

Your First Screening Session

A step-by-step guide for novice researchers, with a matching English-narrated video and captions.

Start safely: use the built-in sample before importing your own CSV. No account is required. Keep an unchanged copy of your original dataset.

Open [PurifyData Pro](#) | Video length: 06:38 | 12 chapters

Follow Along

01	Start without an account	00:00 / p. 2
02	Practise with the sample	00:30 / p. 3
03	Import and select variables	00:55 / p. 4
04	Review the screening settings	01:25 / p. 5
05	Read the overview counts	02:04 / p. 6
06	Inspect one-variable outliers	02:39 / p. 7
07	Inspect score patterns and correlations	03:08 / p. 8
08	Understand the combined workflow	03:43 / p. 9
09	Check normality and the data basis	04:12 / p. 10
10	Use recommendations as guidance	04:48 / p. 11
11	Try optional imputation carefully	05:22 / p. 12
12	Export and document your decisions	05:59 / p. 13

Scope: preliminary data screening and initial normality assessment. Flags are prompts for investigation; they are not automatic reasons to discard observations. This tutorial does not prescribe a final statistical model.

01. Start without an account

VIDEO 00:00 | Follow the numbered steps

1. Open the app in a current desktop browser, such as Edge or Chrome.
2. On Home, find Open workspace and Try sample data below the introductory information.
3. Use Open workspace for your own CSV. Use Try sample data for a safe practice session.
4. Keep your original data file separately; the app is not a permanent data-storage service.

 **Data screening workspace**

PurifyData Pro is a non-profit research tool that is free forever to use. No subscription, payment, or credit card is required. No account is needed. Import CSV files, select variables, screen outliers, assess normality, run imputation, and export reports freely.

 **Protected analysis workspace**

PurifyData Pro acts as a completely self-contained local workspace, it does not collect, monitor, track, or share your proprietary data records. Simply closing or refreshing your browser tab immediately flushes the entire structural data state from your computer's temporary memory, erasing all active records completely.

 **AI assistant** BETA

Ask plain-language questions about outliers, normality, missing data, imputation, and app outputs. You may optionally review and share a privacy-safe results summary; raw dataset records are not sent automatically.

Free forever Non-profit access. No payment, subscription, or credit card required.

 **Open workspace**

 **Try sample data**

 **Support the app**

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

Remember

The menu's Learn section contains Video walkthrough and PDF guide, with download options. This tutorial demonstrates interface operation; it does not prescribe deleting outliers, imputing missingness, or using identical cut-offs in every study.

02. Practise with the sample

VIDEO 00:30 | Follow the numbered steps

1. Click Try sample data on Home.
2. Check the Variables panel: the example has 16 rows, 7 columns and 5 detected numeric variables.
3. Use the left panel for settings and the workspace tabs for results.
4. The example is for learning; its results are not evidence about your own research data.

Variables

Built-in sample dataset: 16 rows, 7 columns, 5 numeric variables detected.

Import CSV

☐

participant

text

☒

age

numeric

☒

stress_score

numeric

☒

sleep_hours

numeric

☒

reaction_time

numeric

☒

exercise_days

numeric

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

Remember

You can reopen the menu using the three-dot button in the top-right corner. Open workspace starts at the top without clearing an existing dataset.

03. Import and select variables

VIDEO 00:55 | Follow the numbered steps

1. Save a plain CSV with column names in the first row. Keep cases in rows and variables in columns.
2. Click Import CSV in the Variables panel, or use Import CSV in the menu.
3. Choose your file from your device. The demonstration reuses the built-in example as a CSV.
4. In Select required variables, select the numeric measures needed for the analysis.
5. Click Analyze selected variables. Inspect the resulting variable names and row count.

CSV VARIABLE SETUP

Select required variables

Cancel import

5 numeric variable(s) were detected in practice-data.csv.

- Select variables for the main workflow: Univar. Outliers, Multivar. Outliers, and the Combined Outliers sequence.
- No variable-count cap is applied to univariate outlier screening or Mahalanobis outlier screening.
- Mahalanobis requires at least two complete cases. If the ordinary covariance matrix is singular or high-dimensional, the app uses a small ridge-regularized covariance inverse.
- Normality assessment is an additional diagnostic and may use responsiveness limits.

Search variables

Type a variable name

Select all

Deselect all

☒ age

☒ stress_score

☒ sleep_hours

☒ reaction_time

☒ exercise_days

5 of 5 selected

Analyze selected variables

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

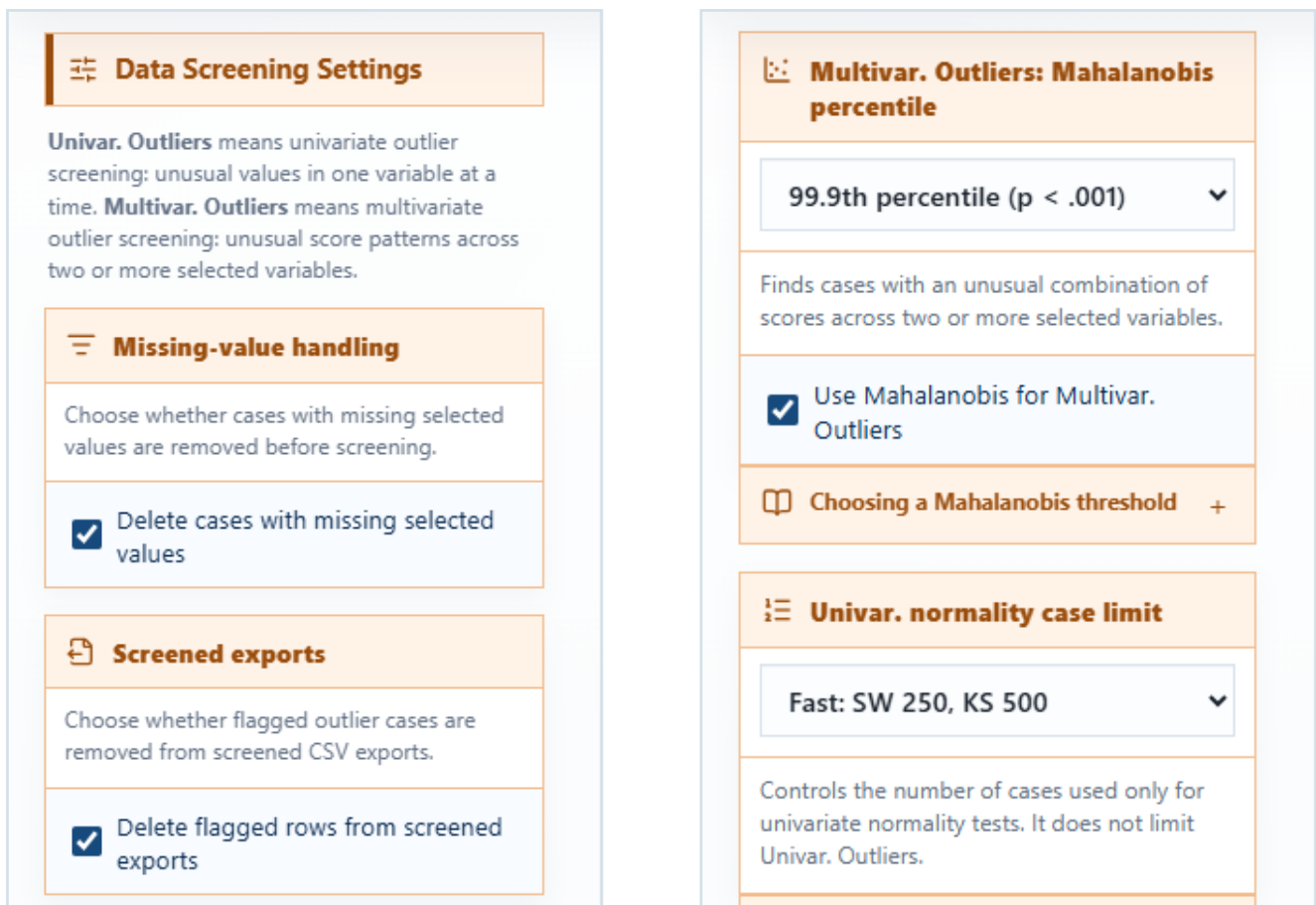
Remember

Exclude identifiers and variables inappropriate for continuous-data screening. If a measure is not detected as numeric, check for text, units, inconsistent decimal formatting, or non-numeric missing-value codes in your source file.

04. Review the screening settings

VIDEO 01:25 | Follow the numbered steps

1. Read the Screening Setting Notice after importing, then close it using its X button.
2. Find the orange Data Screening Settings section; its icon-labelled title bands identify the rules.
3. Review missing-value handling, Z-score and IQR cut-offs, and the export-deletion checkbox.
4. Scroll within the left panel to review Mahalanobis, skewness, kurtosis and normality case limits.
5. Expand a setting's guide for context, then choose settings that fit your research plan.



Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

Remember

If missing-value deletion is off and selected values are missing, the Missing-value decision dialog offers Retain for manual review or Create imputed dataset. These are different choices; do not choose imputation merely to make a warning disappear.

05. Read the overview counts

VIDEO 02:04 | Follow the numbered steps

- 1. Click Overview after choosing your variables and settings.
- 2. Read Rows, Variables and Selected numeric first.
- 3. Compare Complete rows with Multivar. tested and Combined tested.
- 4. Use the corresponding retained cards to understand the final screened counts.
- 5. Expand Understanding overview counts when a label is unfamiliar.

ROWS	VARIABLES	SELECTED NUMERIC	COMPLETE ROWS	UNIVAR. RETAINED
16	7	5	14	13
	MULTIVAR. TESTED	MULTIVAR. RETAINED	COMBINED TESTED	COMBINED RETAINED
	14	14	13	13

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

Remember
A case can be retained without being tested by Mahalanobis distance when missing-value deletion is off. Not applied does not mean zero outliers; it means the corresponding workflow was not applied. Read status wording as well as numbers.

06. Inspect one-variable outliers

VIDEO 02:39 | Follow the numbered steps

1. Click Univar. Outliers.
2. Inspect the screening table for the selected variables and enabled rules.
3. Read Univariate Screening Note to confirm the rules actually used.
4. Scroll to Univariate Deleted and Screened Cases for case-level review.
5. Use Export cases CSV when you need a separate case-review record.

Univariate Screening

Export screened CSV

Variable	N	Missing	Mean	SD	Min	Q1	Median	Q3	Max	IQR low	IQR high	Skew	Kurtosis
age	13	0	41.62	15.61	22	29	41	50	69	-2.5	81.5	0.37	-1.1
stress_score	13	0	27.23	6.99	18	21	27	32	40	4.5	48.5	0.3	-1.1
sleep_hours	13	0	6.22	0.88	4.6	5.5	6.4	6.9	7.4	3.4	9	-0.35	-1.1
reaction_time	13	0	333.92	40.28	284	301	329	355	412	220	436	0.58	-0.9
exercise_days	13	0	2.31	1.44	0	1	2	3	5	-2	6	0.28	-1.1

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state. The table continues horizontally; scroll inside the app to inspect the remaining columns.

Remember

Outlier screening and deleting flagged cases are not the same action. Review the export-deletion checkbox and substantively justify any exclusions. Keep a record of the original values and decisions.

07. Inspect score patterns and correlations

VIDEO 03:08 | Follow the numbered steps

1. Click Multivar. Outliers.
2. Inspect Mahalanobis Distance, the selected cut-off and the number of tested cases.
3. Scroll to Correlation Matrix to review variable relationships and any high-correlation cautions.
4. Use Export matrix CSV if you need the matrix outside the app.
5. Read the screening note and case table before deciding whether exclusions are justified.

Mahalanobis Distance		Export screened CSV
Complete cases tested		14
Multivariate outliers		0
Mahalanobis percentile		99.9%
Degrees of freedom		5
Chi-square cut-off		20.52
Empirical data percentile		11.23
Covariance method		Ordinary inverse

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

Remember

High correlation is a caution about relationships among variables, not an automatic outlier-deletion rule. No Mahalanobis value is different from a calculated non-outlier result.

08. Understand the combined workflow

VIDEO 03:43 | Follow the numbered steps

- 1. Click Combined Outliers.
- 2. Read Combined Outlier Screening and its screening note.
- 3. Remember the order: Univar. Outliers first, then Multivar. Outliers on the remaining cases.
- 4. Compare its tested and retained counts with the standalone workflows.
- 5. Choose the export that matches your documented analysis plan.

Combined Outlier Screening		Export screened CSV
Workflow	Univar. Outliers then Multivar. Outliers	
Cases after Univar. Outliers		13
Complete cases tested		13
Multivariate outliers		0
Mahalanobis percentile		99.9%
Chi-square cut-off		20.52
Rows retained in Combined export		13

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

Remember

Do not mix standalone Multivar. Outliers statistics with a Combined Outliers dataset without checking that they refer to the same cases and settings.

09. Check normality and the data basis

VIDEO 04:12 | Follow the numbered steps

1. Click Normality and wait until the progress indicator finishes.
2. Read Normality Assessment Data Basis before interpreting statistics.
3. Expand Decision rule to inspect the app's criteria and case-limit settings.
4. Scroll to the univariate table, histograms and Q-Q plots.
5. Review standalone multivariate and combined assessments separately.

U UNIVAR. OUTLIERS DATA 13 case(s) After missing-value handling and Univar. Outliers. Supports univariate normality, histograms, Q-Q plots, and before/after boxplots.	M STANDALONE MULTIVAR. OUTLIERS DATA 14 case(s) After missing-value handling and standalone Multivar. Outliers. Supports Mardia, Henze-Zirkler, and the chi-square Q-Q plot.	U+M COMBINED OUTLIERS UNIVARIATE DATA 13 case(s) After missing-value handling, Univar. Outliers, and stage 2 Multivar. Outliers. Supports combined univariate normality, histograms, Q-Q plots, and before/after boxplots.
U+M COMBINED OUTLIERS MULTIVARIATE DATA 13 case(s) After missing-value handling, Univar. Outliers, and stage 2 Multivar. Outliers. Supports Mardia, Henze-Zirkler, and the chi-square Q-Q plot.		

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

Remember

Check the number of cases actually tested, especially when a case-limit mode is used. These are initial screening diagnostics, not a complete assessment of regression residuals or every assumption of a later statistical model.

10. Use recommendations as guidance

VIDEO 04:48 | Follow the numbered steps

- 1. Click Recommendations.
- 2. Read Consolidated Decisions and note the workflow to which each decision refers.
- 3. Expand only the Recommended Statistical Analysis Path cards relevant to your question.
- 4. Read the Plain-language Reporting Note as a starting point for your own checked account.

Recommended Decision - • Univar. Outliers data: Parametric with outlier check. Use a parametric method, then compare results with and without justified flagged cases. • Standalone Multivar. Outliers data: Parametric with outlier check. Use a parametric method, then compare results with and without justified flagged cases. • CFA, CB-SEM, and PLS-SEM: CFA/CB-SEM suitable. Conventional ML CFA/CB-SEM may be reasonable; use PLS-SEM for prediction-oriented or exploratory work. • Final Combined Outliers data: Robust parametric preferred. Prefer robust standard errors, bootstrapping, or model-specific diagnostics. • Outliers were flagged. Compare results with and without justified deleted cases before drawing a conclusion. • Use the Consolidated Decisions cards when you need the detailed evidence behind these recommendations.	First-generation Analyses +
Measurement and Advanced Modelling +	How to Choose +

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

Remember

The app supports initial data screening and initial normality assessment. It does not replace a full research analysis, subject-matter judgement, or independent statistical validation where required.

11. Try optional imputation carefully

VIDEO 05:22 | Follow the numbered steps

- 1. Click Imputation, or find Optional Imputation in the left panel.
- 2. Choose a method appropriate for the purpose; Mean is used only as a demonstration here.
- 3. Click Run imputation and inspect the Imputation Report.
- 4. Use Export imputed CSV to save a separate derived dataset.
- 5. Use imputed data for screening applies the derived copy to screening. Restore original data returns to the original imported dataset.

Imputation Report

Use imputed data for screeningExport imputed CSV

Dataset relationship

The original screening dataset uses listwise deletion for missing selected values. This report describes the separate imputed export; choose Use imputed data for screening to make it the active workspace dataset.

Method	Mean imputation
Imputed cells	2 of 80
age	0 value(s) imputed
stress_score	1 value(s) imputed
sleep_hours	1 value(s) imputed
reaction_time	0 value(s) imputed
exercise_days	0 value(s) imputed

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state.

Remember

Optional Imputation in the left panel also has icon-labelled Method and hot-deck Seed settings. Creating an imputed dataset and using it for screening are separate actions. These are single-imputation methods, not multiple imputation or full-information model estimation; document their limitations.

12. Export and document your decisions

VIDEO 05:59 | Follow the numbered steps

1. Choose the Univar. Outliers, Multivar. Outliers or Combined Outliers workflow you intend to report.
2. Click Export screened CSV on that page to save its screened dataset.
3. Save Export cases CSV and Export matrix CSV when relevant; these are different outputs.
4. On Normality, use Export normality CSV. On Imputation, use Export imputed CSV for the derived data.
5. Record your variables, rules, missing-value handling, exclusions and working-data basis alongside the files.

Univariate Screening

Export screened CSV

Variable	N	Missing	Mean	SD	Min	Q1	Median	Q3	Max	IQR low	IQR high	Skew	Kurtosis
age	13	0	41.62	15.61	22	29	41	50	69	-2.5	81.5	0.37	-1.1
stress_score	13	0	27.23	6.99	18	21	27	32	40	4.5	48.5	0.3	-1.1
sleep_hours	13	0	6.22	0.88	4.6	5.5	6.4	6.9	7.4	3.4	9	-0.35	-1.1
reaction_time	13	0	333.92	40.28	284	301	329	355	412	220	436	0.58	-0.1
exercise_days	13	0	2.31	1.44	0	1	2	3	5	-2	6	0.28	-1.1

Actual app interface. Built-in practice data; the screenshot shows this chapter's example state. The table continues horizontally; scroll inside the app to inspect the remaining columns.

Remember

Do not overwrite the original CSV. Closing or refreshing clears the app's working dataset, but it does not erase files you have downloaded. The optional AI assistant is not required for this workflow; do not put confidential information into its questions.