVE OUTPUT

Group: Scale · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Agreement among 3+ raters. It belongs to the **Scale** family, used to assess whether survey items form a reliable, unidimensional construct. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to validate a measurement instrument before using its scores.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

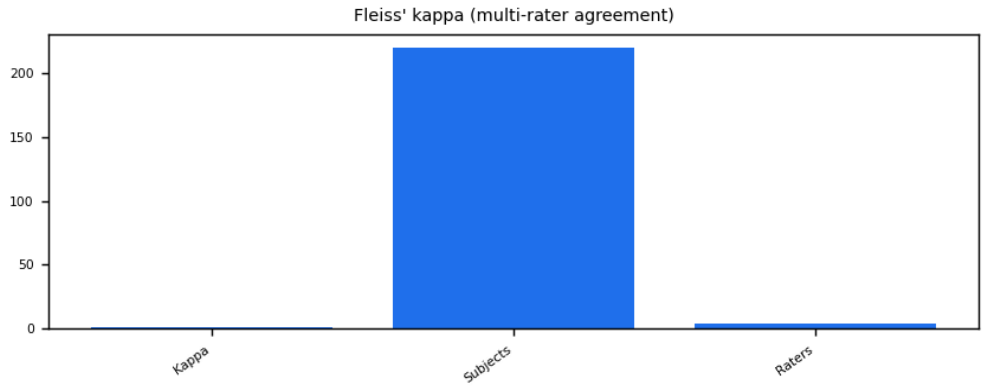
STEPS

1. Statistics → **Scale** → **Fleiss' kappa**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Fleiss' kappa (multi-rater agreement)

Kappa	Subjects	Raters	Agreement
0.2185	220	3	fair



DETAILED INTERPRETATION

Agreement among 3+ raters on categorical ratings. The lead result here reads Kappa = 0.2185; Subjects = 220; Raters = 3; Agreement = fair. A reliability coefficient such as Cronbach's alpha at or above .70, together with clean loadings, indicates the items measure one coherent construct. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst validate a measurement instrument before using its scores.

3.35 Intraclass correlation (ICC)

LIVE OUTPUT

Group: Scale · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

ICC(2,1) and ICC(2,k) reliability across raters/measures. It belongs to the **Scale** family, used to assess whether survey items form a reliable, unidimensional construct. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to validate a measurement instrument before using its scores.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5

peou1	peou2	peou3
3	3	3
4	4	5
5	4	5
2	1	1

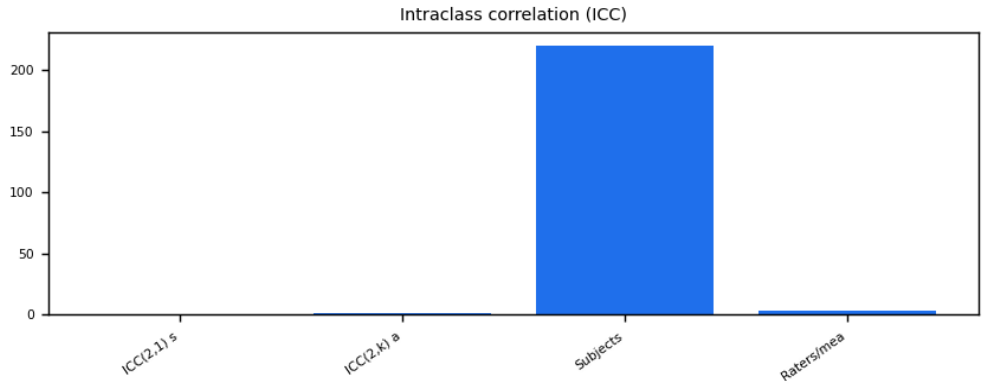
STEPS

1. Statistics → **Scale** → **Intraclass correlation (ICC)**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Intraclass correlation (ICC)

ICC(2,1) single	ICC(2,k) average	Subjects	Raters/measures
0.6719	0.86	220	3



DETAILED INTERPRETATION

ICC: <0.5 poor · 0.5-0.75 moderate · 0.75-0.9 good · >0.9 excellent reliability. The lead result here reads ICC(2,1) single = 0.6719; ICC(2,k) average = 0.86; Subjects = 220; Raters/measures = 3. A reliability coefficient such as Cronbach's alpha at or above .70, together with clean loadings, indicates the items measure one coherent construct. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst validate a measurement instrument before using its scores.

3.36 Item Response Theory (2PL)

LIVE OUTPUT

Group: Scale · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Item discrimination & difficulty with characteristic curves (psychometrics). It belongs to the **Scale** family, used to assess whether survey items form a reliable, unidimensional construct. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to validate a measurement instrument before using its scores.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

STEPS

1. Statistics → **Scale** → **Item Response Theory (2PL)**. 2. Select the demo dataset above. 3. Set parameters — `items=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

IRT (2PL) item parameters

Item	Discrimination (a)	Difficulty (b)	P(correct)	Verdict
peou1	—	—	1	failed
peou2	—	—	1	failed
peou3	—	—	1	failed

DETAILED INTERPRETATION

a = how sharply the item separates ability (≥0.65 good); b = ability level at 50% correct (higher = harder). Two-stage estimation; validate against a full MML model for high-stakes use. The lead result here reads Item = peou1; Discrimination (a) = —; Difficulty (b) = —; P(correct) = 1. A reliability coefficient such as Cronbach's alpha at or above .70, together with clean loadings, indicates the items measure one coherent construct. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst validate a measurement instrument before using its scores.

3.37 Kendall's W (concordance)

LIVE OUTPUT

Group: Scale · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Agreement among rankers. It belongs to the **Scale** family, used to assess whether survey items form a reliable, unidimensional construct. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to validate a measurement instrument before using its scores.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

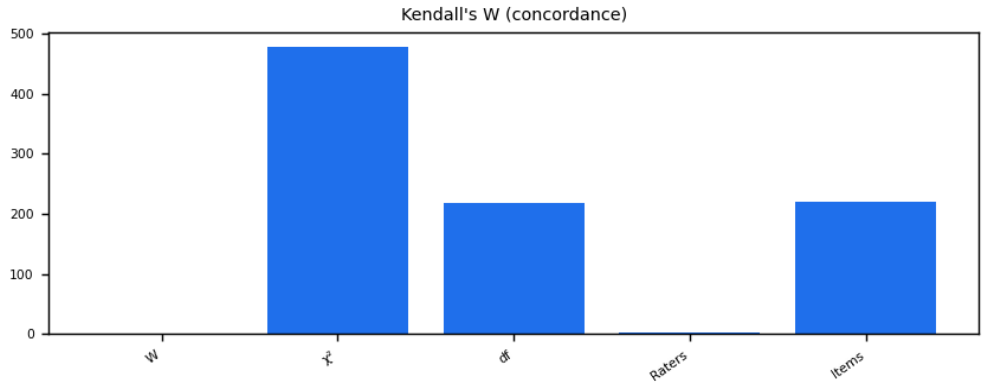
STEPS

1. Statistics → **Scale** → **Kendall's W (concordance)**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Kendall's W (concordance)

W	χ^2	df	p	Raters	Items
0.7289	478.9	219	< .001	3	220



DETAILED INTERPRETATION

Agreement among raters ranking the same items: 0 = none, 1 = perfect concordance. The lead result here reads W = 0.7289; χ^2 = 478.9; df = 219; p = < .001. A reliability coefficient such as Cronbach's alpha at or above .70, together with clean loadings, indicates the items measure one coherent construct. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst validate a measurement instrument before using its scores.

3.38 Lin's concordance (CCC)

LIVE OUTPUT

Group: Scale · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Agreement between two continuous measures. It belongs to the **Scale** family, used to assess whether survey items form a reliable, unidimensional construct. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, pu1. Analysts reach for it to validate a measurement instrument before using its scores.

SAMPLE DATA INPUT

peou1	pu1
3	4
3	3
4	5
5	3
2	3

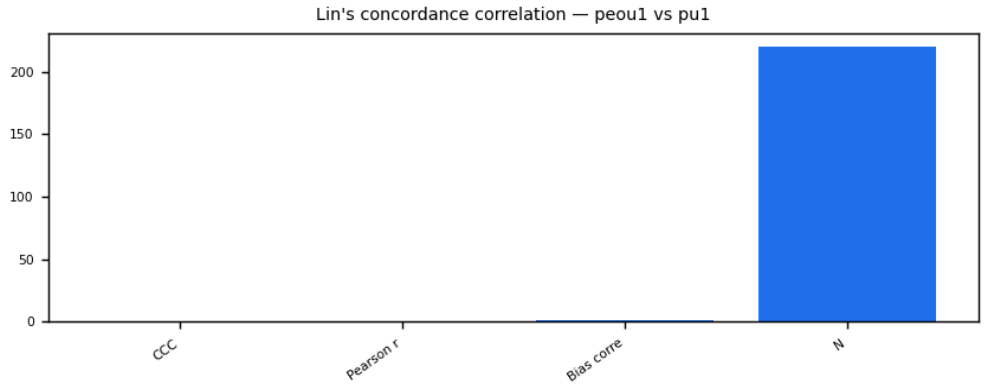
STEPS

1. Statistics → **Scale** → **Lin's concordance (CCC)**. 2. Select the demo dataset above. 3. Set parameters — `column1=peou1, column2=pu1`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Lin's concordance correlation — peou1 vs pu1

CCC	Pearson r	Bias correction (Cb)	N
0.3195	0.3208	0.9959	220



DETAILED INTERPRETATION

CCC measures agreement (accuracy + precision) between two methods; 1 = perfect. The lead result here reads CCC = 0.3195; Pearson r = 0.3208; Bias correction (Cb) = 0.9959; N = 220. A reliability coefficient such as Cronbach's alpha at or above .70, together with clean loadings, indicates the items measure one coherent construct. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst validate a measurement instrument before using its scores.

3.39 Reliability (Cronbach's α)

LIVE OUTPUT

Group: Scale · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Scale reliability with item-total statistics and α-if-deleted. It belongs to the **Scale** family, used to assess whether survey items form a reliable, unidimensional construct. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, peou2, peou3. Analysts reach for it to validate a measurement instrument before using its scores.

SAMPLE DATA INPUT

peou1	peou2	peou3
3	4	5
3	3	3
4	4	5
5	4	5
2	1	1

STEPS

1. Statistics → **Scale** → **Reliability (Cronbach's α)**. 2. Select the demo dataset above. 3. Set parameters — `columns=peou1, peou2, peou3`. 4. Click **Run** and read the tables/chart below.

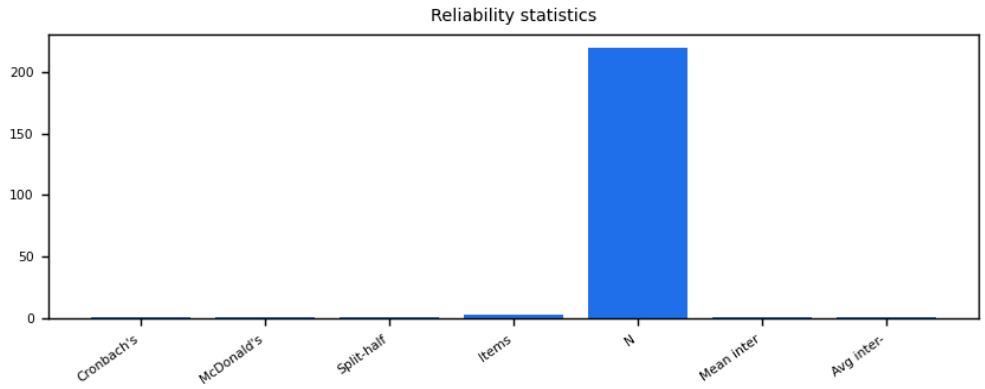
OUTPUT

Reliability statistics

Cronbach's α	McDonald's ω	Split-half (Spearman-Brown)	Items	N	Mean inter-item r	Avg inter-item cov.
0.8605	0.9153	0.8549	3	220	0.6738	0.9806

Item-total statistics

Item	Mean	SD	Corrected item-total r	α if item deleted
peou1	3.914	1.274	0.7512	0.7916
peou2	4.009	1.182	0.7466	0.7948
peou3	3.886	1.163	0.7137	0.825



DETAILED INTERPRETATION

α / ω / split-half ≥ 0.7 acceptable · ≥ 0.8 good · ≥ 0.9 excellent. McDonald's ω (congeneric) is preferred when items are not strictly parallel. The lead result here reads Cronbach's α = 0.8605; McDonald's ω = 0.9153; Split-half (Spearman-Brown) = 0.8549; Items = 3. A reliability coefficient such as Cronbach's alpha at or above .70, together with clean loadings, indicates the items measure one coherent construct. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst validate a measurement instrument before using its scores.

4. Time Series & Forecasting

4.1 Classical decomposition

LIVE OUTPUT

Group: Forecasting · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Trend+seasonal+residual split. It belongs to the **Forecasting** family, used to project a series forward with quantified uncertainty. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to plan capacity, budgets or risk around a projected path.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

STEPS

1. Statistics → **Forecasting** → **Classical decomposition**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D, model=additive, seasonal_periods=7`. 4. Click **Run** and read the tables/chart below.

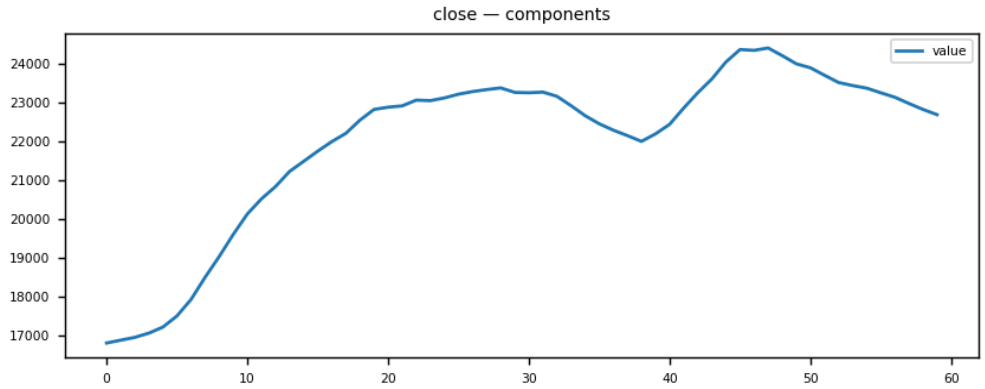
OUTPUT

Classical additive decomposition — close

Model	Seasonal period	Trend SD	Residual SD	Observations
additive	7	2.649e+04	1172	883

Seasonal component (one cycle)

Season position	Effect
1	-103.3
2	15.28
3	-85.01
4	149.7
5	-0.5181
6	53.86
7	-30.04



DETAILED INTERPRETATION

Separates a series into trend + seasonal + residual (moving-average based). The lead result here reads Model = additive; Seasonal period = 7; Trend SD = 2.649e+04; Residual SD = 1172. Compare candidate models on error metrics such as MAPE and RMSE; the forecast path with its prediction interval shows both the expected trajectory and its uncertainty. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst plan capacity, budgets or risk around a projected path.

4.2 Expert modeler (auto-forecast)

LIVE OUTPUT

Group: Forecasting · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Auto-select ARIMA vs. exponential smoothing by AIC, with forecast. It belongs to the **Forecasting** family, used to project a series forward with quantified uncertainty. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to plan capacity, budgets or risk around a projected path.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

STEPS

1. Statistics → **Forecasting** → **Expert modeler (auto-forecast)**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, horizon=6`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Expert modeler — selected model

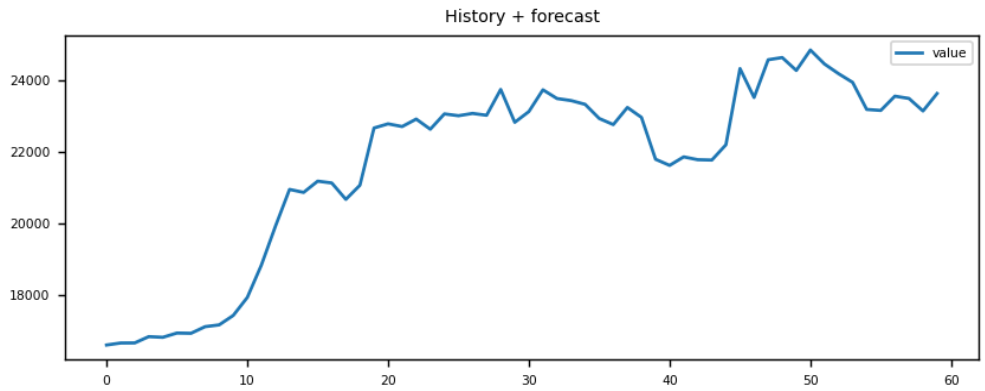
Chosen model	AIC	RMSE	MAPE (%)	Observations	Horizon
ETS(A,N)	1.301e+04	1579	1.781	883	6

Model comparison (by AIC)

Model	AIC
ETS(A,N)	1.301e+04
ARIMA(1, 1, 0)	1.55e+04
ARIMA(0, 1, 1)	1.55e+04
ARIMA(1, 1, 1)	1.55e+04
ARIMA(0, 1, 0)	1.551e+04
ARIMA(2, 1, 1)	1.551e+04

Forecast

Step	Forecast	95% CI low	95% CI high
1	1.057e+05	1.026e+05	1.088e+05
2	1.058e+05	1.027e+05	1.089e+05
3	1.059e+05	1.028e+05	1.09e+05
4	1.06e+05	1.029e+05	1.091e+05
5	1.061e+05	1.03e+05	1.092e+05
6	1.062e+05	1.031e+05	1.093e+05



DETAILED INTERPRETATION

Best model chosen automatically by AIC across ARIMA and exponential-smoothing families. The lead result here reads Chosen model = ETS(A,N); AIC = 1.301e+04; RMSE = 1579; MAPE (%) = 1.781. Compare candidate models on error metrics such as MAPE and RMSE; the forecast path with its prediction interval shows both the expected trajectory and its uncertainty. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst plan capacity, budgets or risk around a projected path.

4.3 Exponential smoothing (Holt-Winters/ETS)

LIVE OUTPUT

Group: Forecasting · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Simple, Holt (trend) and Holt-Winters (seasonal) smoothing with forecast. It belongs to the **Forecasting** family, used to project a series forward with quantified uncertainty. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to plan capacity, budgets or risk around a projected path.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

STEPS

1. Statistics → **Forecasting** → **Exponential smoothing (Holt-Winters/ETS)**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D, trend=add, seasonal=none, seasonal_periods=7, horizon=6`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Holt (double) exponential smoothing — close

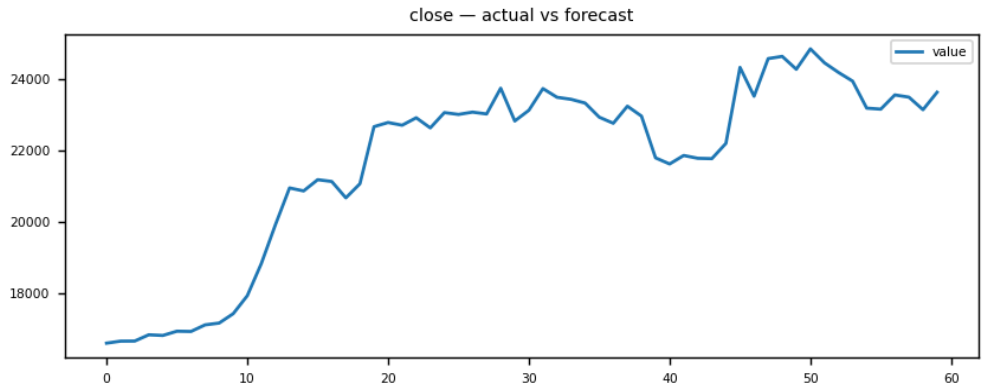
Property	Value
Trend	add
Seasonal	none
Seasonal periods	—
SSE	2.2e+09
AIC	1.301e+04
Observations	883

Smoothing parameters

Parameter	Estimate
α (level)	0.9209
β (trend)	0

Forecast (6 periods)

Period	Forecast
2025-06-02	1.057e+05
2025-06-03	1.058e+05
2025-06-04	1.059e+05
2025-06-05	1.06e+05
2025-06-06	1.061e+05
2025-06-07	1.062e+05



DETAILED INTERPRETATION

Exponential smoothing weights recent observations more heavily; add trend (Holt) and seasonal (Holt-Winters) components as the data needs. The lead result here reads Property = Trend; Value = add. Compare candidate models on error metrics such as MAPE and RMSE; the forecast path with its prediction interval shows both the expected trajectory and its uncertainty. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst plan capacity, budgets or risk around a projected path.

4.4 STL decomposition

LIVE OUTPUT

Group: Forecasting · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Robust seasonal-trend decomposition with strength-of-seasonality metrics. It belongs to the **Forecasting** family, used to project a series forward with quantified uncertainty. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to plan capacity, budgets or risk around a projected path.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

STEPS

1. Statistics → **Forecasting** → **STL decomposition**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D, seasonal_periods=7`. 4. Click **Run** and read the tables/chart below.

OUTPUT

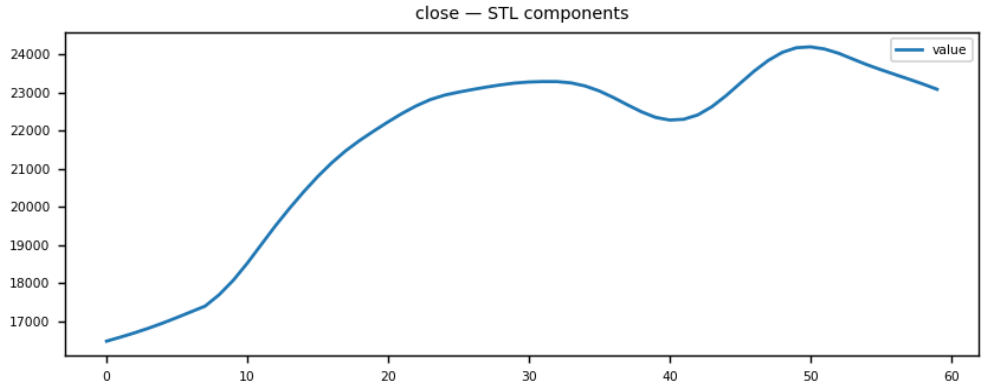
STL decomposition — close

Metric	Value
Seasonal period	7
Strength of trend (0-1)	0.9955
Strength of seasonality (0…	0
Remainder SD	1799
Observations	883

Seasonal component (one cycle)

Season position	Effect
1	82.89
2	37.69
3	-186.5
4	-337.5
5	-186.7
6	230.6

Season position	Effect
7	334.9



DETAILED INTERPRETATION

Strength near 1 = strong trend/seasonality. Robust STL down-weights outliers. The lead result here reads Metric = Seasonal period; Value = 7. Compare candidate models on error metrics such as MAPE and RMSE; the forecast path with its prediction interval shows both the expected trajectory and its uncertainty. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst plan capacity, budgets or risk around a projected path.

4.5 Seasonal ARIMA (SARIMA/SARIMAX)

LIVE OUTPUT

Group: Forecasting · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

ARIMA with a seasonal component (P,D,Q,s) and forecast with 95% CI. It belongs to the **Forecasting** family, used to project a series forward with quantified uncertainty. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to plan capacity, budgets or risk around a projected path.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

STEPS

1. Statistics → **Forecasting** → **Seasonal ARIMA (SARIMA/SARIMAX)**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D, ar=5, diff=5, ma=5, sar=5, sdiff=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

SARIMA(5, 5, 5)×(5, 5, 5, 7) — close

Property	Value
Non-seasonal (p,d,q)	(5, 5, 5)
Seasonal (P,D,Q,s)	(5, 5, 5, 7)
AIC	2.084e+04
BIC	2.094e+04
Log likelihood	-1.04e+04
Observations	883

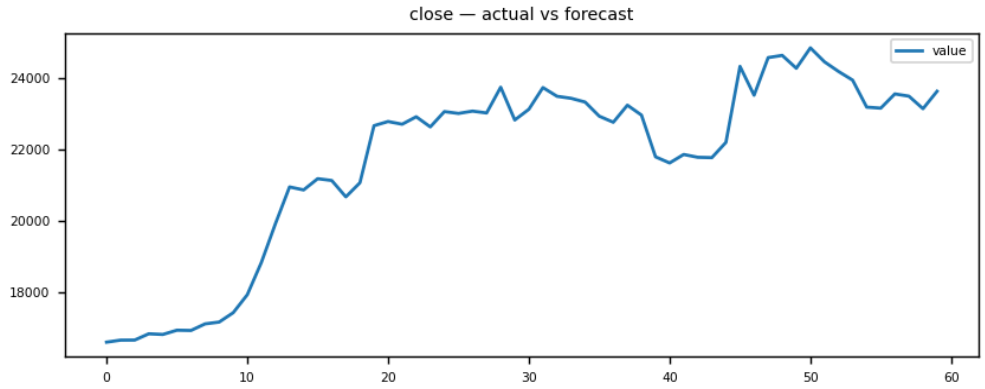
Coefficients

Term	Estimate	SE	p
ar.L1	-1.998	0	< .001

Term	Estimate	SE	p
ar.L2	-2.289	0	< .001
ar.L3	-2.202	0	< .001
ar.L4	-1.746	0	< .001
ar.L5	-0.4924	0	< .001
ma.L1	-0.9249	0.0007	< .001
ma.L2	0.222	0.0279	< .001
ma.L3	-0.055	0.0063	< .001
ma.L4	0.0604	0.2183	0.7819
ma.L5	0.0511	0.1794	0.7756
ar.S.L7	-2.332	0.013	< .001
ar.S.L14	-3.107	0.1648	< .001

Forecast (6 periods, 95% CI)

Period	Forecast	CI low	CI high
2025-06-02	1.106e+05	-2.398e+05	4.609e+05
2025-06-03	1.055e+05	-7.065e+05	9.175e+05
2025-06-04	1.317e+05	-1.549e+06	1.812e+06
2025-06-05	1.376e+05	-2.838e+06	3.113e+06
2025-06-06	1.586e+05	-4.781e+06	5.099e+06
2025-06-07	1.906e+05	-7.706e+06	8.087e+06



DETAILED INTERPRETATION

Lower AIC/BIC = better fit. Check residuals are white noise (Ljung-Box). The lead result here reads Property = Non-seasonal (p,d,q); Value = (5, 5, 5). Compare candidate models on error metrics such as MAPE and RMSE; the forecast path with its prediction interval shows both the expected trajectory and its uncertainty. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst plan capacity, budgets or risk around a projected path.

4.6 Seasonal naïve forecast

LIVE OUTPUT

Group: Forecasting · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Baseline seasonal benchmark forecast. It belongs to the **Forecasting** family, used to project a series forward with quantified uncertainty. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to plan capacity, budgets or risk around a projected path.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

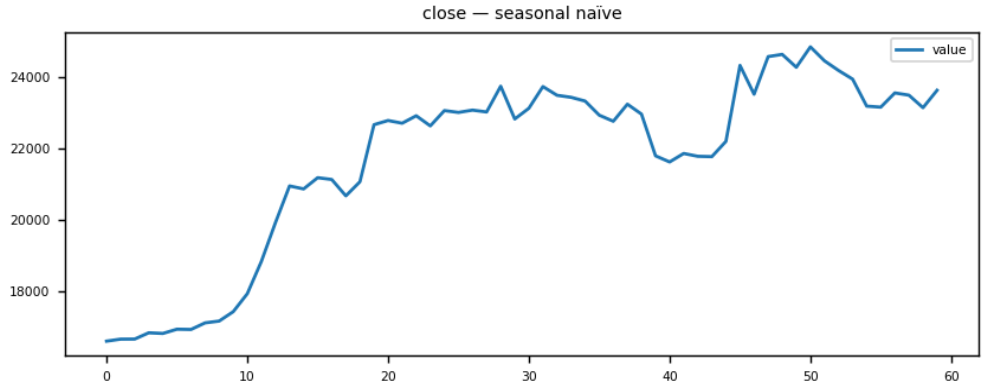
STEPS

1. Statistics → **Forecasting** → **Seasonal naïve forecast**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D, seasonal_periods=7, horizon=6`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Seasonal naïve forecast — close

Horizon	First forecast	Seasonal period
6	1.094e+05	7



DETAILED INTERPRETATION

Baseline forecast that repeats the last season — a benchmark for any model. The lead result here reads Horizon = 6; First forecast = 1.094e+05; Seasonal period = 7. Compare candidate models on error metrics such as MAPE and RMSE; the forecast path with its prediction interval shows both the expected trajectory and its uncertainty. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst plan capacity, budgets or risk around a projected path.

4.7 Time series (ARIMA)

LIVE OUTPUT

Group: Forecasting · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Stationarity test, seasonal decomposition, ARIMA forecast with CI. It belongs to the **Forecasting** family, used to project a series forward with quantified uncertainty. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to plan capacity, budgets or risk around a projected path.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

STEPS

1. Statistics → **Forecasting** → **Time series (ARIMA)**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D, horizon=6`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Time series — close (D-aggregated, 883 periods)

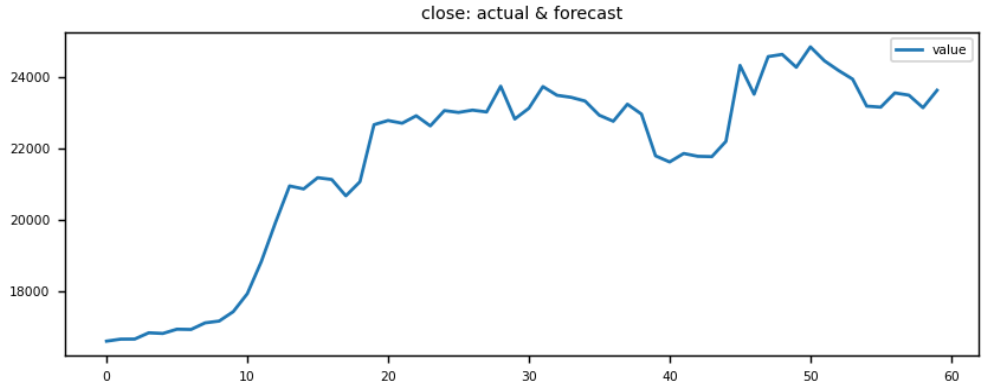
Property	Value
ADF statistic	-0.3824
ADF p (stationarity)	0.913
ARIMA order (p,d,q)	(1, 1, 1)
AIC	1.55e+04
Log likelihood	-7749

Forecast (6 periods, 95% CI)

Period	Forecast	CI low	CI high
2025-06-02	1.056e+05	1.025e+05	1.087e+05
2025-06-03	1.056e+05	1.014e+05	1.098e+05
2025-06-04	1.056e+05	1.004e+05	1.108e+05
2025-06-05	1.056e+05	9.965e+04	1.116e+05
2025-06-06	1.056e+05	9.896e+04	1.122e+05
2025-06-07	1.056e+05	9.833e+04	1.129e+05

Seasonal component (additive decomposition)

Season position	Effect
1	-103.3
2	15.28
3	-85.01
4	149.7
5	-0.5181
6	53.86
7	-30.04



DETAILED INTERPRETATION

ADF p < .05 suggests the series is stationary. The lead result here reads Property = ADF statistic; Value = -0.3824. Compare candidate models on error metrics such as MAPE and RMSE; the forecast path with its prediction interval shows both the expected trajectory and its uncertainty. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst plan capacity, budgets or risk around a projected path.

4.8 ACF / PACF

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Autocorrelation & partial autocorrelation with chart. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close
1.662e+04
1.667e+04
1.668e+04
1.685e+04
1.683e+04

STEPS

1. Statistics → **Time series** → **ACF / PACF**. 2. Select the demo dataset above. 3. Set parameters — `column=close, lags=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Autocorrelation — close

Lag	ACF	PACF
1	0.995	0.9962
2	0.9904	0.0441
3	0.9858	-0.0059
4	0.9811	-0.0125
5	0.9761	-0.0326

DETAILED INTERPRETATION

ACF/PACF identify AR and MA orders. Significant spikes (|r| > 2/√N) suggest structure. The lead result here reads Lag = 1; ACF = 0.995; PACF = 0.9962. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.9 ARCH / EGARCH / GJR-GARCH

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Asymmetric & exponential volatility models with a choice of error distribution. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

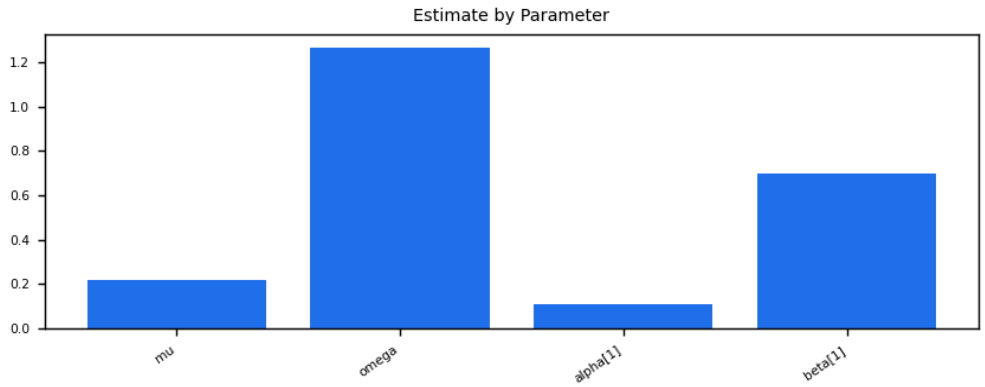
STEPS

1. Statistics → **Time series** → **ARCH / EGARCH / GJR-GARCH**. 2. Select the demo dataset above. 3. Set parameters — `column=close, model=GARCH, dist=normal, date=date, as_returns=no, order=1`. 4. Click **Run** and read the tables/chart below.

OUTPUT

GARCH volatility model — close

Parameter	Estimate	SE	p
mu	0.2206	0.0801	0.0059
omega	1.262	0.5495	0.0216
alpha[1]	0.1097	0.0416	0.0084
beta[1]	0.6995	0.1044	< .001



DETAILED INTERPRETATION

GARCH with normal errors on % returns. Log-likelihood = -2060.2 · AIC = 4128.4. Persistence α+β = 0.809; volatility half-life ≈ 3.3 periods. The lead result here reads Parameter = mu; Estimate = 0.2206; SE = 0.0801; p = 0.0059. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.10 ARCH-LM test

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Test for ARCH heteroskedasticity (volatility clustering). It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close
1.662e+04
1.667e+04
1.668e+04
1.685e+04
1.683e+04

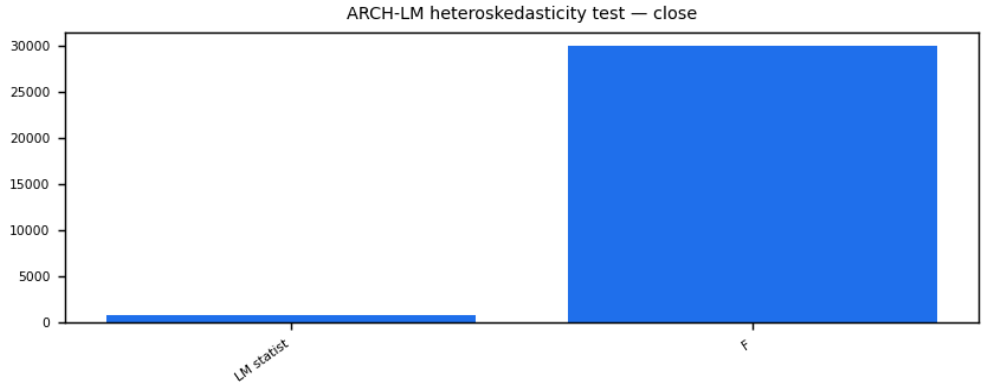
STEPS

1. Statistics → **Time series** → **ARCH-LM test**. 2. Select the demo dataset above. 3. Set parameters — `column=close, lags=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

ARCH-LM heteroskedasticity test — close

LM statistic	LM p	F	F p
872.9	< .001	2.992e+04	< .001



DETAILED INTERPRETATION

p < .05 → volatility clustering (ARCH effects) present — consider a GARCH model. The lead result here reads LM statistic = 872.9; LM p = < .001; F = 2.992e+04; F p = < .001. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.11 ARDL model

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Autoregressive distributed-lag model for mixed I(0)/I(1) series. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, volume. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	volume
1.662e+04	9.693e+04
1.667e+04	1.219e+05
1.668e+04	1.595e+05
1.685e+04	2.204e+05
1.683e+04	1.635e+05

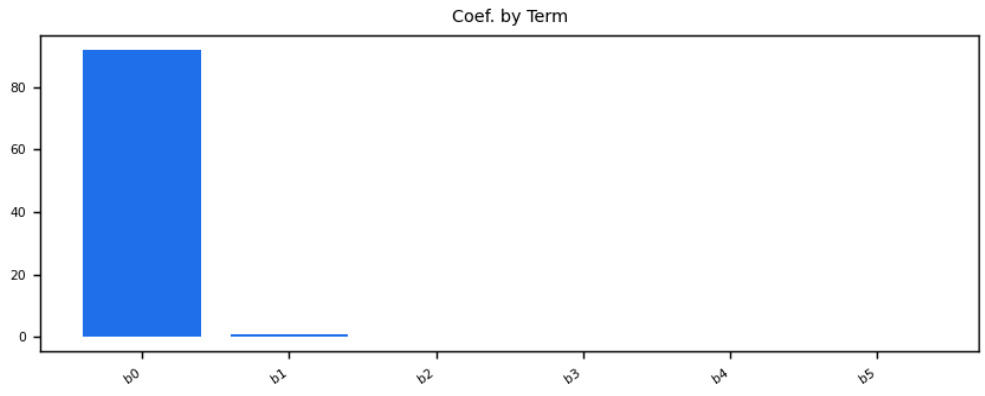
STEPS

1. Statistics → **Time series** → **ARDL model**. 2. Select the demo dataset above. 3. Set parameters — `dependent=close, predictors=volume, lags=3, exog_order=1`. 4. Click **Run** and read the tables/chart below.

OUTPUT

ARDL model — close

Term	Coef.	SE	z	p
b0	92.02	151.1	0.609	0.5425
b1	0.9194	0.0338	27.17	< .001
b2	0.0939	0.0459	2.047	0.0407
b3	-0.0134	0.0339	-0.396	0.6922
b4	0.0011	0.0015	0.731	0.4646
b5	-0.0006	0.0015	-0.425	0.6707



DETAILED INTERPRETATION

ARDL(3,1) with 1 regressor(s). AIC=15474.10, BIC=15507.56. ARDL captures short-run dynamics and (via lags) long-run relationships for mixed I(0)/I(1) series. The lead result here reads Term = b0; Coef. = 92.02; SE = 151.1; z = 0.609. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.12 Augmented Dickey-Fuller

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Unit-root / stationarity test. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

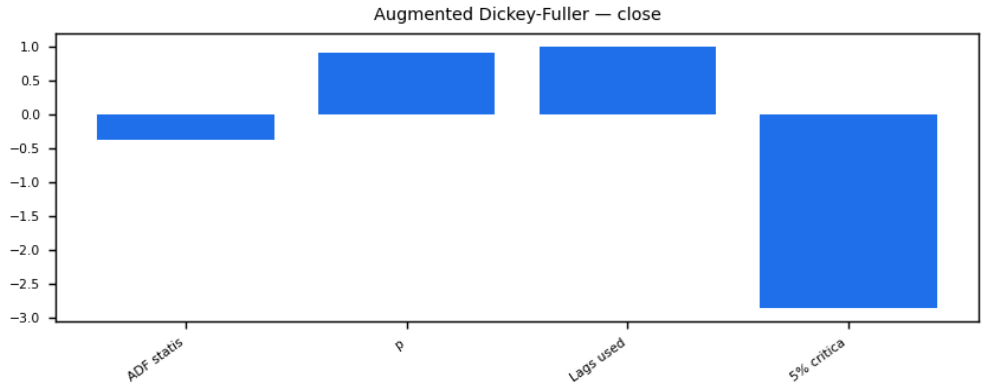
STEPS

1. Statistics → **Time series** → **Augmented Dickey-Fuller**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Augmented Dickey-Fuller — close

ADF statistic	p	Lags used	5% critical
-0.3824	0.913	1	-2.865



DETAILED INTERPRETATION

p < .05 → stationary (reject the unit-root null). The lead result here reads ADF statistic = -0.3824; p = 0.913; Lags used = 1; 5% critical = -2.865. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.13 Connectedness (Diebold-Yilmaz)

LIVE OUTPUT

Group: Time series · Dataset: daily returns of four assets — stock, bond, gold and crypto

WHAT THIS DOES

Volatility/return spillovers across series via VAR generalized FEVD. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily returns of four assets — stock, bond, gold and crypto, drawing on ret_stock, ret_bond, ret_gold, ret_crypto. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

ret_stock	ret_bond	ret_gold	ret_crypto
0.019	0.002	0.01	0.014
0.032	0.036	-0.024	-0.014
-0.046	-0.036	-0	0.018

ret_stock	ret_bond	ret_gold	ret_crypto
-0.002	0.015	-0.02	0.001
0.019	0.029	0.025	0.028

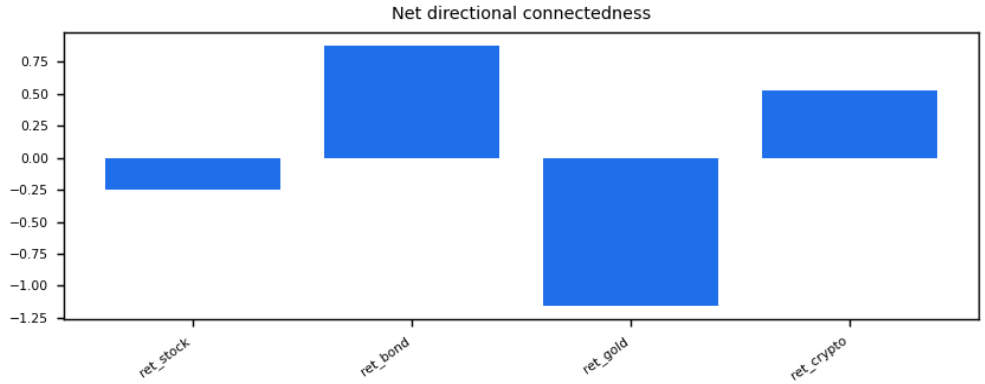
STEPS

1. Statistics → **Time series** → **Connectedness (Diebold-Yilmaz)**. 2. Select the demo dataset above. 3. Set parameters — `columns=ret_stock, ret_bond, ret_gold, ret_crypto, lags=2`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Diebold-Yilmaz connectedness (VAR(2), H = 10)

Series	FROM others (%)	TO others (%)	NET (%)
ret_stock	7.36	7.11	-0.25
ret_bond	8.18	9.06	0.88
ret_gold	3.65	2.5	-1.16
ret_crypto	0.3	0.83	0.53



DETAILED INTERPRETATION

Total connectedness index (spillover) = 19.5%. NET > 0 = a net transmitter of shocks; NET < 0 = a net receiver. The lead result here reads Series = ret_stock; FROM others (%) = 7.36; TO others (%) = 7.11; NET (%) = -0.25. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily returns of four assets — stock, bond, gold and crypto, this lets the analyst understand the dynamics driving a monitored metric.

4.14 Differencing + ADF

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Difference a series to stationarity. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

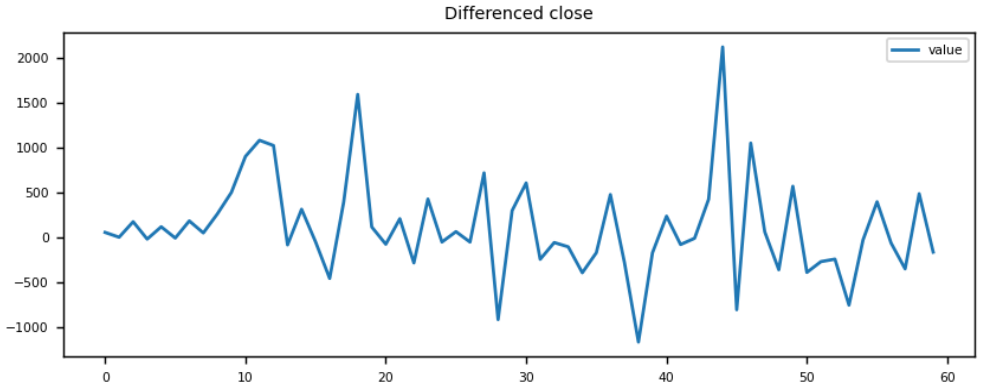
STEPS

1. Statistics → **Time series** → **Differencing + ADF**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D, order=1`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Differencing (d=1) — close

ADF p (original)	ADF p (differenced)	Observations
0.913	< .001	882



DETAILED INTERPRETATION

Differencing removes trends; the differenced series should become stationary (ADF p < .05). The lead result here reads ADF p (original) = 0.913; ADF p (differenced) = < .001; Observations = 882. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.15 Engle-Granger cointegration

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Two-series cointegration test. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, open. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	open
1.662e+04	1.654e+04
1.667e+04	1.662e+04
1.668e+04	1.667e+04
1.685e+04	1.668e+04
1.683e+04	1.685e+04

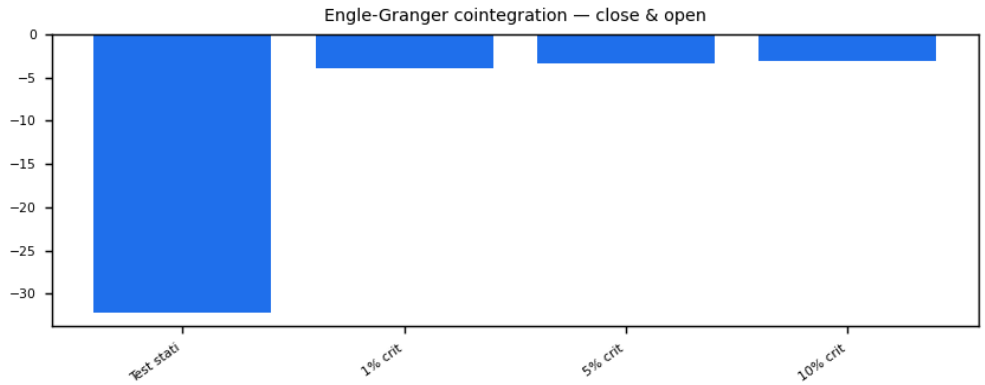
STEPS

1. Statistics → **Time series** → **Engle-Granger cointegration**. 2. Select the demo dataset above. 3. Set parameters — `column1=close, column2=open`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Engle-Granger cointegration — close & open

Test statistic	p	1% crit	5% crit	10% crit
-32.18	< .001	-3.909	-3.343	-3.049



DETAILED INTERPRETATION

p < .05 → the two series are cointegrated (share a long-run equilibrium). The lead result here reads Test statistic = -32.18; p = < .001; 1% crit = -3.909; 5% crit = -3.343. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.16 GARCH volatility

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Model volatility clustering (ARCH/GARCH). It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

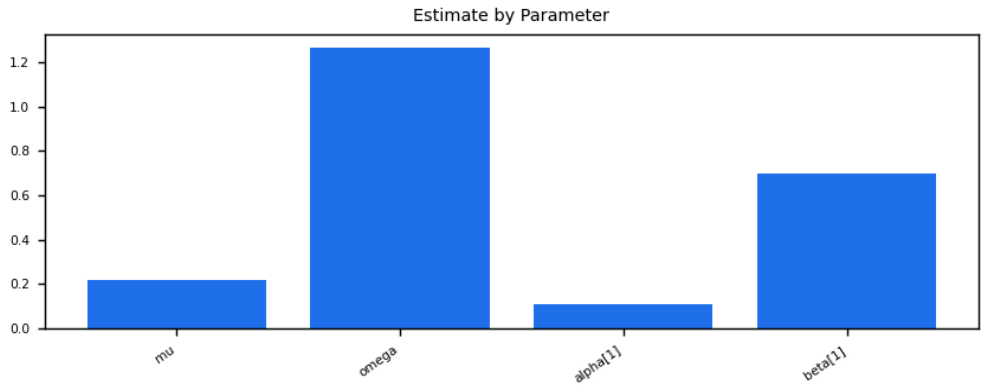
STEPS

1. Statistics → **Time series** → **GARCH volatility**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, as_returns=no, order=1`. 4. Click **Run** and read the tables/chart below.

OUTPUT

GARCH(1,1) volatility model — close

Parameter	Estimate	SE	p
mu	0.2206	0.0801	0.0059
omega	1.262	0.5495	0.0216
alpha[1]	0.1097	0.0416	0.0084
beta[1]	0.6995	0.1044	< .001



DETAILED INTERPRETATION

Modelled on % returns. Log-likelihood = -2060.2 · AIC = 4128.4. Significant ARCH/GARCH terms indicate volatility clustering. The lead result here reads Parameter = mu; Estimate = 0.2206; SE = 0.0801; p = 0.0059. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.17 Granger causality

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Test whether one series helps predict another. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, volume, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	volume	date
1.662e+04	9.693e+04	2023-01-01
1.667e+04	1.219e+05	2023-01-02
1.668e+04	1.595e+05	2023-01-03
1.685e+04	2.204e+05	2023-01-04
1.683e+04	1.635e+05	2023-01-05

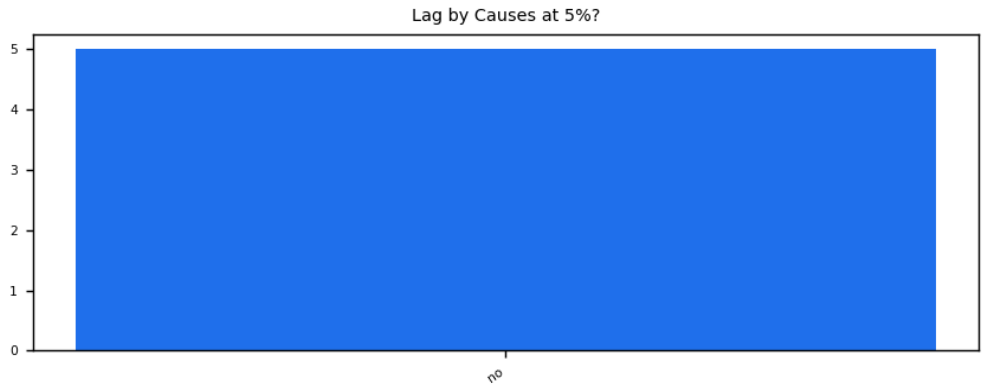
STEPS

1. Statistics → **Time series** → **Granger causality**. 2. Select the demo dataset above. 3. Set parameters — `dependent=close, cause=volume, date=date, maxlag=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Granger causality — does volume cause close?

Lag	F	p	Causes at 5%?
1	0.1969	0.6573	no
2	0.2372	0.7889	no
3	0.1495	0.93	no
4	0.2401	0.9157	no
5	0.3023	0.9116	no



DETAILED INTERPRETATION

Tests whether past values of volume improve prediction of close. Significant p at any lag suggests Granger causality. The lead result here reads Lag = 1; F = 0.1969; p = 0.6573; Causes at 5%? = no. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.18 Hodrick-Prescott filter

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Trend/cycle decomposition. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

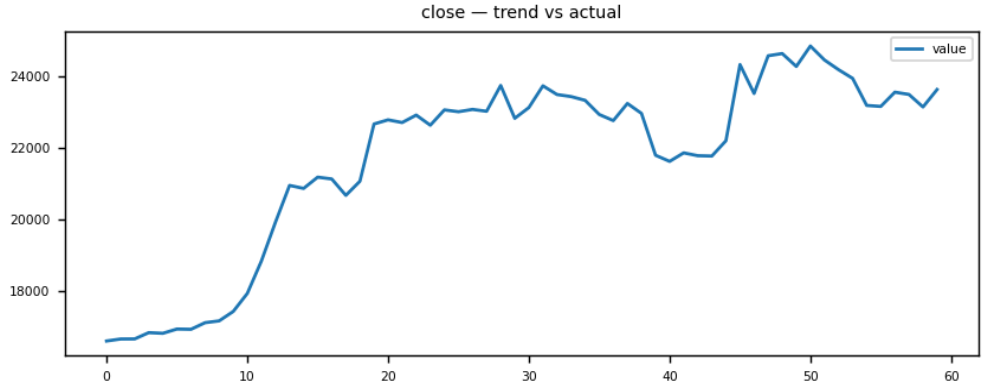
STEPS

1. Statistics → **Time series** → **Hodrick-Prescott filter**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D, lambda=0.2`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Hodrick-Prescott filter — close

Lambda	Trend SD	Cycle SD	Observations
0.2	2.671e+04	467.2	883



DETAILED INTERPRETATION

Splits a series into a smooth trend and a cyclical component (λ=1600 for quarterly macro). The lead result here reads Lambda = 0.2; Trend SD = 2.671e+04; Cycle SD = 467.2; Observations = 883. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.19 Johansen cointegration

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Trace + max-eigenvalue cointegration test. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on open, high, close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

open	high	close	date
1.654e+04	1.663e+04	1.662e+04	2023-01-01
1.662e+04	1.68e+04	1.667e+04	2023-01-02
1.667e+04	1.678e+04	1.668e+04	2023-01-03
1.668e+04	1.699e+04	1.685e+04	2023-01-04
1.685e+04	1.688e+04	1.683e+04	2023-01-05

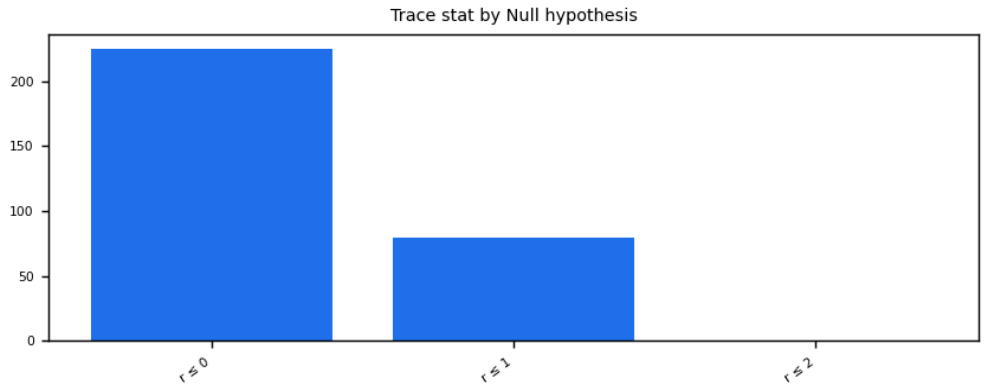
STEPS

1. Statistics → **Time series** → **Johansen cointegration**. 2. Select the demo dataset above. 3. Set parameters — `columns=open, high, close, date=date, lags=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Johansen cointegration test

Null hypothesis	Trace stat	5% crit.	Trace decision	Max-eigen stat	5% crit.
r ≤ 0	225.1	29.8	reject	145.9	21.13
r ≤ 1	79.21	15.49	reject	79.12	14.26
r ≤ 2	0.0881	3.841	cannot reject	0.0881	3.841



DETAILED INTERPRETATION

2 cointegrating relationship(s) at 5%. Reject when the trace statistic exceeds the critical value — indicates a long-run equilibrium. The lead result here reads Null hypothesis = r ≤ 0; Trace stat = 225.1; 5% crit. = 29.8; Trace decision = reject. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.20 KPSS test

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Stationarity test (reversed null). It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

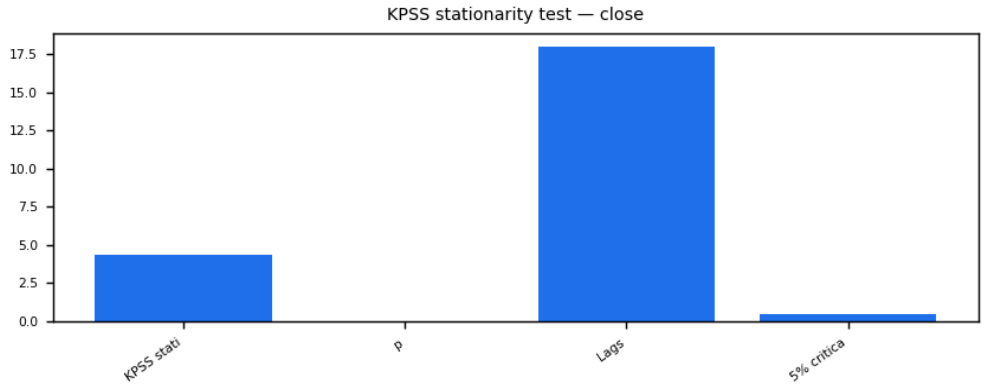
STEPS

1. Statistics → **Time series** → **KPSS test**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D`. 4. Click **Run** and read the tables/chart below.

OUTPUT

KPSS stationarity test — close

KPSS statistic	p	Lags	5% critical
4.373	0.01	18	0.463



DETAILED INTERPRETATION

Null = stationary; p < .05 → non-stationary (complements ADF). The lead result here reads KPSS statistic = 4.373; p = 0.01; Lags = 18; 5% critical = 0.463. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.21 Ljung-Box test

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Test for remaining autocorrelation (white noise). It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close
1.662e+04
1.667e+04

close
1.668e+04
1.685e+04
1.683e+04

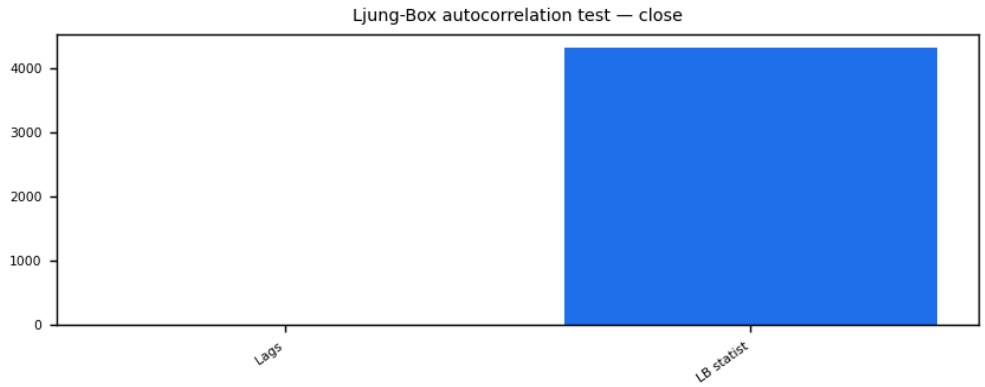
STEPS

1. Statistics → **Time series** → **Ljung-Box test**. 2. Select the demo dataset above. 3. Set parameters — `column=close, lags=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Ljung-Box autocorrelation test — close

Lags	LB statistic	p
5	4314	< .001



DETAILED INTERPRETATION

p < .05 → significant autocorrelation remains (series/residuals not white noise). The lead result here reads Lags = 5; LB statistic = 4314; p = < .001. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.22 Local projections (IRF)

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Jordà impulse responses — response of outcome to a shock by horizon. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, volume, open. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	volume	open
1.662e+04	9.693e+04	1.654e+04
1.667e+04	1.219e+05	1.662e+04
1.668e+04	1.595e+05	1.667e+04
1.685e+04	2.204e+05	1.668e+04
1.683e+04	1.635e+05	1.685e+04

STEPS

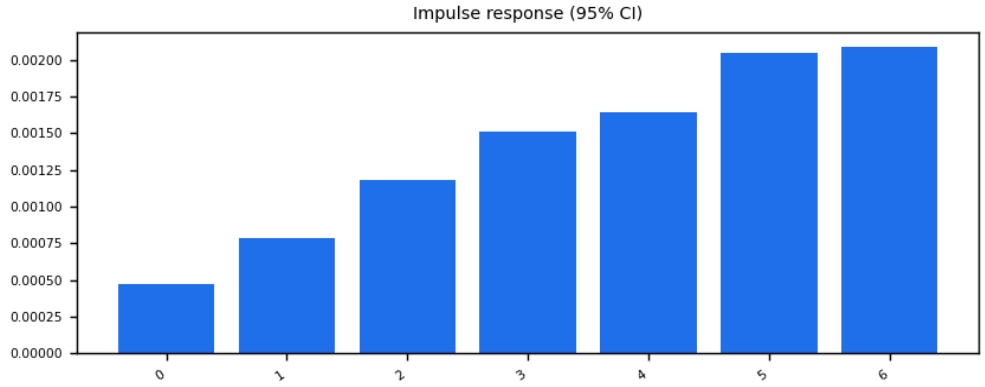
1. Statistics → **Time series** → **Local projections (IRF)**. 2. Select the demo dataset above. 3. Set parameters — `outcome=close, shock=volume, controls=open, horizon=6`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Local-projection impulse response — close to volume

Horizon	Response	SE	p
0	0.0005	0.0004	0.2751

Horizon	Response	SE	p
1	0.0008	0.0007	0.2326
2	0.0012	0.0009	0.1954
3	0.0015	0.0012	0.215
4	0.0016	0.0015	0.2842
5	0.002	0.0017	0.2317
6	0.0021	0.002	0.2934



DETAILED INTERPRETATION

Jordà (2005) local projections: the response of the outcome to the shock at each horizon (HAC/Newey–West SE). Increasingly preferred over VAR impulse responses. The lead result here reads Horizon = 0; Response = 0.0005; SE = 0.0004; p = 0.2751. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.23 Mann-Kendall trend test

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Monotonic-trend test. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close
1.662e+04
1.667e+04
1.668e+04
1.685e+04
1.683e+04

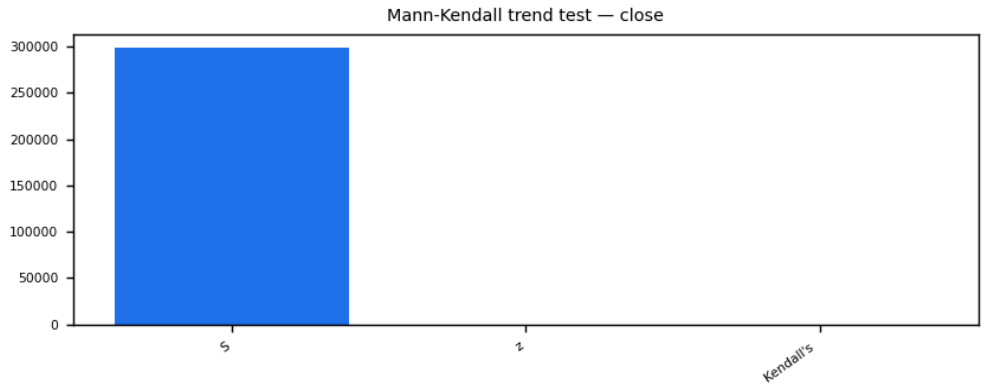
STEPS

1. Statistics → **Time series** → **Mann-Kendall trend test**. 2. Select the demo dataset above. 3. Set parameters — `column=close`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Mann-Kendall trend test — close

S	z	p	Kendall's τ	Trend
2.987e+05	34.12	< .001	0.767	increasing



DETAILED INTERPRETATION

Nonparametric monotonic-trend test for a time-ordered series. The lead result here reads S = 2.987e+05; z = 34.12; p = < .001; Kendall's τ = 0.767. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.24 Markov-switching regression

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Regime-switching model (bull/bear, high/low volatility states). It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close
1.662e+04
1.667e+04
1.668e+04
1.685e+04
1.683e+04

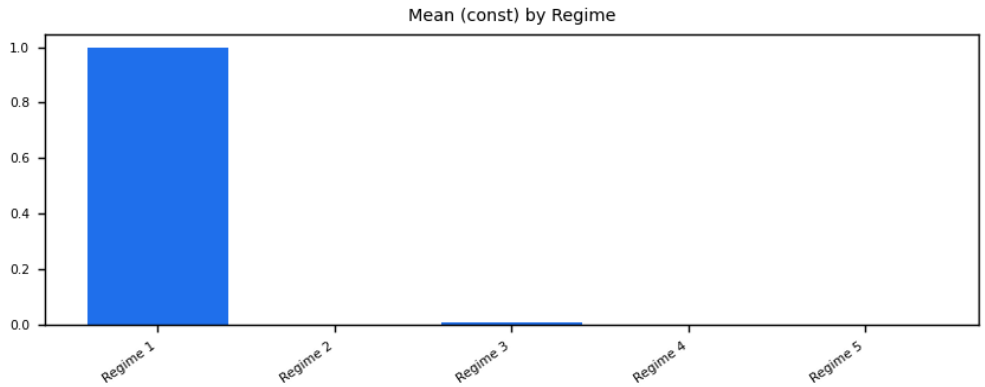
STEPS

1. Statistics → **Time series** → **Markov-switching regression**. 2. Select the demo dataset above. 3. Set parameters — `series=close, regimes=5, switching_variance=yes`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Markov-switching regimes — close

Regime	Mean (const)	Expected duration
Regime 1	0.9968	309.5
Regime 2	0.0001	201.2
Regime 3	0.0063	96.15
Regime 4	0	245.8
Regime 5	0	642.9



DETAILED INTERPRETATION

5-regime Markov-switching model. Regimes capture distinct states (e.g. bull/bear, high/low volatility); expected duration = average persistence in periods. Log-likelihood -8626.68, AIC 17313.36. The lead result here reads Regime = Regime 1; Mean (const) = 0.9968; Expected duration = 309.5. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.25 Moving averages (SMA/EWMA)

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Smoothing with SMA & EWMA. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

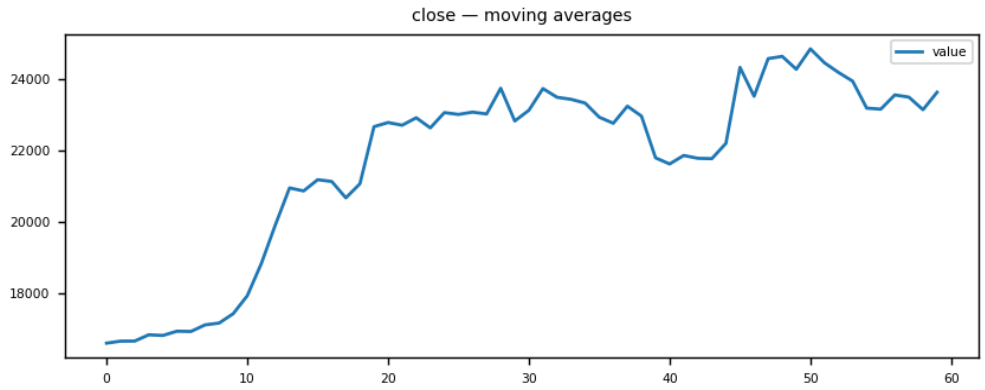
STEPS

1. Statistics → **Time series** → **Moving averages (SMA/EWMA)**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date, frequency=D, window=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Moving averages — close

Window	Last MA	Last EWMA	Observations
5	1.055e+05	1.057e+05	883



DETAILED INTERPRETATION

Simple and exponentially-weighted moving averages for smoothing/trend. The lead result here reads Window = 5; Last MA = 1.055e+05; Last EWMA = 1.057e+05; Observations = 883. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.26 Rolling correlation

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Time-varying correlation + chart. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, volume. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	volume
1.662e+04	9.693e+04
1.667e+04	1.219e+05
1.668e+04	1.595e+05
1.685e+04	2.204e+05
1.683e+04	1.635e+05

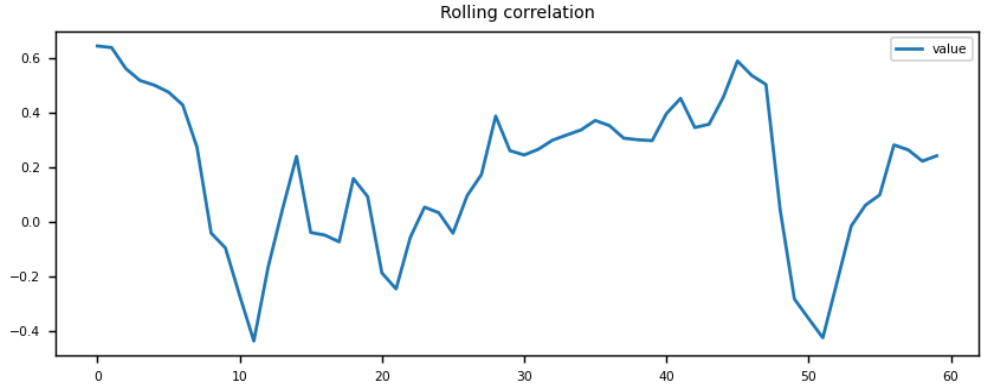
STEPS

1. Statistics → **Time series** → **Rolling correlation**. 2. Select the demo dataset above. 3. Set parameters — `column1=close, column2=volume, window=20`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Rolling correlation — close vs volume

Window	Mean rolling r	Min	Max	N
20	0.0153	-0.8407	0.8178	883



DETAILED INTERPRETATION

Tracks how the correlation between two series changes over time. The lead result here reads Window = 20; Mean rolling r = 0.0153; Min = -0.8407; Max = 0.8178. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.27 Spectral periodogram

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Dominant cycles / power spectrum. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close
1.662e+04
1.667e+04
1.668e+04
1.685e+04
1.683e+04

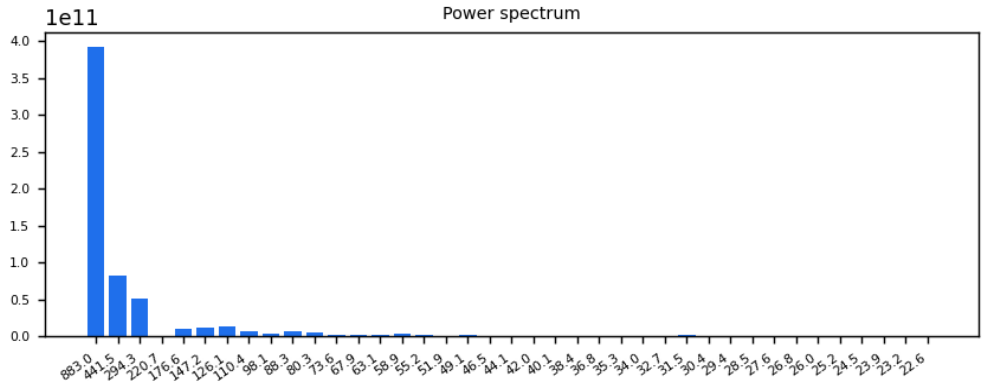
STEPS

1. Statistics → **Time series** → **Spectral periodogram**. 2. Select the demo dataset above. 3. Set parameters — `column=close`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Spectral periodogram — close

Period	Frequency	Power
883	0.0011	3.926e+11
441.5	0.0023	8.252e+10
294.3	0.0034	5.231e+10
126.1	0.0079	1.413e+10
147.2	0.0068	1.159e+10
176.6	0.0057	1.128e+10
110.4	0.0091	7.499e+09
88.3	0.0113	7.289e+09



DETAILED INTERPRETATION

Dominant cycles: the period with the highest power is the strongest cyclicity. The lead result here reads Period = 883; Frequency = 0.0011; Power = 3.926e+11. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a

monitored metric.

4.28 Structural VAR (SVAR)

LIVE OUTPUT

Group: Time series · Dataset: daily returns of four assets — stock, bond, gold and crypto

WHAT THIS DOES

Cholesky-identified structural VAR with impulse responses & variance decomposition. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily returns of four assets — stock, bond, gold and crypto, drawing on ret_stock, ret_bond, ret_gold. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

ret_stock	ret_bond	ret_gold
0.019	0.002	0.01
0.032	0.036	-0.024
-0.046	-0.036	-0
-0.002	0.015	-0.02
0.019	0.029	0.025

STEPS

1. Statistics → **Time series** → **Structural VAR (SVAR)**. 2. Select the demo dataset above. 3. Set parameters — `columns=ret_stock, ret_bond, ret_gold, lags=2, horizon=8`. 4. Click **Run** and read the tables/chart below.

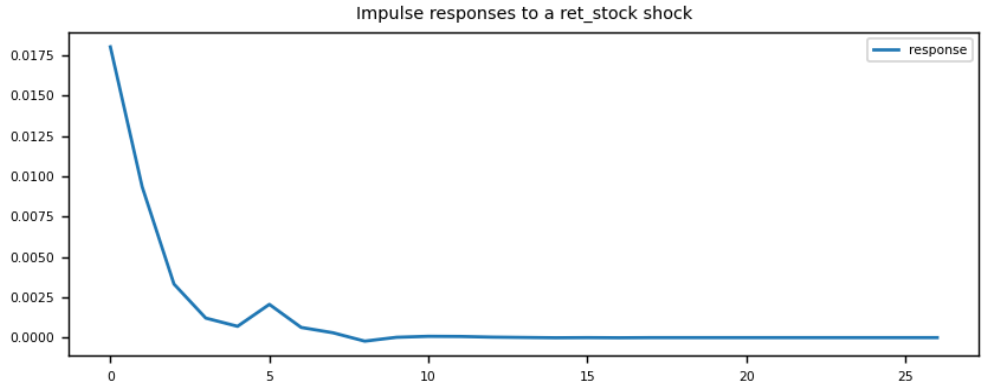
OUTPUT

Structural VAR (Cholesky, lags=2) — impulse responses to a shock in ret_stock

Horizon	→ ret_stock	→ ret_bond	→ ret_gold
0	0.018	0.0094	0.0033
1	0.0012	0.0007	0.0021
2	0.0006	0.0003	-0.0002
3	0	0.0001	0.0001
4	0	0	-0
5	-0	-0	0
6	0	0	-0
7	0	0	0
8	0	0	0

Forecast-error variance decomposition (h=8, %)

Variable	from ret_stock	from ret_bond	from ret_gold
ret_stock	99.23	0.53	0.25
ret_bond	36.53	62.63	0.84
ret_gold	3.72	7.51	88.77



DETAILED INTERPRETATION

Orthogonalised impulse responses: how each variable reacts over time to a one-standard-deviation structural shock in the first variable (recursive/Cholesky identification). The lead result here reads Horizon = 0; → ret_stock =

0.018; → ret_bond = 0.0094; → ret_gold = 0.0033. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily returns of four assets — stock, bond, gold and crypto, this lets the analyst understand the dynamics driving a monitored metric.

4.29 Unit-root tests

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

ADF, Phillips-Perron, DF-GLS and KPSS stationarity tests. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

close	date
1.662e+04	2023-01-01
1.667e+04	2023-01-02
1.668e+04	2023-01-03
1.685e+04	2023-01-04
1.683e+04	2023-01-05

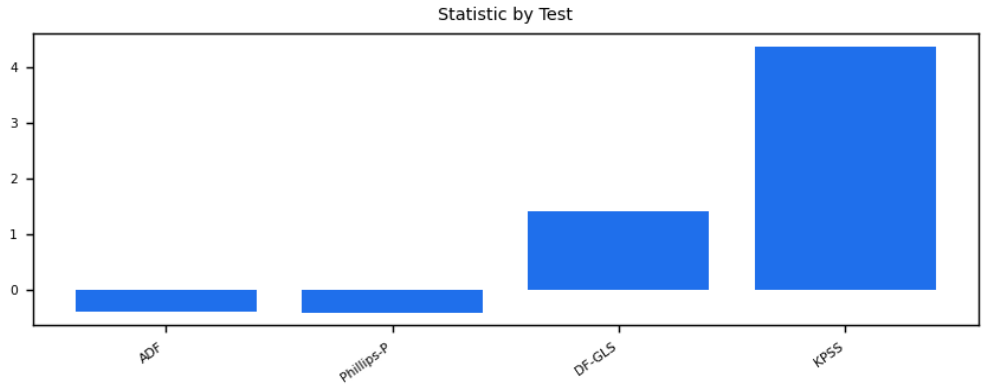
STEPS

1. Statistics → **Time series** → **Unit-root tests**. 2. Select the demo dataset above. 3. Set parameters — `column=close, date=date`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Unit-root / stationarity tests — close

Test	Statistic	p	Verdict
ADF	-0.3824	0.913	unit root
Phillips-Perron	-0.4018	0.9098	unit root
DF-GLS	1.414	0.963	unit root
KPSS	4.373	< .001	unit root



DETAILED INTERPRETATION

ADF/PP/DF-GLS: H■ = unit root (p < .05 → stationary). KPSS: H■ = stationary (p < .05 → unit root). Agreement across tests is reassuring. The lead result here reads Test = ADF; Statistic = -0.3824; p = 0.913; Verdict = unit root. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.30 VAR (vector autoregression)

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Multivariate VAR with variance decomposition + Granger causality. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on open, high, close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

open	high	close	date
1.654e+04	1.663e+04	1.662e+04	2023-01-01
1.662e+04	1.68e+04	1.667e+04	2023-01-02
1.667e+04	1.678e+04	1.668e+04	2023-01-03
1.668e+04	1.699e+04	1.685e+04	2023-01-04
1.685e+04	1.688e+04	1.683e+04	2023-01-05

STEPS

1. Statistics → **Time series** → **VAR (vector autoregression)**. 2. Select the demo dataset above. 3. Set parameters — `columns=open, high, close, date=date, maxlags=4, horizon=6`. 4. Click **Run** and read the tables/chart below.

OUTPUT

VAR(2) model — open, high, close

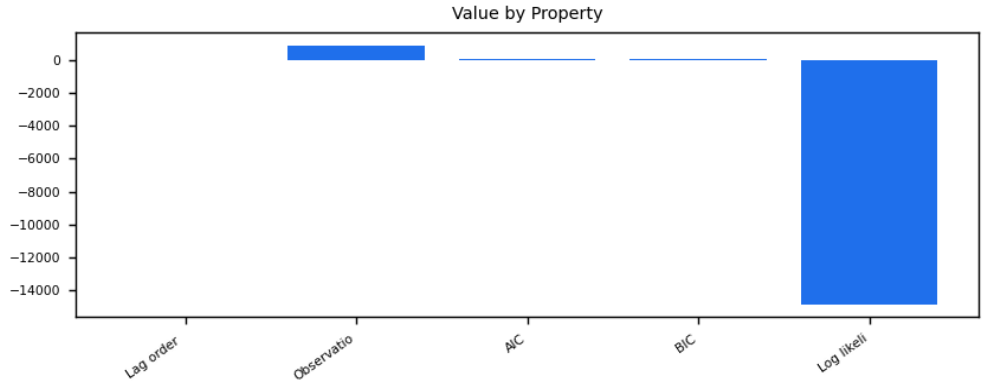
Property	Value
Lag order (AIC)	2
Observations	881
AIC	25.2
BIC	25.32
Log likelihood	-1.483e+04

Variance decomposition of open (h=6)

Shock from	% of variance
open	0.0211
high	67.75
close	32.23

Variance decomposition of high (h=6)

Shock from	% of variance
open	0.0288
high	74.16
close	25.81



DETAILED INTERPRETATION

Vector autoregression — each variable regressed on lags of all variables. Lag chosen by AIC. The lead result here reads Property = Lag order (AIC); Value = 2. Read the actual-vs-fitted overlay and the diagnostic statistics for

autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.31 VECM (error correction)

LIVE OUTPUT

Group: Time series · Dataset: daily Bitcoin (BTC-USD) prices — open/high/low/close and volume

WHAT THIS DOES

Vector error-correction model for cointegrated series. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, drawing on open, high, close, date. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

open	high	close	date
1.654e+04	1.663e+04	1.662e+04	2023-01-01
1.662e+04	1.68e+04	1.667e+04	2023-01-02
1.667e+04	1.678e+04	1.668e+04	2023-01-03
1.668e+04	1.699e+04	1.685e+04	2023-01-04
1.685e+04	1.688e+04	1.683e+04	2023-01-05

STEPS

1. Statistics → **Time series** → **VECM (error correction)**. 2. Select the demo dataset above. 3. Set parameters — `columns=open, high, close, date=date, lags=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

VECM — open, high, close

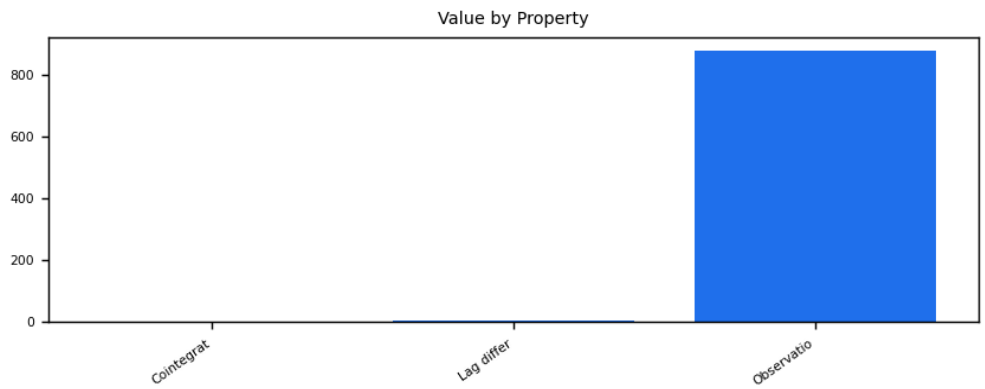
Property	Value
Cointegration rank	2
Lag differences	5
Observations	877

Cointegrating vectors (β, normalized)

Variable	CE1	CE2
open	1	-0
high	-0	1
close	-1	-1.019

Adjustment coefficients (α — speed of error correction)

Equation	CE1	CE2
open	-1.105	0
high	-170.1	-0.4657
close	-110.2	0.1681



DETAILED INTERPRETATION

Vector error-correction model — captures long-run cointegrating relationships plus short-run dynamics. The lead result here reads Property = Cointegration rank; Value = 2. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily Bitcoin (BTC-USD) prices — open/high/low/close and volume, this lets the analyst understand the dynamics driving a monitored metric.

4.32 Value at Risk & Expected Shortfall

LIVE OUTPUT

Group: Time series · Dataset: daily returns of four assets — stock, bond, gold and crypto

WHAT THIS DOES

Historical, Gaussian & Cornish-Fisher VaR plus CVaR for a return series. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily returns of four assets — stock, bond, gold and crypto, drawing on ret_crypto. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

ret_crypto
0.014
-0.014
0.018
0.001
0.028

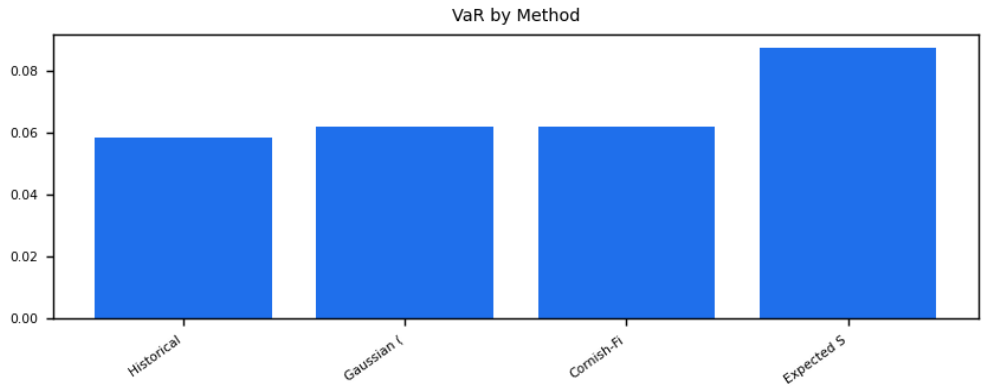
STEPS

1. Statistics → **Time series** → **Value at Risk & Expected Shortfall**. 2. Select the demo dataset above. 3. Set parameters — `returns=ret_crypto, confidence=0.95`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Value at Risk & Expected Shortfall (95% confidence)

Method	VaR	Interpretation
Historical	0.0586	Worst 5% empirical loss
Gaussian (parametric)	0.0621	Normal-distribution loss
Cornish-Fisher	0.0622	Adjusted for skew & kurtos…
Expected Shortfall (CVaR)	0.0876	Mean loss beyond VaR



DETAILED INTERPRETATION

VaR is the loss not exceeded with 95% probability (as a positive number). n = 300, skew = -0.11, excess kurtosis = 1.40. The lead result here reads Method = Historical; VaR = 0.0586; Interpretation = Worst 5% empirical los. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily returns of four assets — stock, bond, gold and crypto, this lets the analyst understand the dynamics driving a monitored metric.

4.33 Wavelet coherence

LIVE OUTPUT

Group: Time series · Dataset: daily returns of four assets — stock, bond, gold and crypto

WHAT THIS DOES

Time-frequency co-movement of two series across short/medium/long horizons. It belongs to the **Time series** family, used to characterise temporal structure, trend and dependence in a series. In this example it is applied to daily returns of four assets — stock, bond, gold and crypto, drawing on ret_stock, ret_bond. Analysts reach for it to understand the dynamics driving a monitored metric.

SAMPLE DATA INPUT

ret_stock	ret_bond
0.019	0.002
0.032	0.036
-0.046	-0.036
-0.002	0.015
0.019	0.029

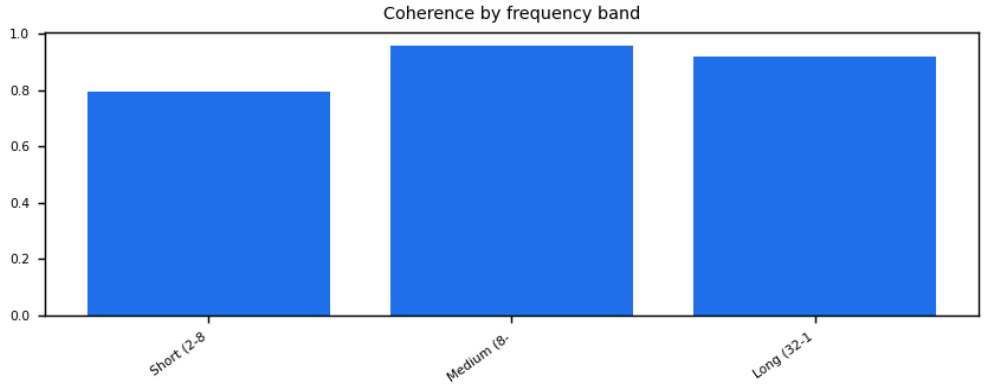
STEPS

1. Statistics → **Time series** → **Wavelet coherence**. 2. Select the demo dataset above. 3. Set parameters — `series1=ret_stock, series2=ret_bond`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Wavelet coherence — ret_stock vs ret_bond

Frequency band (period)	Mean coherence
Short (2-8)	0.7924
Medium (8-32)	0.9575
Long (32-128)	0.9176



DETAILED INTERPRETATION

Overall mean coherence = 0.919. Coherence (0-1) measures local co-movement at each frequency; high long-period coherence = the series move together over long horizons (Morlet continuous wavelet, 300 observations). The lead result here reads Frequency band (period) = Short (2-8); Mean coherence = 0.7924. Read the actual-vs-fitted overlay and the diagnostic statistics for autocorrelation, stationarity or lead-lag structure before trusting anything built on the series. For daily returns of four assets — stock, bond, gold and crypto, this lets the analyst understand the dynamics driving a monitored metric.

5. Panel & Longitudinal Data

5.1 Dynamic-panel GMM (Arellano-Bond)

LIVE OUTPUT

Group: Panel data · Dataset: a firm-year panel — return on assets, leverage and size

WHAT THIS DOES

One-step difference GMM for dynamic panels with Sargan & AR(2) tests. It belongs to the **Panel data** family, used to exploit repeated observations on the same units over time. In this example it is applied to a firm-year panel — return on assets, leverage and size, drawing on firm, year, roa, leverage, firm_size. Analysts reach for it to separate genuine drivers from fixed unit differences.

SAMPLE DATA INPUT

firm	year	roa	leverage	firm_size
Firm000	2015	-0.141	-0.248	8.42
Firm000	2016	0.036	0.11	7.447
Firm000	2017	1.214	0.749	9.635
Firm000	2018	0.687	-1.233	7.042
Firm000	2019	0.678	0.203	6.268

STEPS

1. Statistics → **Panel data** → **Dynamic-panel GMM (Arellano-Bond)**. 2. Select the demo dataset above. 3. Set parameters — `entity=firm, time=year, dependent=roa, predictors=leverage, firm_size`. 4. Click **Run** and read the tables/chart below.

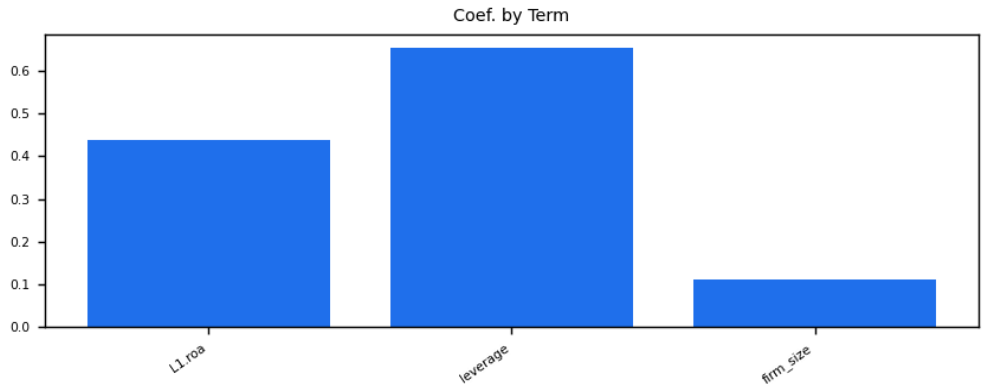
OUTPUT

Dynamic-panel GMM (Arellano-Bond, one-step difference)

Term	Coef.	Robust SE	z	p
L1.roa	0.4389	0.0526	8.341	< .001
leverage	0.654	0.0195	33.53	< .001
firm_size	0.1105	0.0172	6.432	< .001

Specification diagnostics

Test	Value
Instruments	23
Groups (entities)	120
Periods	8
Sargan over-identification…	3.091
Sargan df	20
Sargan p	1
AR(2) residual test p	< .001



DETAILED INTERPRETATION

y_it = ρ·y_i,t-1 + β·x_it + η_i + ε_it, first-differenced with GMM-style lagged-level instruments. n = 120 entities × 8 periods. The lead result here reads Term = L1.roa; Coef. = 0.4389; Robust SE = 0.0526; z = 8.341. Panel

estimators net out unit- and time-specific effects; use the Hausman test to choose fixed vs random effects and read coefficients as within-unit relationships. For a firm-year panel — return on assets, leverage and size, this lets the analyst separate genuine drivers from fixed unit differences.

5.2 Fama-MacBeth regression

LIVE OUTPUT

Group: Panel data · Dataset: a state-year policy panel with staggered reform adoption

WHAT THIS DOES

Two-pass cross-sectional regression for priced factor premia (asset pricing). It belongs to the **Panel data** family, used to exploit repeated observations on the same units over time. In this example it is applied to a state-year policy panel with staggered reform adoption, drawing on stock_return, risk_factor, period. Analysts reach for it to separate genuine drivers from fixed unit differences.

SAMPLE DATA INPUT

stock_return	risk_factor	period
0.026	-1.656	1
-0.01	-1.656	2
-0.051	-1.656	3
-0.133	-1.656	4
-0.034	-1.656	5

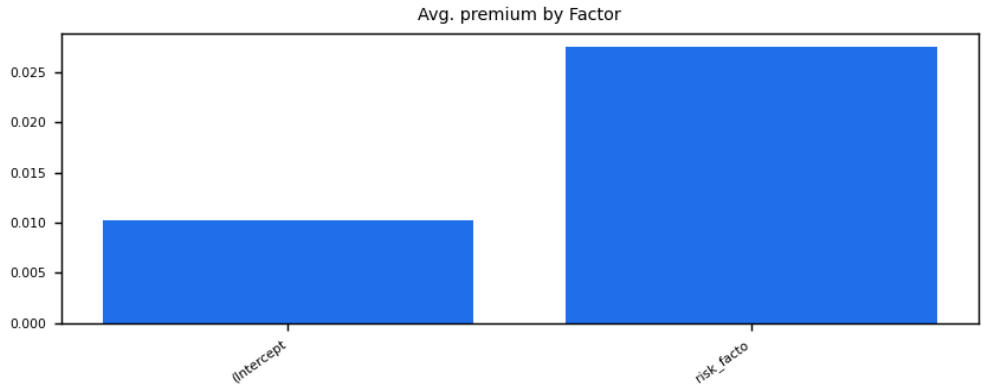
STEPS

1. Statistics → **Panel data** → **Fama-MacBeth regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=stock_return, predictors=risk_factor, time=period`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Fama-MacBeth two-pass regression

Factor	Avg. premium	FM SE	t	p
(Intercept / avg return)	0.0102	0.0018	5.834	< .001
risk_factor	0.0275	0.0014	19.14	< .001



DETAILED INTERPRETATION

Period-by-period cross-sectional regressions averaged over 8 periods; standard errors from the time-series of the coefficients. A significant factor premium is priced. The lead result here reads Factor = (Intercept / avg retur; Avg. premium = 0.0102; FM SE = 0.0018; t = 5.834. Panel estimators net out unit- and time-specific effects; use the Hausman test to choose fixed vs random effects and read coefficients as within-unit relationships. For a state-year policy panel with staggered reform adoption, this lets the analyst separate genuine drivers from fixed unit differences.

5.3 Panel: Hausman test

LIVE OUTPUT

Group: Panel data · Dataset: a firm-year panel — return on assets, leverage and size

WHAT THIS DOES

Fixed vs random effects specification test. It belongs to the **Panel data** family, used to exploit repeated observations on the same units over time. In this example it is applied to a firm-year panel — return on assets, leverage and size, drawing on firm, year, roa, leverage, firm_size. Analysts reach for it to separate genuine drivers from fixed unit differences.

SAMPLE DATA INPUT

firm	year	roa	leverage	firm_size
Firm000	2015	-0.141	-0.248	8.42
Firm000	2016	0.036	0.11	7.447
Firm000	2017	1.214	0.749	9.635
Firm000	2018	0.687	-1.233	7.042
Firm000	2019	0.678	0.203	6.268

STEPS

1. Statistics → **Panel data** → **Panel: Hausman test**. 2. Select the demo dataset above. 3. Set parameters — `entity=firm, time=year, dependent=roa, predictors=leverage, firm_size`. 4. Click **Run** and read the tables/chart below.

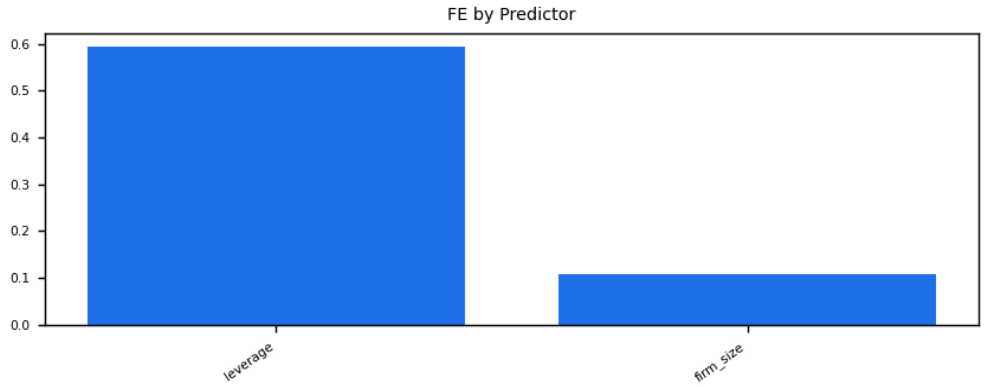
OUTPUT

Coefficient comparison

Predictor	FE	RE	Difference
leverage	0.5934	0.5989	-0.0056
firm_size	0.1078	0.1117	-0.0039

Hausman specification test (FE vs RE)

Chi-square	df	p
41.32	2	< .001



DETAILED INTERPRETATION

The lead result here reads Predictor = leverage; FE = 0.5934; RE = 0.5989; Difference = -0.0056. Panel estimators net out unit- and time-specific effects; use the Hausman test to choose fixed vs random effects and read coefficients as within-unit relationships. For a firm-year panel — return on assets, leverage and size, this lets the analyst separate genuine drivers from fixed unit differences.

5.4 Panel: fixed effects

LIVE OUTPUT

Group: Panel data · Dataset: a firm-year panel — return on assets, leverage and size

WHAT THIS DOES

Within (fixed-effects) regression with entity/time effects (Stata xtreg,fe). It belongs to the **Panel data** family, used to exploit repeated observations on the same units over time. In this example it is applied to a firm-year panel — return on assets, leverage and size, drawing on firm, year, roa, leverage, firm_size. Analysts reach for it to separate genuine drivers from fixed unit differences.

SAMPLE DATA INPUT

firm	year	roa	leverage	firm_size
Firm000	2015	-0.141	-0.248	8.42
Firm000	2016	0.036	0.11	7.447
Firm000	2017	1.214	0.749	9.635
Firm000	2018	0.687	-1.233	7.042
Firm000	2019	0.678	0.203	6.268

STEPS

1. Statistics → **Panel data** → **Panel: fixed effects**. 2. Select the demo dataset above. 3. Set parameters — `entity=firm, time=year, dependent=roa, predictors=leverage, firm_size`. 4. Click **Run** and read the tables/chart below.

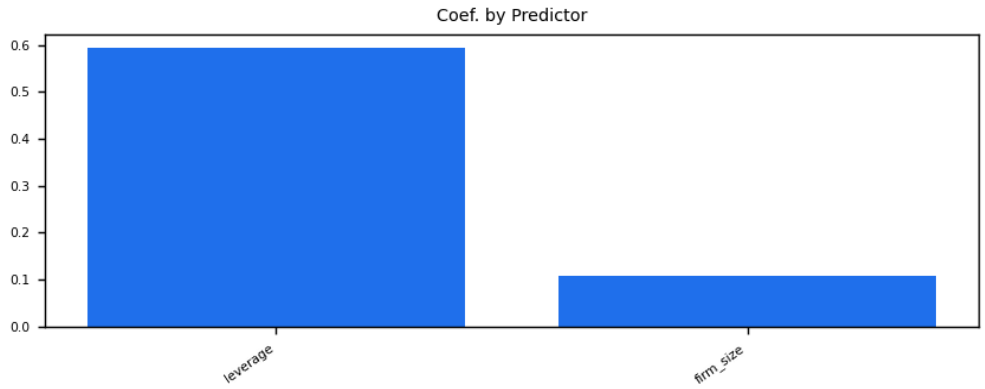
OUTPUT

Fixed-effects (within) regression — roa

Predictor	Coef.	Clustered SE	t	p	CI low	CI high
leverage	0.5934	0.0215	27.57	< .001	0.5511	0.6356
firm_size	0.1078	0.0223	4.826	< .001	0.064	0.1517

Panel structure

Property	Value
Observations	960
Entities	120
Time periods	8
F-statistic	404.1
F p-value	< .001



DETAILED INTERPRETATION

Entity effects; SE clustered by entity. Within R^2 = 0.4909 · overall R^2 = 0.5882. The lead result here reads Predictor = leverage; Coef. = 0.5934; Clustered SE = 0.0215; t = 27.57. Panel estimators net out unit- and time-specific effects; use the Hausman test to choose fixed vs random effects and read coefficients as within-unit relationships. For a firm-year panel — return on assets, leverage and size, this lets the analyst separate genuine drivers from fixed unit differences.

5.5 Panel: random effects

LIVE OUTPUT

Group: Panel data · Dataset: a firm-year panel — return on assets, leverage and size

WHAT THIS DOES

Random-effects GLS panel regression (Stata xtreg,re). It belongs to the **Panel data** family, used to exploit repeated observations on the same units over time. In this example it is applied to a firm-year panel — return on assets, leverage and size, drawing on firm, year, roa, leverage, firm_size. Analysts reach for it to separate genuine drivers from fixed unit differences.

SAMPLE DATA INPUT

firm	year	roa	leverage	firm_size
Firm000	2015	-0.141	-0.248	8.42
Firm000	2016	0.036	0.11	7.447
Firm000	2017	1.214	0.749	9.635
Firm000	2018	0.687	-1.233	7.042
Firm000	2019	0.678	0.203	6.268

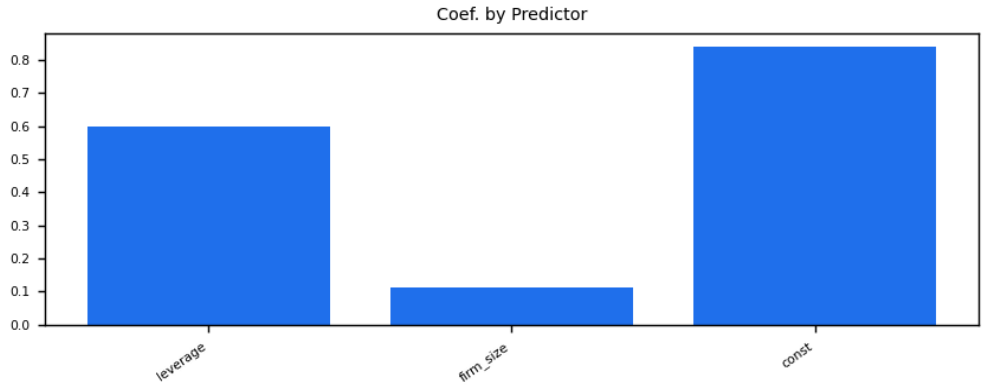
STEPS

1. Statistics → **Panel data** → **Panel: random effects**. 2. Select the demo dataset above. 3. Set parameters — `entity=firm, time=year, dependent=roa, predictors=leverage, firm_size`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Random-effects (GLS) regression — roa

Predictor	Coef.	SE	z	p	CI low	CI high
leverage	0.5989	0.0214	27.96	< .001	0.5569	0.641
firm_size	0.1117	0.0205	5.444	< .001	0.0714	0.1519
const	0.8383	0.1826	4.59	< .001	0.4799	1.197



DETAILED INTERPRETATION

Overall R^2 = 0.2846 · between R^2 = 0.1601. Random effects assume the entity effect is uncorrelated with the regressors (check with Hausman). The lead result here reads Predictor = leverage; Coef. = 0.5989; SE = 0.0214; z = 27.96. Panel estimators net out unit- and time-specific effects; use the Hausman test to choose fixed vs random effects and read coefficients as within-unit relationships. For a firm-year panel — return on assets, leverage and size, this lets the analyst separate genuine drivers from fixed unit differences.

6. Causal Inference & Policy Evaluation

6.1 Difference-in-Differences

INPUT REQUIREMENT

Group: Causal inference · Dataset: a state-year policy panel with staggered reform adoption

WHAT THIS DOES

Causal effect from treated×post interaction (parallel-trends design). It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a state-year policy panel with staggered reform adoption, drawing on gdp_growth, period, state, reform_year. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

gdp_growth	period	state	reform_year
2.046	1	State00	3
1.891	2	State00	3
4.443	3	State00	3
4.286	4	State00	3
5.823	5	State00	3

STEPS

1. Statistics → **Causal inference** → **Difference-in-Differences**. 2. Select the demo dataset above. 3. Set parameters — `outcome=gdp_growth, time=period, treatment_time=5, unit=state, cohort=reform_year`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Missing required parameter: treated

DETAILED INTERPRETATION

This technique has a specific data requirement: *Missing required parameter: treated*. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion.. Supplied with a suitable dataset for a state-year policy panel with staggered reform adoption, it returns the standard output for a causal inference method and supports the decision to attribute an outcome change to an intervention, not confounding.

6.2 Double-LASSO (causal effect)

LIVE OUTPUT

Group: Causal inference · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Post-double-selection LASSO — unbiased treatment effect with high-dimensional controls. It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, monthly_charges, tenure_months, support_calls. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

monetary	monthly_charges	tenure_months	support_calls
3197	52.94	68	1
1233	113.7	45	5
2501	35.51	49	1
1468	71.45	64	2
4195	29.16	42	1

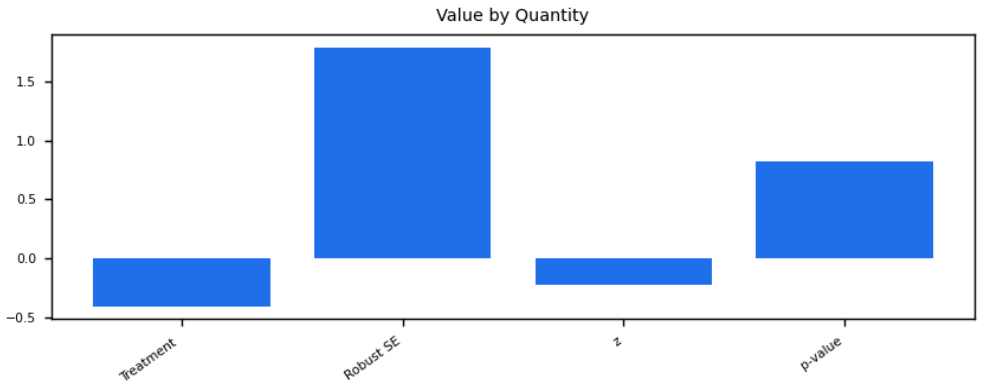
STEPS

1. Statistics → **Causal inference** → **Double-LASSO (causal effect)**. 2. Select the demo dataset above. 3. Set parameters — `outcome=monetary, treatment=monthly_charges, controls=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Post-double-selection LASSO — effect of monthly_charges on monetary

Quantity	Value
Treatment effect	-0.4078
Robust SE	1.79
z	-0.228
p-value	0.8198
95% CI	[-3.9159, 3.1003]
Controls selected	1 of 2



DETAILED INTERPRETATION

Belloni–Chernozhukov–Hansen double-selection: LASSO selects controls predictive of BOTH the outcome and the treatment, then OLS estimates an unbiased treatment effect with valid inference (robust to model-selection error). The lead result here reads Quantity = Treatment effect; Value = -0.4078. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst attribute an outcome change to an intervention, not confounding.

6.3 Event study (dynamic DiD)

LIVE OUTPUT

Group: Causal inference · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Leads/lags around treatment — tests parallel trends & dynamic effects. It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, monthly_charges, tenure_months, support_calls. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

monetary	monthly_charges	tenure_months	support_calls
3197	52.94	68	1
1233	113.7	45	5
2501	35.51	49	1
1468	71.45	64	2
4195	29.16	42	1

STEPS

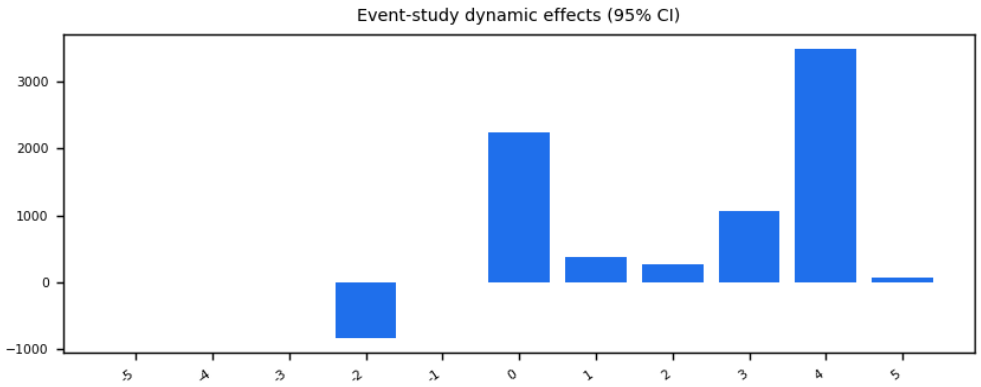
1. Statistics → **Causal inference** → **Event study (dynamic DiD)**. 2. Select the demo dataset above. 3. Set parameters — `outcome=monetary, unit=monthly_charges, time=tenure_months, cohort=support_calls, window=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Event study — monetary

Rel. period	Coef.	SE	p	CI low	CI high
-5	0				

Rel. period	Coef.	SE	p	CI low	CI high
-4	0				
-3	0				
-2	-842.8				
0	2249				
1	377.2				
2	262.2				
3	1067				
4	3501				
5	71.14				



DETAILED INTERPRETATION

Coefficients are relative to period −1 (reference). Leads (negative periods) should be ≈0 for the parallel-trends assumption to hold; pre-period effects look flat — supports parallel trends. The lead result here reads Rel. period = -5; Coef. = 0; SE = ; p = . The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst attribute an outcome change to an intervention, not confounding.

6.4 Mediation analysis (bootstrap)

LIVE OUTPUT

Group: Causal inference · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Hayes model 4 — indirect effect a·b with bootstrap CI. It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, pu1, int1. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

peou1	pu1	int1
3	4	4
3	3	3
4	5	5
5	3	5
2	3	2

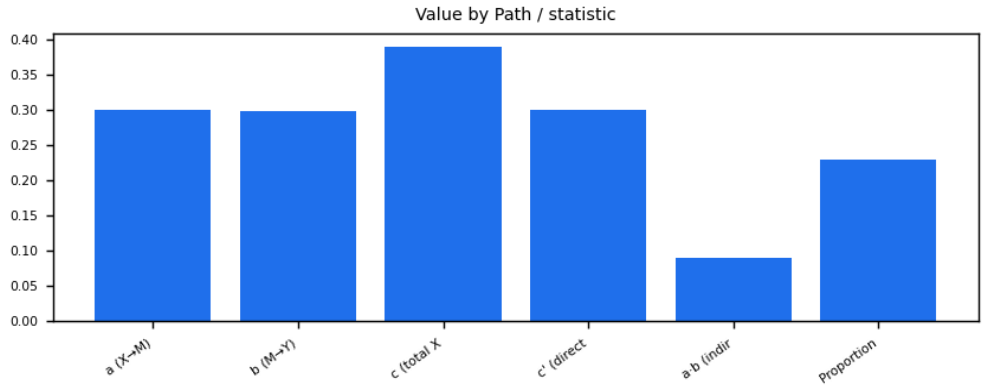
STEPS

1. Statistics → **Causal inference** → **Mediation analysis (bootstrap)**. 2. Select the demo dataset above. 3. Set parameters — `x=peou1, mediator=pu1, y=int1, covariates=, bootstrap=200`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Mediation: peou1 → pu1 → int1

Path / statistic	Value
a (X→M)	0.3006
b (M→Y)	0.2977
c (total X→Y)	0.3897
c' (direct X→Y)	0.3002
a·b (indirect)	0.0895
Indirect 95% bootstrap CI	[0.0407, 0.1408]
Proportion mediated	0.23



DETAILED INTERPRETATION

Simple mediation (Hayes model 4), 200 bootstrap resamples. The indirect effect a·b is significant (bias-corrected percentile CI excludes 0). The lead result here reads Path / statistic = a (X→M); Value = 0.3006. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst attribute an outcome change to an intervention, not confounding.

6.5 Moderation analysis

LIVE OUTPUT

Group: Causal inference · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Interaction X×W with simple slopes (mean-centred). It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, pu1, int1. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

peou1	pu1	int1
3	4	4
3	3	3
4	5	5
5	3	5
2	3	2

STEPS

1. Statistics → **Causal inference** → **Moderation analysis**. 2. Select the demo dataset above. 3. Set parameters — `x=peou1, moderator=pu1, y=int1, covariates=`. 4. Click **Run** and read the tables/chart below.

OUTPUT

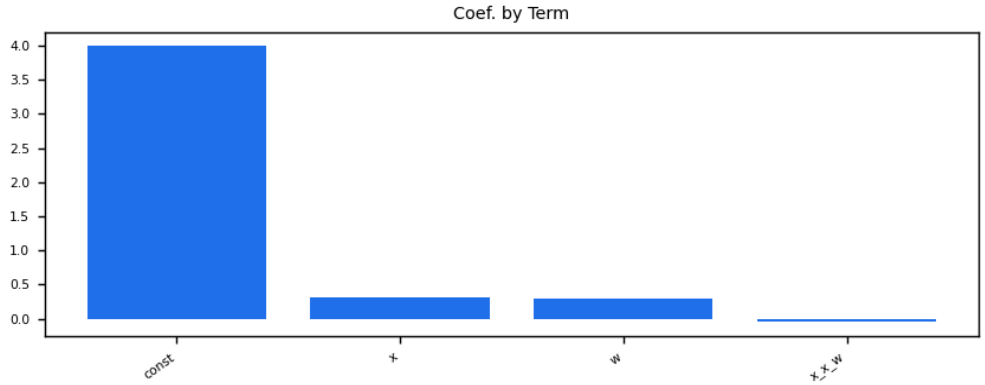
Moderation: peou1 × pu1 → int1

Term	Coef.	SE	t	p
const	3.995	0.0746	53.52	< .001
x	0.3088	0.065	4.751	< .001
w	0.2973	0.0698	4.258	< .001

Term	Coef.	SE	t	p
x_x_w	-0.0453	0.0505	-0.896	0.3703

Simple slopes (effect of X)

Moderator level	Slope
Low moderator (–1 SD)	0.3628
Mean moderator	0.3088
High moderator (+1 SD)	0.2548



DETAILED INTERPRETATION

Moderation: interaction peou1×pu1 coef = -0.0453, p=0.3703. Predictors mean-centred. Simple slopes show the effect of X at low/mean/high moderator. The lead result here reads Term = const; Coef. = 3.995; SE = 0.0746; t = 53.52. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst attribute an outcome change to an intervention, not confounding.

6.6 Propensity-score matching / IPW

LIVE OUTPUT

Group: Causal inference · Dataset: a real-world patient cohort — treatment, recovery and confounders

WHAT THIS DOES

ATT via nearest-neighbour matching + IPW ATE + covariate balance. It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a real-world patient cohort — treatment, recovery and confounders, drawing on recovered, treated, age, comorbidities, baseline_severity. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

recovered	treated	age	comorbidities	baseline_severity
0	1	65	1	6.38
0	0	76	2	3.94
1	1	51	2	5.46
0	0	30	3	4.65
0	0	59	1	3.59

STEPS

1. Statistics → **Causal inference** → **Propensity-score matching / IPW**. 2. Select the demo dataset above. 3. Set parameters — `outcome=recovered, treatment=treated, covariates=age, comorbidities, baseline_severity`. 4. Click **Run** and read the tables/chart below.

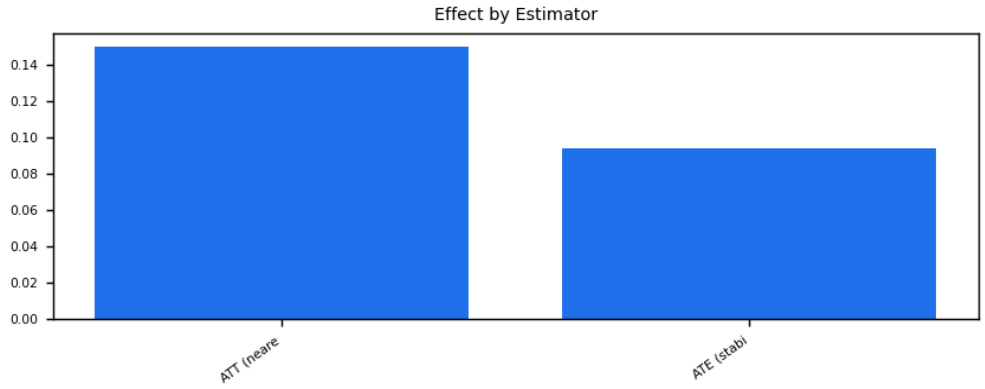
OUTPUT

Treatment effect on recovered

Estimator	Effect	n treated	n matched
ATT (nearest-neighbour PSM…	0.1497	170	167
ATE (stabilized IPW)	0.0935	170	130

Covariate balance (standardised mean difference)

Covariate	SMD before	SMD after	Verdict
age	0.638	-0.028	balanced
comorbidities	0.323	-0.133	imbalanced
baseline_severity	0.33	0.166	imbalanced



DETAILED INTERPRETATION

Propensity scores from logistic regression on the covariates. ATT compares each treated unit to its nearest-propensity control (caliper 0.2·SD of the logit). The lead result here reads Estimator = ATT (nearest-neighbour; Effect = 0.1497; n treated = 170; n matched = 167. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion. For a real-world patient cohort — treatment, recovery and confounders, this lets the analyst attribute an outcome change to an intervention, not confounding.

6.7 Regression discontinuity

INPUT REQUIREMENT

Group: Causal inference · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Local-linear LATE at a cutoff (sharp RD, triangular kernel). It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, monthly_charges. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

monetary	monthly_charges
3197	52.94
1233	113.7
2501	35.51
1468	71.45
4195	29.16

STEPS

1. Statistics → **Causal inference** → **Regression discontinuity**. 2. Select the demo dataset above. 3. Set parameters — `outcome=monetary, running=monthly_charges, cutoff=5, bandwidth=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Too few observations within the bandwidth — widen it

DETAILED INTERPRETATION

This technique has a specific data requirement: *Too few observations within the bandwidth — widen it*. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a causal inference method and supports the decision to attribute an outcome change to an intervention, not confounding.

6.8 Staggered DiD (Callaway-Sant'Anna)

LIVE OUTPUT

Group: Causal inference · Dataset: a state-year policy panel with staggered reform adoption

WHAT THIS DOES

Heterogeneity-robust staggered difference-in-differences with event study. It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a state-year policy panel with staggered reform adoption, drawing on gdp_growth, state, period, reform_year. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

gdp_growth	state	period	reform_year
2.046	State00	1	3
1.891	State00	2	3
4.443	State00	3	3
4.286	State00	4	3
5.823	State00	5	3

STEPS

1. Statistics → **Causal inference** → **Staggered DiD (Callaway-Sant'Anna)**. 2. Select the demo dataset above. 3. Set parameters — `outcome=gdp_growth, unit=state, time=period, cohort=reform_year`. 4. Click **Run** and read the tables/chart below.

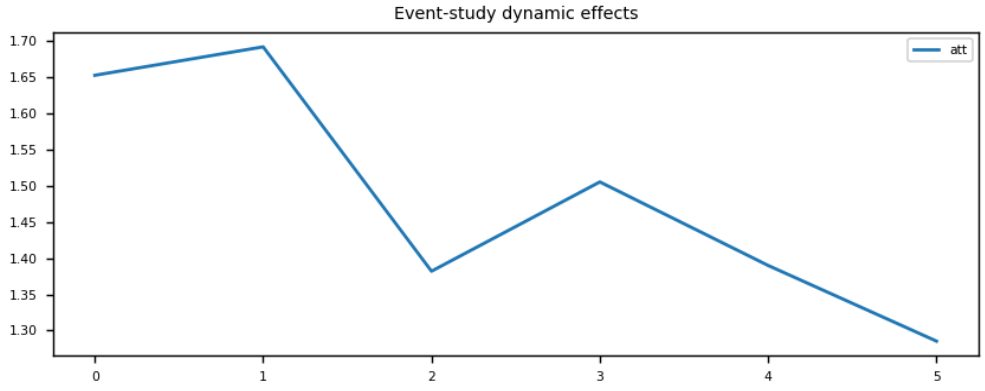
OUTPUT

Staggered DiD (Callaway-Sant'Anna) — overall effect

Overall ATT	Bootstrap SE	z	p	Cohorts	Group-time cells
1.535	0.1059	14.5	< .001	3	12

Event study (dynamic effects by time relative to treatment)

Relative period (e)	ATT(e)	Cells
0	1.653	3
1	1.692	3
2	1.382	2
3	1.506	2
4	1.39	1
5	1.285	1



DETAILED INTERPRETATION

Aggregated ATT across treatment cohorts using never-treated (or not-yet-treated) controls — robust to the negative-weighting problem of two-way fixed-effects DiD. The lead result here reads Overall ATT = 1.535; Bootstrap SE = 0.1059; z = 14.5; p = < .001. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion. For a state-year policy panel with staggered reform adoption, this lets the analyst attribute an outcome change to an intervention, not confounding.

6.9 Synthetic control

INPUT REQUIREMENT

Group: Causal inference · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Comparative case study: weighted donor pool vs treated unit. It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, monthly_charges, tenure_months. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

monetary	monthly_charges	tenure_months
3197	52.94	68
1233	113.7	45
2501	35.51	49
1468	71.45	64
4195	29.16	42

STEPS

1. Statistics → **Causal inference** → **Synthetic control**. 2. Select the demo dataset above. 3. Set parameters — `outcome=monetary, unit=monthly_charges, time=tenure_months, treated_unit=, treatment_time=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Missing required parameter: treated_unit

DETAILED INTERPRETATION

This technique has a specific data requirement: *Missing required parameter: treated_unit*. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a causal inference method and supports the decision to attribute an outcome change to an intervention, not confounding.

6.10 Target-trial emulation (RWE / IPTW)

LIVE OUTPUT

Group: Causal inference · Dataset: a real-world patient cohort — treatment, recovery and confounders

WHAT THIS DOES

Real-world-evidence treatment effect via inverse-probability weighting + balance. It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a real-world patient cohort — treatment, recovery and confounders, drawing on recovered, treated, age, comorbidities, baseline_severity. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

recovered	treated	age	comorbidities	baseline_severity
0	1	65	1	6.38
0	0	76	2	3.94
1	1	51	2	5.46
0	0	30	3	4.65
0	0	59	1	3.59

STEPS

1. Statistics → **Causal inference** → **Target-trial emulation (RWE / IPTW)**. 2. Select the demo dataset above. 3. Set parameters — `outcome=recovered, treatment=treated, covariates=age, comorbidities, baseline_severity`. 4. Click **Run** and read the tables/chart below.

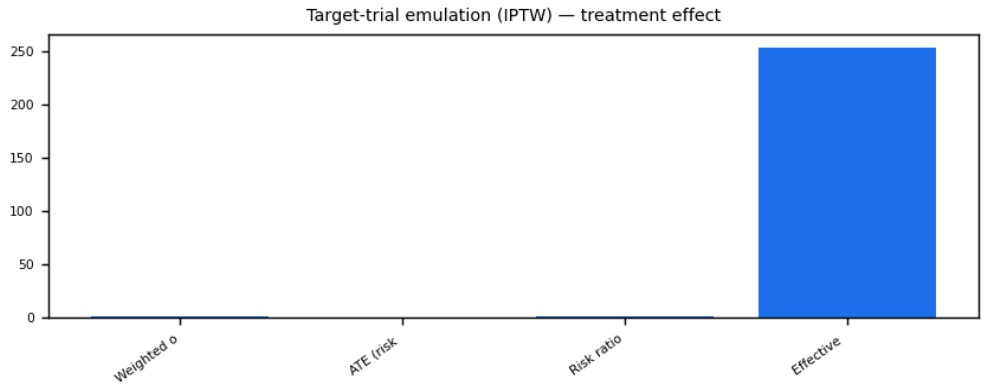
OUTPUT

Target-trial emulation (IPTW) — treatment effect

Weighted outcome (treated)	Weighted outcome (control)	ATE (risk difference)	Risk ratio	Effective N
0.3714	0.2778	0.0935	1.337	253.4

Covariate balance (standardised mean difference)

Covariate	SMD before	SMD after	Status
age	0.6405	0.0135	balanced
comorbidities	0.3246	-0.0072	balanced
baseline_severity	0.3311	0.0718	balanced



DETAILED INTERPRETATION

Inverse-probability-of-treatment weighting estimates the average treatment effect from observational (real-world) data by balancing measured confounders. The lead result here reads Weighted outcome (treated) = 0.3714; Weighted outcome (control) = 0.2778; ATE (risk difference) = 0.0935; Risk ratio = 1.337. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion. For a real-world patient cohort — treatment, recovery and confounders, this lets the analyst attribute an outcome change to an intervention, not confounding.

6.11 Uplift modeling (treatment effect)

LIVE OUTPUT

Group: Causal inference · Dataset: a real-world patient cohort — treatment, recovery and confounders

WHAT THIS DOES

Two-model uplift: incremental response of a treatment per unit. It belongs to the **Causal inference** family, used to estimate the causal effect of a treatment or policy net of confounding. In this example it is applied to a real-world patient cohort — treatment, recovery and confounders, drawing on recovered, treated, age, comorbidities, baseline_severity. Analysts reach for it to attribute an outcome change to an intervention, not confounding.

SAMPLE DATA INPUT

recovered	treated	age	comorbidities	baseline_severity
0	1	65	1	6.38
0	0	76	2	3.94
1	1	51	2	5.46
0	0	30	3	4.65
0	0	59	1	3.59

STEPS

1. Statistics → **Causal inference** → **Uplift modeling (treatment effect)**. 2. Select the demo dataset above. 3. Set parameters — `outcome=recovered, treatment=treated, predictors=age, comorbidities, baseline_severity`. 4. Click **Run** and read the tables/chart below.

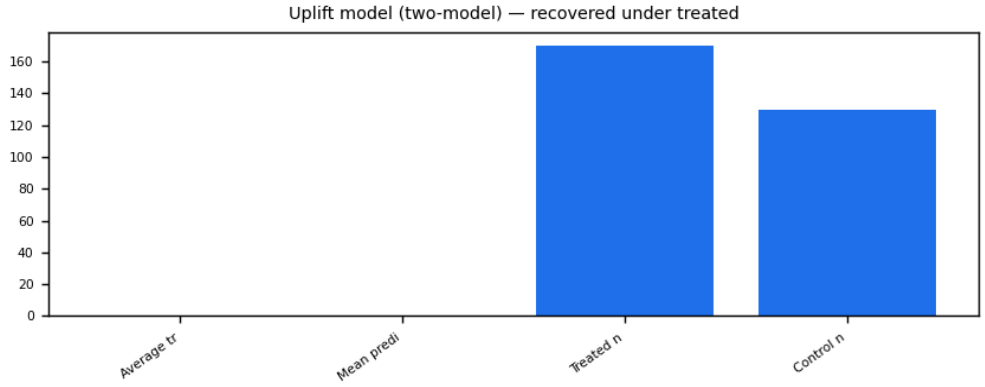
OUTPUT

Uplift model (two-model) — recovered under treated

Average treatment effect (observed)	Mean predicted uplift	Treated n	Control n
0.0104	0.1018	170	130

Uplift by decile (1 = highest predicted uplift)

Decile	N	Mean predicted uplift	Observed uplift
1	30	0.1326	-0.0667
2	30	0.1217	0.1111
3	30	0.1137	0.086
4	30	0.1088	0.3333
5	30	0.1036	-0.0452
6	30	0.0985	0.267
7	30	0.094	-0.4167
8	30	0.0892	-0.0833
9	30	0.0837	0.0476
10	30	0.0724	-0.08



DETAILED INTERPRETATION

Two-model (T-learner) uplift: separate outcome models for treated and control; uplift = P(y|treat) – P(y|control) per unit. The lead result here reads Average treatment effect (observed) = 0.0104; Mean predicted uplift = 0.1018; Treated n = 170; Control n = 130. The estimate is the causal impact after adjusting for confounding - inspect balance, parallel-trends or robustness diagnostics before drawing a policy conclusion. For a real-world patient cohort — treatment, recovery and confounders, this lets the analyst attribute an outcome change to an intervention, not confounding.

7. Bayesian Methods

7.1 Bayesian SEM (MCMC)

LIVE OUTPUT

Group: Bayesian · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Full Gibbs-MCMC path/mediation model with posterior credible intervals. It belongs to the **Bayesian** family, used to estimate parameters as posterior distributions with credible intervals. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, pu1, int1. Analysts reach for it to reason about uncertainty directly in probability terms.

SAMPLE DATA INPUT

peou1	pu1	int1
3	4	4
3	3	3
4	5	5
5	3	5
2	3	2

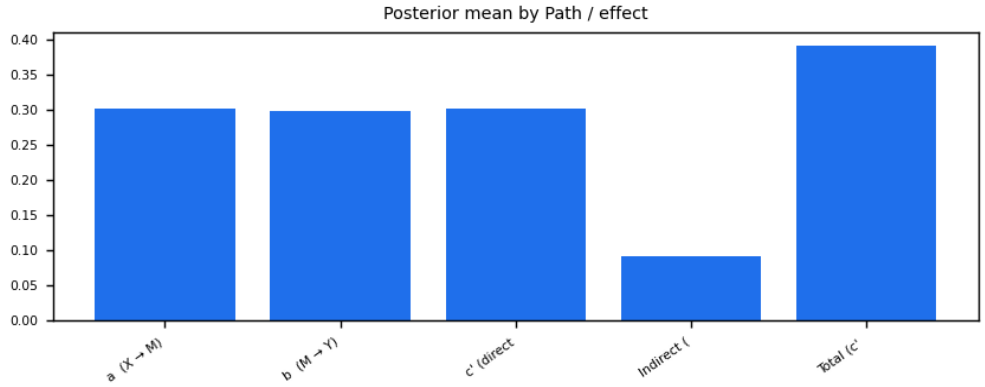
STEPS

1. Statistics → **Bayesian** → **Bayesian SEM (MCMC)**. 2. Select the demo dataset above. 3. Set parameters — `independent=peou1, mediator=pu1, outcome=int1`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Bayesian SEM / mediation — full MCMC (Gibbs sampler)

Path / effect	Posterior mean	SD	95% credible interval	P(>0)
a (X → M)	0.3019	0.0605	[0.1781, 0.4240]	1
b (M → Y)	0.2987	0.0597	[0.1843, 0.4143]	1
c' (direct X → Y)	0.3007	0.0583	[0.1824, 0.4091]	1
Indirect (a·b)	0.0902	0.0256	[0.0439, 0.1432]	1
Total (c' + a·b)	0.3909	0.0582	[0.2763, 0.5026]	1



DETAILED INTERPRETATION

Coefficients estimated by Gibbs MCMC (conjugate priors, 3,000 draws, 1,000 burn-in). A credible interval for the indirect effect (a·b) that excludes 0 provides Bayesian evidence for mediation. The lead result here reads Path / effect = a (X → M); Posterior mean = 0.3019; SD = 0.0605; 95% credible interval = [0.1781, 0.4240]. Posterior means and credible intervals summarise the plausible parameter values; a 95% credible interval that excludes zero is the Bayesian counterpart of significance. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst reason about uncertainty directly in probability terms.

7.2 Bayesian SEM / mediation

LIVE OUTPUT

Group: Bayesian · Dataset: a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items

WHAT THIS DOES

Path model with a Bayesian indirect-effect posterior & credible interval. It belongs to the **Bayesian** family, used to estimate parameters as posterior distributions with credible intervals. In this example it is applied to a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, drawing on peou1, pu1, int1. Analysts reach for it to reason about uncertainty directly in probability terms.

SAMPLE DATA INPUT

peou1	pu1	int1
3	4	4
3	3	3
4	5	5
5	3	5
2	3	2

STEPS

1. Statistics → **Bayesian** → **Bayesian SEM / mediation**. 2. Select the demo dataset above. 3. Set parameters — `independent=peou1, mediator=pu1, outcome=int1`. 4. Click **Run** and read the tables/chart below.

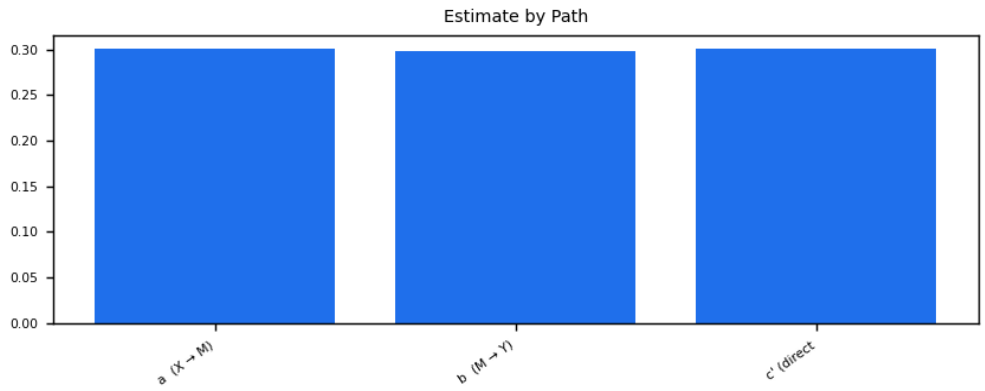
OUTPUT

Bayesian SEM / mediation — paths (peou1 → pu1 → int1)

Path	Estimate	SE	p
a (X → M)	0.3006	0.0601	< .001
b (M → Y)	0.2977	0.0614	< .001
c' (direct X → Y)	0.3002	0.0575	< .001

Effects with 95% credible interval

Effect	Posterior mean	95% credible interval	P(effect > 0)
Indirect (a·b)	0.0896	[0.0437, 0.1454]	1
Direct (c')	0.3002	—	—
Total (c' + a·b)	0.3897	—	—



DETAILED INTERPRETATION

A path model estimated equation-by-equation; posterior for the indirect effect below. The lead result here reads Path = a (X → M); Estimate = 0.3006; SE = 0.0601; p = < .001. Posterior means and credible intervals summarise the plausible parameter values; a 95% credible interval that excludes zero is the Bayesian counterpart of significance. For a technology-acceptance survey — perceived ease-of-use (peou), usefulness (pu) and intention (int) items, this lets the analyst reason about uncertainty directly in probability terms.

7.3 Bayesian contingency table

LIVE OUTPUT

Group: Bayesian · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Bayes factor for association vs. independence in a two-way table. It belongs to the **Bayesian** family, used to estimate parameters as posterior distributions with credible intervals. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on contract, monthly_charges. Analysts reach for it to reason about uncertainty directly in probability terms.

SAMPLE DATA INPUT

contract	monthly_charges
Month-to-month	52.94
One year	113.7
Month-to-month	35.51
Two year	71.45
Two year	29.16

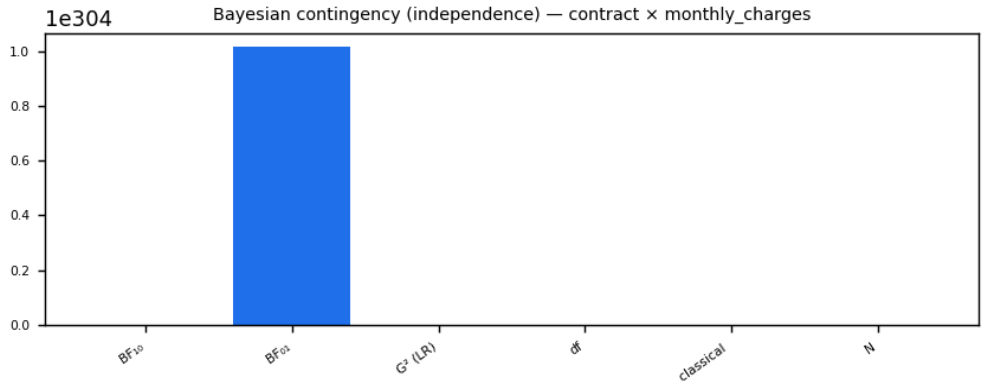
STEPS

1. Statistics → **Bayesian** → **Bayesian contingency table**. 2. Select the demo dataset above. 3. Set parameters — `row=contract, column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Bayesian contingency (independence) — contract × monthly_charges

BF■■	BF■■	Evidence	G² (LR)	df	classical p	N
0	1.014e+304	extreme evidence for H■ (n…	1556	1550	0.4489	800



DETAILED INTERPRETATION

BF■■ = evidence for association vs. independence (BIC approximation to the loglinear Bayes factor). The lead result here reads BF■■ = 0; BF■■ = 1.014e+304; Evidence = extreme evidence for H; G² (LR) = 1556. Posterior means and credible intervals summarise the plausible parameter values; a 95% credible interval that excludes zero is the Bayesian counterpart of significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reason about uncertainty directly in probability terms.

7.4 Bayesian correlation (JZS)

LIVE OUTPUT

Group: Bayesian · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

JZS Bayes factor for a non-zero Pearson correlation. It belongs to the **Bayesian** family, used to estimate parameters as posterior distributions with credible intervals. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, monetary. Analysts reach for it to reason about uncertainty directly in probability terms.

SAMPLE DATA INPUT

monthly_charges	monetary
52.94	3197
113.7	1233
35.51	2501
71.45	1468
29.16	4195

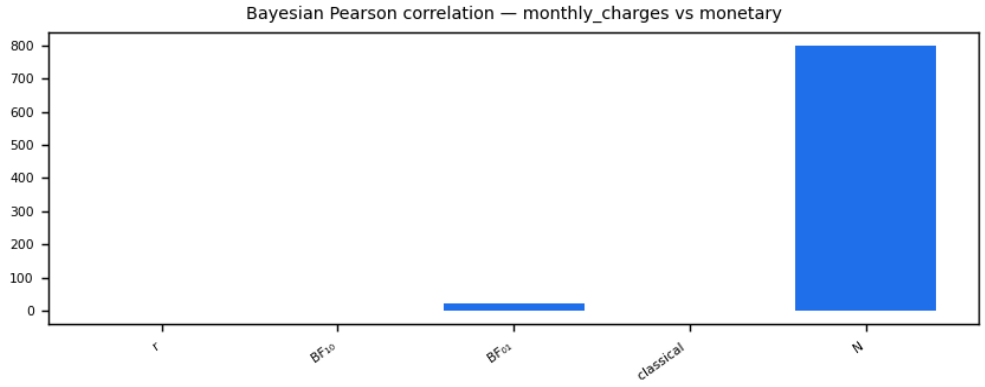
STEPS

1. Statistics → **Bayesian** → **Bayesian correlation (JZS)**. 2. Select the demo dataset above. 3. Set parameters — `column1=monthly_charges, column2=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Bayesian Pearson correlation — monthly_charges vs monetary

r	BF■■	BF■■	Evidence	classical p	N
-0.0103	0.0462	21.64	strong evidence for H■ (nu…	0.7707	800



DETAILED INTERPRETATION

JZS Bayes factor for a non-zero correlation (pingouin, default r-scale). The lead result here reads r = -0.0103; BF■■ = 0.0462; BF■■ = 21.64; Evidence = strong evidence for H■. Posterior means and credible intervals summarise the plausible parameter values; a 95% credible interval that excludes zero is the Bayesian counterpart of significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reason about uncertainty directly in probability terms.

7.5 Bayesian linear regression

LIVE OUTPUT

Group: Bayesian · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Bayes factor that predictors improve on the null model. It belongs to the **Bayesian** family, used to estimate parameters as posterior distributions with credible intervals. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to reason about uncertainty directly in probability terms.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

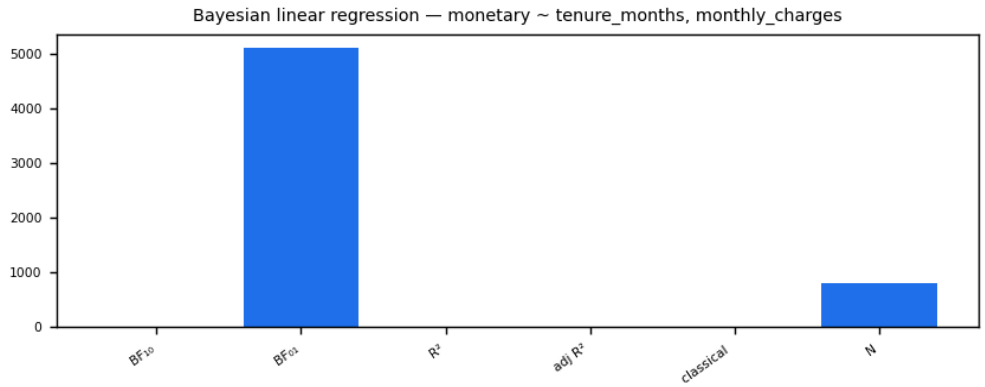
STEPS

1. Statistics → **Bayesian** → **Bayesian linear regression**. 2. Select the demo dataset above. 3. Set parameters — `dependent=monetary, predictors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Bayesian linear regression — monetary ~ tenure_months, monthly_charges, support_calls

BF■■	BF■■	Evidence	R^2	adj R^2	classical p	N
0.0002	5105	extreme evidence for H■ (n…	0.0037	-0	0.3971	800



DETAILED INTERPRETATION

BF■■ = evidence that the predictors jointly improve on an intercept-only model (BIC approximation). The lead result here reads BF■■ = 0.0002; BF■■ = 5105; Evidence = extreme evidence for H; R^2 = 0.0037. Posterior means and credible intervals summarise the plausible parameter values; a 95% credible interval that excludes zero is the Bayesian counterpart of significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reason about uncertainty directly in probability terms.

7.6 Bayesian one-way ANOVA

LIVE OUTPUT

Group: Bayesian · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Bayes factor for a group-mean difference (BIC/JZS approximation). It belongs to the **Bayesian** family, used to estimate parameters as posterior distributions with credible intervals. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to reason about uncertainty directly in probability terms.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

STEPS

1. Statistics → **Bayesian** → **Bayesian one-way ANOVA**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

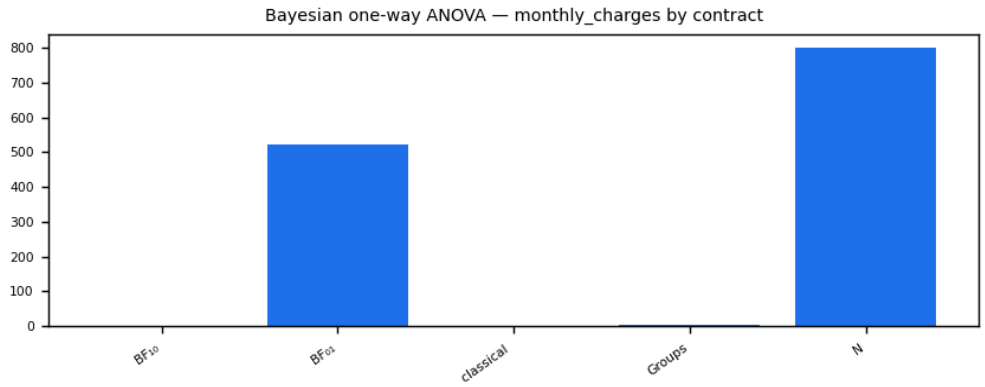
OUTPUT

Bayesian one-way ANOVA — monthly_charges by contract

BF■■	BF■■	Evidence	classical F	classical p	Groups	N
0.0019	521.3	extreme evidence for H■ (n…	0.4269	0.6527	3	800

Group means

contract	N	Mean
Month-to-month	447	68.26
One year	190	70.33
Two year	163	69.91



DETAILED INTERPRETATION

BF■■ = evidence for a group difference vs. the null of equal means (Schwarz/BIC approximation to the JZS Bayes factor). The lead result here reads BF■■ = 0.0019; BF■■ = 521.3; Evidence = extreme evidence for H; classical F = 0.4269. Posterior means and credible intervals summarise the plausible parameter values; a 95% credible interval that excludes zero is the Bayesian counterpart of significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reason about uncertainty directly in probability terms.

7.7 Bayesian repeated-measures ANOVA

LIVE OUTPUT

Group: Bayesian · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Bayes factor for a within-subject condition effect. It belongs to the **Bayesian** family, used to estimate parameters as posterior distributions with credible intervals. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to reason about uncertainty directly in probability terms.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

STEPS

1. Statistics → **Bayesian** → **Bayesian repeated-measures ANOVA**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

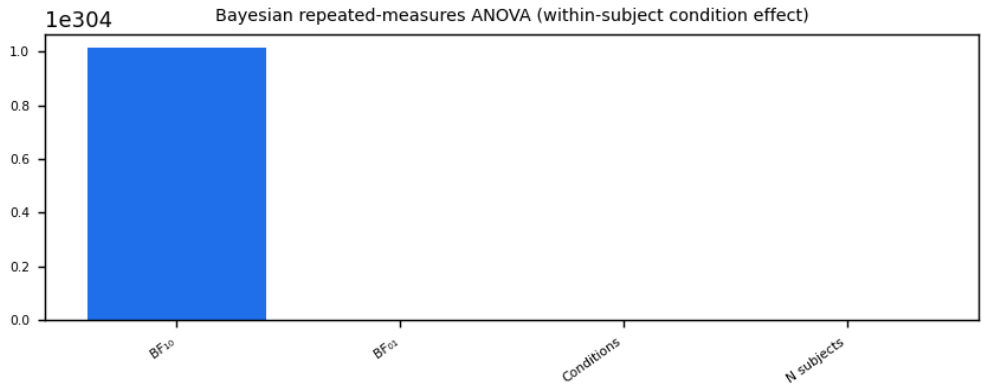
OUTPUT

Bayesian repeated-measures ANOVA (within-subject condition effect)

BF■■	BF■■	Evidence	Conditions	N subjects
1.014e+304	0	extreme evidence for H■ (e…	3	800

Condition means

Condition	Mean	SD
tenure_months	37.1	20.72
monthly_charges	69.09	28.74
support_calls	1.954	1.401



DETAILED INTERPRETATION

BF■■ = evidence for a condition effect (BIC approximation, subject held as a blocking factor). The lead result here reads BF■■ = 1.014e+304; BF■■ = 0; Evidence = extreme evidence for H; Conditions = 3. Posterior means and credible intervals summarise the plausible parameter values; a 95% credible interval that excludes zero is the Bayesian counterpart of significance. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst reason about uncertainty directly in probability terms.

8. Survival & Reliability Engineering

8.1 Accelerated life test (ALT)

LIVE OUTPUT

Group: Reliability · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Weibull AFT of life vs stress (Arrhenius/inverse-power) — extrapolate to use conditions. It belongs to the **Reliability** family, used to model product life and failure behaviour for engineering decisions. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, temp_c, event. Analysts reach for it to set warranty, maintenance and mission-life targets.

SAMPLE DATA INPUT

failure_hours	temp_c	event
1.013e+04	85	1
2214	125	1
2507	105	1
2572	105	1
2366	105	1

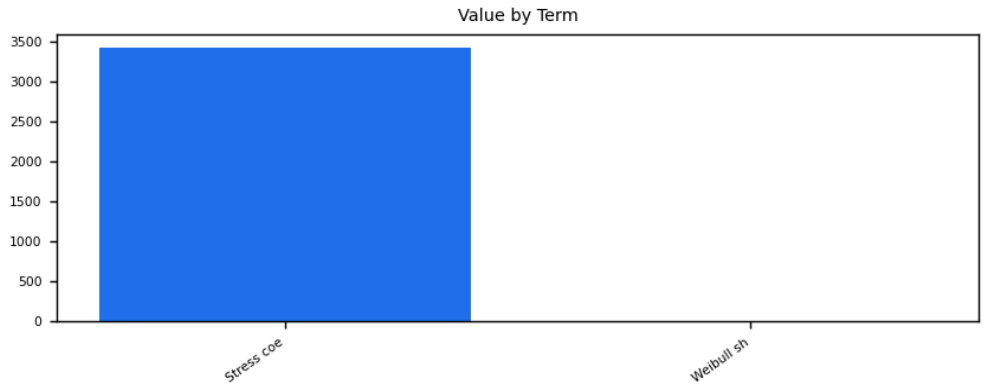
STEPS

1. Statistics → **Reliability** → **Accelerated life test (ALT)**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, stress=temp_c, event=event`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Accelerated life test — failure_hours vs temp_c

Term	Value
Model	arrhenius
Covariate	1/T (Arrhenius)
Stress coefficient (log-li…	3429
Weibull shape ρ	0.8819



DETAILED INTERPRETATION

Weibull accelerated-failure-time model regressing log-life on the transformed stress; extrapolate to use conditions with 'use_stress'. The lead result here reads Term = Model; Value = arrhenius. Fitted life-distribution parameters convert into reliability at a mission time and a characteristic life; a shape parameter above 1 indicates wear-out (rising hazard). For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst set warranty, maintenance and mission-life targets.

8.2 Degradation analysis

INPUT REQUIREMENT

Group: Reliability · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Extrapolate degradation paths to a failure threshold, then fit a Weibull life model. It belongs to the **Reliability** family, used to model product life and failure behaviour for engineering decisions. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on unit_id, temp_c, failure_hours. Analysts reach for it to set warranty, maintenance and mission-life targets.

SAMPLE DATA INPUT

unit_id	temp_c	failure_hours
U000	85	1.013e+04
U001	125	2214
U002	105	2507
U003	105	2572
U004	105	2366

STEPS

1. Statistics → **Reliability** → **Degradation analysis**. 2. Select the demo dataset above. 3. Set parameters — `unit=unit_id, time=temp_c, measurement=failure_hours, threshold=9000`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Could not extrapolate enough pseudo-failure times — check the threshold/direction

DETAILED INTERPRETATION

This technique has a specific data requirement: *Could not extrapolate enough pseudo-failure times — check the threshold/direction*. Fitted life-distribution parameters convert into reliability at a mission time and a characteristic life; a shape parameter above 1 indicates wear-out (rising hazard).. Supplied with a suitable dataset for an accelerated component life test — failure hours, censoring and temperature stress, it returns the standard output for a reliability method and supports the decision to set warranty, maintenance and mission-life targets.

8.3 Life distribution (multi-fit)

LIVE OUTPUT

Group: Reliability · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Fit & compare Weibull/lognormal/exponential/log-logistic by AIC, with profiler. It belongs to the **Reliability** family, used to model product life and failure behaviour for engineering decisions. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, event. Analysts reach for it to set warranty, maintenance and mission-life targets.

SAMPLE DATA INPUT

failure_hours	event
1.013e+04	1
2214	1
2507	1
2572	1
2366	1

STEPS

1. Statistics → **Reliability** → **Life distribution (multi-fit)**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, event=event`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Life-distribution comparison — failure_hours

Distribution	AIC	ΔAIC	log-likelihood	
Lognormal	2878	0	-1437	■ best fit

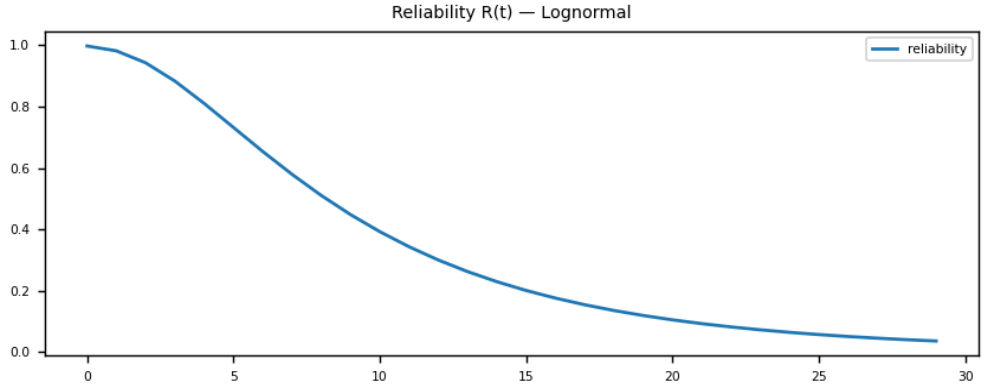
Distribution	AIC	ΔAIC	log-likelihood	
Log-logistic	2881	2.837	-1438	
Weibull	2884	5.675	-1440	
Exponential	2949	71.01	-1474	

Best fit: Lognormal — parameters (95% profile-likelihood CI)

Parameter	Estimate	CI low	CI high
mu_	8.43	8.331	8.528
sigma_	0.6331	0.5613	0.7049

Distribution profiler — Lognormal life quantiles

Quantile	Time
B1 (t where 1% fail)	1051
B10 (t where 10% fail)	2036
B50 (t where 50% fail)	4582
B90 (t where 90% fail)	1.031e+04



DETAILED INTERPRETATION

n = 160 (153 failures, 7 censored). Lower AIC = better. ΔAIC < 2 ≈ comparable fit. The lead result here reads Distribution = Lognormal; AIC = 2878; ΔAIC = 0; log-likelihood = -1437. Fitted life-distribution parameters convert into reliability at a mission time and a characteristic life; a shape parameter above 1 indicates wear-out (rising hazard). For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst set warranty, maintenance and mission-life targets.

8.4 Parametric survival regression (AFT)

LIVE OUTPUT

Group: Reliability · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Accelerated-failure-time regression with a choice of life distribution. It belongs to the **Reliability** family, used to model product life and failure behaviour for engineering decisions. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, event, temp_c. Analysts reach for it to set warranty, maintenance and mission-life targets.

SAMPLE DATA INPUT

failure_hours	event	temp_c
1.013e+04	1	85
2214	1	125
2507	1	105
2572	1	105
2366	1	105

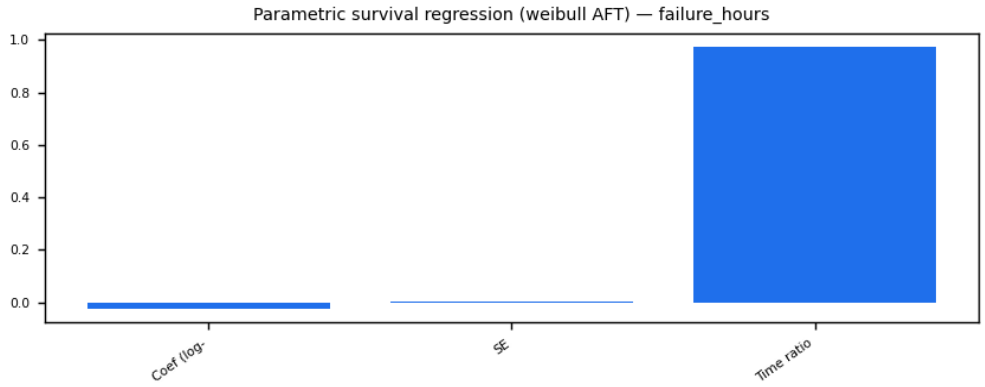
STEPS

1. Statistics → **Reliability** → **Parametric survival regression (AFT)**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, event=event, predictors=temp_c`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Parametric survival regression (weibull AFT) — failure_hours

Covariate	Coef (log-time)	SE	p	Time ratio exp(coef)
temp_c	-0.0241	0.0022	< .001	0.9762



DETAILED INTERPRETATION

n = 160 (153 events). AFT: coef > 0 lengthens life; time-ratio > 1 multiplies survival time. Median C-index = 0.704. The lead result here reads Covariate = temp_c; Coef (log-time) = -0.0241; SE = 0.0022; p = < .001. Fitted life-distribution parameters convert into reliability at a mission time and a characteristic life; a shape parameter above 1 indicates wear-out (rising hazard). For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst set warranty, maintenance and mission-life targets.

8.5 Recurrent events (MCF)

LIVE OUTPUT

Group: Reliability · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Mean cumulative function for repairable-system / warranty events. It belongs to the **Reliability** family, used to model product life and failure behaviour for engineering decisions. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on unit_id, failure_hours. Analysts reach for it to set warranty, maintenance and mission-life targets.

SAMPLE DATA INPUT

unit_id	failure_hours
U000	1.013e+04
U001	2214
U002	2507
U003	2572
U004	2366

STEPS

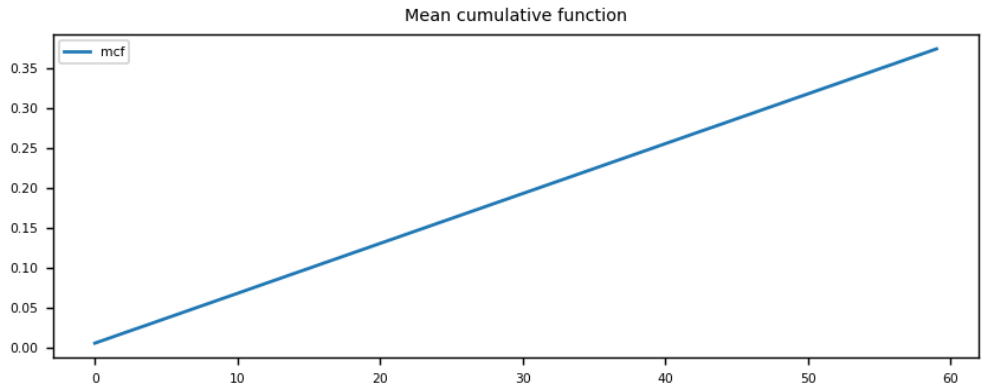
1. Statistics → **Reliability** → **Recurrent events (MCF)**. 2. Select the demo dataset above. 3. Set parameters — `unit=unit_id, time=failure_hours`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Mean cumulative function (MCF) — recurrent events per unit

Time	Events	MCF
740.4	1	0.0063
878.4	1	0.0125
994.8	1	0.0188
1074	1	0.025
1282	1	0.0312
1339	1	0.0375
1434	1	0.0437
1532	1	0.05
1549	1	0.0562

Time	Events	MCF
1669	1	0.0625
1750	1	0.0687
1850	1	0.075



DETAILED INTERPRETATION

160 units, 160 events. MCF(t) = expected cumulative number of events per unit by time t (repairable-systems / warranty analysis). The lead result here reads Time = 740.4; Events = 1; MCF = 0.0063. Fitted life-distribution parameters convert into reliability at a mission time and a characteristic life; a shape parameter above 1 indicates wear-out (rising hazard). For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst set warranty, maintenance and mission-life targets.

8.6 Reliability growth (Crow-AMSAA)

LIVE OUTPUT

Group: Reliability · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

NHPP power-law growth model with instantaneous MTBF and Duane plot. It belongs to the **Reliability** family, used to model product life and failure behaviour for engineering decisions. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours. Analysts reach for it to set warranty, maintenance and mission-life targets.

SAMPLE DATA INPUT

failure_hours
1.013e+04
2214
2507
2572
2366

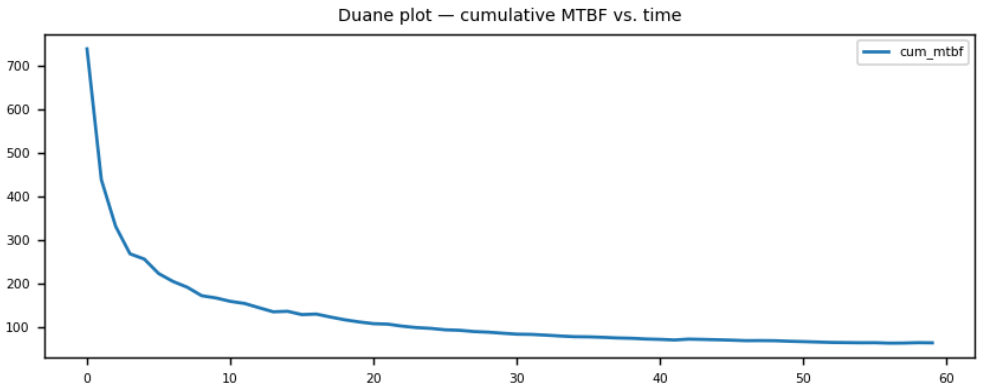
STEPS

1. Statistics → **Reliability** → **Reliability growth (Crow-AMSAA)**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Reliability growth — Crow-AMSAA (NHPP power law)

Shape β	Scale λ	Failures	Total time T	Instantaneous MTBF	Trend
1.026	0.0105	160	1.2e+04	73.11	stable (β≈1)



DETAILED INTERPRETATION

β < 1 means the time between failures is growing (reliability improving); β > 1 means it is shrinking. Instantaneous MTBF is the current mean time between failures. The lead result here reads Shape β = 1.026; Scale λ = 0.0105; Failures = 160; Total time T = 1.2e+04. Fitted life-distribution parameters convert into reliability at a mission time and a characteristic life; a shape parameter above 1 indicates wear-out (rising hazard). For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst set warranty, maintenance and mission-life targets.

8.7 Weibull reliability (life data)

LIVE OUTPUT

Group: Reliability · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Weibull fit with shape/scale, MTTF, B10 and reliability curve (censoring supported). It belongs to the **Reliability** family, used to model product life and failure behaviour for engineering decisions. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, event. Analysts reach for it to set warranty, maintenance and mission-life targets.

SAMPLE DATA INPUT

failure_hours	event
1.013e+04	1
2214	1
2507	1
2572	1
2366	1

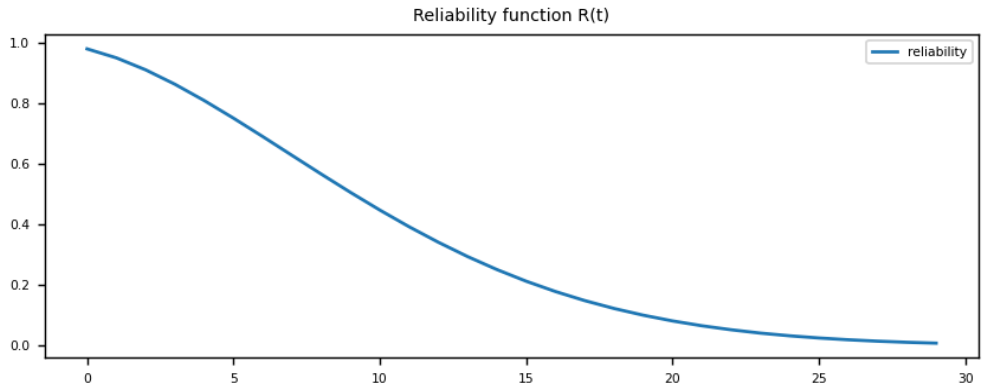
STEPS

1. Statistics → **Reliability** → **Weibull reliability (life data)**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, event=event`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Weibull reliability — failure_hours

Parameter	Value
Shape β	1.822
Scale η (characteristic li…	6144
MTTF (mean life)	5461
B10 life (10% fail)	1786
Failure pattern	increasing failure rate — …



DETAILED INTERPRETATION

Weibull fit by maximum likelihood (right-censored where event=0). β<1 infant mortality, β≈1 random, β>1 wear-out. The lead result here reads Parameter = Shape β; Value = 1.822. Fitted life-distribution parameters convert into reliability at a mission time and a characteristic life; a shape parameter above 1 indicates wear-out (rising hazard). For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst set warranty, maintenance and mission-life targets.

8.8 Cox diagnostics (PH test + baseline)

LIVE OUTPUT

Group: Survival · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Cox model with Schoenfeld proportional-hazards test and baseline survival curve. It belongs to the **Survival** family, used to model time-to-event data with censoring. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, event, temp_c. Analysts reach for it to plan warranties, retention or treatment around time-to-event risk.

SAMPLE DATA INPUT

failure_hours	event	temp_c
1.013e+04	1	85
2214	1	125
2507	1	105
2572	1	105
2366	1	105

STEPS

1. Statistics → **Survival** → **Cox diagnostics (PH test + baseline)**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, event=event, predictors=temp_c`. 4. Click **Run** and read the tables/chart below.

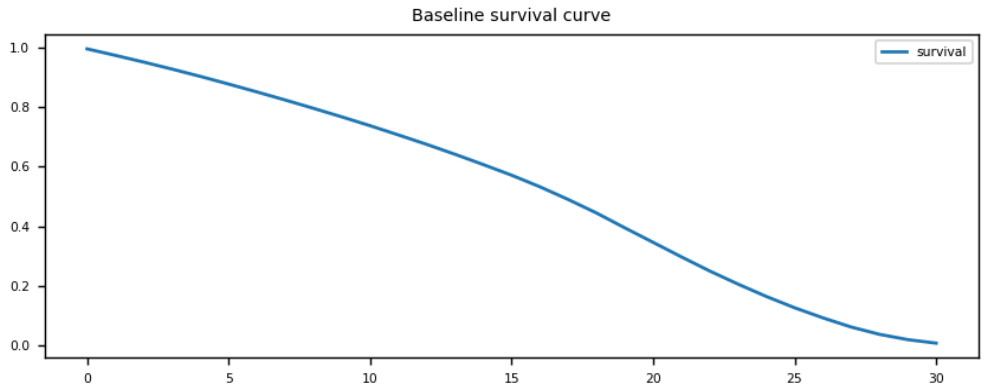
OUTPUT

Cox model — hazard ratios (failure_hours)

Predictor	Coef	Hazard ratio	p
temp_c	0.057	1.059	< .001

Proportional-hazards test (Schoenfeld residuals)

Covariate	χ^2	p	Verdict
temp_c	1.084	0.2978	PH ok



DETAILED INTERPRETATION

n = 160, concordance = 0.704. The lead result here reads Predictor = temp_c; Coef = 0.057; Hazard ratio = 1.059; p = < .001. Survival curves give the probability of lasting past each time and hazard ratios above 1 flag elevated risk, all while using censored records rather than discarding them. For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst plan warranties, retention or treatment around time-to-event risk.

8.9 Cox proportional hazards

LIVE OUTPUT

Group: Survival · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Semi-parametric survival regression with hazard ratios. It belongs to the **Survival** family, used to model time-to-event data with censoring. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, event, temp_c. Analysts reach for it to plan warranties, retention or treatment around time-to-event risk.

SAMPLE DATA INPUT

failure_hours	event	temp_c
1.013e+04	1	85
2214	1	125
2507	1	105
2572	1	105
2366	1	105

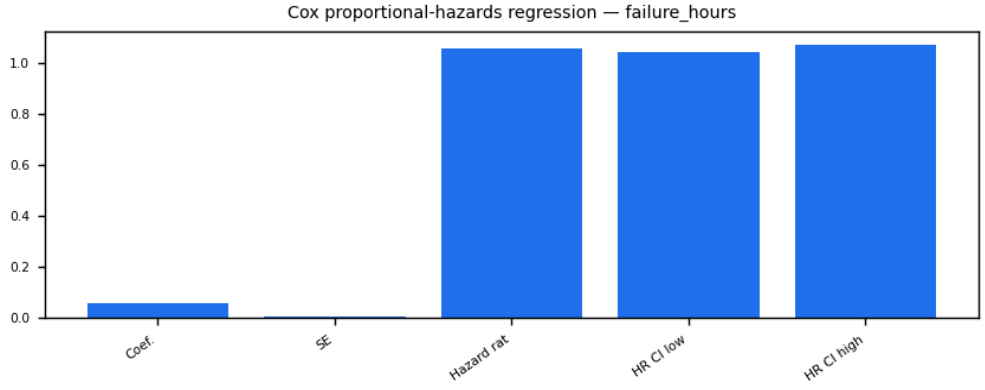
STEPS

1. Statistics → **Survival** → **Cox proportional hazards**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, event=event, predictors=temp_c`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Cox proportional-hazards regression — failure_hours

Predictor	Coef.	SE	p	Hazard ratio	HR CI low	HR CI high
temp_c	0.057	0.0068	< .001	1.059	1.045	1.073



DETAILED INTERPRETATION

n = 160. HR > 1 = higher hazard (shorter survival). Assumes proportional hazards over time. The lead result here reads Predictor = temp_c; Coef. = 0.057; SE = 0.0068; p = < .001. Survival curves give the probability of lasting past each time and hazard ratios above 1 flag elevated risk, all while using censored records rather than discarding them. For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst plan warranties, retention or treatment around time-to-event risk.

8.10 Kaplan-Meier survival

LIVE OUTPUT

Group: Survival · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Survival function + median survival. It belongs to the **Survival** family, used to model time-to-event data with censoring. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, event. Analysts reach for it to plan warranties, retention or treatment around time-to-event risk.

SAMPLE DATA INPUT

failure_hours	event
1.013e+04	1
2214	1
2507	1
2572	1
2366	1

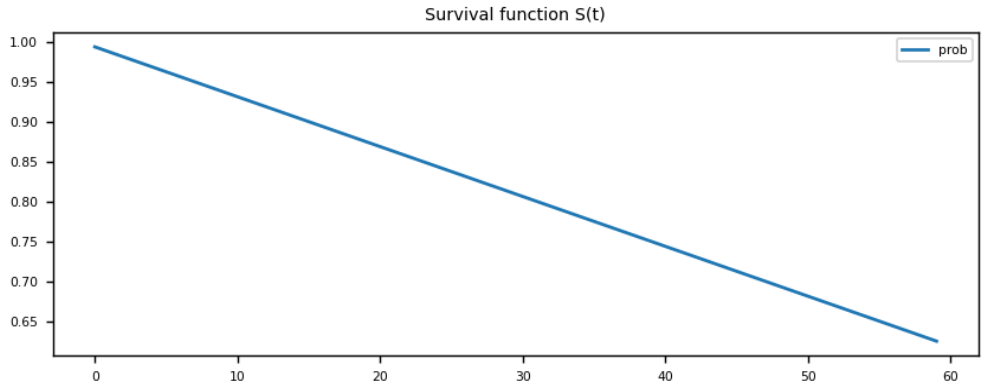
STEPS

1. Statistics → **Survival** → **Kaplan-Meier survival**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, event=event`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Kaplan-Meier survival — failure_hours

Metric	Value
75% survival time	2896
50% survival time	4646
25% survival time	7555
Events	153
Observations	160



DETAILED INTERPRETATION

Median survival = 4645.7. Step function of survival probability over time. The lead result here reads Metric = 75% survival time; Value = 2896. Survival curves give the probability of lasting past each time and hazard ratios above 1 flag elevated risk, all while using censored records rather than discarding them. For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst plan warranties, retention or treatment around time-to-event risk.

8.11 Life table (actuarial)

LIVE OUTPUT

Group: Survival · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Interval survival, conditional death probability & survival curve. It belongs to the **Survival** family, used to model time-to-event data with censoring. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, event. Analysts reach for it to plan warranties, retention or treatment around time-to-event risk.

SAMPLE DATA INPUT

failure_hours	event
1.013e+04	1
2214	1
2507	1
2572	1
2366	1

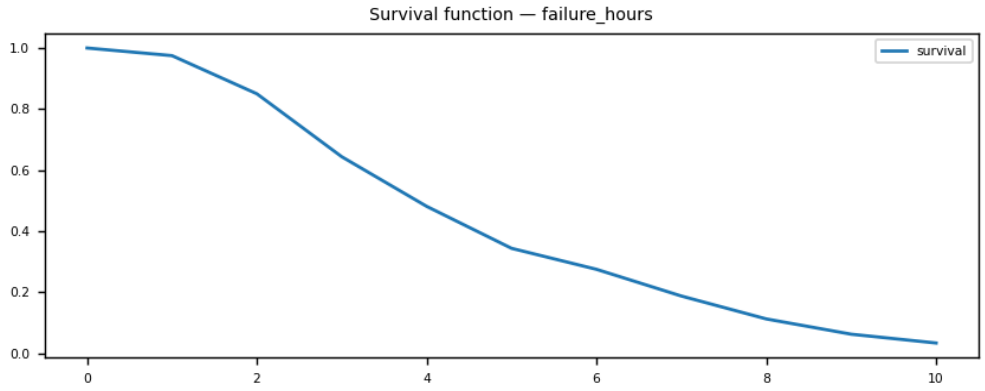
STEPS

1. Statistics → **Survival** → **Life table (actuarial)**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, event=event`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Life table — failure_hours (event = event), 10 intervals

Interval	N at risk	Deaths	Censored	Cond. P(death)	Cum. survival
0.00–1200.00	160	4	0	0.025	1
1200.00–2400.00	156	20	0	0.1282	0.975
2400.00–3600.00	136	33	0	0.2426	0.85
3600.00–4800.00	103	26	0	0.2524	0.6437
4800.00–6000.00	77	22	0	0.2857	0.4812
6000.00–7200.00	55	11	0	0.2	0.3438
7200.00–8400.00	44	14	0	0.3182	0.275
8400.00–9600.00	30	12	0	0.4	0.1875
9600.00–10800.00	18	8	0	0.4444	0.1125
10800.00–12000.00	10	3	7	0.4615	0.0625



DETAILED INTERPRETATION

Actuarial method (censored cases contribute ½ to the risk set). The lead result here reads Interval = 0.00–1200.00; N at risk = 160; Deaths = 4; Censored = 0. Survival curves give the probability of lasting past each time and hazard ratios above 1 flag elevated risk, all while using censored records rather than discarding them. For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst plan warranties, retention or treatment around time-to-event risk.

8.12 Log-rank test

LIVE OUTPUT

Group: Survival · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Compare survival curves across groups. It belongs to the **Survival** family, used to model time-to-event data with censoring. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, event, temp_c. Analysts reach for it to plan warranties, retention or treatment around time-to-event risk.

SAMPLE DATA INPUT

failure_hours	event	temp_c
1.013e+04	1	85
2214	1	125
2507	1	105
2572	1	105
2366	1	105

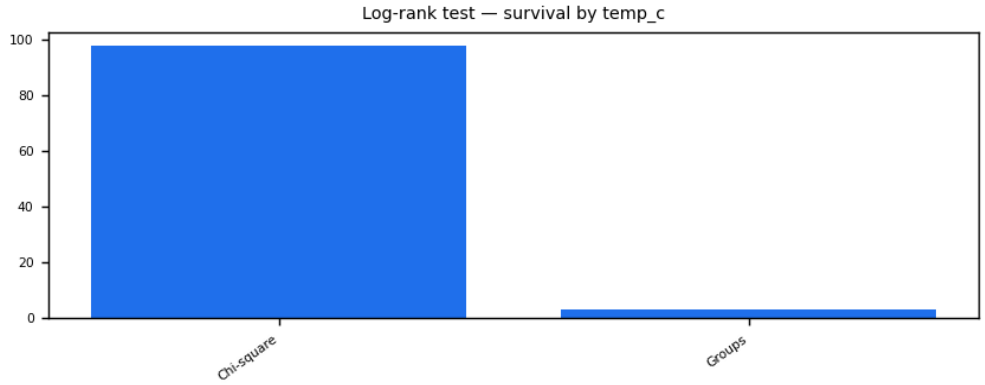
STEPS

1. Statistics → **Survival** → **Log-rank test**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, event=event, group=temp_c`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Log-rank test — survival by temp_c

Chi-square	p	Groups
97.84	< .001	3



DETAILED INTERPRETATION

Compares survival curves across groups (p < .05 → curves differ). The lead result here reads Chi-square = 97.84; p = < .001; Groups = 3. Survival curves give the probability of lasting past each time and hazard ratios above 1 flag elevated risk, all while using censored records rather than discarding them. For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst plan warranties, retention or treatment around time-to-event risk.

8.13 Nelson-Aalen cumulative hazard

LIVE OUTPUT

Group: Survival · Dataset: an accelerated component life test — failure hours, censoring and temperature stress

WHAT THIS DOES

Nonparametric hazard estimate + chart. It belongs to the **Survival** family, used to model time-to-event data with censoring. In this example it is applied to an accelerated component life test — failure hours, censoring and temperature stress, drawing on failure_hours, event. Analysts reach for it to plan warranties, retention or treatment around time-to-event risk.

SAMPLE DATA INPUT

failure_hours	event
1.013e+04	1

failure_hours	event
2214	1
2507	1
2572	1
2366	1

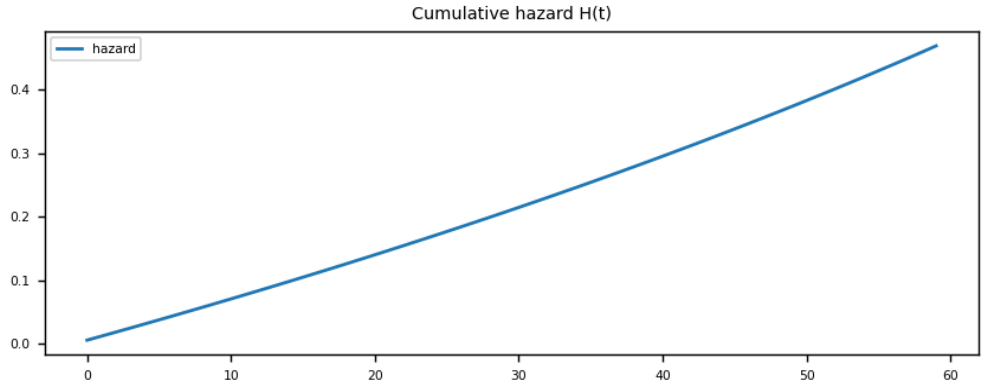
STEPS

1. Statistics → **Survival** → **Nelson-Aalen cumulative hazard**. 2. Select the demo dataset above. 3. Set parameters — `time=failure_hours, event=event`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Nelson-Aalen cumulative hazard — failure_hours

Metric	Value
Final cumulative hazard	3.063
Events	153
Observations	160



DETAILED INTERPRETATION

Nonparametric estimate of the cumulative hazard H(t); steeper = higher event rate. The lead result here reads Metric = Final cumulative hazar; Value = 3.063. Survival curves give the probability of lasting past each time and hazard ratios above 1 flag elevated risk, all while using censored records rather than discarding them. For an accelerated component life test — failure hours, censoring and temperature stress, this lets the analyst plan warranties, retention or treatment around time-to-event risk.

9. Clinical, Epidemiology & Pharmacometrics

9.1 Cochran-Mantel-Haenszel

LIVE OUTPUT

Group: Clinical trials · Dataset: a real-world patient cohort — treatment, recovery and confounders

WHAT THIS DOES

Stratified association test. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a real-world patient cohort — treatment, recovery and confounders, drawing on treated, recovered, comorbidities. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

treated	recovered	comorbidities
1	0	1
0	0	2
1	1	2
0	0	3
0	0	1

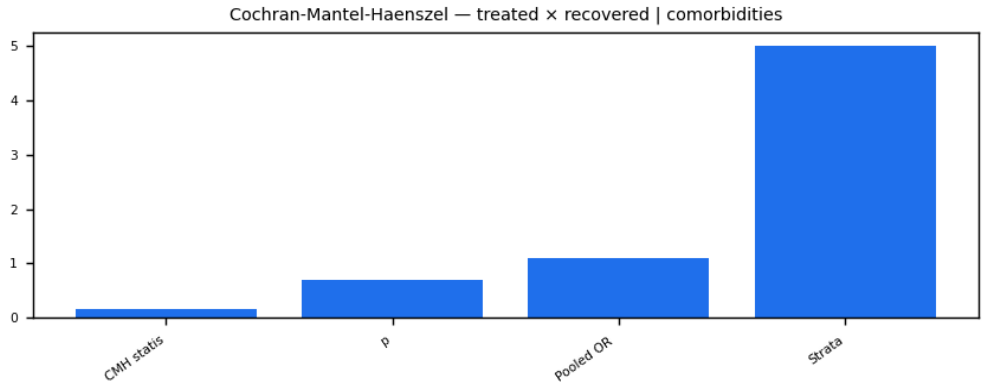
STEPS

1. Statistics → **Clinical trials** → **Cochran-Mantel-Haenszel**. 2. Select the demo dataset above. 3. Set parameters — `exposure=treated, outcome=recovered, stratum=comorbidities`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Cochran-Mantel-Haenszel — treated × recovered | comorbidities

CMH statistic	p	Pooled OR (M-H)	Strata
0.1529	0.6958	1.104	5



DETAILED INTERPRETATION

Tests exposure-outcome association controlling for the stratifying variable. The lead result here reads CMH statistic = 0.1529; p = 0.6958; Pooled OR (M-H) = 1.104; Strata = 5. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a real-world patient cohort — treatment, recovery and confounders, this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.2 Diagnostic test accuracy

INPUT REQUIREMENT

Group: Clinical trials · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Sensitivity/specificity/PPV/NPV/LR. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, tenure_months. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

monthly_charges	tenure_months
52.94	68
113.7	45
35.51	49
71.45	64
29.16	42

STEPS

1. Statistics → **Clinical trials** → **Diagnostic test accuracy**. 2. Select the demo dataset above. 3. Set parameters — `test=monthly_charges, truth=tenure_months`. 4. Click **Run** and read the tables/chart below.

OUTPUT

'monthly_charges' × 'tenure_months' must be a 2×2 table (both binary)

DETAILED INTERPRETATION

This technique has a specific data requirement: *'monthly_charges' × 'tenure_months' must be a 2×2 table (both binary)*. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a clinical trials method and supports the decision to weigh treatment benefit against risk on defensible evidence.

9.3 Dissolution / release kinetics

LIVE OUTPUT

Group: Clinical trials · Dataset: a tablet dissolution test — cumulative % released over time

WHAT THIS DOES

Fit zero/first-order, Higuchi, Korsmeyer-Peppas, Hixson-Crowell + f2 similarity. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a tablet dissolution test — cumulative % released over time, drawing on time_min, pct_released, reference_pct. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

time_min	pct_released	reference_pct
5	23.72	24.74
10	35.55	37.62
15	43.54	45.08
30	60.05	63.77
45	72.07	78.04

STEPS

1. Statistics → **Clinical trials** → **Dissolution / release kinetics**. 2. Select the demo dataset above. 3. Set parameters — `time=time_min, released=pct_released, reference=reference_pct`. 4. Click **Run** and read the tables/chart below.

OUTPUT

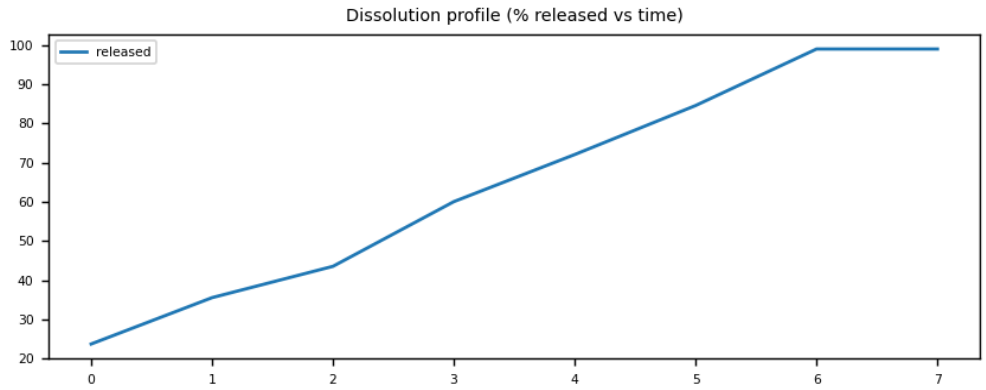
Drug-release kinetics — model fits

Model	R^2	Parameter
Zero-order	0.8841	k0 = 0.6594

Model	R^2	Parameter
First-order	0.9281	k1 = 0.0421
Higuchi	0.9706	kH = 9.1715
Korsmeyer-Peppas	0.9878	n = 0.4632
Hixson-Crowell	0.9604	kHC = 0.0307

Similarity factor f2 (vs reference)

f2	Verdict
75.36	similar (f2 ≥ 50)



DETAILED INTERPRETATION

Best-fit model (highest R^2): Korsmeyer-Peppas. Korsmeyer-Peppas release exponent n = 0.463 → anomalous (non-Fickian) transport. The lead result here reads Model = Zero-order; R^2 = 0.8841; Parameter = k0 = 0.6594. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a tablet dissolution test — cumulative % released over time, this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.4 Incidence rate

LIVE OUTPUT

Group: Clinical trials · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Events per person-time with CI. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, tenure_months. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

monthly_charges	tenure_months
52.94	68
113.7	45
35.51	49
71.45	64
29.16	42

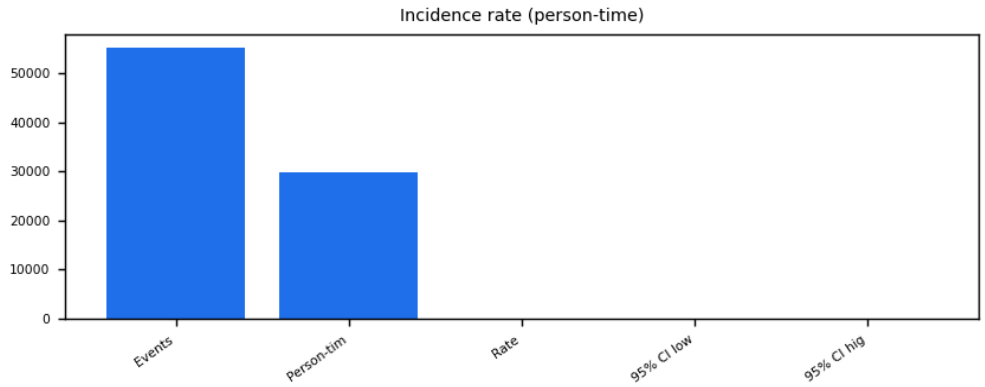
STEPS

1. Statistics → **Clinical trials** → **Incidence rate**. 2. Select the demo dataset above. 3. Set parameters — `events=monthly_charges, person_time=tenure_months`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Incidence rate (person-time)

Events	Person-time	Rate	95% CI low	95% CI high
5.527e+04	2.968e+04	1.862	1.847	1.878



DETAILED INTERPRETATION

Events per unit person-time with an exact Poisson confidence interval. The lead result here reads Events = 5.527e+04; Person-time = 2.968e+04; Rate = 1.862; 95% CI low = 1.847. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.5 Meta-analysis (fixed-effect)

LIVE OUTPUT

Group: Clinical trials · Dataset: a set of studies with effect sizes and standard errors (evidence synthesis)

WHAT THIS DOES

Pool study effects + heterogeneity. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a set of studies with effect sizes and standard errors (evidence synthesis), drawing on effect_logOR, se. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

effect_logOR	se
0.411	0.291
0.681	0.244
0.848	0.272
0.505	0.104
0.032	0.25

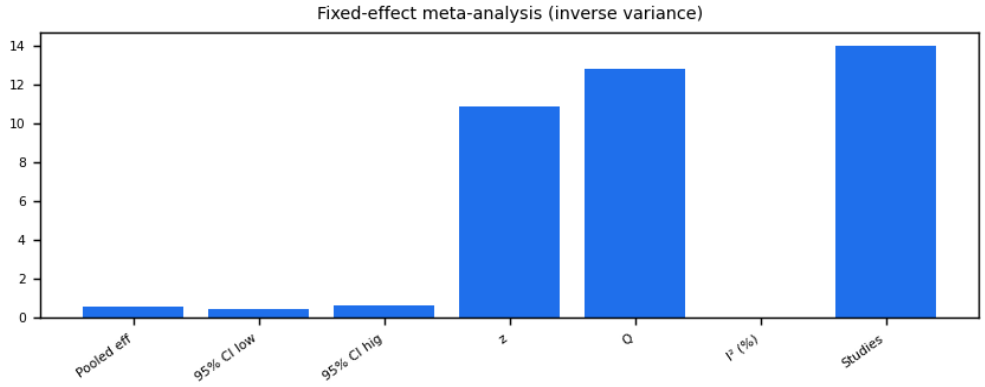
STEPS

1. Statistics → **Clinical trials** → **Meta-analysis (fixed-effect)**. 2. Select the demo dataset above. 3. Set parameters — `effect=effect_logOR, se=se`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Fixed-effect meta-analysis (inverse variance)

Pooled effect	95% CI low	95% CI high	z	p	Q	I² (%)	Studies
0.5267	0.4316	0.6219	10.85	< .001	12.79	0	14



DETAILED INTERPRETATION

Combines study effect sizes weighted by precision. I² > 50% suggests high heterogeneity. The lead result here reads Pooled effect = 0.5267; 95% CI low = 0.4316; 95% CI high = 0.6219; z = 10.85. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a set of studies with effect sizes and standard errors (evidence synthesis), this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.6 Meta-analysis (random effects)

LIVE OUTPUT

Group: Clinical trials · Dataset: a set of studies with effect sizes and standard errors (evidence synthesis)

WHAT THIS DOES

DerSimonian-Laird pooling with forest, funnel, Egger & trim-and-fill. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a set of studies with effect sizes and standard errors (evidence synthesis), drawing on effect_logOR, se, study. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

effect_logOR	se	study
0.411	0.291	Kumar et al. 2016
0.681	0.244	Silva et al. 2021
0.848	0.272	Okafor et al. 2004
0.505	0.104	Rossi et al. 2017
0.032	0.25	Nguyen et al. 2011

STEPS

1. Statistics → **Clinical trials** → **Meta-analysis (random effects)**. 2. Select the demo dataset above. 3. Set parameters — `effect=effect_logOR, se=se, label=study`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Random-effects meta-analysis (DerSimonian-Laird)

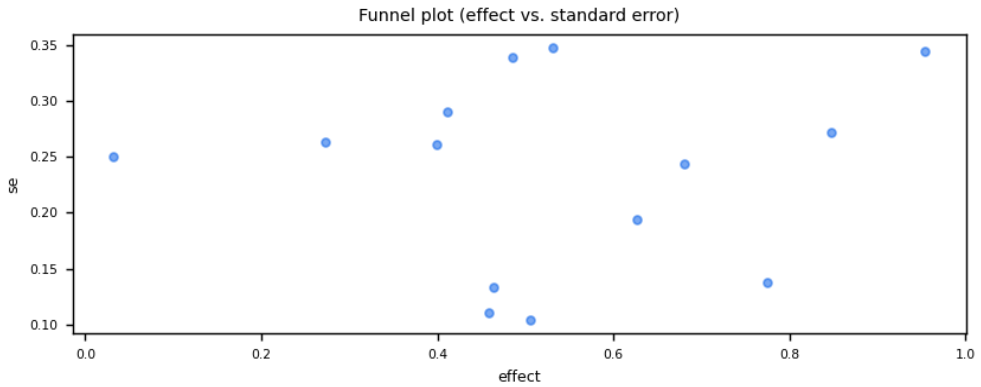
Pooled effect	95% CI low	95% CI high	z	p	τ²	Q	I² (%)	Studies
0.5267	0.4316	0.6219	10.85	< .001	0	12.79	0	14

Forest plot (per study)

Study	Effect	CI low	CI high	Weight
Kumar et al. 2016	0.411	-0.1594	0.9814	2.8%
Silva et al. 2021	0.681	0.2028	1.159	4.0%
Okafor et al. 2004	0.848	0.3149	1.381	3.2%
Rossi et al. 2017	0.505	0.3012	0.7088	21.8%
Nguyen et al. 2011	0.032	-0.458	0.522	3.8%
Cohen et al. 2015	0.954	0.2778	1.63	2.0%
Ivanov et al. 2022	0.626	0.2458	1.006	6.3%
Park et al. 2011	0.458	0.2424	0.6736	19.5%
Tan et al. 2006	0.485	-0.1794	1.149	2.1%
Chen et al. 2019	0.273	-0.2425	0.7885	3.4%
Haas et al. 2020	0.464	0.2033	0.7247	13.3%
Diallo et al. 2009	0.399	-0.1126	0.9106	3.5%

Publication-bias diagnostics

Test	Value
Egger intercept p (asymmet…	0.0011
Trim-and-fill missing stud…	0
Trim-and-fill adjusted eff…	0.5267



DETAILED INTERPRETATION

Between-study variance τ^2 via DerSimonian-Laird. $I^2 > 50\%$ = substantial heterogeneity. The lead result here reads Pooled effect = 0.5267; 95% CI low = 0.4316; 95% CI high = 0.6219; z = 10.85. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a set of studies with effect sizes and standard errors (evidence synthesis), this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.7 Meta-regression

LIVE OUTPUT

Group: Clinical trials · Dataset: a set of studies with effect sizes and standard errors (evidence synthesis)

WHAT THIS DOES

Regress study effect sizes on a moderator (random-effects WLS). It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a set of studies with effect sizes and standard errors (evidence synthesis), drawing on effect_logOR, se, year. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

effect_logOR	se	year
0.411	0.291	2016
0.681	0.244	2021
0.848	0.272	2004
0.505	0.104	2017
0.032	0.25	2011

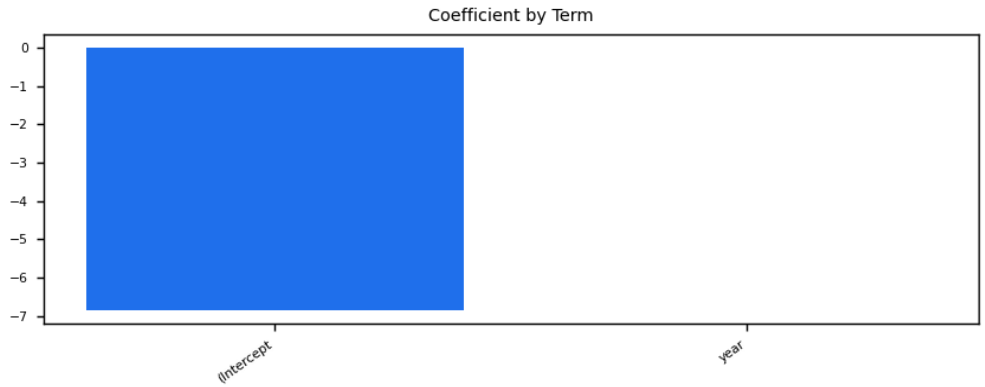
STEPS

1. Statistics → **Clinical trials** → **Meta-regression**. 2. Select the demo dataset above. 3. Set parameters — `effect=effect_logOR, se=se, moderator=year`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Meta-regression — effect ~ year (random-effects WLS)

Term	Coefficient	SE	p
(Intercept)	-6.84	22.65	0.7678
year	0.0037	0.0112	0.7506



DETAILED INTERPRETATION

τ^2 = 0.0000. A significant moderator slope explains part of the between-study heterogeneity. The lead result here reads Term = (Intercept); Coefficient = -6.84; SE = 22.65; p = 0.7678. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a set of studies with effect sizes and standard errors (evidence synthesis), this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.8 Network meta-analysis

LIVE OUTPUT

Group: Clinical trials · Dataset: head-to-head trials comparing several treatments

WHAT THIS DOES

Contrast-based frequentist NMA with league table and P-score ranking. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to head-to-head trials comparing several treatments, drawing on study, treatment, comparator, effect_logOR, se. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

study	treatment	comparator	effect_logOR	se
Trial1_DrugAvPlacebo	DrugA	Placebo	0.433	0.154
Trial1_DrugBvPlacebo	DrugB	Placebo	0.996	0.1
Trial1_DrugAvDrugB	DrugA	DrugB	-0.511	0.138
Trial1_DrugBvDrugC	DrugB	DrugC	0.681	0.226
Trial1_DrugCvPlacebo	DrugC	Placebo	0.305	0.207

STEPS

1. Statistics → **Clinical trials** → **Network meta-analysis**. 2. Select the demo dataset above. 3. Set parameters — `study=study, treat_a=treatment, treat_b=comparator, effect=effect_logOR, se=se`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Network meta-analysis — summary

Treatments	Comparisons	Reference	Model	τ^2	Q	I^2 (%)
4	24	DrugA	random	0	9.495	0

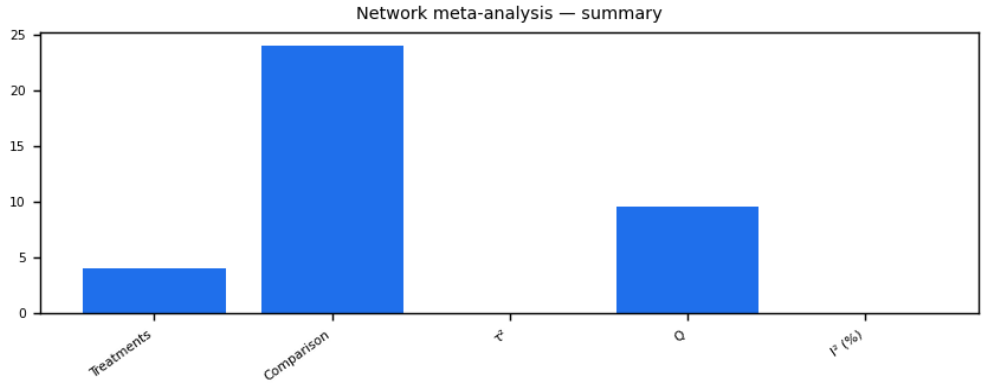
Basic estimates (each treatment vs. reference)

Contrast	Effect	95% CI low	95% CI high	p
DrugB vs DrugA	0.4492	0.3424	0.556	< .001
DrugC vs DrugA	-0.2842	-0.4004	-0.1679	< .001
Placebo vs DrugA	-0.542	-0.6541	-0.4299	< .001

League table (row vs. column, effect [95% CI])

vs →	DrugA	DrugB	DrugC	Placebo
DrugA	—	-0.45 [-0.56, -0.34]	0.28 [0.17, 0.40]	0.54 [0.43, 0.65]
DrugB	0.45 [0.34, 0.56]	—	0.73 [0.62, 0.84]	0.99 [0.89, 1.09]
DrugC	-0.28 [-0.40, -0.17]	-0.73 [-0.84, -0.62]	—	0.26 [0.14, 0.37]

vs →	DrugA	DrugB	DrugC	Placebo
Placebo	-0.54 [-0.65, -0.43]	-0.99 [-1.09, -0.89]	-0.26 [-0.37, -0.14]	—



DETAILED INTERPRETATION

Contrast-based frequentist NMA. Effects are on the input scale (e.g. log-OR / mean diff). The lead result here reads Treatments = 4; Comparisons = 24; Reference = DrugA; Model = random. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For head-to-head trials comparing several treatments, this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.9 Non-inferiority test (means)

LIVE OUTPUT

Group: Clinical trials · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

NI test with a margin. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

monthly_charges	churn
52.94	No
113.7	Yes
35.51	No
71.45	No
29.16	No

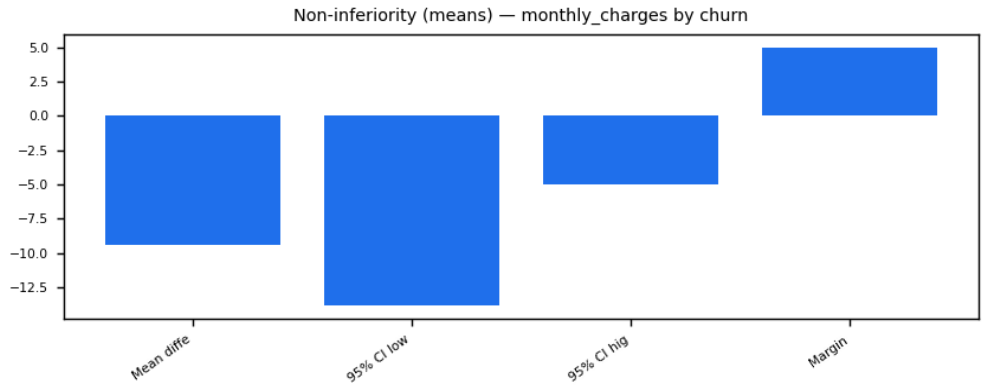
STEPS

1. Statistics → **Clinical trials** → **Non-inferiority test (means)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=churn, margin=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Non-inferiority (means) — monthly_charges by churn

Mean difference	95% CI low	95% CI high	Margin	Conclusion
-9.428	-13.82	-5.031	5	not established



DETAILED INTERPRETATION

Non-inferiority is shown if the CI lower bound stays above –margin (No vs Yes). The lead result here reads Mean difference = -9.428; 95% CI low = -13.82; 95% CI high = -5.031; Margin = 5. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.10 Number needed to treat

LIVE OUTPUT

Group: Clinical trials · Dataset: a real-world patient cohort — treatment, recovery and confounders

WHAT THIS DOES

ARR and NNT. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a real-world patient cohort — treatment, recovery and confounders, drawing on treated, recovered. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

treated	recovered
1	0
0	0
1	1
0	0
0	0

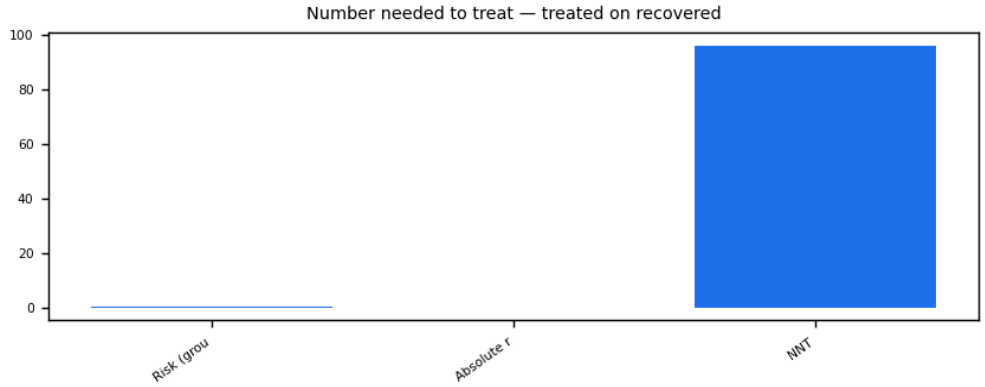
STEPS

1. Statistics → **Clinical trials** → **Number needed to treat**. 2. Select the demo dataset above. 3. Set parameters — `exposure=treated, outcome=recovered`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Number needed to treat — treated on recovered

Risk (group 1)	Risk (group 0)	Absolute risk reduction	NNT
0.3412	0.3308	-0.0104	96.09



DETAILED INTERPRETATION

NNT = patients to treat to prevent one event (lower is better). The lead result here reads Risk (group 1) = 0.3412; Risk (group 0) = 0.3308; Absolute risk reduction = -0.0104; NNT = 96.09. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a real-world patient cohort — treatment, recovery and confounders, this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.11 Odds ratio

LIVE OUTPUT

Group: Clinical trials · Dataset: a real-world patient cohort — treatment, recovery and confounders

WHAT THIS DOES

Odds ratio with 95% CI (2×2). It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a real-world patient cohort — treatment, recovery and confounders, drawing on treated, recovered. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

treated	recovered
1	0
0	0
1	1
0	0
0	0

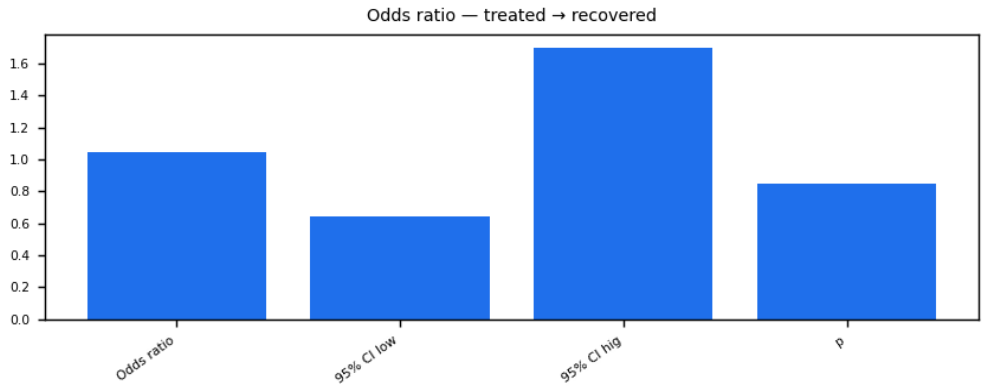
STEPS

1. Statistics → **Clinical trials** → **Odds ratio**. 2. Select the demo dataset above. 3. Set parameters — `exposure=treated, outcome=recovered`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Odds ratio — treated → recovered

Odds ratio	95% CI low	95% CI high	p
1.048	0.6459	1.7	0.8501



DETAILED INTERPRETATION

OR > 1 = higher odds of the outcome with exposure; CI excluding 1 = significant. The lead result here reads Odds ratio = 1.048; 95% CI low = 0.6459; 95% CI high = 1.7; p = 0.8501. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a real-world patient cohort — treatment, recovery and confounders, this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.12 PK non-compartmental analysis

LIVE OUTPUT

Group: Clinical trials · Dataset: a drug concentration-time profile after dosing

WHAT THIS DOES

Cmax, Tmax, AUC, half-life, clearance & volume from a concentration-time profile. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a drug concentration-time profile after dosing, drawing on time_h, concentration_ng_ml. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

time_h	concentration_ng_ml
0.25	13.02
0.5	25.32
1	43.11
1.5	53.52
2	55.94

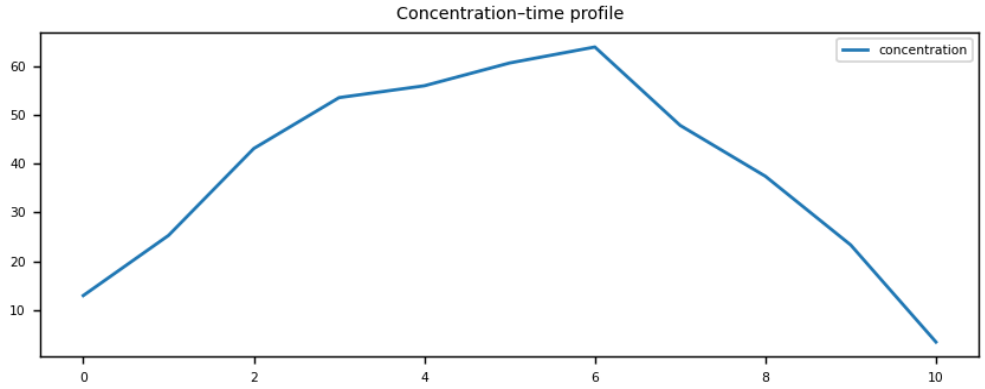
STEPS

1. Statistics → **Clinical trials** → **PK non-compartmental analysis**. 2. Select the demo dataset above. 3. Set parameters — `time=time_h, concentration=concentration_ng_ml, dose=100`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Pharmacokinetics — non-compartmental analysis

Parameter	Value
Cmax	63.88
Tmax	4
AUC(0–t) (trapezoidal)	673.9
AUC(0–∞)	698.1
Terminal rate constant kel	0.1448
Half-life t½	4.787
MRT	8.148
Clearance CL/F	0.1432
Volume Vz/F	0.9892



DETAILED INTERPRETATION

AUC by the linear trapezoidal rule; kel/t½ from log-linear regression of the terminal phase; CL/F and Vz/F require the dose. The lead result here reads Parameter = Cmax; Value = 63.88. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a drug concentration-time profile after dosing, this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.13 ROC curve / AUC

LIVE OUTPUT

Group: Clinical trials · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Discrimination of a predictor. It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, churn_bin. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

monthly_charges	churn_bin
52.94	0
113.7	1
35.51	0
71.45	0
29.16	0

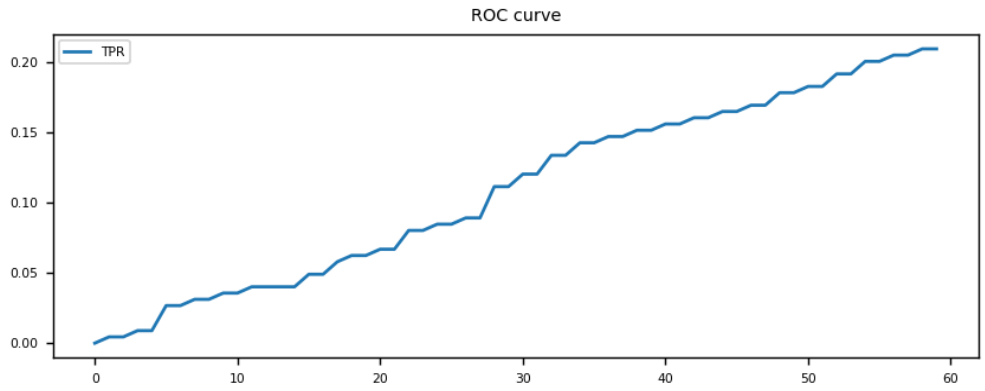
STEPS

1. Statistics → **Clinical trials** → **ROC curve / AUC**. 2. Select the demo dataset above. 3. Set parameters — `predictor=monthly_charges, outcome=churn_bin`. 4. Click **Run** and read the tables/chart below.

OUTPUT

ROC / AUC — monthly_charges predicting churn_bin

AUC	Discrimination	N
0.5942	poor	800



DETAILED INTERPRETATION

AUC: 0.5 = chance · 0.7 fair · 0.8 good · 0.9 excellent discrimination. The lead result here reads AUC = 0.5942; Discrimination = poor; N = 800. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst weigh treatment benefit against risk on defensible evidence.

9.14 Relative risk

LIVE OUTPUT

Group: Clinical trials · Dataset: a real-world patient cohort — treatment, recovery and confounders

WHAT THIS DOES

Risk ratio with 95% CI (2×2). It belongs to the **Clinical trials** family, used to quantify treatment benefit/risk and synthesise clinical evidence. In this example it is applied to a real-world patient cohort — treatment, recovery and confounders, drawing on treated, recovered. Analysts reach for it to weigh treatment benefit against risk on defensible evidence.

SAMPLE DATA INPUT

treated	recovered
1	0

treated	recovered
0	0
1	1
0	0
0	0

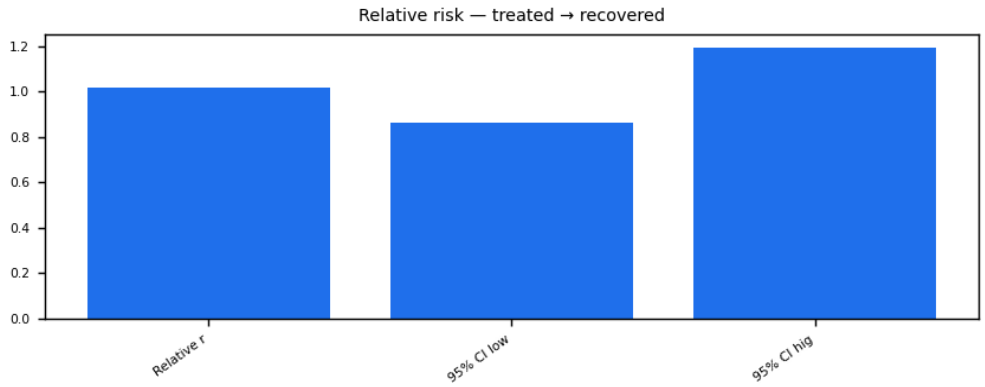
STEPS

1. Statistics → **Clinical trials** → **Relative risk**. 2. Select the demo dataset above. 3. Set parameters — `exposure=treated, outcome=recovered`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Relative risk — treated → recovered

Relative risk	95% CI low	95% CI high
1.016	0.8637	1.195



DETAILED INTERPRETATION

RR > 1 = higher risk with exposure; CI excluding 1 = significant. The lead result here reads Relative risk = 1.016; 95% CI low = 0.8637; 95% CI high = 1.195. Effect measures (risk ratio, odds ratio or pooled effect) with confidence intervals quantify benefit or risk - intervals excluding 1 for ratios (or 0 for differences) are significant. For a real-world patient cohort — treatment, recovery and confounders, this lets the analyst weigh treatment benefit against risk on defensible evidence.

10. Design of Experiments & Power

10.1 One-way ANOVA power

LIVE OUTPUT

Group: Design · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Observed (post-hoc) power and N for 80% power at the sample effect size. It belongs to the **Design** family, used to determine the sample size or power for a planned study. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract. Analysts reach for it to size a study so it is neither under- nor over-powered.

SAMPLE DATA INPUT

monthly_charges	contract
52.94	Month-to-month
113.7	One year
35.51	Month-to-month
71.45	Two year
29.16	Two year

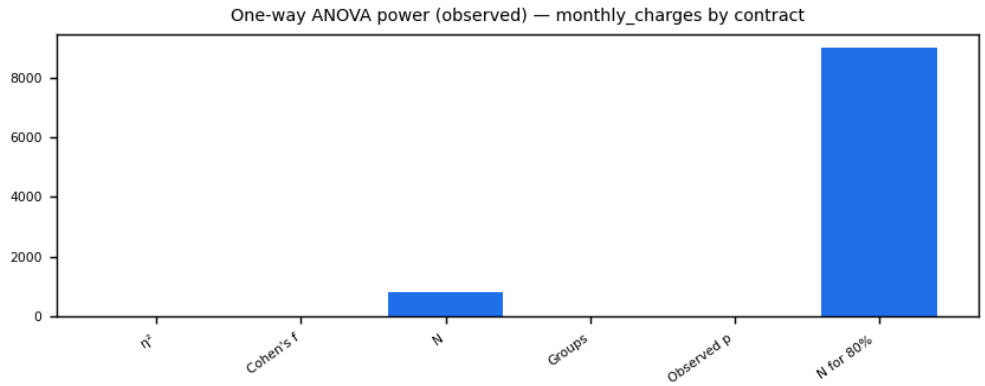
STEPS

1. Statistics → **Design** → **One-way ANOVA power**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, group=contract`. 4. Click **Run** and read the tables/chart below.

OUTPUT

One-way ANOVA power (observed) — monthly_charges by contract

η²	Cohen's f	N	Groups	Observed power	N for 80% power
0.0011	0.0327	800	3	0.12	8997



DETAILED INTERPRETATION

Observed (post-hoc) power at the sample effect size, and the total N that would give 80% power for the same effect. The lead result here reads η² = 0.0011; Cohen's f = 0.0327; N = 800; Groups = 3. The output states the probability of detecting the specified effect and the sample size required to reach the target power. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst size a study so it is neither under- nor over-powered.

10.2 Power analysis

LIVE OUTPUT

Group: Design · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

A-priori sample size for t-tests and ANOVA. It belongs to the **Design** family, used to determine the sample size or power for a planned study. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to size a study so it is neither under- nor over-powered.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

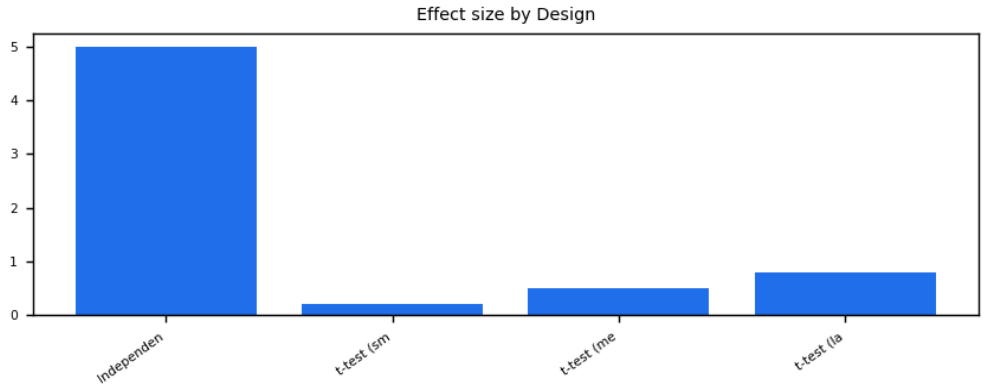
STEPS

1. Statistics → **Design** → **Power analysis**. 2. Select the demo dataset above. 3. Set parameters — `kind=t-test, effect_size=5, alpha=0.1, power=5, groups=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

A priori power analysis (required sample size)

Design	Effect size	α	Power	Required n	Unit
Independent t-test	5	0.1	5	7.379e+17	per group
t-test (small d=0.2)	0.2	0.1	5	3.875e+04	per group
t-test (medium d=0.5)	0.5	0.1	5	7.379e+19	per group
t-test (large d=0.8)	0.8	0.1	5	360	per group



DETAILED INTERPRETATION

Effect size: Cohen's d for t-tests, Cohen's f for ANOVA. The lead result here reads Design = Independent t-test; Effect size = 5; α = 0.1; Power = 5. The output states the probability of detecting the specified effect and the sample size required to reach the target power. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst size a study so it is neither under- nor over-powered.

10.3 Power — correlation

INPUT REQUIREMENT

Group: Design · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Sample size to detect a correlation. It belongs to the **Design** family, used to determine the sample size or power for a planned study. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to size a study so it is neither under- nor over-powered.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

STEPS

1. Statistics → **Design** → **Power — correlation**. 2. Select the demo dataset above. 3. Set parameters — `r=5, alpha=0.1, power=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Expected correlation r must be between -1 and 1 (and non-zero)

DETAILED INTERPRETATION

This technique has a specific data requirement: *Expected correlation r must be between -1 and 1 (and non-zero)*. The output states the probability of detecting the specified effect and the sample size required to reach the target power.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a design method and supports the decision to size a study so it

is neither under- nor over-powered.

10.4 Power — two proportions

INPUT REQUIREMENT

Group: Design · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Sample size for two proportions. It belongs to the **Design** family, used to determine the sample size or power for a planned study. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to size a study so it is neither under- nor over-powered.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

STEPS

1. Statistics → **Design** → **Power — two proportions**. 2. Select the demo dataset above. 3. Set parameters — `prop1=5, prop2=5, alpha=0.1, power=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Proportions must be between 0 and 1

DETAILED INTERPRETATION

This technique has a specific data requirement: *Proportions must be between 0 and 1*. The output states the probability of detecting the specified effect and the sample size required to reach the target power.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a design method and supports the decision to size a study so it is neither under- nor over-powered.

10.5 Box-Behnken design

INPUT REQUIREMENT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

3-level RSM design (3 factors). It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

STEPS

1. Statistics → **Design of experiments** → **Box-Behnken design**. 2. Select the demo dataset above. 3. Set parameters — `n_factors=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Box-Behnken implemented for exactly 3 factors

DETAILED INTERPRETATION

This technique has a specific data requirement: *Box-Behnken implemented for exactly 3 factors*. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a design of experiments method and supports the decision to learn the most from the fewest experimental runs.

10.6 Central composite design

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Response-surface (RSM) design. It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

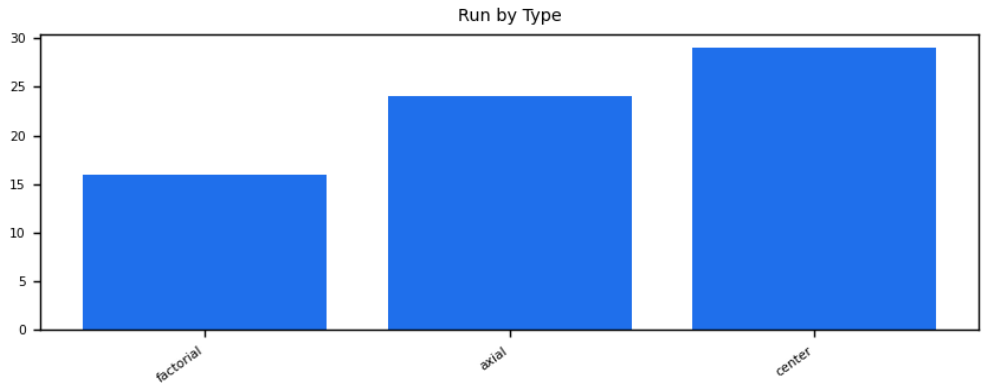
STEPS

1. Statistics → **Design of experiments** → **Central composite design**. 2. Select the demo dataset above. 3. Set parameters — `n_factors=5, center_points=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Central composite design — 4 factors (29 runs, α=2.0)

Run	Type	A	B	C	D
1	factorial	-1	-1	-1	-1
2	factorial	-1	-1	-1	1
3	factorial	-1	-1	1	-1
4	factorial	-1	-1	1	1
5	factorial	-1	1	-1	-1
6	factorial	-1	1	-1	1
7	factorial	-1	1	1	-1
8	factorial	-1	1	1	1
9	factorial	1	-1	-1	-1
10	factorial	1	-1	-1	1
11	factorial	1	-1	1	-1
12	factorial	1	-1	1	1



DETAILED INTERPRETATION

Response-surface design (factorial + axial + center) for fitting a quadratic model. The lead result here reads Run = 1; Type = factorial; A = -1; B = -1. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

10.7 D-optimal custom design

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Model-driven optimal design when a classical design is impractical. It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

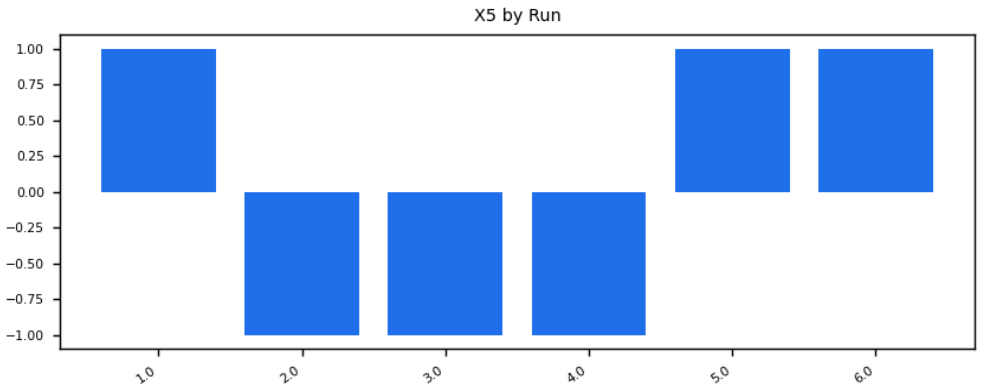
STEPS

1. Statistics → **Design of experiments** → **D-optimal custom design**. 2. Select the demo dataset above. 3. Set parameters — `n_factors=5, model=main, runs=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

D-optimal design — 6 runs, 5 factors (main)

Run	X1	X2	X3	X4	X5
1	1	1	-1	1	1
2	1	-1	1	-1	-1
3	-1	-1	1	1	-1
4	-1	1	-1	-1	-1
5	-1	1	1	-1	1
6	-1	-1	-1	-1	1



DETAILED INTERPRETATION

D-optimal design for a main model in 5 factors, 6 runs (model has 6 parameters). Coordinate-exchange algorithm; relative D-efficiency ≈ 0.9048. Factors are coded −1/+1 (0 = centre); map to your real low/high settings. The lead result here reads Run = 1; X1 = 1; X2 = 1; X3 = -1. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

10.8 Definitive screening design (DSD)

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

3-level screening design (Jones-Nachtsheim), ~2m+1 runs, estimable curvature. It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

STEPS

1. Statistics → **Design of experiments** → **Definitive screening design (DSD)**. 2. Select the demo dataset above. 3. Set parameters — `num_factors=4`. 4. Click **Run** and read the tables/chart below.

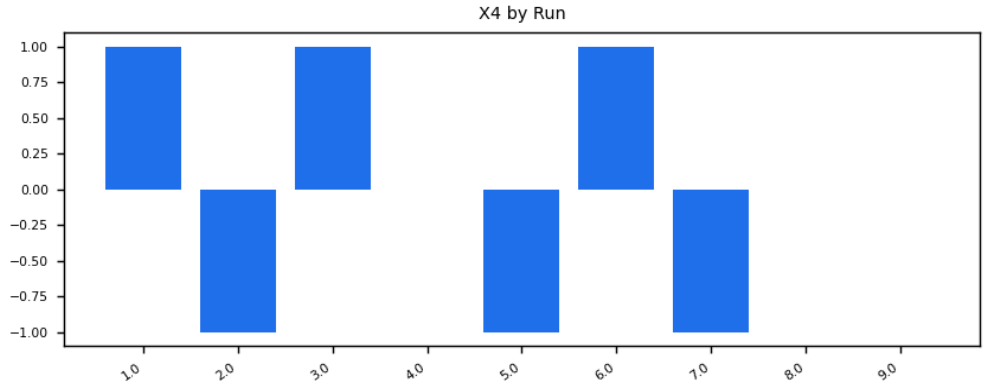
OUTPUT

Definitive Screening Design — 4 factors, 9 runs

Run	X1	X2	X3	X4
1	0	1	1	1
2	-1	0	1	-1
3	-1	-1	0	1
4	-1	1	-1	0
5	0	-1	-1	-1
6	1	0	-1	1
7	1	1	0	-1
8	1	-1	1	0
9	0	0	0	0

How to use

Step	Action
1	Map –1/0/+1 to each factor…
2	Run the experiment and rec…
3	Analyse with 'Fit definiti…



DETAILED INTERPRETATION

Three-level design (–1/0/+1). Runs = 9. Main effects are orthogonal to each other and to two-factor interactions; quadratic effects are estimable. Max |off-diagonal main-effect correlation| = 0.000. The lead result here reads Run = 1; X1 = 0; X2 = 1; X3 = 1. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

10.9 Factorial ANOVA

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Analyze a factorial experiment. It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, contract, region. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

monetary	contract	region
3197	Month-to-month	East
1233	One year	North

monetary	contract	region
2501	Month-to-month	North
1468	Two year	West
4195	Two year	West

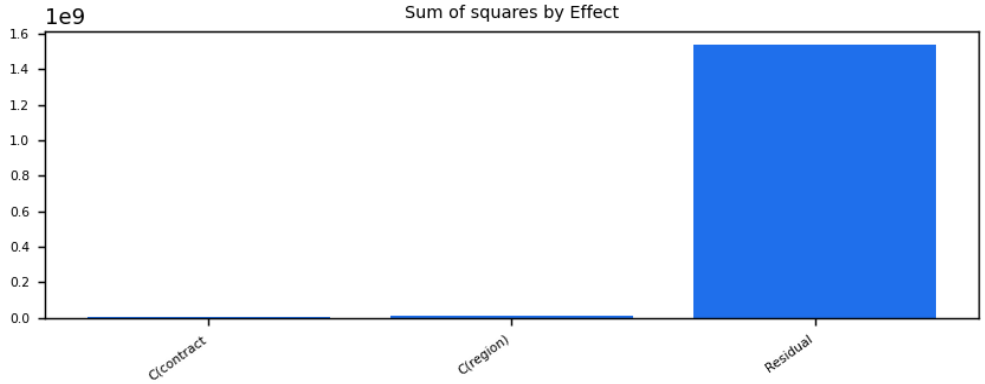
STEPS

1. Statistics → **Design of experiments** → **Factorial ANOVA**. 2. Select the demo dataset above. 3. Set parameters — `response=monetary, factors=contract, region`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Factorial ANOVA — monetary

Effect	Sum of squares	df	F	p
C(contract)	2.607e+06	2	0.6682	0.5129
C(region)	1.254e+07	3	2.143	0.0933
C(contract):C(region)	3.927e+06	6	0.3356	0.9182
Residual	1.537e+09	788		



DETAILED INTERPRETATION

Model R² = 0.012. Main effects and two-way interactions of the design factors. The lead result here reads Effect = C(contract); Sum of squares = 2.607e+06; df = 2; F = 0.6682. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

10.10 Fit definitive screening

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Fit main effects + quadratic curvature from a DSD run's response. It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

STEPS

1. Statistics → **Design of experiments** → **Fit definitive screening**. 2. Select the demo dataset above. 3. Set parameters — `response=monetary, factors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

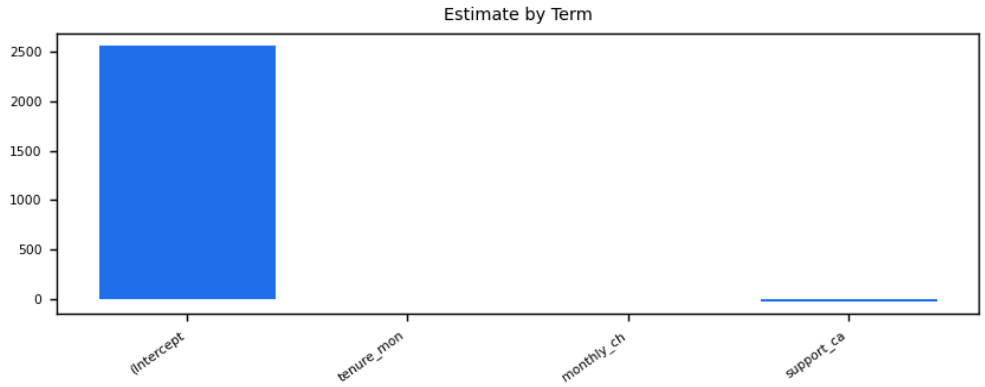
OUTPUT

Fit definitive screening — main effects (monetary)

Term	Estimate	SE	p
(Intercept)	2554	169.4	< .001
tenure_months	3.913	2.391	0.1021
monthly_charges	-0.3864	1.719	0.8222
support_calls	-20.12	35.35	0.5694

Curvature scan (each quadratic added to the main-effects model)

Quadratic term	Estimate	p	ΔR²
tenure_months²	-0.2353	0.0812	0.0038
monthly_charges²	-0.0058	0.9305	0
support_calls²	-19.13	0.314	0.0013



DETAILED INTERPRETATION

R^2 = 0.0037 · n = 800. Main effects are orthogonal in a DSD. Active (p<.05): none. The lead result here reads Term = (Intercept); Estimate = 2554; SE = 169.4; p = < .001. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

10.11 Fractional factorial design

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Half-fraction screening design. It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

STEPS

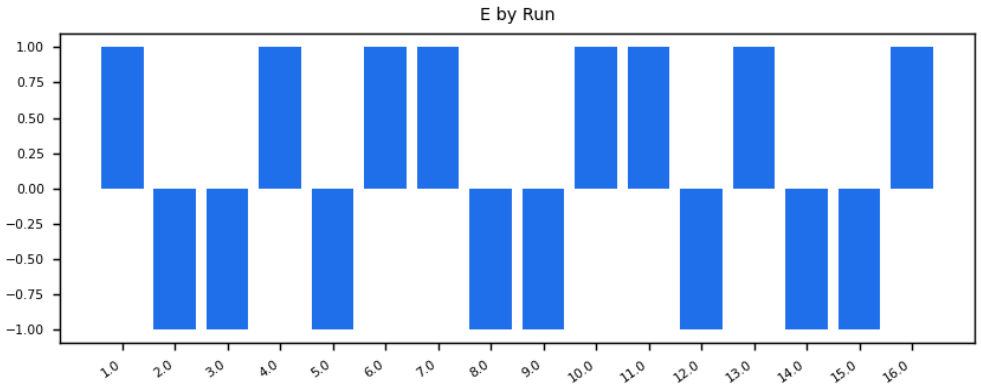
1. Statistics → **Design of experiments** → **Fractional factorial design**. 2. Select the demo dataset above. 3. Set parameters — `n_factors=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Fractional factorial 2^(5-1) — 16 runs

Run	A	B	C	D	E
1	-1	-1	-1	-1	1
2	-1	-1	-1	1	-1
3	-1	-1	1	-1	-1
4	-1	-1	1	1	1
5	-1	1	-1	-1	-1

Run	A	B	C	D	E
6	-1	1	-1	1	1
7	-1	1	1	-1	1
8	-1	1	1	1	-1
9	1	-1	-1	-1	-1
10	1	-1	-1	1	1
11	1	-1	1	-1	1
12	1	-1	1	1	-1



DETAILED INTERPRETATION

Half-fraction with defining relation I = ABCDE. Estimates main effects with far fewer runs (aliased with high-order interactions). The lead result here reads Run = 1; A = -1; B = -1; C = -1. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

10.12 Full factorial design

INPUT REQUIREMENT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Generate a full factorial DOE. It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

STEPS

1. Statistics → **Design of experiments** → **Full factorial design**. 2. Select the demo dataset above. 3. Set parameters — `n_factors=5, levels=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Design too large (>256 runs) — reduce factors/levels

DETAILED INTERPRETATION

This technique has a specific data requirement: *Design too large (>256 runs) — reduce factors/levels*. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a design of experiments method and supports the decision to learn the most from the fewest experimental runs.

10.13 Mixture design (simplex-lattice)

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Simplex-lattice design for formulation/mixture experiments (proportions sum to 1). It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

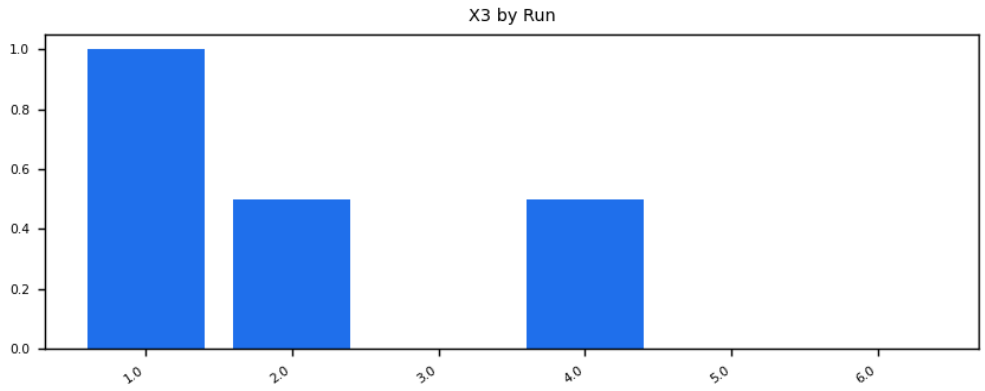
STEPS

1. Statistics → **Design of experiments** → **Mixture design (simplex-lattice)**. 2. Select the demo dataset above. 3. Set parameters — `num_components=3, degree=2`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Simplex-lattice mixture design {3,2} — 6 runs

Run	X1	X2	X3
1	0	0	1
2	0	0.5	0.5
3	0	1	0
4	0.5	0	0.5
5	0.5	0.5	0
6	1	0	0



DETAILED INTERPRETATION

Component proportions sum to 1 in every run. A {q,m} lattice places points at all proportions that are multiples of 1/2. Fit with a Scheffé mixture model. The lead result here reads Run = 1; X1 = 0; X2 = 0; X3 = 1. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

10.14 Response surface (RSM)

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Fit a quadratic optimization model. It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monetary, tenure_months, monthly_charges, support_calls. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

monetary	tenure_months	monthly_charges	support_calls
3197	68	52.94	1
1233	45	113.7	5
2501	49	35.51	1
1468	64	71.45	2
4195	42	29.16	1

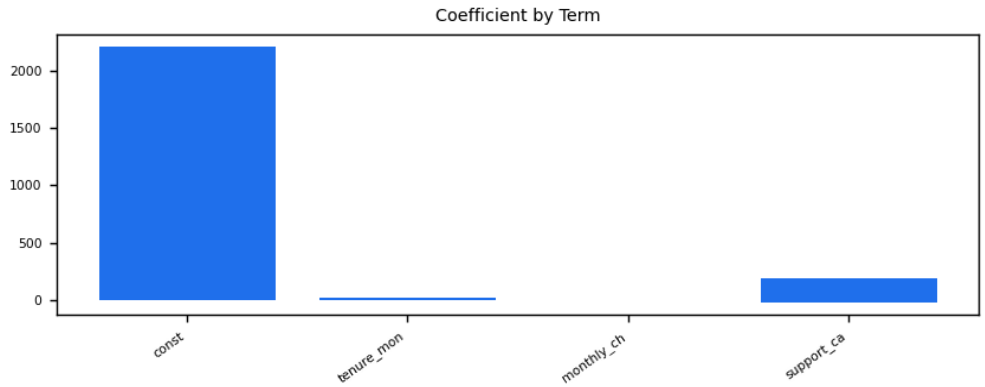
STEPS

1. Statistics → **Design of experiments** → **Response surface (RSM)**. 2. Select the demo dataset above. 3. Set parameters — `response=monetary, factors=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Response surface (quadratic) — monetary

Term	Coefficient	p
const	2205	< .001
tenure_months	17.15	0.1601
tenure_months^2	-0.2519	0.0643
monthly_charges	0.9333	0.9268
monthly_charges^2	0.0013	0.9844
support_calls	187	0.2054
support_calls^2	-20.93	0.2719
tenure_months:monthly_char…	0.0587	0.4788
tenure_months:support_call…	0.5352	0.7634
monthly_charges:support_ca…	-1.905	0.1267



DETAILED INTERPRETATION

R^2 = 0.013. Quadratic model with interactions for optimization/RSM. The lead result here reads Term = const; Coefficient = 2205; p = < .001. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

10.15 Space-filling design (Latin Hypercube)

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Latin Hypercube design for computer experiments / simulation. It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

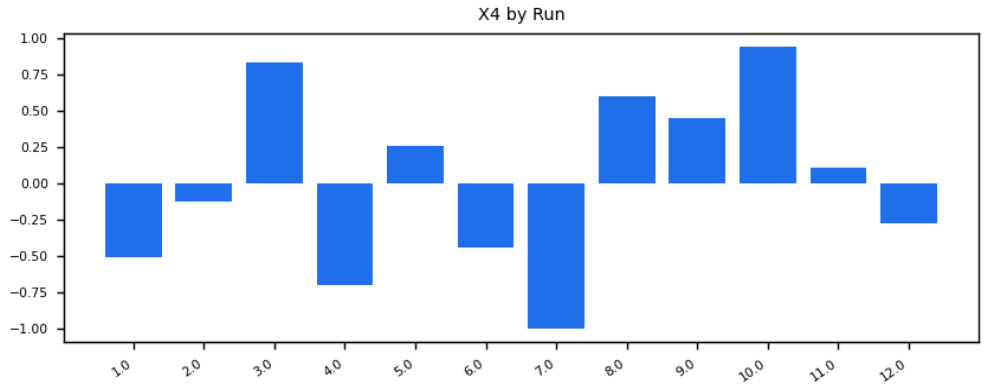
STEPS

1. Statistics → **Design of experiments** → **Space-filling design (Latin Hypercube)**. 2. Select the demo dataset above. 3. Set parameters — `num_factors=4, num_runs=12`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Latin Hypercube space-filling design — 4 factors, 12 runs

Run	X1	X2	X3	X4
1	0.7272	0.955	-0.1735	-0.5028
2	0.3645	0.5145	0.2322	-0.1216
3	-0.4239	-0.6558	-0.136	0.8329
4	-0.1429	-0.0056	0.7117	-0.6959
5	0.0228	-0.9236	-0.7166	0.2629
6	-0.5047	-0.1874	0.0549	-0.4412
7	-0.2692	-0.7306	0.3338	-0.9968
8	0.5524	0.2249	-0.4481	0.6018
9	-0.8558	0.7131	-0.5876	0.4483
10	0.919	0.0184	-0.989	0.9404
11	0.2381	-0.387	0.9009	0.1103
12	-0.7319	0.3516	0.6288	-0.2705



DETAILED INTERPRETATION

Space-filling design for computer experiments / simulation. Coded to [-1, 1]; map to real ranges. Centered L2-discrepancy = 0.0237 (lower = more uniform coverage). The lead result here reads Run = 1; X1 = 0.7272; X2 = 0.955; X3 = -0.1735. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

10.16 Taguchi orthogonal array

LIVE OUTPUT

Group: Design of experiments · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Robust-design screening arrays (L4/L8/L16/L32, L9/L27). It belongs to the **Design of experiments** family, used to plan efficient experiments that span the factor space. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to learn the most from the fewest experimental runs.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

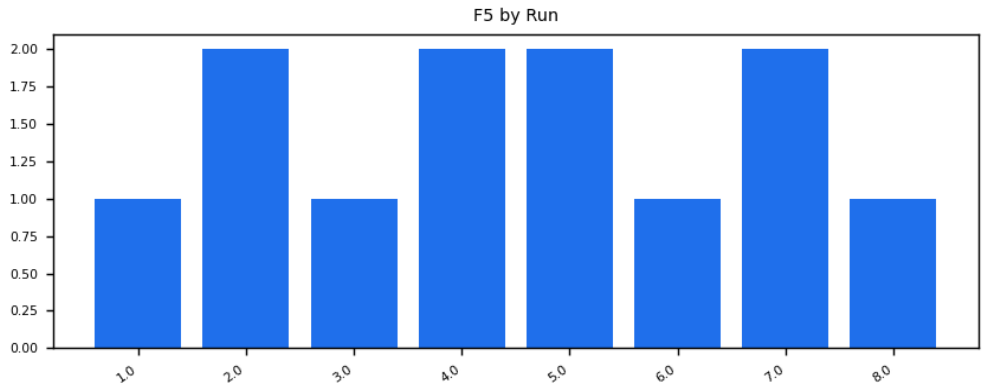
STEPS

1. Statistics → **Design of experiments** → **Taguchi orthogonal array**. 2. Select the demo dataset above. 3. Set parameters — `n_factors=5, levels=2`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Taguchi L8 design — 8 runs, 5 factors, 2 levels

Run	F1	F2	F3	F4	F5
1	1	1	1	1	1
2	2	1	2	1	2
3	1	2	2	1	1
4	2	2	1	1	2
5	1	1	1	2	2
6	2	1	2	2	1
7	1	2	2	2	2
8	2	2	1	2	1



DETAILED INTERPRETATION

Taguchi L8 orthogonal array: 8 runs for 5 factors at 2 levels (a highly fractionalised design for screening main effects). Levels are coded 1…L; map to your real settings. Analyse the response with a factorial ANOVA + S/N ratios. The lead result here reads Run = 1; F1 = 1; F2 = 1; F3 = 1. The design spans the factor space efficiently; the fitted effects rank which factors and interactions drive the response so it can be optimised. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst learn the most from the fewest experimental runs.

11. Quality & Statistical Process Control

11.1 CUSUM control chart

LIVE OUTPUT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Cumulative-sum shift detection. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

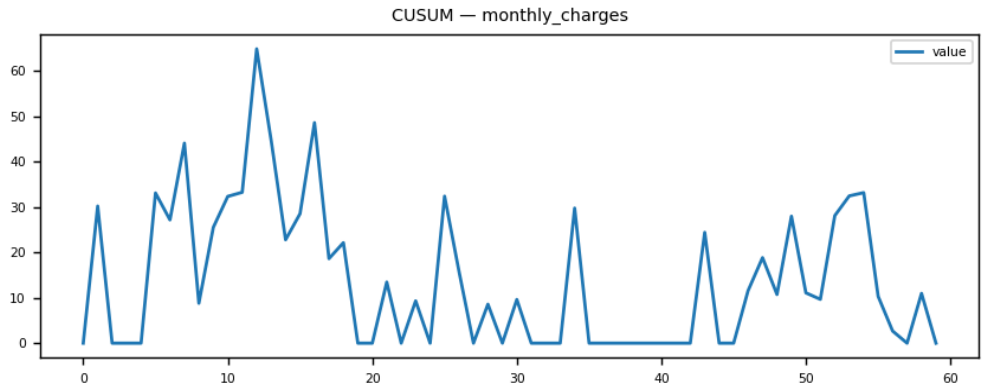
STEPS

1. Statistics → **Quality & SPC** → **CUSUM control chart**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

CUSUM control chart — monthly_charges

Reference k	Decision interval h	# signals	N
14.37	143.7	1	800



DETAILED INTERPRETATION

Cumulative-sum chart is highly sensitive to small persistent process shifts. The lead result here reads Reference k = 14.37; Decision interval h = 143.7; # signals = 1; N = 800. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst catch process drift before it reaches the customer.

11.2 Control chart builder (auto/phased)

LIVE OUTPUT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Auto-select I-MR/X-bar-R/p/c chart with optional baseline-phase limits. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, tenure_months. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges	tenure_months
52.94	68
113.7	45
35.51	49
71.45	64
29.16	42

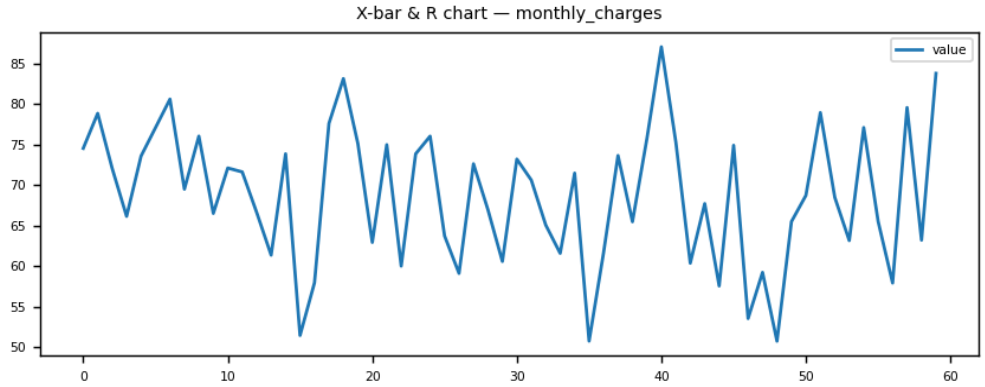
STEPS

1. Statistics → **Quality & SPC** → **Control chart builder (auto/phased)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, chart_type=auto, subgroup_size=30, phase=tenure_months`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Control chart builder — X-bar & R — monthly_charges

Chart	Center	UCL	LCL	# out of control	Points	Limits from
X-bar & R	69.09	94.16	44.03	0	80	phase '68'



DETAILED INTERPRETATION

Auto-selected the chart from the data type. With a phase column, limits are computed from the baseline phase and applied throughout. The lead result here reads Chart = X-bar & R; Center = 69.09; UCL = 94.16; LCL = 44.03. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst catch process drift before it reaches the customer.

11.3 EWMA control chart

LIVE OUTPUT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Detects small sustained shifts. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

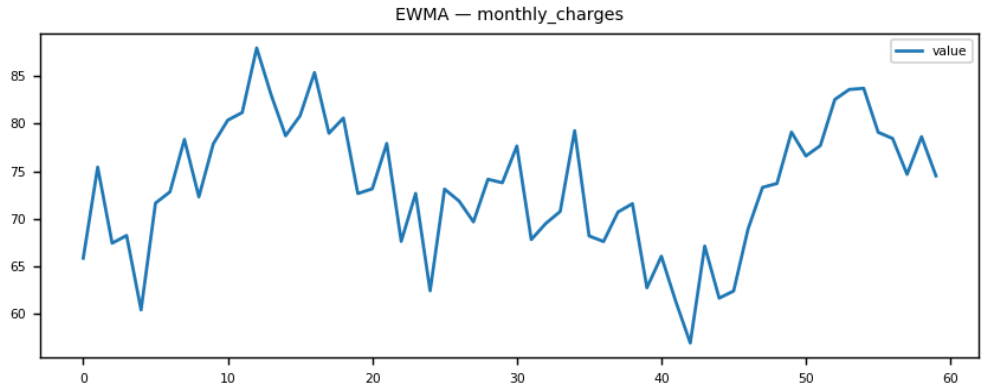
STEPS

1. Statistics → **Quality & SPC** → **EWMA control chart**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, lambda=0.2`. 4. Click **Run** and read the tables/chart below.

OUTPUT

EWMA control chart — monthly_charges

Lambda	Center	# out of control	N
0.2	69.09	0	800



DETAILED INTERPRETATION

Exponentially-weighted moving-average chart detects small, sustained shifts. The lead result here reads Lambda = 0.2; Center = 69.09; # out of control = 0; N = 800. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst catch process drift before it reaches the customer.

11.4 Gage R&R (MSA)

INPUT REQUIREMENT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Measurement systems analysis. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, tenure_months, support_calls. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges	tenure_months	support_calls
52.94	68	1
113.7	45	5
35.51	49	1
71.45	64	2
29.16	42	1

STEPS

1. Statistics → **Quality & SPC** → **Gage R&R (MSA)**. 2. Select the demo dataset above. 3. Set parameters — `part=monthly_charges, operator=tenure_months, measurement=support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Gage R&R expects a small set of parts and operators — got 776 parts × 71 operators. Choose the part-ID and operator columns (categorical), n

DETAILED INTERPRETATION

This technique has a specific data requirement: *Gage R&R expects a small set of parts and operators — got 776 parts × 71 operators. Choose the part-ID and operator columns (categorical), n*. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a quality & spc method and supports the decision to catch process drift before it reaches the customer.

11.5 I-MR control chart

LIVE OUTPUT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Individuals & moving-range chart. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

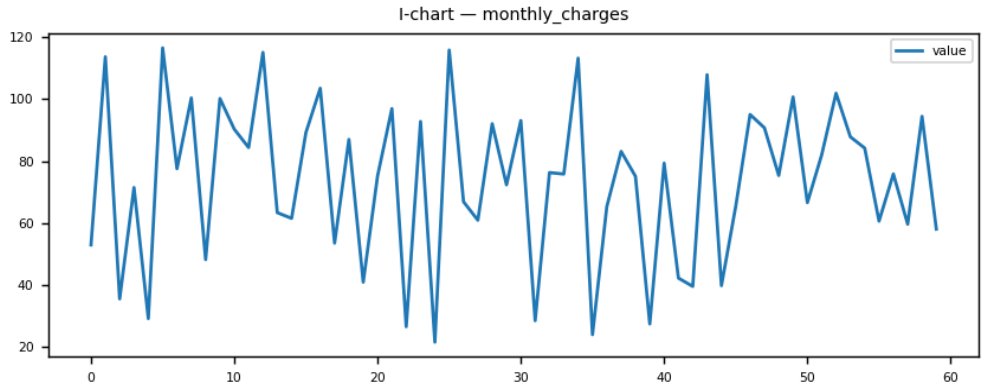
STEPS

1. Statistics → **Quality & SPC** → **I-MR control chart**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Individuals (I-MR) control chart — monthly_charges

Center line	UCL	LCL	MR-bar	# out of control	N
69.09	161.6	-23.45	34.79	0	800



DETAILED INTERPRETATION

Points beyond the control limits signal a special-cause (out-of-control) process. The lead result here reads Center line = 69.09; UCL = 161.6; LCL = -23.45; MR-bar = 34.79. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst catch process drift before it reaches the customer.

11.6 Multivariate SPC (PCA T²/SPE)

LIVE OUTPUT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Model-driven multivariate control charts (Hotelling T^2 and SPE/Q). It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on tenure_months, monthly_charges, support_calls. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

tenure_months	monthly_charges	support_calls
68	52.94	1
45	113.7	5
49	35.51	1
64	71.45	2
42	29.16	1

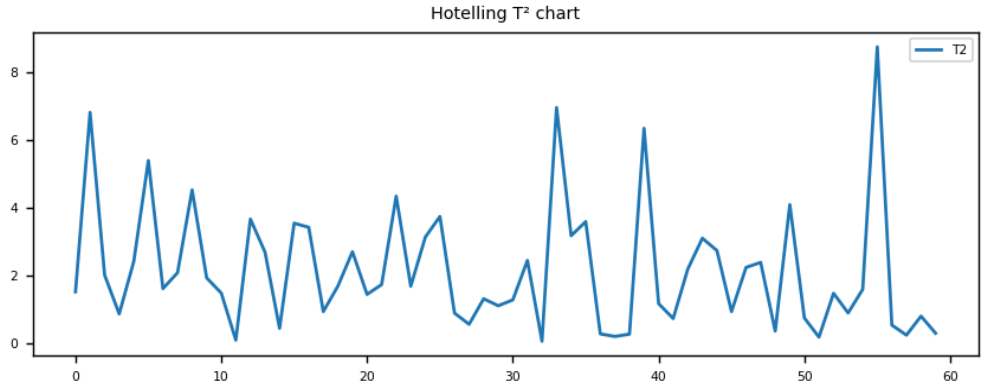
STEPS

1. Statistics → **Quality & SPC** → **Multivariate SPC (PCA T²/SPE)**. 2. Select the demo dataset above. 3. Set parameters — `columns=tenure_months, monthly_charges, support_calls, variance=0.9`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Model-driven multivariate SPC (PCA)

Components	Variance kept	T² limit (99%)	SPE limit (99%)	T² out-of-control	SPE out-of-control	N
2	69.6%	9.275	5.776	1	7	800



DETAILED INTERPRETATION

Hotelling T² tracks variation inside the PCA model; SPE (Q) tracks residual variation outside it. Points above a limit are multivariate out-of-control signals. The lead result here reads Components = 2; Variance kept = 69.6%; T² limit (99%) = 9.275; SPE limit (99%) = 5.776. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst catch process drift before it reaches the customer.

11.7 Pareto analysis

LIVE OUTPUT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Vital-few frequency chart. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

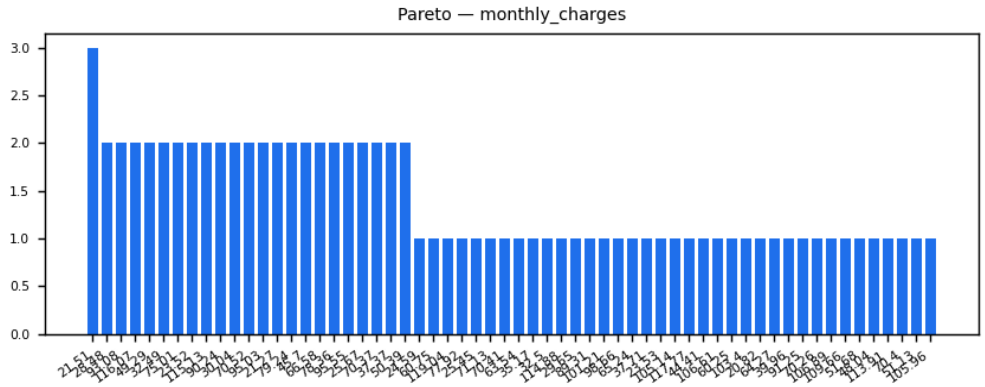
STEPS

1. Statistics → **Quality & SPC** → **Pareto analysis**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Pareto analysis — monthly_charges

Category	Count	%	Cumulative %
21.51	3	0.375	0.375
28.48	2	0.25	0.625
93.08	2	0.25	0.875
116.07	2	0.25	1.125
49.29	2	0.25	1.375
32.49	2	0.25	1.625
75.01	2	0.25	1.875
23.52	2	0.25	2.125
115.13	2	0.25	2.375
90.24	2	0.25	2.625
30.04	2	0.25	2.875
70.52	2	0.25	3.125



DETAILED INTERPRETATION

The 'vital few': categories are ordered by frequency (80/20 rule). The lead result here reads Category = 21.51; Count = 3; % = 0.375; Cumulative % = 0.375. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst catch process drift before it reaches the customer.

11.8 Process capability (Cp/Cpk)

LIVE OUTPUT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Capability indices vs spec limits. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

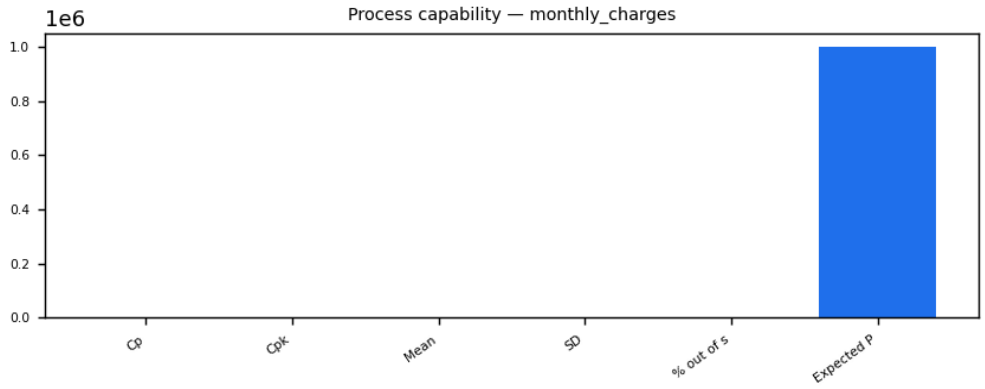
STEPS

1. Statistics → **Quality & SPC** → **Process capability (Cp/Cpk)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, lsl=5, usl=5`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Process capability — monthly_charges

Cp	Cpk	Mean	SD	% out of spec	Expected PPM defective
0	-0.7433	69.09	28.74	100	1e+06



DETAILED INTERPRETATION

Cp/Cpk ≥ 1.33 is generally capable; ≥ 1.67 is excellent. Cpk accounts for centering. The lead result here reads Cp = 0; Cpk = -0.7433; Mean = 69.09; SD = 28.74. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst catch process drift before it reaches the customer.

11.9 X-bar & R chart

LIVE OUTPUT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Subgroup mean/range control chart. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

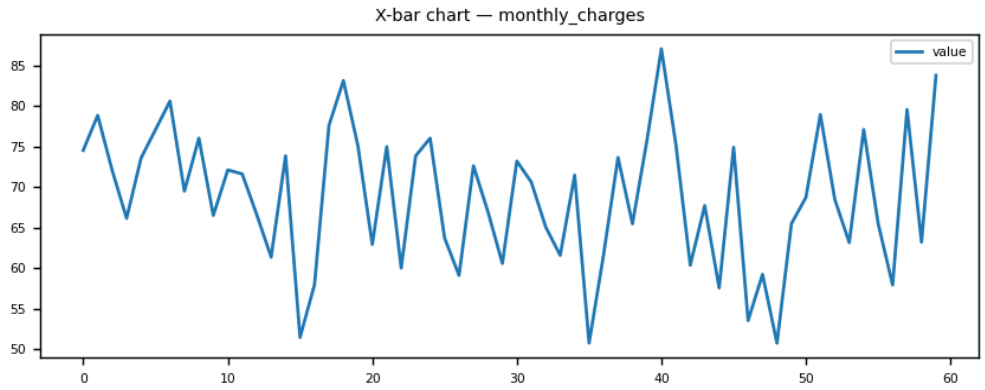
STEPS

1. Statistics → **Quality & SPC** → **X-bar & R chart**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, subgroup_size=30`. 4. Click **Run** and read the tables/chart below.

OUTPUT

X-bar & R control chart — monthly_charges (subgroup n=10)

Grand mean	R-bar	X-bar UCL	X-bar LCL	R UCL	# out of control	Subgroups
69.09	81.38	94.16	44.03	144.6	0	80



DETAILED INTERPRETATION

Monitors the subgroup mean (X-bar) and range (R) of a process over time. The lead result here reads Grand mean = 69.09; R-bar = 81.38; X-bar UCL = 94.16; X-bar LCL = 44.03. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst catch process drift before it reaches the customer.

11.10 c-chart (defect count)

LIVE OUTPUT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Defects-per-unit control chart. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

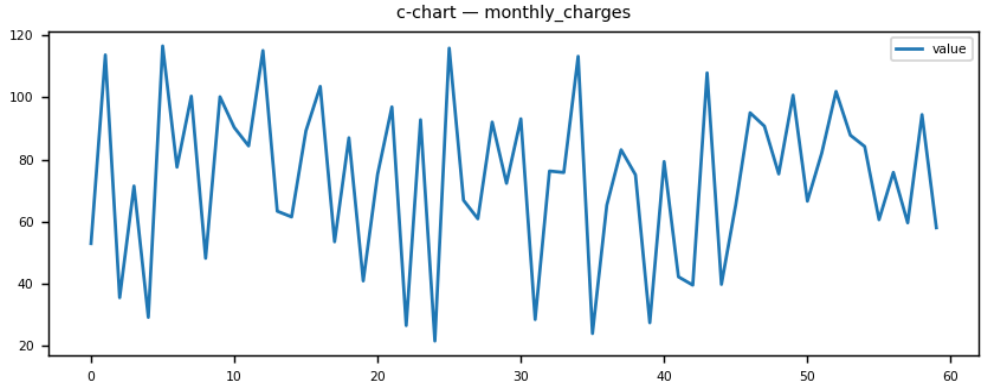
STEPS

1. Statistics → **Quality & SPC** → **c-chart (defect count)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges`. 4. Click **Run** and read the tables/chart below.

OUTPUT

c-chart (defects per unit) — monthly_charges

Center (c-bar)	UCL	LCL	# out of control	N
69.09	94.03	44.15	389	800



DETAILED INTERPRETATION

Monitors the count of defects per inspection unit (Poisson). The lead result here reads Center (c-bar) = 69.09; UCL = 94.03; LCL = 44.15; # out of control = 389. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst catch process drift before it reaches the customer.

11.11 p-chart (proportion)

INPUT REQUIREMENT

Group: Quality & SPC · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Fraction-defective control chart. It belongs to the **Quality & SPC** family, used to monitor a process for stability and capability. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges. Analysts reach for it to catch process drift before it reaches the customer.

SAMPLE DATA INPUT

monthly_charges
52.94
113.7
35.51
71.45
29.16

STEPS

1. Statistics → **Quality & SPC** → **p-chart (proportion)**. 2. Select the demo dataset above. 3. Set parameters — `column=monthly_charges, subgroup_size=30`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Need at least 2 subgroups

DETAILED INTERPRETATION

This technique has a specific data requirement: *Need at least 2 subgroups*. Points beyond the control limits or non-random runs flag an out-of-control process, while capability indices compare the process spread against the specification.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a quality & spc method and supports the decision to catch process drift before it reaches the customer.

---

12. Consumer & Market Research

12.1 Conjoint analysis (part-worths)

LIVE OUTPUT

Group: Consumer research · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Ratings-based conjoint — attribute utilities & importance. It belongs to the **Consumer research** family, used to segment customers and quantify preferences for business decisions. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, contract, region. Analysts reach for it to target the right customers with the right offer.

SAMPLE DATA INPUT

monthly_charges	contract	region
52.94	Month-to-month	East
113.7	One year	North
35.51	Month-to-month	North
71.45	Two year	West
29.16	Two year	West

STEPS

1. Statistics → **Consumer research** → **Conjoint analysis (part-worths)**. 2. Select the demo dataset above. 3. Set parameters — `rating=monthly_charges, attributes=contract, region`. 4. Click **Run** and read the tables/chart below.

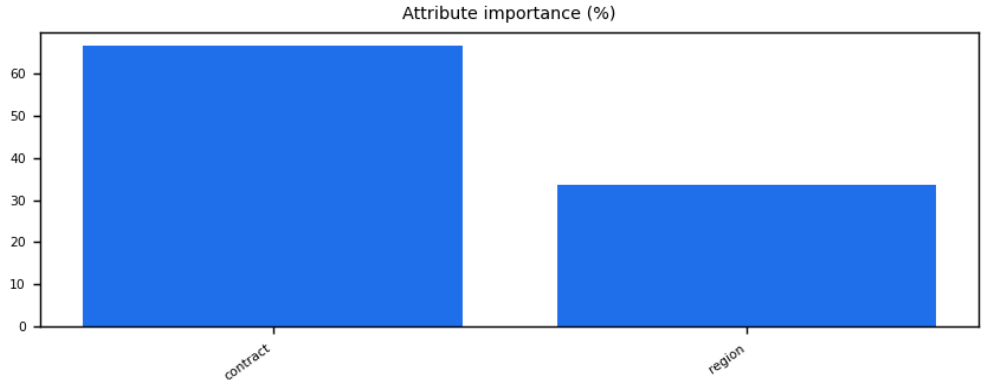
OUTPUT

Conjoint part-worth utilities — monthly_charges

Attribute	Level	Part-worth	Note
contract	Month-to-month	0	(reference)
contract	One year	2.092	
contract	Two year	1.635	
region	East	0	(reference)
region	North	0.6966	
region	South	-0.29	
region	West	0.7706	

Attribute importance

Attribute	Utility range	Importance %
contract	2.092	66.4
region	1.061	33.6



DETAILED INTERPRETATION

Utilities from OLS of monthly_charges on attribute levels (first level = reference = 0). Model R^2 = 0.001. Higher part-worth = more preferred level. The lead result here reads Attribute = contract; Level = Month-to-month; Part-worth = 0; Note = (reference). The result segments customers or quantifies preference/part-worths, feeding

directly into targeting, pricing and product-design decisions. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst target the right customers with the right offer.

12.2 Conjoint design generator

LIVE OUTPUT

Group: Consumer research · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

D-efficient orthogonal profile plan for a conjoint study. It belongs to the **Consumer research** family, used to segment customers and quantify preferences for business decisions. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on the selected variables. Analysts reach for it to target the right customers with the right offer.

SAMPLE DATA INPUT

Uses the full demo dataset described above.

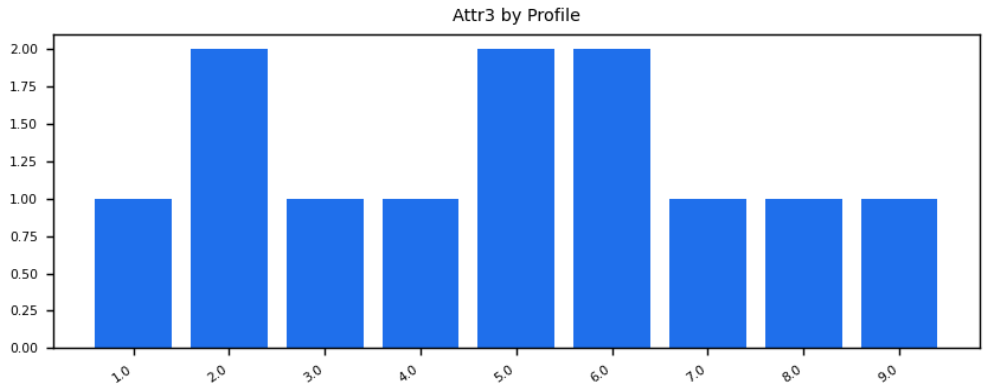
STEPS

1. Statistics → **Consumer research** → **Conjoint design generator**. 2. Select the demo dataset above. 3. Set parameters — `levels=3,3,2, num_profiles=0`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Conjoint design — 3 attributes (3, 3, 2), 9 profiles

Profile	Attr1	Attr2	Attr3
1	3	2	1
2	2	3	2
3	2	1	1
4	1	3	1
5	3	1	2
6	1	2	2
7	2	2	1
8	1	1	1
9	3	3	1



DETAILED INTERPRETATION

D-efficient main-effects plan selected from 18 full-factorial combinations. Levels shown 1..k per attribute. Relative D-efficiency = 0.259. Present these profiles to respondents and analyse ratings with the conjoint analysis. The lead result here reads Profile = 1; Attr1 = 3; Attr2 = 2; Attr3 = 1. The result segments customers or quantifies preference/part-worths, feeding directly into targeting, pricing and product-design decisions. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst target the right customers with the right offer.

12.3 Discrete choice (conditional logit)

INPUT REQUIREMENT

Group: Consumer research · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Choice-based conjoint — utility weights from choice data. It belongs to the **Consumer research** family, used to segment customers and quantify preferences for business decisions. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, tenure_months, support_calls. Analysts reach for it to target the right customers with the right offer.

SAMPLE DATA INPUT

monthly_charges	tenure_months	support_calls
52.94	68	1
113.7	45	5
35.51	49	1
71.45	64	2
29.16	42	1

STEPS

1. Statistics → **Consumer research** → **Discrete choice (conditional logit)**. 2. Select the demo dataset above. 3. Set parameters — `choice=monthly_charges, choice_set=tenure_months, attributes=tenure_months, monthly_charges, support_calls`. 4. Click **Run** and read the tables/chart below.

OUTPUT

Requires discrete-choice data — one row per alternative within each choice set, a chosen indicator, a choice-set id and alternative attributes (e.g. a conjoint or transport-mode study).

DETAILED INTERPRETATION

This technique has a specific data requirement: *Requires discrete-choice data — one row per alternative within each choice set, a chosen indicator, a choice-set id and alternative attributes (e.g. a conjoint or transport-mode study)*. The result segments customers or quantifies preference/part-worths, feeding directly into targeting, pricing and product-design decisions.. Supplied with a suitable dataset for a telecom customer base — tenure, monthly charges, support calls, spend and churn, it returns the standard output for a consumer research method and supports the decision to target the right customers with the right offer.

12.4 MaxDiff (best-worst scaling)

LIVE OUTPUT

Group: Consumer research · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Preference scores from best/worst choices. It belongs to the **Consumer research** family, used to segment customers and quantify preferences for business decisions. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on monthly_charges, tenure_months. Analysts reach for it to target the right customers with the right offer.

SAMPLE DATA INPUT

monthly_charges	tenure_months
52.94	68
113.7	45
35.51	49
71.45	64
29.16	42

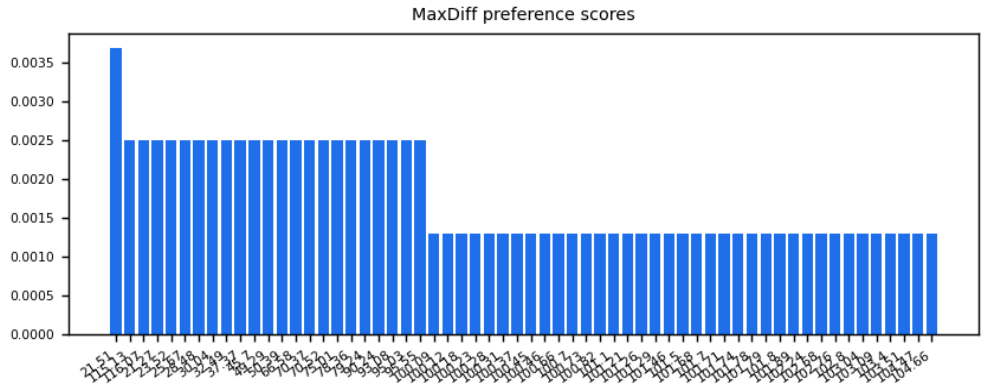
STEPS

1. Statistics → **Consumer research** → **MaxDiff (best-worst scaling)**. 2. Select the demo dataset above. 3. Set parameters — `best=monthly_charges, worst=tenure_months`. 4. Click **Run** and read the tables/chart below.

OUTPUT

MaxDiff (best-worst) scores — 847 items

Item	Best count	Worst count	Score
21.51	3	0	0.0037
115.13	2	0	0.0025
116.07	2	0	0.0025
21.27	2	0	0.0025
23.52	2	0	0.0025
25.67	2	0	0.0025
28.48	2	0	0.0025
30.04	2	0	0.0025
32.49	2	0	0.0025
37.37	2	0	0.0025
45.7	2	0	0.0025
49.29	2	0	0.0025



DETAILED INTERPRETATION

Score = (best – worst) / tasks; higher = more preferred. Best-worst scaling spreads preferences more sharply than rating scales. The lead result here reads Item = 21.51; Best count = 3; Worst count = 0; Score = 0.0037. The result segments customers or quantifies preference/part-worths, feeding directly into targeting, pricing and product-design decisions. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst target the right customers with the right offer.

12.5 RFM analysis (direct marketing)

LIVE OUTPUT

Group: Consumer research · Dataset: a telecom customer base — tenure, monthly charges, support calls, spend and churn

WHAT THIS DOES

Recency/Frequency/Monetary scoring into customer value tiers. It belongs to the **Consumer research** family, used to segment customers and quantify preferences for business decisions. In this example it is applied to a telecom customer base — tenure, monthly charges, support calls, spend and churn, drawing on recency_days, frequency, monetary. Analysts reach for it to target the right customers with the right offer.

SAMPLE DATA INPUT

recency_days	frequency	monetary
297	46	3197
179	33	1233
291	6	2501
112	6	1468
135	38	4195

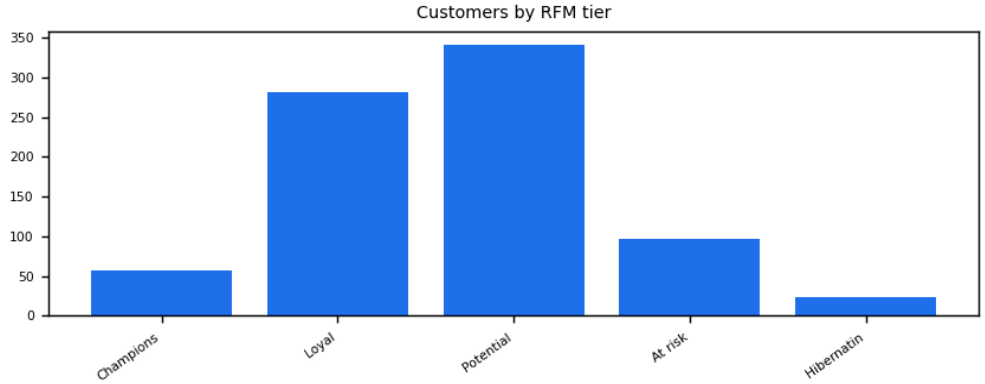
STEPS

1. Statistics → **Consumer research** → **RFM analysis (direct marketing)**. 2. Select the demo dataset above. 3. Set parameters — `recency=recency_days, frequency=frequency, monetary=monetary`. 4. Click **Run** and read the tables/chart below.

OUTPUT

RFM analysis — customer value tiers

Segment	Customers	% of base	Avg monetary	Avg frequency
Champions	57	7.1%	4180	38.58
Loyal	281	35.1%	3187	30.89
Potential	341	42.6%	2289	21.66
At risk	97	12.1%	1760	13.95
Hibernating	24	3.0%	911.1	6.92



DETAILED INTERPRETATION

Recency/Frequency/Monetary quintile scores (5 = best) combined into value tiers. Champions = recent, frequent, high-spend customers. The lead result here reads Segment = Champions; Customers = 57; % of base = 7.1%; Avg monetary = 4180. The result segments customers or quantifies preference/part-worths, feeding directly into targeting, pricing and product-design decisions. For a telecom customer base — tenure, monthly charges, support calls, spend and churn, this lets the analyst target the right customers with the right offer.