JASP for Audit User Manual

Statistical Auditing Group

“The best things in life are free.”

Contents

Preface

The **JASP for Audit User Manual** provides detailed instructions and best practices for working with the Audit module in the free and open-source software JASP. It covers various aspects, including data import and export, analysis techniques, and interpretation of results.

The Statistical Auditing Group at Nyenrode Business University, which develops and maintains JASP for Audit, curates the manual to ensure users have accurate and up-to-date information.

Getting Started

Statistical theory is fundamental to many auditing procedures. To perform these procedures effectively, auditors need user-friendly software for statistical analyses and the knowledge to interpret the results. JASP (JASP Team, 2025) is an open-source, free-of-charge, cross-platform statistical software program that supports statistical auditing through its Audit module (Derks et al., 2021, 2023).

The Audit module (i.e., JASP for Audit) enables auditors to plan, execute, and interpret a wide range of statistical auditing procedures using state-of-the-art statistical methods, thereby reducing programming errors and simplifying the process. Tailored for auditors, the module features an intuitive interface that aligns with audit processes and international standards on auditing. In addition to standard frequentist methods, the Audit module incorporates Bayesian methods to enhance audit transparency and efficiency by utilizing existing information.

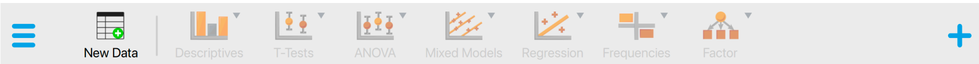
In summary, the Audit module takes care of the complex statistical work, enabling you to concentrate on interpreting the results of your analysis. The remaining paragraphs in this chapter discuss how to get started using JASP for Audit.

Downloading JASP

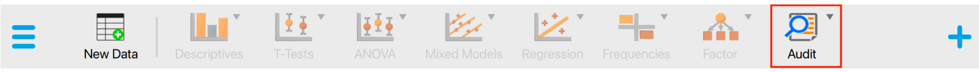
JASP for Audit is part of JASP, which can be freely downloaded from www.jasp-stats.org. Click the 'Download JASP' button on the homepage to access the download page and choose your preferred installation. JASP is available for Windows, MacOS, Linux, and Chrome OS.

Enabling the Audit module

After opening JASP, you will see the following main menu bar at the top of the screen.



To find the Audit module, click the '+' icon on the right of this menu bar. A different menu will appear on the right side which shows all available modules. Check the box next to 'Audit' to make the module visible in the main menu bar. You can now access the Audit module and its analyses by clicking its module icon in the menu bar (see image below).



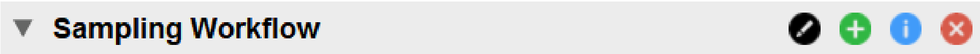
You can find detailed instructions for each analysis in the Audit module in the corresponding chapter of this manual.

Miscellaneous

The following paragraphs detail miscellaneous features, including where to locate help files and how to the reliability of the statistical results is ensured.

Help files

Once you open an analysis in the Audit module, you can click the blue 'i' icon next to the analysis title to access a help file that explains its functionality. Additional help files for certain settings can be accessed by clicking the blue 'i' icon next to those settings.



Validation of statistical results

The statistical results generated by the Audit module are based on the R package jfa (Derks, 2025). For comprehensive documentation and information on the benchmarks used for validation, please visit the package website at https://koenderks.github.io/jfa/.

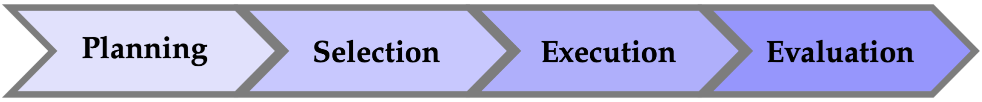
Part I

Audit Sampling

Chapter 1

Sampling Workflow

The goal of statistical audit sampling is to infer the misstatement in a population based on a representative sample. This can be challenging, but the Audit module simplifies the process into four stages: planning, selection, execution, and evaluation.



More detailed information about the individual stages in the audit sampling workflow is provided below.

1.1 The four stages of the sampling workflow

In the planning stage, you determine the sample size needed to support the assertion that the population's misstatement is below the performance materiality. This involves using prior audit outcomes and information about inherent risk and control risk. Expectations about error rates also influence the sample size required to maintain statistical confidence.

Using the sample size from the planning stage, you select a statistically representative sample. Each sampling unit receives an inclusion probability, and units are selected based on these probabilities. Monetary unit sampling assigns probabilities to individual monetary units, making higher-value items more likely to be selected. Record sampling assigns equal probabilities to all items.

In the execution stage, you assess the correctness of selected items. The simplest method categorizes items as correct or incorrect, while a more accurate method considers the true value (audit value) of items. Annotating samples with audit values provides a more precise estimate of misstatement. If book values are unavailable, use the correct/incorrect method.

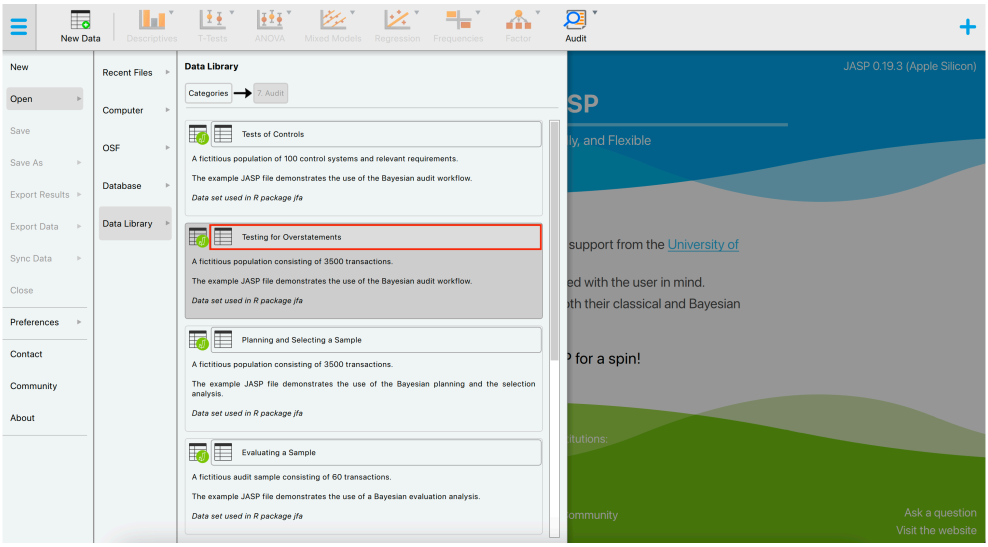
In the evaluation stage, you use the annotated sample to infer the total misstatement in the population. Statistical techniques calculate a projected maximum misstatement, and the population is approved if this is below the performance materiality.

This manual emphasizes the practical application of the audit sampling workflow in JASP. For a deeper understanding of the statistical theory behind the four stages of the audit sampling workflow, read the free online book Statistical Audit Sampling with R.

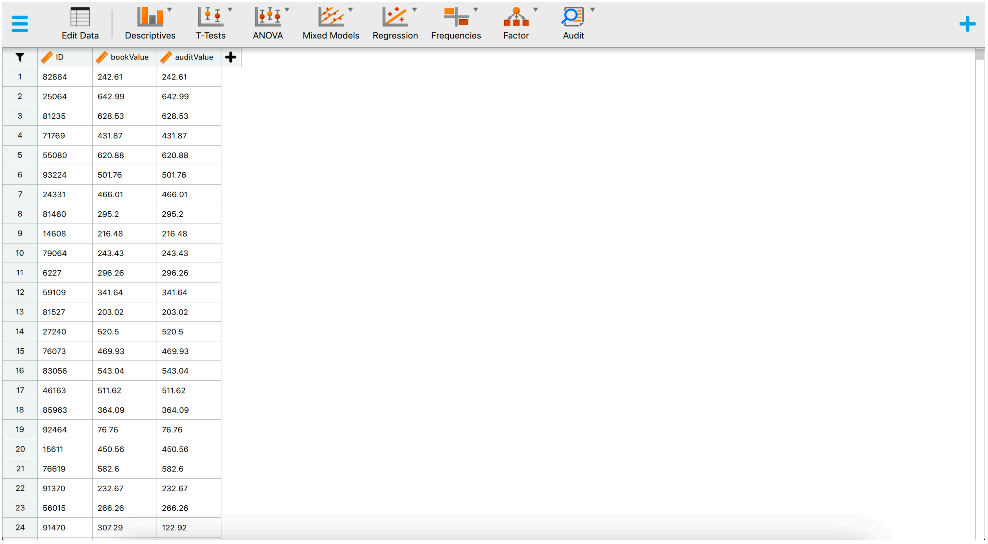
1.2 Practical example

The Audit module in JASP offers two ways to navigate the audit sampling workflow: the Sampling Workflow analysis, which guides you through all four stages, and individual analyses for Planning, Selection, and Evaluation. This chapter uses the classical sampling workflow analysis to explain the Audit module's core functionality. Note that a Bayesian variant of the sampling workflow is also available.

Let's explore an example of the audit sampling workflow. To follow along, open the 'Testing for Overstatements' dataset from the Data Library. Navigate to the top-left menu, click 'Open', then 'Data Library', select '7. Audit', and finally click on the text 'Testing for Overstatements' (not the green JASP-icon button).

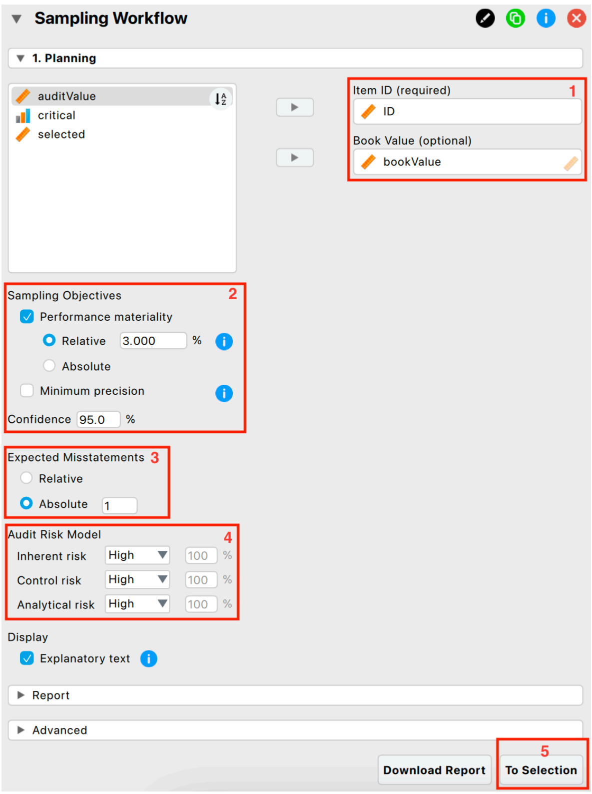


This will open a dataset with 3500 rows and three columns: 'ID', 'bookValue', and 'auditValue'. The 'ID' column represents the identification number of the items in the population. The 'bookValue' column shows the recorded values of the items, while the 'auditValue' column displays the true values. The 'auditValue' column is included for illustrative purposes, as auditors typically know the true values only for the audited sample, not for all items in the population.



1.2.1 Stage 1: Planning

To start the sampling workflow, click on the Audit module icon and select 'Sampling Workflow'. This will open the following interface, where you need to specify the settings for the statistical analysis.



The following five settings are required:

1. **Indicate the variables:** First, enter the variable indicating the identification numbers of the items in the corresponding box. Optionally, if you have access to the book values of the items, you can enter this variable as well.

2. **Sampling objectives:** Next, formulate your sampling objectives. Enable the 'Performance materiality' objective if you want to test whether the total misstatement in the population exceeds a certain limit (i.e., the performance materiality). This approach enables you to plan a sample such that, when the sample meets your expectations, the maximum error is said to be below performance materiality. Enable the 'Minimum precision' objective if you want to obtain a required minimum precision when estimating the total misstatement in the population. This approach enables you to plan a sample such that, when the sample meets expectations, the uncertainty of your estimate is within a tolerable percentage. In the example, we choose a performance materiality of 3.5%.

3. **Expected misstatement:** Then, indicate how many misstatements are tolerable in the sample. In the example, we choose to tolerate one full misstatement in the sample.

4. **Prior information:** Additionally, indicate the risks of material misstatement via the audit risk model. According to the Audit Risk Model, audit risk can be divided into three constituents: inherent risk, control risk, and detection risk. Inherent risk is the risk posed by an error in a financial statement due to a factor other than a failure of internal controls. Control risk is the probability that a material misstatement is not prevented or detected by the internal control systems of the company (e.g., computer-managed databases). Both these risks are commonly assessed by the auditor on a 3-point scale consisting of low, medium, and high. Detection risk is the probability that an auditor will fail to find material misstatements in an organization's financial statements. For a given level of audit risk, the tolerable level of detection risk bears an inverse relationship to the other two assessed risks. Intuitively, a greater risk of material misstatement should require a lower tolerable detection risk and, accordingly, more persuasive audit evidence. In this example, we choose to set all risks to 'High' and solely rely on evidence from substantive testing.

The primary output from the planning stage, shown below, indicates that a minimum sample size of 134 sampling units is required to achieve 95% assurance that the misstatement in the population is below 3.5%, while allowing for one misstatement in the sample.

Table 1. Planning Summary

	Value
Performance materiality	0.035
Inherent risk	1.000
Control risk	1.000
Analytical risk	1.000
Detection risk	0.050
Tolerable misstatements	1.000
Minimum sample size [a]	134

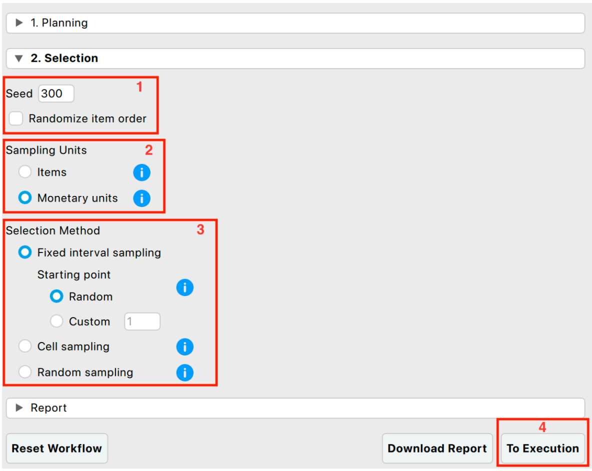
Note. The minimum sample size is based on the binomial distribution (p = 0.035)
[a] Based on this sample size, the selection interval spans 10471.8 units.

5. **Next stage:** Finally, progress to the selection stage by clicking the 'To Selection' button.

For a more detailed explanation of the settings and output in the planning stage, see Chapter 2.

1.2.2 Stage 2: Selection

In the selection stage, you must select the 134 sampling units from the population. Once the 'To Selection' button is pressed, the interface from the selection stage opens.



The following four settings are required:

1. **Randomness:** Begin by selecting the settings related to randomness in the selection procedure. The seed setting is important as it ensures that random procedures are reproducible, allowing for consistent results across multiple runs. A random number will be chosen each time you start the analysis. Additionally, the 'Randomize item order' setting is available to randomly shuffle the rows in the dataset, which can help mitigate any biases that might arise from the original order of the data.

2. **Sampling units:** Next, specify the sampling units for the selection process. These units can either be items or monetary units. If no book value variable

is provided, the sampling units default to 'Items', enabling attribute sampling. Conversely, if a book value variable was indicated during the planning stage, the sampling units default to 'Monetary units', facilitating monetary unit sampling (MUS). MUS is particularly useful for auditing financial data as it considers the monetary value of each unit.

3. **Sampling method:** Then, choose the selection method to be used in the sampling process. The available algorithms include:

- **Fixed interval sampling:** This method selects units at regular intervals from the dataset, ensuring a systematic sampling approach.
- **Cell sampling:** This technique involves dividing the dataset into cells and randomly selecting units from each cell, promoting a systematic sampling approach with a bit of randomness.
- **Random sampling:** This approach randomly selects units from the entire dataset, providing a simple yet effective method for ensuring randomness.

The primary output from the selection stage, as shown in the first table below, reveals that 134 sampling units were selected from 134 items. The sample's total value amounts to €67,821.22, representing 4.8% of the total population value. The second table provides details specific to interval selection using monetary unit sampling. It indicates the number of items selected in the 'Top stratum', which includes all items larger than a single interval (for fixed interval selection). In this instance, there were 0 items in the top stratum.

Table 3. *Selection Summary* ▼

No. units	No. items	Selection value	% of population value
134	134	€67,821.22	4.8%

Note. From each of the intervals of size 10471.8, unit 9584 is selected using seed 300.

Table 4. *Information about Monetary Interval Selection*

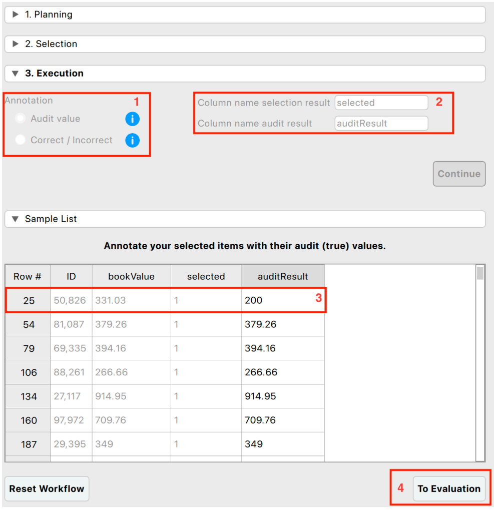
	Items	Value	Selected items	Selected units	Selection value	% of total value
Total	3,500	€1,403,220.82	134	134	€67,821.22	4.8%
Top stratum	0	€0	0	0	€0	0%
Bottom stratum	3,500	€1,403,220.82	134	134	€67,821.22	4.8%

Note. The top stratum consists of all items with a book value larger than a single interval.

4. **Next stage:** Finally, progress to the execution stage by clicking the 'To Execution' button.

1.2.3 Stage 3: Execution

In the execution stage, you must judge the fairness of the 134 sampled items. Once the 'To Execution' button is pressed, the interface from the execution stage opens.



The following four settings are required:

1. **Annotation method:** First, decide how to annotate the selected items. You have two choices:

   - Audit value: Annotate the items with their audit (true) values. This method is recommended (and automatically selected) when the items have a monetary value.
   - Correct / Incorrect: Annotate the items as correct (0) or incorrect (1). This method is recommended (and automatically selected) when the items do not have a monetary value.

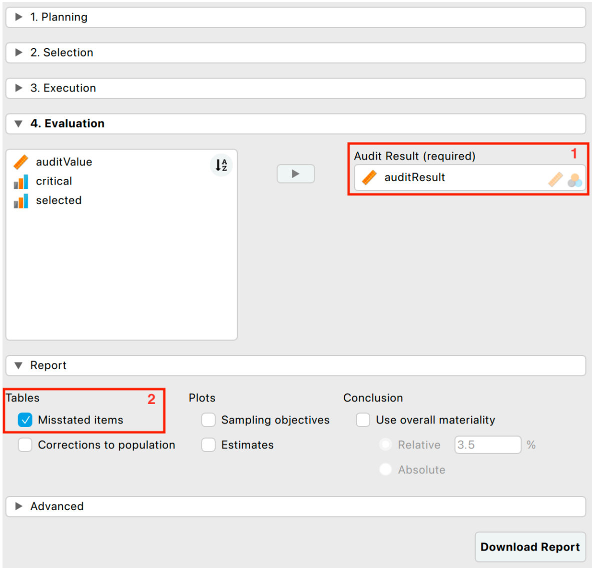
2. **Column names:** Next, specify the names of the two columns that will be added to the dataset. The first column name will indicate the result of the selection, while the second column name will contain the annotation of the items. Click the 'Continue' button to confirm the settings and open the data viewer.

3. **Annotating items:** Then, use the data viewer to annotate the selected items with their book value. For example, in this case, item 50826 (row 25, highlighted in red) had a book value of €333.03 but a true value of €200. The remaining items have correctly reported book values.

4. **Next stage:** Finally, progress to the evaluation stage by clicking the 'To Evaluation' button.

1.2.4 Stage 4: Evaluation

In the evaluation stage, you assess the misstatement in the sample and extrapolate it to the entire population. Once you press the 'To Evaluation' button, the interface for the evaluation stage will open.



The following setting is required:

1. **Annotation variable:** Specify the variable that contains the annotation of the items in the corresponding box.

The following setting is optional:

2. **Additional tables:** It is recommended to request the 'Misstated items' table from the 'Report' section. This table displays the items in the sample where the book value did not match the true value. Additional tables and figures to clarify the output, which will be discussed in Chapter 4, can be requested here as well.

The primary output from the evaluation stage, as shown in the first table below, indicates that the most likely misstatement in the population is estimated to be 0.003, or 0.3%. The 95% upper bound for this estimate is 0.027, or 2.7%. This upper bound is lower than the performance materiality of 3.5%, meaning the auditor has achieved at least 95% assurance that the population misstatement is below the performance materiality.

Table 4. *Evaluation Summary*

	Value
Performance materiality	0.035
Sample size	134
Misstatements	1
Taint	0.396
Most likely misstatement	0.003
95% Upper bound	0.027
Precision	0.025
p-value	0.019

Note. The results are computed using the binomial distribution.

Table 5. *Misstated Items*

ID	Book value	Audit value	Difference	Taint	Counted
50,826	€331.03	€200	€131.03	0.396	x1
Total			€131.03	0.396	

Based on the results of this statistical analysis, the auditor concludes that the population is free of material misstatement.

Chapter 2

Planning

This chapter is about the 'Planning' analysis in the 'Audit Sampling' section of the module.

2.1 Purpose of the analysis

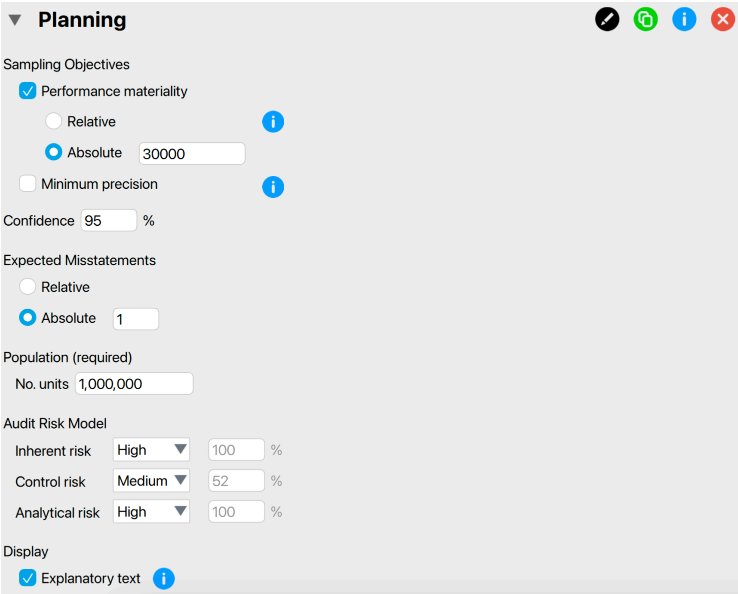
The goal of the planning analysis is to determine the minimum sample size needed to meet the audit's objectives. For example, a common audit objective is to obtain a specific level of confidence that the misstatement in the population is below the tolerable misstatement rate. This rate can be expressed as a monetary amount, known as performance materiality.

2.2 Practical example

Let's consider an example of a planning analysis. Imagine we are auditing a population of 1,000 items with a total value of €1,000,000. In this scenario, we aim to determine the minimum sample size required to conclude, with 95% confidence, that the population does not contain misstatements exceeding the performance materiality of €30,000, which is 3% of the total value. Furthermore, we aim to incorporate a buffer and approve the population if a single misstatement is identified in the sample.

2.2.1 Main settings

To plan the minimum sample size for this audit objective, we open the 'Planning' analysis within the Audit module. The interface for the planning analysis is displayed below.



These are the main settings for the analysis:

- **Sampling objectives: Performance materiality:** In this section, we can input the performance materiality either as a percentage (relative) or as a monetary amount (absolute). If we choose to enter it as a monetary amount, we must also specify the number of units in the population. Here, we enter €30,000 as the absolute performance materiality.

- **Sampling objectives: Minimum precision:** We can choose this setting if we want to identify the misstatement in the population with a specified minimum uncertainty (i.e., the difference between the most likely misstatement and the upper limit for the misstatement). However, since this is not relevant to our audit objective, we leave this box unchecked.

- **Confidence:** Specify the confidence level for your analysis. This level, which complements the significance level, dictates when to reject the null hypothesis and, consequently, the amount of work needed to approve the population. A higher confidence level necessitates more audit evidence to conclude that the population is free of material misstatement. In this example, we use a confidence level of 95%.

- **Expected misstatements:** Specify the number of misstatements tolerated in the sample. This means that if you find the specified number of misstatements in the sample, you can still approve the population. In this example, we tolerate a single misstatement, so we specify this setting to an absolute value of 1.

- **Population: No. units:** Specify the number of sampling units in the population. If you intend to select monetary units, this represents the total value of the population. If you plan to select items, this refers to the number of items in the population. In this case, we intend to use monetary unit sampling and hence we fill in the total population value of €1,000,000 here.

- **Audit risk model:** Indicate the risks of material misstatement using the audit risk model. This model helps reduce the required confidence level for the audit sampling procedure (1 - detection risk) by assessing inherent risk, control risk, and analytical risk. This results in less persuasive audit evidence being required. The model is expressed as:

$$\text{Audit risk} = \text{Inherent risk} \times \text{Control risk} \times \text{Analytical risk} \times \text{Detection risk}$$

. Inherent risk, control risk, and analytical risk are typically evaluated on a 3-point scale: high, medium, and low. These assessments are mapped onto percentages based on professional judgment. The standard percentages used by JASP for Audit are based on those used by the Dutch independent government auditor and are provided in the output table below.

Table 1. *Default Settings Audit Risk Model*

	Inherent risk	Control risk	Analytical risk
High	100 %✓	100 %	100 %✓
Medium	63 %	52 %✓	50 %
Low	40 %	34 %	25 %

✓ = Selected

In this example, let's assume we have conducted internal control testing, enabling us to set the internal control risk to 'Medium', which corresponds to 52%. Consequently, the detection risk can be calculated as $\frac{0.05}{1 \times 0.52 \times 1} = 9.6\%$.

- **Display: Explanatory text:** Finally, select whether to show explanatory text in the output.

2.2.2 Main output

The main table in the output below displays the performance materiality as a proportion, along with the probabilities for the audit risk model. In this scenario, the detection risk is 9.6%. The second-to-last row indicates the tolerable misstatements as a number, showing that only a single misstatement is allowed in the sample. The final row presents the minimum sample size required to meet the sampling objectives, which is 130 units in this case. The note below the table clarifies that this sample size is determined using the binomial distribution (check out the 'Advanced' section for alternative methods).

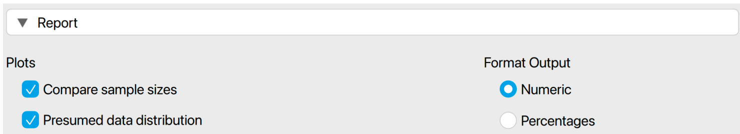
Table 2. *Planning Summary* ▼

	Value
Performance materiality	0.030
Inherent risk	1.000
Control risk	0.520
Analytical risk	1.000
Detection risk	0.096
Tolerable misstatements	1.000
Minimum sample size [a]	130

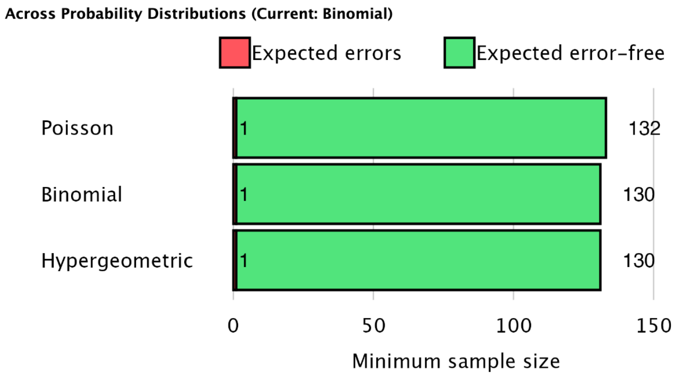
Note. The minimum sample size is based on the binomial distribution (p = 0.03)
[a] Based on this sample size, the selection interval spans 7692.31 units.

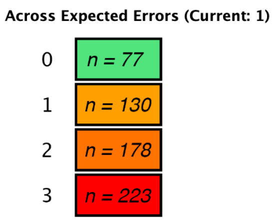
2.2.3 Report

The following settings enable you to expand the report with additional output, such as tables and figures.

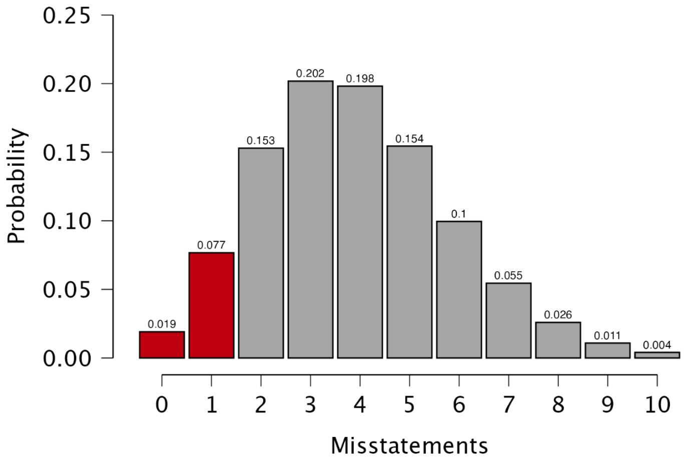


- **Plots: Compare sample sizes:** This setting generates two figures. The first figure illustrates the minimum sample size under three statistical distributions commonly used in statistical auditing: the Poisson distribution, the binomial distribution, and the hypergeometric distribution. The second figure displays the minimum sample size for various tolerable misstatements.





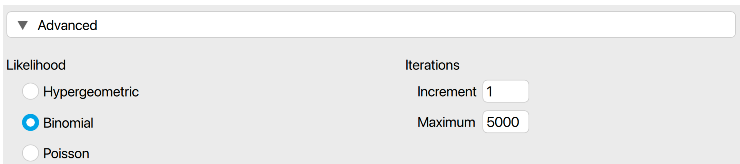
- **Plots: Presumed data distribution:** This figure illustrates the presumed distribution of misstatements in the sample under the hypothesis of material misstatement in the population. The red bar highlights the tolerable misstatements, which together have a probability lower than the detection risk. In this scenario, the figure visualizes that if the population contains material misstatement, there is a 1.9% + 7.7% = 9.6% probability of observing zero or one misstatements in the sample of 130 units. This probability is sufficiently low to reject the hypothesis of tolerable misstatement.



- **Format output:** This setting lets you choose whether certain numbers in the tables are displayed as proportions or percentages.

2.2.4 Advanced

The following advanced settings enable you to customize the statistical computations according to your preferences.



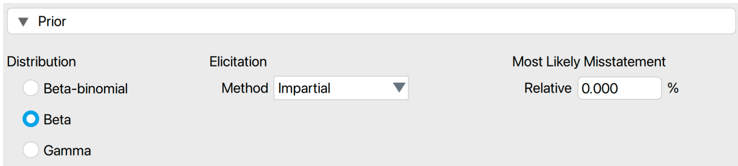
- **Likelihood:** The likelihood is the distribution used to calculate the probabilities of observing a certain number of misstatements. The hypergeometric likelihood (available only if 'No. units' is filled in) assumes a finite population and results in smaller sample sizes for small populations. The binomial and Poisson distributions yield similar sample sizes when the population is large.
- **Iterations: Increment:** Select the step size for the sample sizes to be considered. For example, a value of 5 will include sample sizes of 5, 10, 15, etc., while a value of 20 will include sample sizes of 20, 40, 60. The default value for this setting is 1, which considers all possible sample sizes.
- **Iterations: Maximum:** Choose the maximum sample size to be considered. The analysis will stop if the sample size exceeds this value. The default value is 5000.

2.3 Bayesian planning

The Audit module includes an analysis called 'Bayesian Planning,' which is the Bayesian variant of the planning analysis. This enhanced analysis offers additional options beyond those available in the classical planning analysis, emphasizing the integration of various types of pre-existing audit information.

2.3.1 Prior

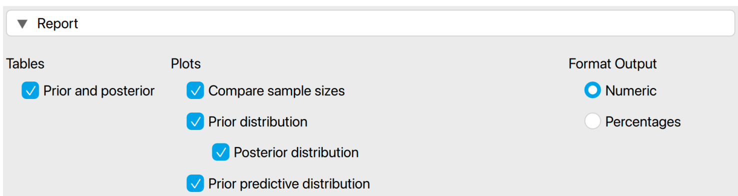
These settings enable you to customize how different types of pre-existing audit information are integrated into the statistical analysis. For more details on the theory behind Bayesian planning and the types of prior distributions, read the corresponding section in Statistical Audit Sampling with R.



- **Distribution:** Select the functional form of the prior distribution. The default is the beta distribution, which is conjugate to the binomial likelihood. Other options include the gamma distribution (conjugate to the Poisson likelihood) and the beta-binomial prior distribution (conjugate to the hypergeometric likelihood).
- **Elicitation: Method:** Choose the type of pre-existing information to be included in the prior distribution. By default, an 'uninformative' prior distribution is used, which incorporates negligible information. Alternatively, the prior distribution can be based on an earlier sample, risk assessments from the Audit Risk Model, or the assumption of impartiality.
- **Most likely misstatement:** Indicate the mode of the prior distribution, which represents the expected most likely misstatement in the population. Keep in mind that this differs from the tolerable deviation rate in the sample. This option is necessary only when the 'Impartial' or 'Risk assessments' elicitation method is chosen.

2.3.2 Report

The following settings enable you to expand the report in the Bayesian planning analysis with additional output, such as tables and figures.



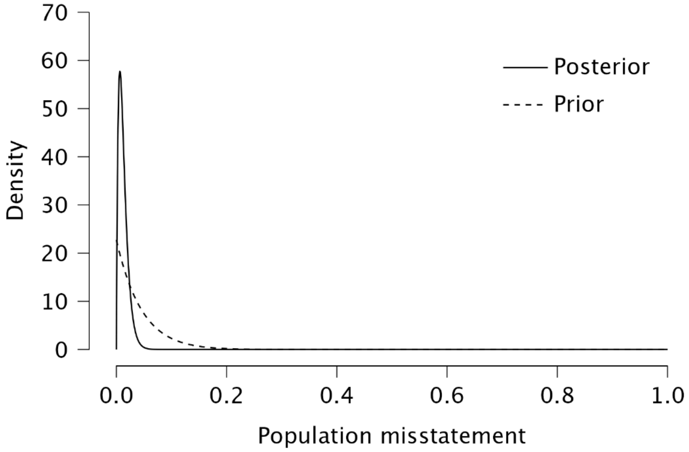
- **Tables: Prior and posterior:** Check this box to generate a table displaying descriptive statistics of the prior distribution and the expected posterior distribution, which represents the posterior distribution if the planned sample is observed.

Table 2. Descriptive Statistics for Prior and Expected Posterior Distribution ▼

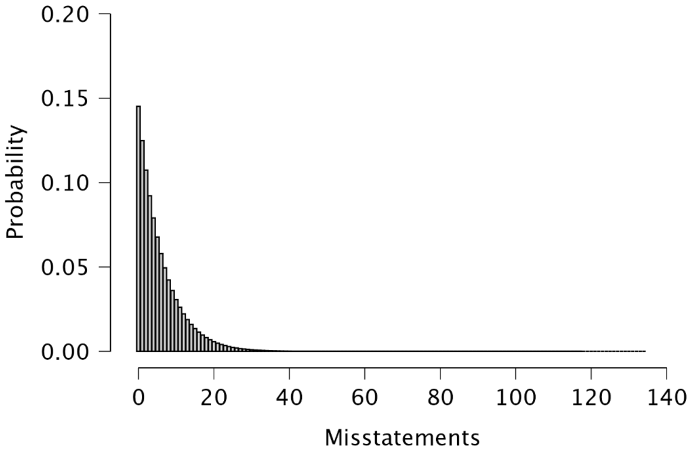
	Prior	Posterior	Shift
Functional form	beta(α = 1, β = 22.757)	beta(α = 2, β = 155.757)	
Support H_-	0.500	0.951	1.901
Support H_+	0.500	0.049	0.099
Ratio $^{H_-}/_{H_+}$	1.000	19.258	19.258
Mean	0.042	0.013	−0.029
Median	0.030	0.011	−0.019
Mode	0.000	0.006	0.006
95% Upper bound	0.123	0.030	−0.093
Precision	0.123	0.023	

Note. H_-: $\theta < 0.03$ vs. H_+: $\theta > 0.03$.

- **Plots: Prior (and posterior) distribution:** Check this box to generate a figure displaying the prior distribution. If the box for the posterior distribution is also checked, the figure will include the posterior distribution after observing the expected sample.



- **Plots: Prior predictive distribution:** Check this box to generate a figure displaying the prior predictive distribution, which illustrates the probabilities of a certain number of misstatements in the sample based on the prior distribution. This can help you verify if the prior distribution is reasonable at the data level.



Chapter 3

Selection

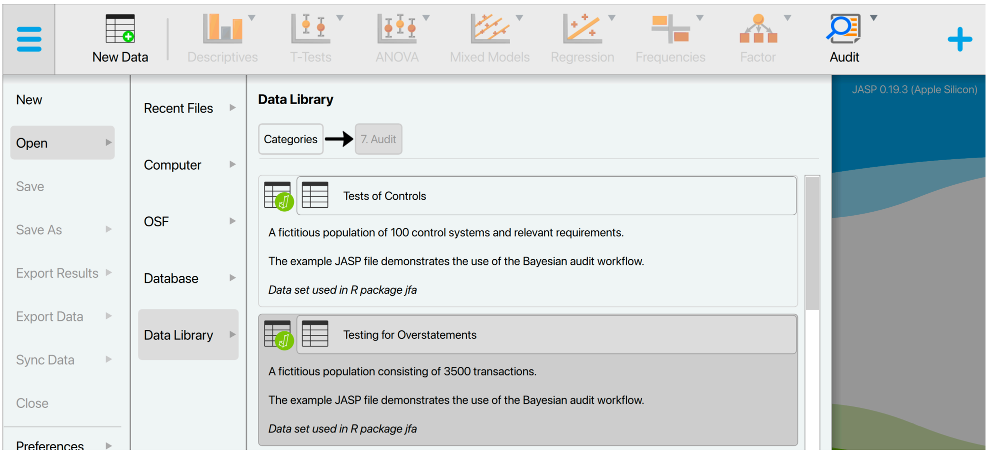
This chapter is about the 'Selection' analysis in the 'Audit Sampling' section of the module.

3.1 Purpose of the analysis

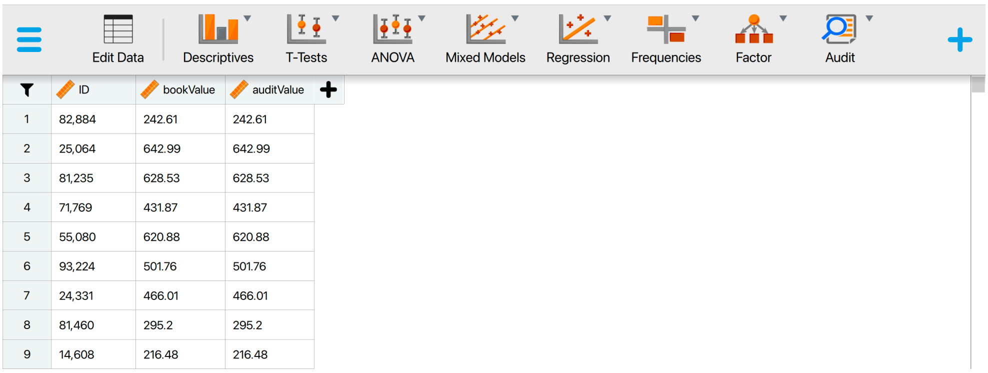
The main goal of the selection analysis is to draw a representative sample of items from the population. These items can then be marked in the population file so they can be easily identified and tested. Particularly in an audit context, special sampling methods, such as monetary unit sampling, are used to ensure the sample has specific characteristics or meets certain criteria, such as always including items with a high book value.

3.2 Practical example

Let's explore an example of a selection analysis. To follow along, open the 'Testing for Overstatements' dataset from the Data Library. Navigate to the top-left menu, click 'Open', then 'Data Library', select '7. Audit', and finally click on the text 'Testing for Overstatements' (not the green JASP-icon button).



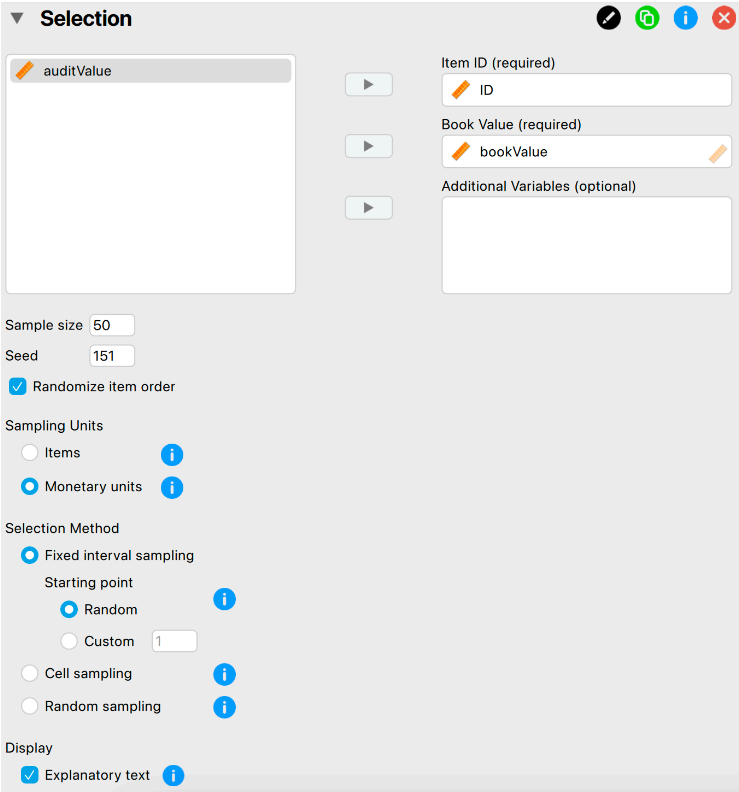
This will open a dataset with 3500 rows and three columns: 'ID', 'bookValue', and 'auditValue'. The 'ID' column represents the identification number of the items in the population. The 'bookValue' column shows the recorded values of the items, while the 'auditValue' column displays the true values. The 'auditValue' column is included for illustrative purposes, as auditors typically know the true values only for the audited sample, not for all items in the population.



	ID	bookValue	auditValue
1	82,884	242.61	242.61
2	25,064	642.99	642.99
3	81,235	628.53	628.53
4	71,769	431.87	431.87
5	55,080	620.88	620.88
6	93,224	501.76	501.76
7	24,331	466.01	466.01
8	81,460	295.2	295.2
9	14,608	216.48	216.48

3.2.1 Main settings

In this example, we aim to select a sample of 50 monetary units from the population using monetary unit sampling with a fixed interval. To draw this sample, we open the 'Selection' analysis within the Audit module. The interface for the selection analysis is displayed below.



These are the main settings for the analysis:

- **Variables:** Start by entering the variable that holds the identification numbers for the items into the 'Item ID' field. Additionally, since we are performing monetary unit sampling, enter the variable 'bookValue' into the 'Book Value' field. Any variables you enter into the 'Additional Variables' field will be displayed along with the selected items in any output tables.
- **Sample size:** Specify the number of sampling units you want to select from the population. In this example, we aim to test a sample of 100 monetary units, so we enter the value 50 in this field.
- **Seed:** A seed in computing is a starting point for generating random numbers. By setting a seed, you ensure that the results of the selection procedure can be reproduced across computers, which is useful for sharing your analysis.
- **Randomize item order:** Choose whether to randomly shuffle the items in the population before starting the selection process. This can help eliminate any patterns that may exist in the dataset. It's generally a good idea to use this setting, so we enable it in this example.
- **Sampling units:** Choose the type of sampling units you want to select. Selecting 'Items' will perform attribute sampling, while 'Monetary units' will perform monetary unit sampling. Since we have access to book values in this example, we select 'Monetary units'.
- **Selection method:** Choose the selection algorithm. Since we want to sample monetary units using a fixed interval, we select 'Fixed interval sampling'.

  - **Fixed interval sampling: Starting point:** This setting determines the starting point in the first interval. To enhance randomness, we set it to 'Random'. Alternatively, we could choose a specific starting point by selecting the 'Custom' option.
- **Display: Explanatory text:** Finally, select whether to show explanatory text in the output.

3.2.2 Main output

The first main table in the output, shown below, displays the number of selected units and the number of items from which these units were chosen. In this example, 50 sampling units have been selected across 50 items. Additionally, the table shows the total value of the items in the sample and the percentage of the population value that these sample items represent. The 50 items have a total value of €27,998.55, which is 2% of the total population value of €1,403,220.82, as calculated by 27,998.55 / 1,403,220.82 = 0.01995. The note under the table shows that the length of a single interval is €28,064.42.

Table 1. *Selection Summary*

No. units	No. items	Selection value	% of population value
50	50	€27,998.55	2 %

Note. From each of the intervals of size 28064.42, unit 13456 is selected using seed 151.

The second main table in the output provides details specific to interval selection methods. It divides the population into two strata: the top stratum, which includes all items with a book value greater than a single interval of €28,064.42 (the top stratum limit would be two interval lenghts for cell sampling), and the bottom stratum, which contains items with a book value smaller than €28,064.42. In this example, there are no items with a book value exceeding €28,064.42, so the top stratum is empty.

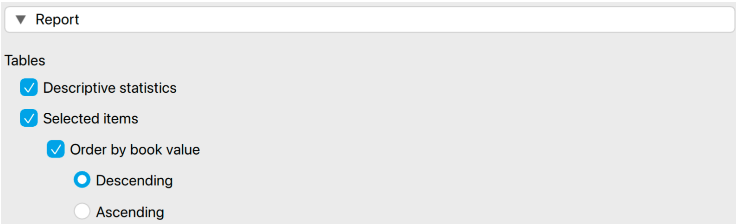
Table 2. *Information about Monetary Interval Selection* ▼

	Items	Value	Selected items	Selected units	Selection value	% of total value
Total	3,500	€1,403,220.82	50	50	€27,998.55	2 %
Top stratum	0	€0	0	0	€0	0 %
Bottom stratum	3,500	€1,403,220.82	50	50	€27,998.55	2 %

Note. The top stratum consists of all items with a book value larger than a single interval.

3.2.3 Report

The following settings enable you to expand the report with additional output, such as tables and figures.



- **Tables: Descriptive statistics:** Checking this box generates a table of descriptive statistics (e.g., mean, median, standard deviation) for the variable in the 'Book Value' field and all variables in the 'Additional Variables' field. This can be used to gain insights into the distribution and characteristics of the sample.

Table 3. *Descriptive Statistics for Sample*

	Items	bookValue
Valid cases	50	50
Mean		559.971
Median		533.135
Std. deviation		274.372
Variance		75,280.232
Range		1,321.310
Minimum		104.480
Maximum		1,425.790

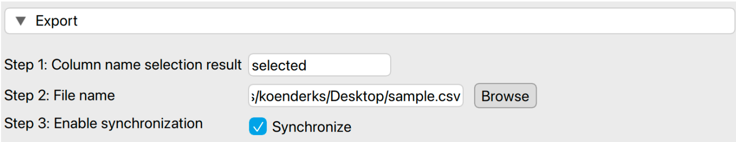
- **Tables: Selected items:** Checking this box generates a table that lists all the selected items in the sample along with their corresponding book values, if this variable is provided.
  - **Order by book value:** This setting enables you to sort the items in the table based on their book value, with the option to arrange them in either ascending or descending order. In this example, we sorted the book values in descending order.

Table 4. *Selected Items*

Row	Selected	ID	bookValue
386	1	7,650	€1,425.79
2,311	1	85,014	€1,189.66
1,371	1	68,134	€1,109.19
306	1	11,569	€1,025.03
705	1	21,900	€919
2,178	1	81,326	€839.21
579	1	5,712	€833.33
85	1	4,437	€758.71
1,009	1	97,578	€754.48
1,171	1	39,619	€747.26
224	1	10,375	€746.81
1,282	1	53,993	€744.55
2,521	1	20,853	€734.19

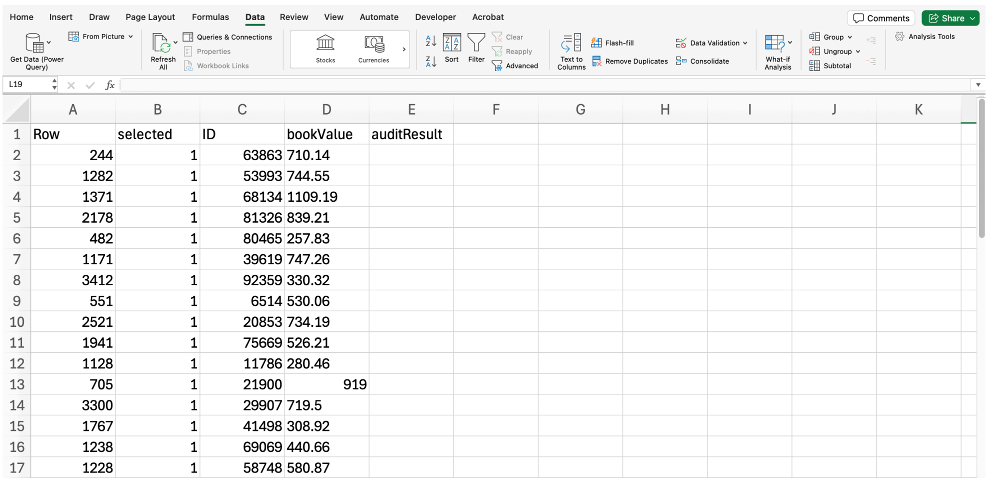
3.2.4 Export

The following settings enable you to isolate and export the selected items to a .csv file.



- **Column name selection result:** Enter the name of the column that will be added to the population file. This column will contain the results of the selection procedure, indicating whether the item is selected for the sample and how many times it is included.
- **File name:** Click 'Browse' to choose a location on your computer where you want to save the sample list.
- **Enable synchronization:** Finally, click on this setting to create the .csv file on your computer. When this setting is enabled, any changes you make to the sample by adjusting settings in the interface will be immediately reflected in the .csv file. If you prefer not to have this automatic update, uncheck this box after enabling it initially.

After applying these settings, you should find the resulting .csv file saved on your computer.



Row	selected	ID	bookValue	auditResult
244	1	63863	710.14	
1282	1	53993	744.55	
1371	1	68134	1109.19	
2178	1	81326	839.21	
482	1	80465	257.83	
1171	1	39619	747.26	
3412	1	92359	330.32	
551	1	6514	530.06	
2521	1	20853	734.19	
1941	1	75669	526.21	
1128	1	11786	280.46	
705	1	21900	919	
3300	1	29907	719.5	
1767	1	41498	308.92	
1238	1	69069	440.66	
1228	1	58748	580.87	

Chapter 4

Evaluation

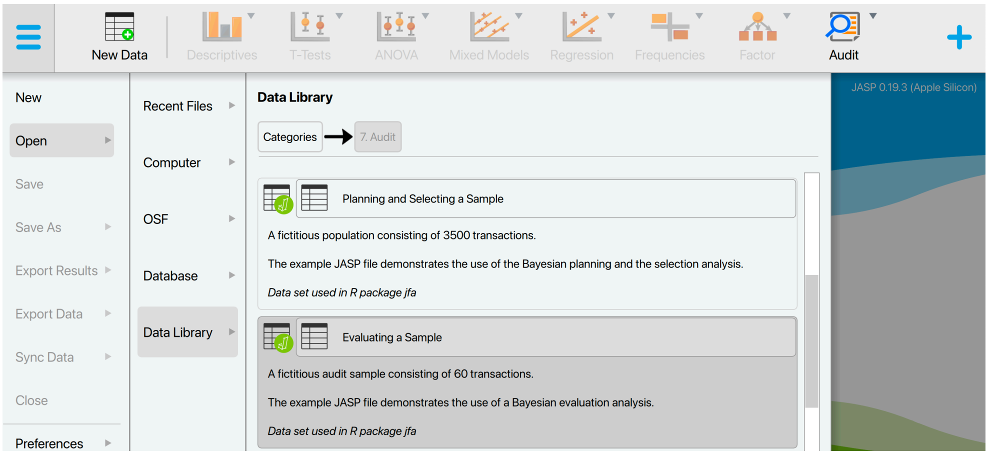
This chapter is about the ‘Evaluation’ analysis in the ‘Audit Sampling’ section of the module.

4.1 Purpose of the analysis

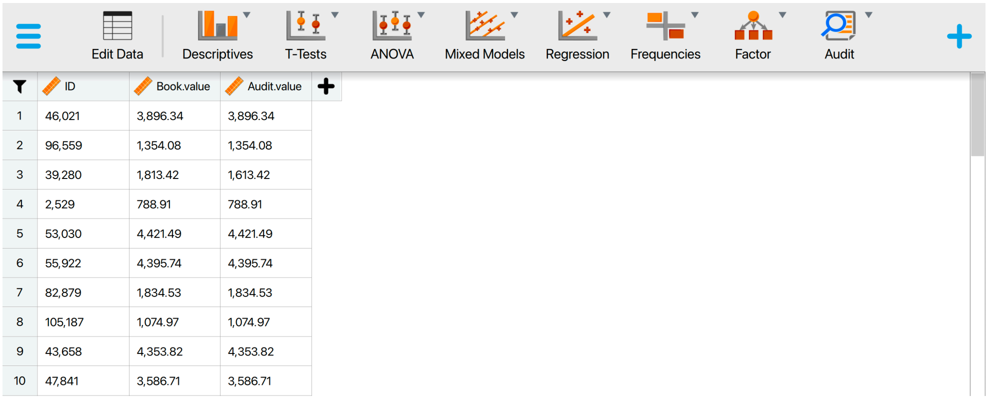
The purpose of the evaluation analysis is to estimate the misstatement in the population from an audited sample and, if necessary, determine if the misstatement is below the performance materiality threshold. This enables auditors to conclude, with a certain level of assurance, whether the population is free of material misstatement.

4.2 Practical example

Let’s explore an example of an evaluation analysis. To follow along, open the ‘Evaluating a Sample’ dataset from the Data Library. Navigate to the top-left menu, click ‘Open’, then ‘Data Library’, select ‘7. Audit’, and finally click on the text ‘Evaluating a Sample’ (not the green JASP-icon button).

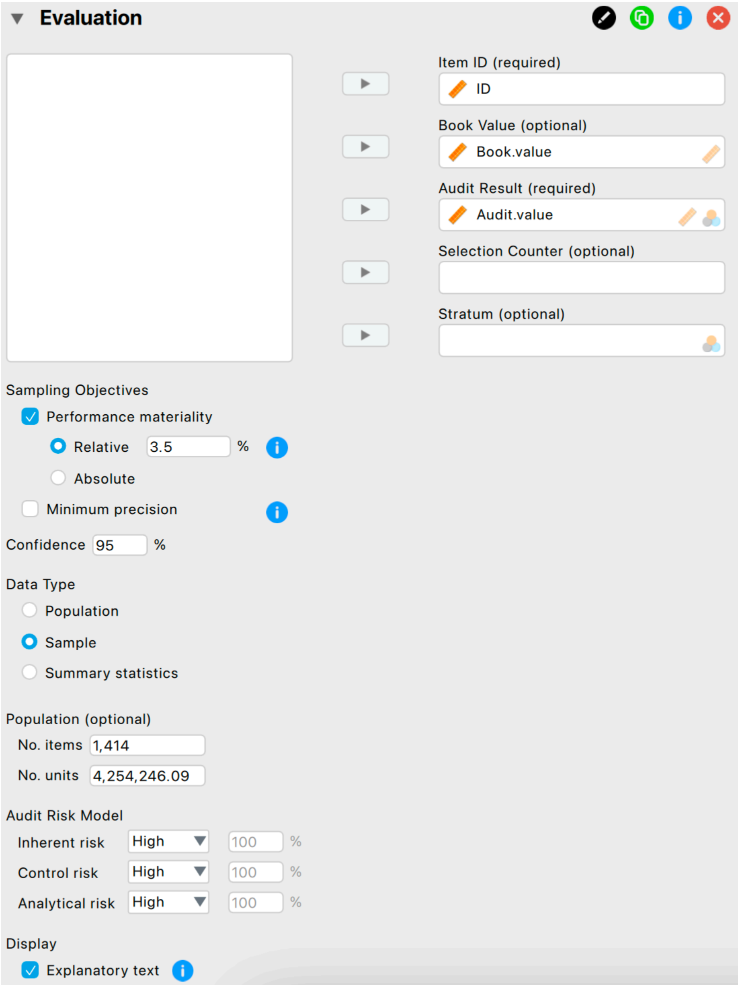


This will open a dataset with 90 rows and three columns: 'ID', 'Book.value', 'Audit.value'. The 'ID' column represents the identification number of the items in the population. The 'Book.value' and 'Audit.value' columns show the recorded and true values of the items, respectively. The sample is drawn from a population of 1,414 items. In this scenario, we seek to determine, with 95% confidence, whether the population contains no misstatements exceeding the performance materiality threshold of 3.5% of the total population value, which amounts to €4,254,246.09.



4.2.1 Main settings

To evaluate this audit sample, we open the 'Evaluation' analysis within the Audit module. The interface for the evaluation analysis is displayed below.



These are the main settings for the analysis:

- **Variables:** Begin by entering the variable that contains the identification numbers for the items into the 'Item ID' field. Then, input the variables that hold the book values and audit (true) values of the items into their respective fields. If your data includes an indicator for which items are part of the sample, drag this to the 'Selection Counter' box. Similarly, if there's an indicator identifying the stratum to which an item belongs, drag this to the 'Stratum' box.
- **Sampling objectives: Performance materiality:** In this section, you can input the performance materiality either as a percentage (relative) or as a monetary amount (absolute). For this example, we enter the performance materiality as a relative value of 3.5%.
- **Sampling objectives: Minimum precision:** This objective requires that the misstatement in the population is estimated with a specified minimum uncertainty (the difference between the most likely misstatement and the upper limit for the misstatement). Since this is not relevant to our audit objective, we leave this box unchecked.

- **Confidence:** Specify the confidence level for your analysis. This level, which complements the significance level, dictates when to reject the null hypothesis and the amount of work needed to approve the population. A higher confidence level requires more audit evidence to conclude that the population is free of material misstatement. In this example, we use a confidence level of 95%.
- **Data type:** Indicate the type of data you are working with. The 'Population' data type assumes that the loaded data file is a full population, with selected items indicated via the 'Selection Counter' variable. This removes the need to manually enter the number of items and units in the population. The 'Sample' data type assumes that the loaded data file is a sample list and requires entering the number of items and units in the population manually. The 'Summary statistics' data type eliminates the need to load a data file and enter variables, assuming the data comes in the form of two values: the sample size and the number of misstatements.
- **Population: No. items:** Enter the number of items in the population. In this example, the population consists of 1,414 items, so we input the value 1,414 here.
- **Population: No. units:** Enter the total value of the population. In this example, the population has a total value of €4,254,246.09, so we input the value 4,254,246.09 here.
- **Audit risk model:** Input the assessed risks of material misstatement into the Audit Risk Model here. For further details on this setting, refer to Chapter 2.
- **Display: Explanatory text:** Finally, select whether to show explanatory text in the output.

4.2.2 Main output

The main table in the output below shows the performance materiality (and minimum precision if enabled), along with the sample size and the number of identified misstatements in the sample. The 'Taint' row displays the sum of the taints, which are the fractional misstatements of the items. Finally, the table presents the estimated most likely misstatement in the population, the 95% upper bound, and the associated precision (the difference between the most likely misstatement and the upper bound).

Table 1. *Evaluation Summary*

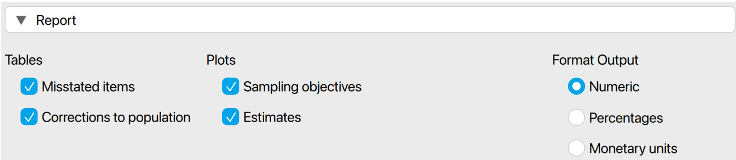
	Value
Performance materiality	0.035
Sample size	90
Misstatements	1
Taint	0.110
Most likely misstatement	0.001
95% Upper bound	0.035
Precision	0.034

Note. The results are computed using the Stringer method.

In this example, the sample consisted of 90 items, with one misstatement identified. This misstatement had a taint of 0.110. Consequently, the most likely misstatement in the population is estimated to be 0.001, or 0.1%. The 95% upper bound for this estimate is 0.035, or 3.5%, and the precision is 3.4%. This upper bound matches the performance materiality of 3.5%, indicating that the auditor has achieved at least 95% assurance that the population is free of material misstatement.

4.2.3 Report

The following settings enable you to expand the report with additional output, such as tables and figures.



- **Tables: Misstated items:** Check this box to generate a table displaying the misstated items in the sample. In this instance, the single misstated item had a book value of €1,813.42 and an audit (true) value of €1,613.42, resulting in a misstatement of €200 and a taint of 0.110.

Table 2. *Misstated Items*

ID	Book value	Audit value	Difference	Taint	Counted
39,280	€1,813.42	€1,613.42	€200	0.110	x1
Total			€200	0.110	

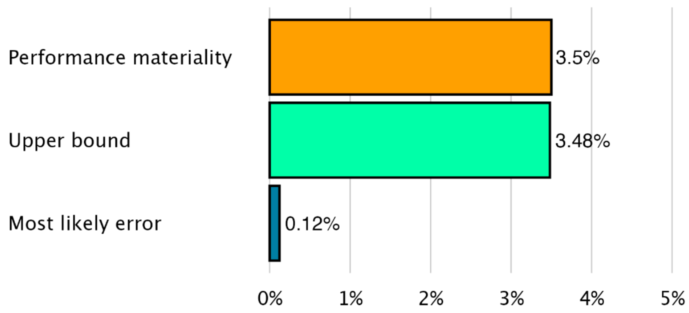
- **Tables: Corrections to population:** Check this box to generate a table indicating the necessary corrections to the population to meet a specific objective. For example, to ensure the population is free of misstatements with 95% confidence, a correction of the upper limit to 3.5% of the population value is required.

Table 4. *Corrections to Population*

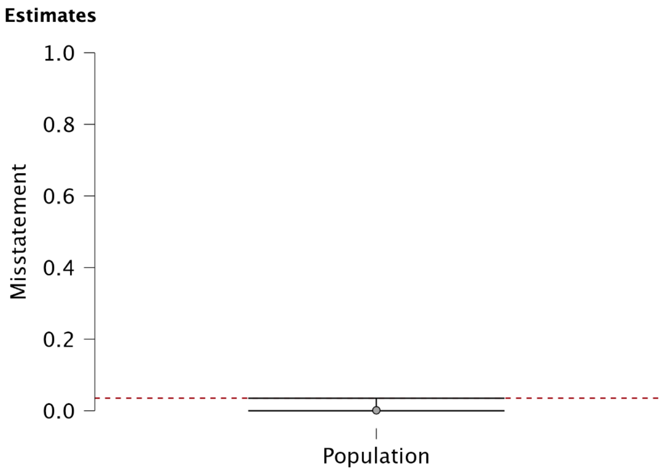
	Correction
No misstatements with 95% confidence	0.035

Note. The correction to achieve no misstatements is the upper bound.

- **Plots: Sampling objectives:** Check this box to generate a figure displaying the sampling objectives, the most likely error, and the upper bound. In this case, the sole sampling objective was the performance materiality. Since the upper bound is lower than the performance materiality, it is highlighted in green.

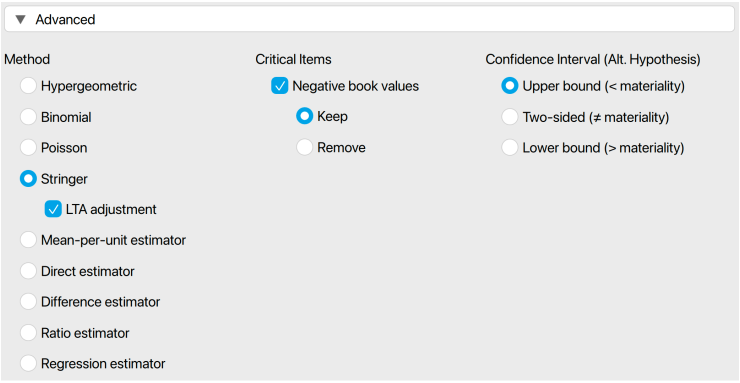


- **Plots: Estimates:** Check this box to generate a figure displaying the most likely misstatement along with the upper and lower limits. This figure is generally useful only if you have entered a variable in the 'Stratum' box, as it provides a quick visual overview of the magnitude of the misstatement in the various strata.



4.2.4 Advanced

The following advanced settings enable you to customize the statistical computations according to your preferences.



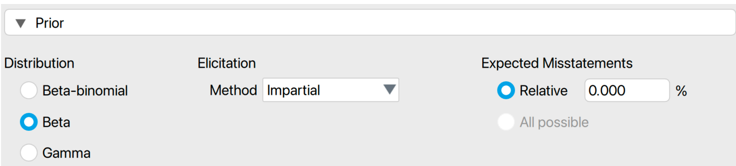
- **Method:** Choose the statistical method to calculate the upper limit of the misstatement. In this example, we selected the Stringer bound as the evaluation method because it considers the taints of the items, making it less conservative than the Poisson, binomial, and hypergeometric distributions. Note that the default setting is 'Binomial', so you must manually select the Stringer bound if you wish to use it.
- **Critical items:** Choose which items are excluded from the statistical evaluation and designated as critical items. Currently, the only option is to mark negative values as critical items, which are kept by default and subtracted from the most likely misstatement and upper bound.
- **Confidence interval (Alt. hypothesis):** Choose whether to calculate a one-sided confidence interval (upper bound or lower bound) or a two-sided confidence interval for the population misstatement. This selection determines the alternative hypothesis being tested.

4.3 Bayesian evaluation

The Audit module includes an analysis called 'Bayesian Evaluation,' which is the Bayesian variant of the evaluation analysis. This enhanced analysis offers additional options beyond those available in the classical evaluation analysis, emphasizing the integration of various types of pre-existing audit information.

4.3.1 Prior

These settings enable you to customize how different types of pre-existing audit information are integrated into the statistical analysis. For more details on the theory behind Bayesian evaluation and the types of prior distributions, read the corresponding section in Statistical Audit Sampling with R.

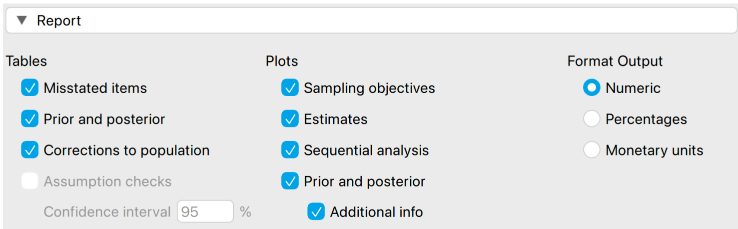


- **Distribution:** Select the functional form of the prior distribution. The default is the beta distribution, which is conjugate to the binomial likelihood. Other options include the gamma distribution (conjugate to the Poisson likelihood) and the beta-binomial prior distribution (conjugate to the hypergeometric likelihood).
- **Elicitation: Method:** Choose the type of pre-existing information to be included in the prior distribution. By default, an 'uninformative' prior distribution is used, which incorporates negligible information. Alternatively, the prior distribution can be based on an earlier sample, risk assessments from the Audit Risk Model, or the assumption of impartiality.
- **Most likely misstatement:** Indicate the mode of the prior distribution, which represents the expected most likely misstatement in the population. Keep in mind that this differs from the tolerable deviation rate in the sample. This

option is necessary only when the 'Impartial' or 'Risk assessments' elicitation method is chosen.

4.3.2 Report

The following settings enable you to expand the report in the Bayesian evaluation analysis with additional output, such as tables and figures.



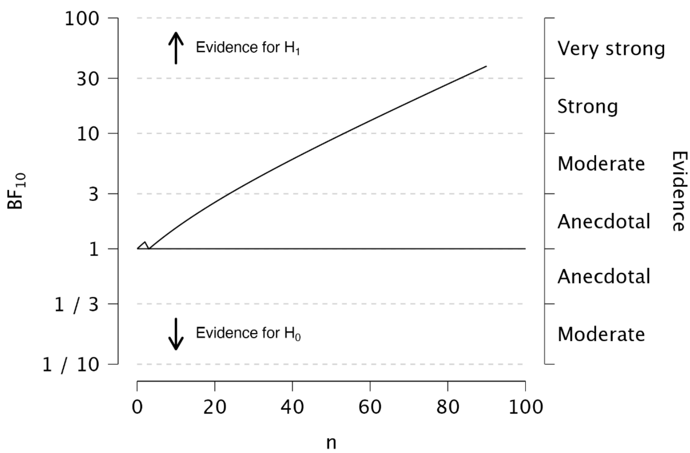
- **Tables: Prior and posterior:** Check this box to generate a table displaying descriptive statistics of the prior distribution and the realized posterior distribution.

Table 3. *Descriptive Statistics for Prior and Posterior Distribution*

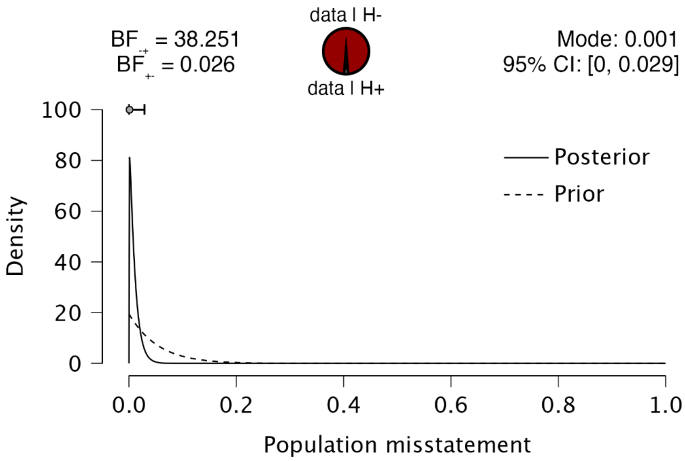
	Prior	Posterior	Shift
Functional form	beta(α = 1, β = 19.456)	beta(α = 1.11, β = 109.345)	
Support H_-	0.500	0.975	1.949
Support H_+	0.500	0.025	0.051
Ratio $^{H_-}/_{H_+}$	1.000	38.251	38.251
Mean	0.049	0.010	−0.039
Median	0.035	0.007	−0.028
Mode	0.000	0.001	0.001
95% Upper bound	0.143	0.029	−0.114
Precision	0.143	0.028	

Note. H_-: $\theta < 0.035$ vs. H_+: $\theta > 0.035$.

- **Plots: Sequential analysis:** Check this box to produce a figure showing the Bayes factor as a function of the sample size.

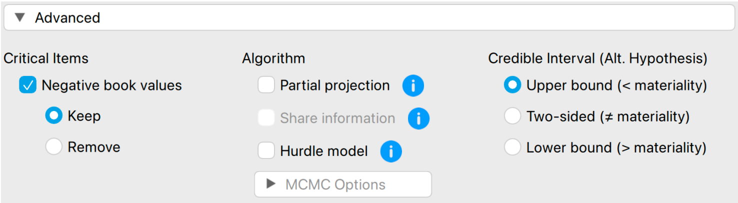


- **Plots: Prior and posterior:** Check this box to generate a figure displaying the prior and posterior distribution. If the box for additional information distribution is also checked, the figure will include information about the posterior distribution and the Bayes factor.



4.3.3 Advanced

The following advanced settings enable you to customize the statistical computations in the Bayesian evaluation analysis according to your preferences.



- **Algorithm: Partial projection:** Check this box to separate the observed misstatement from the unobserved misstatement during evaluation, projecting the uncertainty only onto the unobserved portion of the population.
- **Algorithm: Share information:** Check this box to apply a hierarchical model when analyzing a stratified sample. To enable this option, you must specify a variable in the 'Stratum' box.
- **Algorithm: Hurdle model:** Select this option to apply the hurdle model, an alternative evaluation method. This approach is a Bayesian alternative to the Stringer bound because it properly accounts for the partial misstatements in the items.

Chapter 5

True Value Estimation

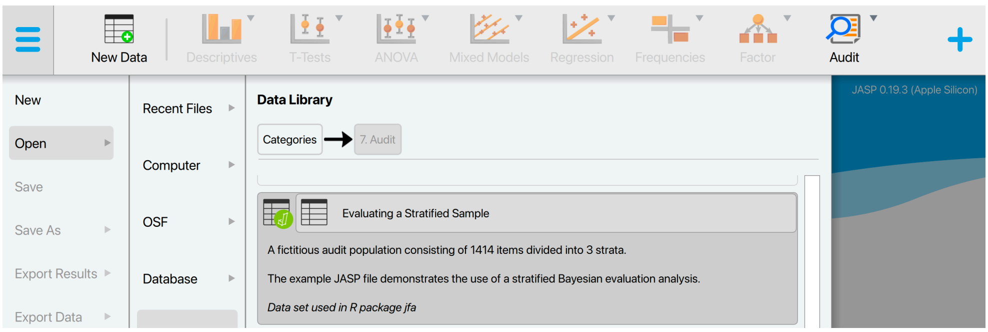
This chapter is about the 'True Value Estimation' analysis in the 'Audit Sampling' section of the module.

5.1 Purpose of the analysis

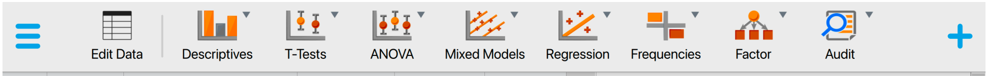
The objective of the true value estimation analysis is to estimate the true value of the population based on a sample. This procedure is commonly used when an audit sample contains many misstatements. In such cases, the auditor cannot conclude that the population is free of material misstatement but aims to estimate its true value. The estimation procedures in this analysis assume a minimum of 30 misstatements in the sample.

5.2 Practical example

Let's explore an example analysis of a true value estimation analysis. To follow along, open the 'Evaluating a Stratified Sample' dataset from the Data Library. Navigate to the top-left menu, click 'Open', then 'Data Library', select '7. Audit', and finally click on the text 'Evaluating a Stratified Sample' (not the green JASP-icon button).



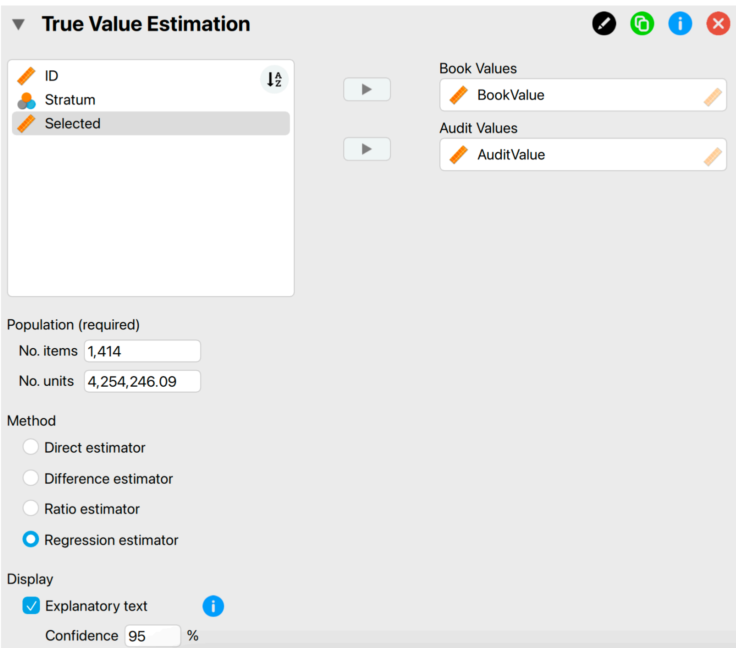
This will open a dataset with 1414 rows and five columns: 'ID', 'Stratum', 'BookValue', 'AuditValue', and 'Selected', which represents a population. The 'ID' column represents the identification number of the items in the population. The 'Stratum' column shows the location from which the item was retrieved. The 'BookValue' and 'AuditValue' columns show the recorded and true values of the items, respectively. Finally, the 'Selected' column shows which items were selected to be included in the sample. The total value of the population (i.e., the sum of the 'BookValue' column) is €4,254,246,09. Note that the audit values of all items that were not selected in the sample (the value of 'Selected' is 0) are empty (NA).



	ID	Stratum	BookValue	AuditValue	Selected
1	1	Distribution center	402.94	362.65	1
2	2	Distribution center	1,954.69	1,856.96	4
3	3	Distribution center	319.99	319.99	1
4	4	Distribution center	195.59	195.59	1
5	5	Distribution center	304.98	NA	0
6	6	Distribution center	599.22	599.22	1
7	7	Distribution center	960.11	NA	0
8	8	Distribution center	96.89	96.89	1

5.2.1 Main settings

In this example, we want to estimate the true value of the population based on the audite sample. To do this, we open the 'True Value Estimation' analysis within the Audit module. The interface for this analysis is displayed below.



These are the main settings for the analysis:

- **Variables:** First, enter the variables indicating the book values and audit (i.e., true) values of the sample items in the corresponding box.
- **Population: No. items:** Enter the number of items in the population. In this example, the population consists of 1,414 items, so we input the value 1,414 here.
- **Population: No. units:** Enter the total value of the population. In this example, the population has a total value of €4,254,246.09, so we input the value 4,254,246.09 here.
- **Method:** Select the statistical method for estimating the true value (Touw & Hoogduin, 2012). The regression estimator is typically the most accurate method, so we choose this method here.
- **Display: Explanatory text:** Finally, select whether to show explanatory text in the output.
- **Display: Confidence:** Set the confidence level used in the explanatory text. In this example, we use a confidence level of 95%.

5.2.2 Main output

The main table in the output below presents the point estimate for the true population value, along with the uncertainty of the estimate and its 95% confidence interval. In this example, the true value of the population is estimated to be €2,512,392.17, with an uncertainty of €551,398.32. Therefore, we can be 95% confident that the true value of the population lies between €1,960,993.85 and €3,063,790.49. The confidence interval does not include the recorded population value of €4,254,246.09.

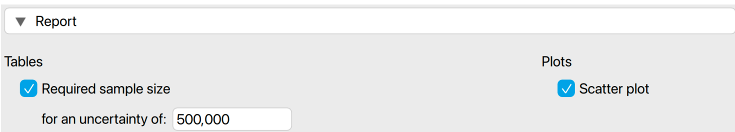
Regression estimator

		95% Confidence interval	
Estimate $\hat{W}$	Uncertainty	Lower	Upper
€2,512,392.17	€551,398.32	€1,960,993.85	€3,063,790.49

Note. Displayed numbers may differ from exact outcomes due to rounding in the calculations.

5.2.3 Report

The following settings enable you to expand the report with additional output, such as tables and figures.

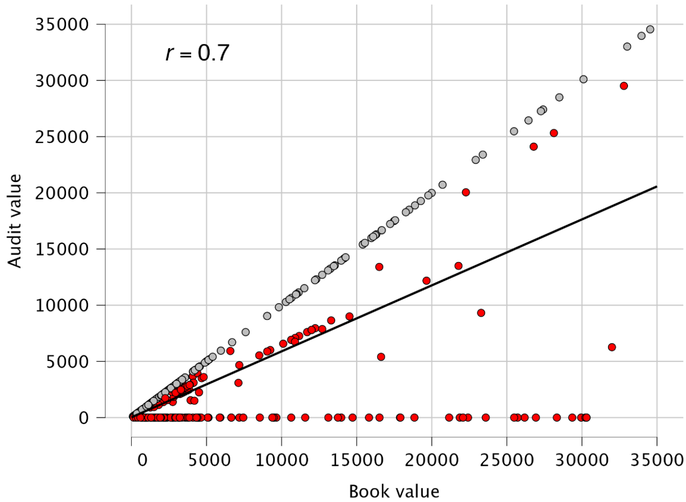


- **Tables: Required sample size:** Checking this box calculates the required sample size to achieve a specific level of uncertainty in the estimate. For example, the current uncertainty in the estimate is €551,398.32. The table below indicates that to reduce this uncertainty to €500,000, a total sample of 459 items is needed. Since we have already sampled 400 items, only 59 additional samples are required.

Required Sample Size

Estimator	Uncertainty	Required *n*	Additional *n*
Regression	€500,000	459	59

- **Plots: Scatter plot:** Checking this box generates a figure that compares the book values of the items in the sample against their true values. Points on the diagonal, shown in gray, represent items where the book value matches the true value. Points in red indicate items where the book value does not match the true value. The black line represents the Pearson correlation between the book values and audit values, which in this case is $r = 0.7$.



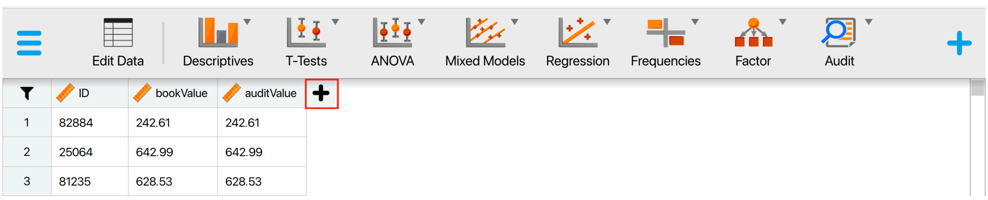
Chapter 6

How-to's

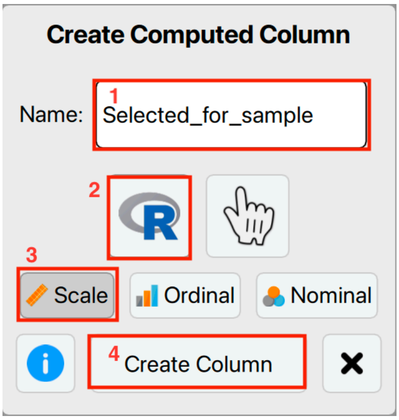
This chapter contains various step-by-step instructions on how to perform best practices using the analyses in the 'Audit Sampling' section of the module.

6.1 Extending a sample selected with a fixed interval

If you have selected a sample using a fixed interval and need to extend it because your population has become larger, follow these steps to manually create a selection counter column. Click the '+' icon next to the dataset and select 'Compute column'.



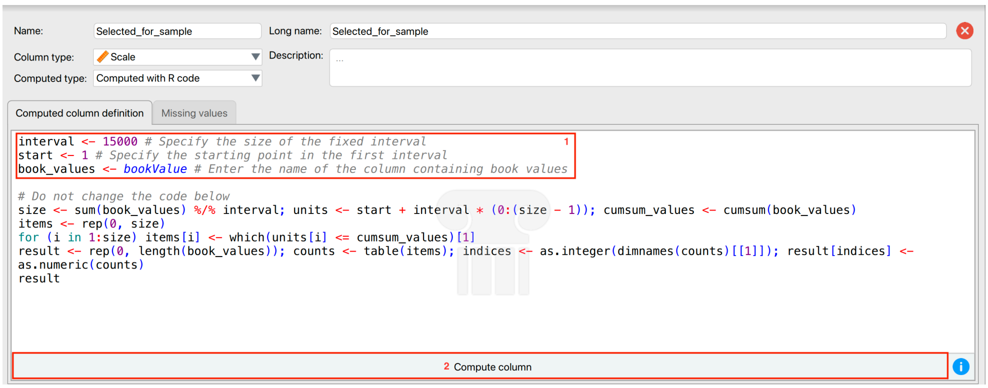
Next, name the new column (for example, 'Selected_for_sample'), then click the R icon, followed by the 'scale' icon, and click on 'Create column'. This will open an editor where you can input the necessary R code.



Copy and paste the following R script into the editor, ensuring that you replace the first three lines with the appropriate values for your dataset. The `interval` represents the fixed interval size, `start` defines the starting point within the first interval, and `book_values` should be replaced with the column name containing your book values.

```
interval <- ... # Specify the size of the fixed interval
start <- ... # Specify the starting point in the first interval
book_values <- ... # Enter the name of the column containing book
 ↪ values

# Do not change the code below
size <- sum(book_values) %/% interval; units <- start + interval *
 ↪ (0:(size - 1)); cumsum_values <- cumsum(book_values)
items <- rep(0, size)
for (i in 1:size) items[i] <- which(units[i] <= cumsum_values)[1]
result <- rep(0, length(book_values)); counts <- table(items); indices
 ↪ <- as.integer(dimnames(counts)[[1]]); result[indices] <-
 ↪ as.numeric(counts)
result
```

For example, if you are selecting with an interval of €15,000 and a starting point of 1, you should enter the following values.



If you later extend your dataset by adding new data, you must re-run the computed column to update the selection while maintaining the same interval (e.g., €15,000).

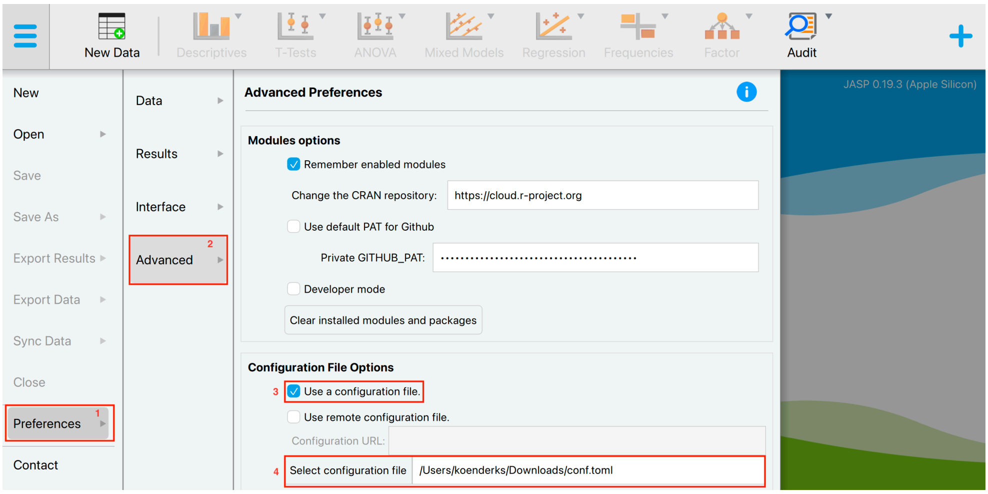
6.2 Configurating and locking default options

You can launch JASP with a configuration file to set restrictions, disable, or hide certain options from users. This is particularly useful for deploying JASP for Audit within your organization, ensuring control over the options available to less statistically experienced users.

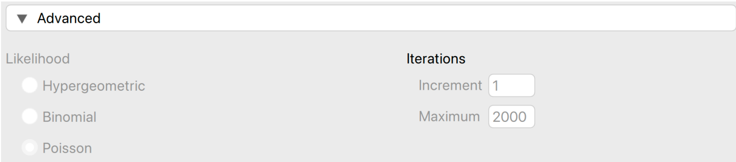
The configuration file type is .toml. The example .toml file below sets the likelihood in the classical planning analysis to the Poisson distribution and locks the option, preventing it from being changed. Additionally, it specifies the increment (the option name in the QML file is 'by') as 1 and the maximum sample size (the option name in the QML file is 'max') as 2,000 instead of the default 5,000.

```
Format = "0.1.0"
JASPVersion = "0.19.3"

EnabledModules = ["jaspAudit"]

[Modules.jaspAudit.Analyses.auditClassicalPlanning.Options]
likelihood = {Value = "poisson", Lock = true}
by = {Value = 1, Lock = true}
max = {Value = 2000, Lock = true}
```

To load this config file into JASP, go to 'Preferences' in the left menu, then 'Advanced', and click 'Use a configuration file' in the 'Configuration File Options' box. Select the location of your config file and restart JASP for the changes to take effect.



When you start the planning analysis, the options under advanced settings will be set to the corresponding defaults and cannot be changed by the user.



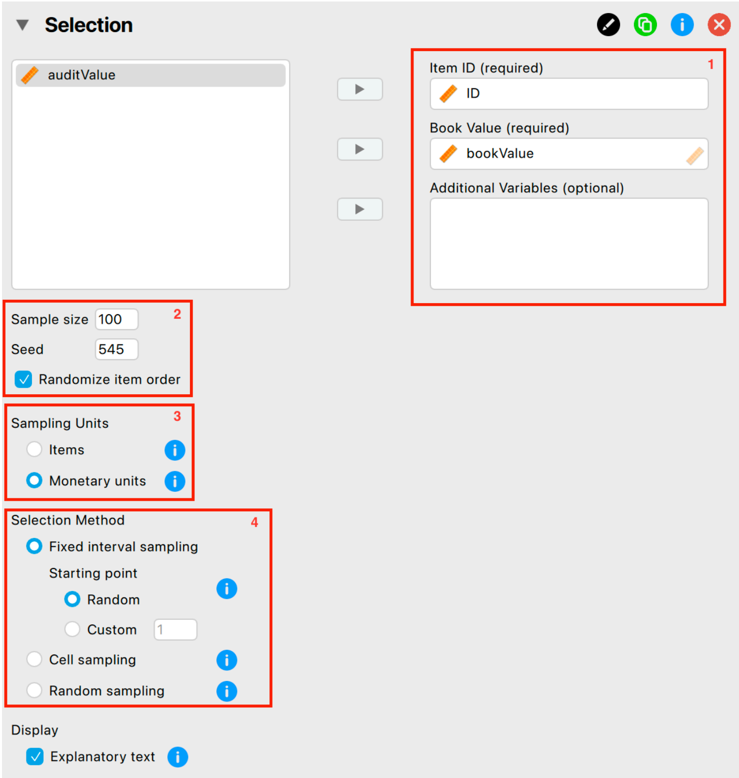
6.3 Bayesian optional stopping

The goal of Bayesian optional stopping is to monitor the evidence as it comes in and decide whether or not to continue data collection based on the evidence that has been observed. Instead of specifying a fixed sample size in advance, the auditor can monitor the evidence as it comes in. This means that you can sequentially expand and evaluate your sample, until you have sufficient evidence to draw a conclusion or until you are out of resources.

Let's explore an example on how to perform Bayesian optional stopping in JASP. To follow along, open the 'Testing for Overstatements' dataset from the Data Library. Navigate to the top-left menu, click 'Open', then 'Data Library', select '7. Audit', and finally click on the text 'Testing for Overstatements' (do not click the green JASP-icon).

6.3.1 Settings

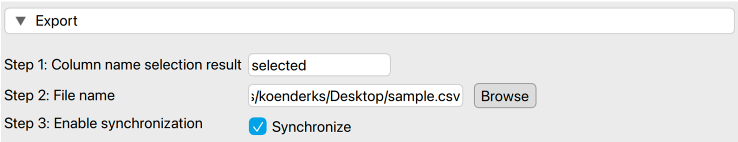
First, you are going to select the maximum sample that you can afford to check based on your resources. Note that it is unlikely that you will need to check the entire sample. To select the sample, use the 'Selection' analysis from the 'Audit Sampling' menu.



The following settings are required:

1. **Indicate the variables:** First, enter the variable indicating the identification numbers of the items in the corresponding box. Optionally, if you have access to the book values of the items, you can enter this variable as well.
2. **Specify maximum sample size:** Specify the number of sampling units you want to select from the population. Make sure to check the box for 'Randomize item order' as well to mitigate any biases that might arise from the original order of the data.
3. **Sampling units:** Specify the sampling units for the selection process. In this example, we are dealing with book values, so we choose 'Monetary Units'.
4. **Selection method:** Choose a selection algorithm. In this example, we choose 'Fixed Interval' with a random starting point.

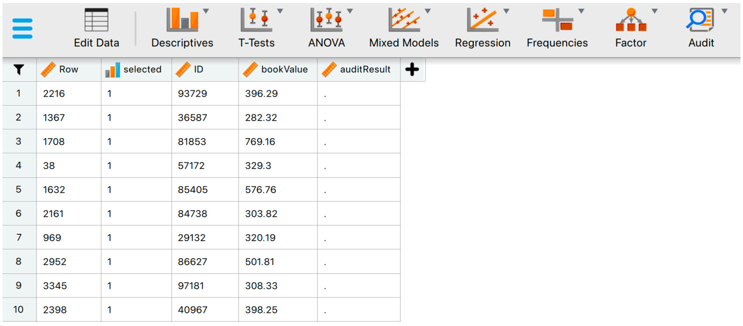
Now, we need to export the sample to a separate file. This way, we can fill in the true audit values for the sample one by one. The following settings enable you to isolate and export the selected items to a .csv file.



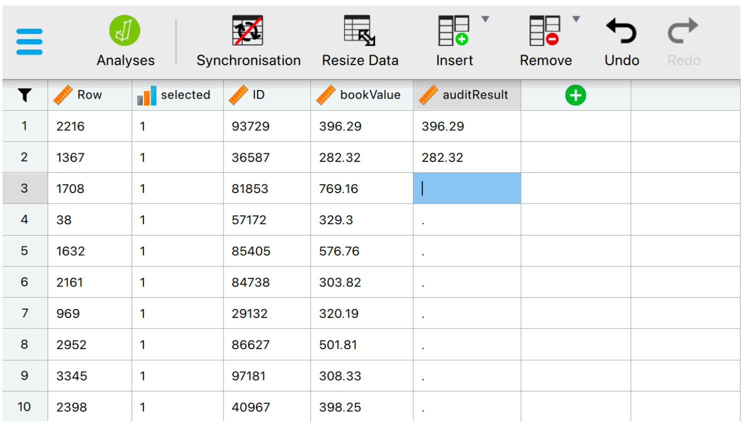
- **Column name selection result:** Enter the name of the column that will be added to the population file. This column will contain the results of the selection procedure, indicating whether the item is selected for the sample and how many times it is included.
- **File name:** Click 'Browse' to choose a location on your computer where you want to save the sample list.
- **Enable synchronization:** Finally, click on this setting to create the .csv file on your computer. When this setting is enabled, any changes you make to the sample by adjusting settings in the interface will be immediately reflected in the .csv file. If you prefer not to have this automatic update, uncheck this box after enabling it initially.

After applying these settings, you should find the resulting .csv file saved on your computer.

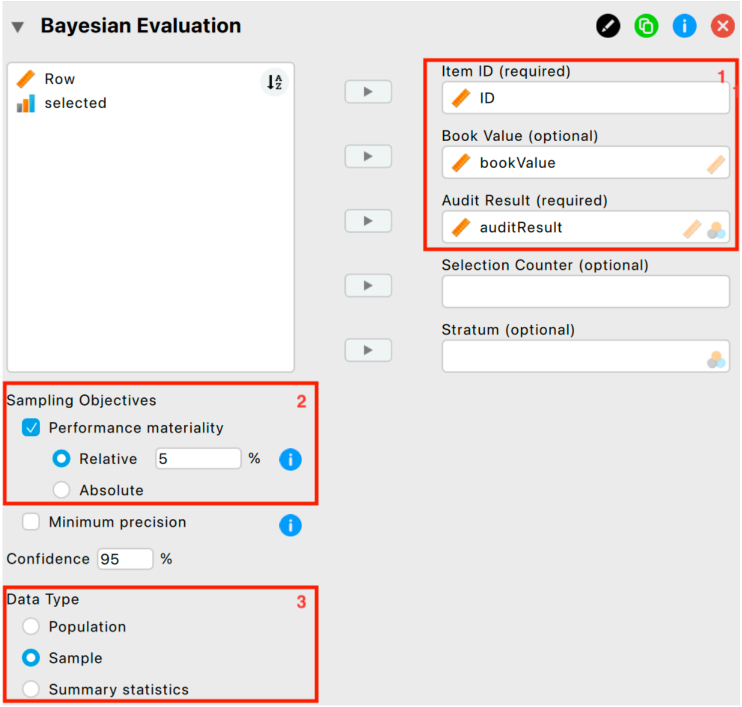
Now, you can start filling in the audit values that correspond to the bookvalues selected in your sample. Load the exported sample file by navigating to 'File' > 'Open' and selecting the .csv file you saved on your computer. This should look as follows:



The data shows all the selected transactions in the sample, including their row numbers, IDs and bookvalues. The column for auditResult is still empty. This column is to be filled by the auditor. To start filling in this column, you can click on 'Edit Data'. This brings you into the following panel.



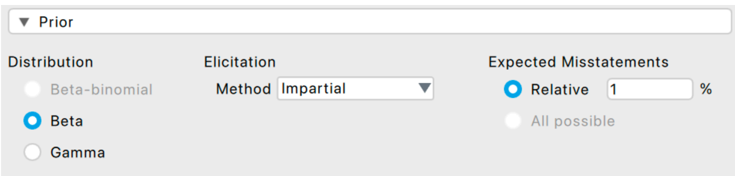
In this panel, you can manually fill in the correct audit value. Bayesian optional stopping allows you to decide, at any moment, whether to continue or stop filling in correct audit values. To make this decision, you can use the 'Bayesian Evaluation' section:



The following settings are required:

1. **Variables**: Begin by entering the variable that contains the identification numbers for the items into the 'Item ID' field. Then, input the variables that hold the book values and audit (true) values of the items into their respective fields.
2. **Performance materiality**: Input the performance materiality either as a percentage (relative) or as a monetary amount (absolute). For this example, we choose a performance materiality of 5%
3. **Data type**: Indicate the type of data you are working with. Here, we are working with sample data.

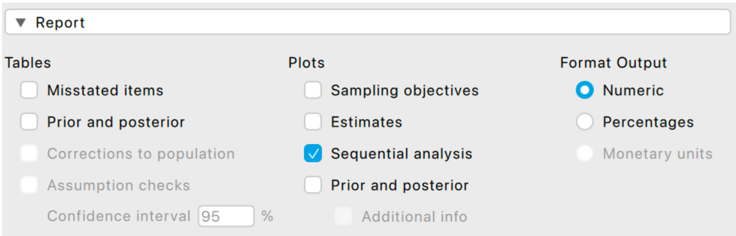
Furthermore, the prior needs to be specified:



- **Impartial prior:** Select the impartial prior as a prior for the evaluation by expanding the 'Prior' section and selecting 'Impartial' under 'Elicitation method'.

- **Expected misstatement rate:** Specify the expected misstatement rate to underlying the impartial prior distribution. For this example, we assume that the expected misstatement rate is 1%.

Finally, we want the output to show the sequantial analysis plot:



Sequential analysis: Expand the 'Report' section and select the checkbox for sequential analysis. This will produce a figure showing the Bayes factor as a function of the sample size in the output.

6.3.2 Output

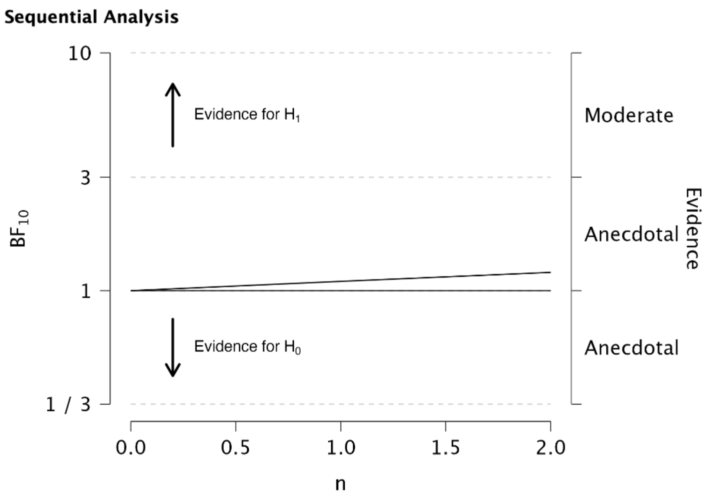
The output shows the results of the audit values you have filled out so far. Hence, we see that the sample size is 2, and no misstatements were found in this sample. The Bayes factor provides an indication of the amount of evidence has been observed. A BF_{-+} of 1.194 indicates that the data are 1.194 times more likely under H_{-} (the misstatement in the population is lower than the performance materiality) than under H_{+} (the misstatement in the population is higher than the performance materiality).

***Table 1.** Evaluation Summary*

	Value
Performance materiality	0.05
Sample size	2
Misstatements	0
Taint	0.000
Most likely misstatement	0.009
95% Upper bound	0.163
Precision	0.154
BF_{-+}	1.194

Note. The results are computed using the beta distribution.

Furthermore, the output shows a figure that depicts the Bayes factor (BF_{-+}) as a function of the sample size (n). This figure illustrates how the evidence for the hypothesis H_{-} versus the hypothesis H_{+} accumulates. Here, we see that over the course of the small sample that we have collected, the Bayes factor has increased slightly.



Based on the output of the Bayesian evaluation, you can decide whether you want to continue or terminate data collection. Mensink et al. (2024) describe an approach to formulate Bayes factor thresholds for terminating data collection. For example, if you use an impartial prior and your allowable risk of incorrect acceptance is 5%, this leads to a Bayes factor threshold of 19. As soon as the Bayes factor exceeds 19, you can approve the population of book values, with no more than 5% risk that this conclusion is incorrect. As this threshold is not yet reached for our example, we need to continue data collection.

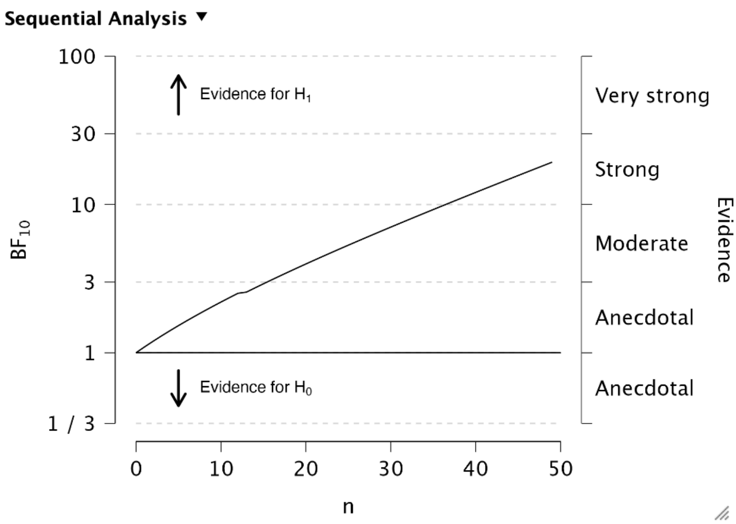
To continue data collection, go back to the 'Edit Data' menu, check more items and fill out more audit values. You can switch between checking items and evaluating your sample as often as you like. When you have sufficient evidence to support your conclusion, you can terminate data collection.

In this example, after checking 45 book values that were selected in the sample and filling in their corresponding audit values (all were correct, except for the 13th invoice which contained a small taint) the output looks as follows.

Table 1. *Evaluation Summary*

	Value
Performance materiality	0.05
Sample size	49
Misstatements	1
Taint	0.02795
Most likely misstatement	0.003
95% Upper bound	0.05
Precision	0.047
BF_{-+}	19.29

Note. The results are computed using the beta distribution.



The Bayes factor is higher than 19, which means we can approve the population of book values with no more than 5% risk that this conclusion is incorrect.

Part II

Data Auditing

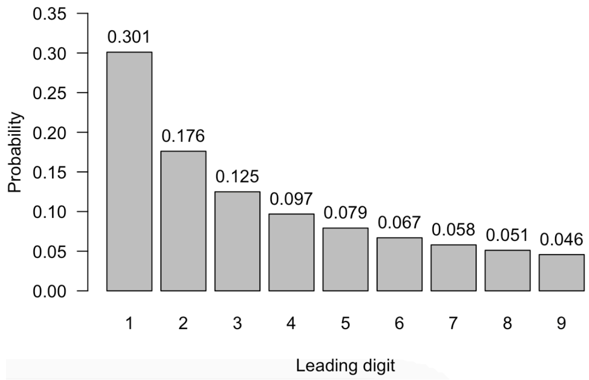
Chapter 7

Benford's Law

This chapter is about the 'Benford's Law' analysis in the 'Data Auditing' section of the module.

7.1 Purpose of the analysis

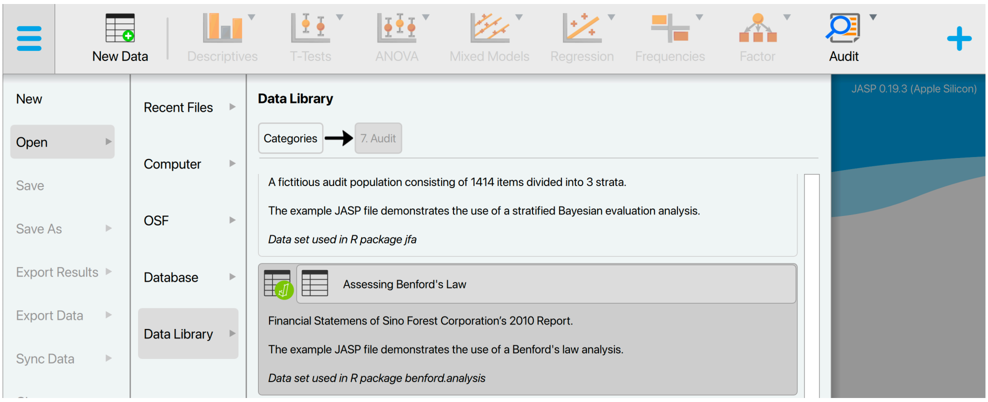
Benford's law states that the distribution of leading digits in a population naturally follows a certain distribution. Specifically, the frequencies of each leading digit d are defined by $p(d) = \log_{10}(1 + \frac{1}{d})$, see the figure below. For instance, the probablity of observing a 1 as a leading digit is 0.301, or 30.1%. This can be tested in a statistical manner. That is, the null hypothesis, H_0, states that the distribution of first digits follows Benford's law, while the alternative hypothesis, H_1, states that it does not.



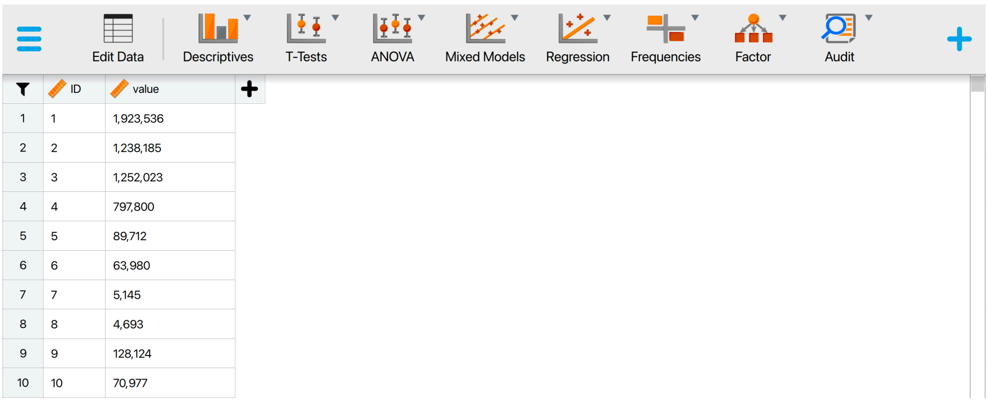
The purpose of the analysis in JASP is to investigate whether the distribution of first, second, or last digits in a set of numbers follows Benford's law. In auditing, this may provide evidence that certain items or transactions in a population might warrant further investigation.

7.2 Practical example

Let's explore an example analysis of Benford's law. To follow along, open the 'Assessing Benford's Law' dataset from the Data Library. Navigate to the top-left menu, click 'Open', then 'Data Library', select '7. Audit', and finally click on the text 'Assessing Benford's Law' (not the green JASP-icon button).

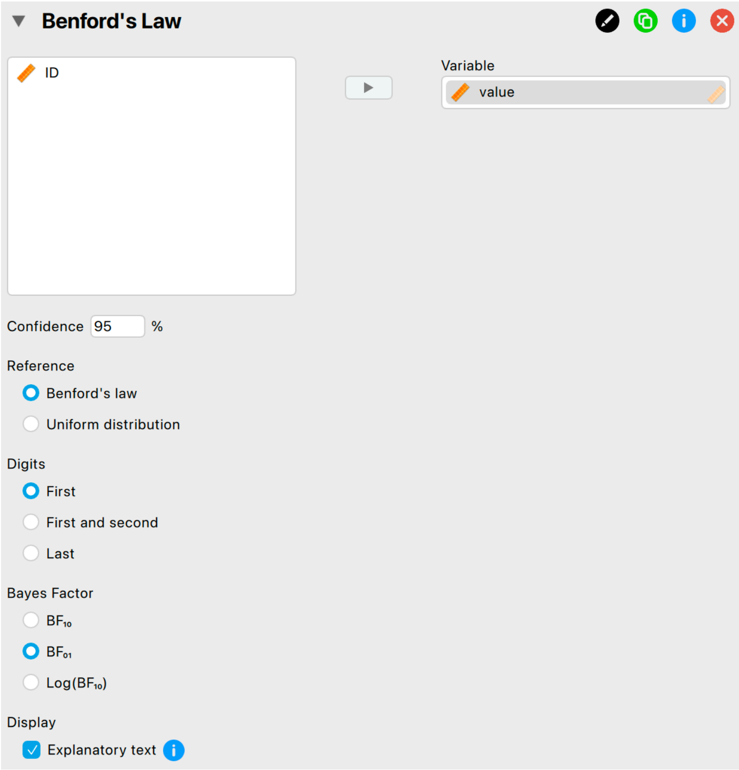


This will open a dataset with 772 rows and two columns: 'ID' and 'value'. The 'ID' column represents the identification number of the items in the population. The 'value' column shows the recorded values of the items.



7.2.1 Main settings

In this example, we will investigate whether the distribution of first digits in the variable 'value', which represents the recorded values of transactions in a financial population, adheres to Benford's law. That is, the null hypothesis, H_0, states that the distribution of first digits follows Benford's law, while the alternative hypothesis, H_1, states that it does not. To test this, we open the 'Benford's Law' analysis from the Audit module. The interface of the Benford's law analysis is shown below.



These are the main settings for the analysis:

- **Variable:** Begin by entering the variable whose digit distribution you wish to test in the designated box. In the example, this is the variable 'value', so we drag this variable to the field on the right.
- **Confidence:** Indicate the confidence level for your analysis. This level, which complements the significance level, determines when to reject the null hypothesis. In the example, we use a confidence level of 95%.
- **Reference:** Select a reference distribution to compare the chosen digits against. By default, this is set to 'Benford's law,' but you can also opt for a uniform distribution. In the example, we select 'Benford's law'.
- **Digits:** Choose which digits to compare against the reference distribution. You can select the first digits (default), the first two digits, or the last digits. Benford's law typically applies to the first or first two digits, while the uniform distribution is usually applied to the last digits. In the example, we choose to test the first digits against Benford's law.
- **Bayes factor:** Select which Bayes factor is displayed in the main output table. 'BF_{10}' represents the Bayes factor in favor of the alternative hypothesis over the null hypothesis, 'BF_{01}' represents the Bayes factor in favor of the null hypothesis over the alternative hypothesis, and '$\text{Log}(BF_{10})$' represents the logarithm of BF_{10}.
- **Display: Explanatory text:** Finally, select whether to show explanatory text in the output.

7.2.2 Main output

The main table in the output, shown below, shows the sample size (n), the mean absolute deviation (MAD), the chi-square value (X^2) and its degrees of freedom (df). The table shows a p-value of 0.478, indicating that H_0 should not be rejected at a significance level of 5%. Furthermore, the table presents the Bayes factor in favor of the null hypothesis, BF_{01}, which is 6.9x10^6. This suggests that the data provide very strong evidence supporting H_0 over H_1.

Table 1. *Omnibus Test – Benford's Law* ▼

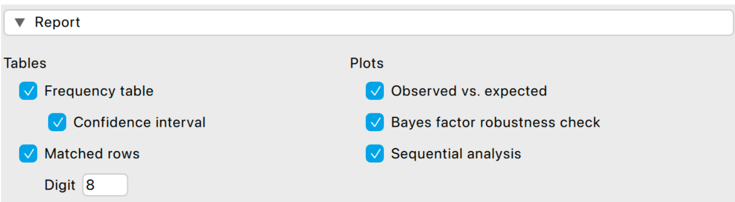
	n	MAD	X^2	df	p	BF_{01} [a]
value	772	0.007	7.652	8	0.468	$6.900 \times 10^{+6}$

Note. The null hypothesis specifies that the first digits (1 – 9) in the data set are distributed according to Benford's law.
[a] The Bayes factor is computed using a Dirichlet($\alpha_1,...,\alpha_9$) prior with α = 1.

Note that non-conformity to Benford's law does not necessarily indicate fraud. A Benford's law analysis should therefore only be used to acquire insight into whether a population might need further investigation.

7.2.3 Report

The following settings enable you to expand the report with additional output, such as tables and figures.



- **Tables: Frequency table:** Check this box to display a table of the observed and expected frequencies of the digits. Clicking the 'Confidence interval' option shows confidence intervals for the observed relative frequencies in the table.

  The frequency table displays the observed count for each leading digit in the second column. Adjacent to this, it shows the expected relative frequency under Benford's law alongside the observed relative frequency in the data. Additionally, p-values and Bayes factors are provided to test whether the observed relative frequencies differ from the expected ones. In this case, only the digit 8 has a p-value smaller than 0.05, indicating a significant deviation from the expected relative frequency under Benford's law.

Table 2. Frequency Table

Leading digit	Count	Benford's law	Relative frequency	95% Confidence Interval Lower	95% Confidence Interval Upper	p[a]	BF_{01}[b]
1	231	0.301	0.299	0.267	0.333	0.937	24.083
2	124	0.176	0.161	0.135	0.188	0.277	15.737
3	97	0.125	0.126	0.103	0.151	0.957	33.397
4	70	0.097	0.091	0.071	0.113	0.626	32.411
5	64	0.079	0.083	0.064	0.105	0.689	37.398
6	54	0.067	0.070	0.053	0.090	0.719	41.126
7	40	0.058	0.052	0.037	0.070	0.537	37.795
8	54	0.051	0.070	0.053	0.090	0.022	3.449
9	38	0.046	0.049	0.035	0.067	0.606	46.150

Note. The null hypothesis specifies that the relative frequency of a digit is equal to its expected relative frequency under Benford's law.
[a] Confidence intervals and *p*-values are based on independent binomial distributions.
[b] Bayes factors are computed using a beta(1, 1) prior.

- **Tables: Matched rows:** Check this box to display a table showing the rows that have a certain number as their leading/last digit(s).

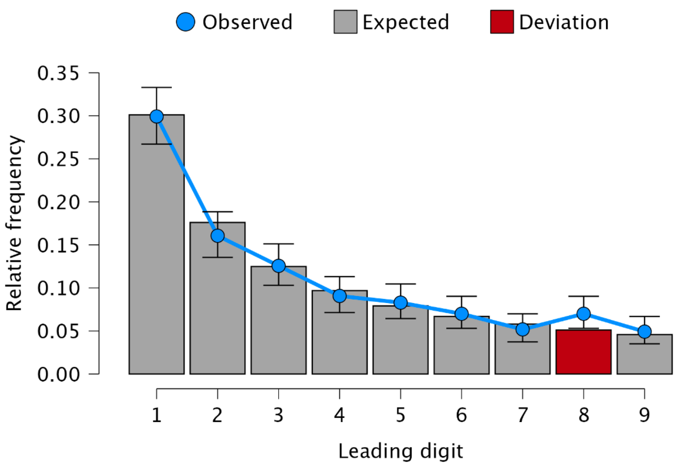
In the example, we request a table of rows that match the digit 8. The first column displays the row number where the digit is found, and the second column shows the matched value. Using this table, you can identify the transactions that may warrant further investigation.

Table 3. Rows Matched to Leading Digit 8 ▼

Row	Value
5	89,712.000
21	8,179.000
33	8,498.000
34	8,756.000
59	87,670.000
84	8,179.000
103	826.000
125	8,555.000
138	82,447.000
192	845.000
199	8,000.000
210	837.000
215	8,000.000
218	8,340.000
225	826.000

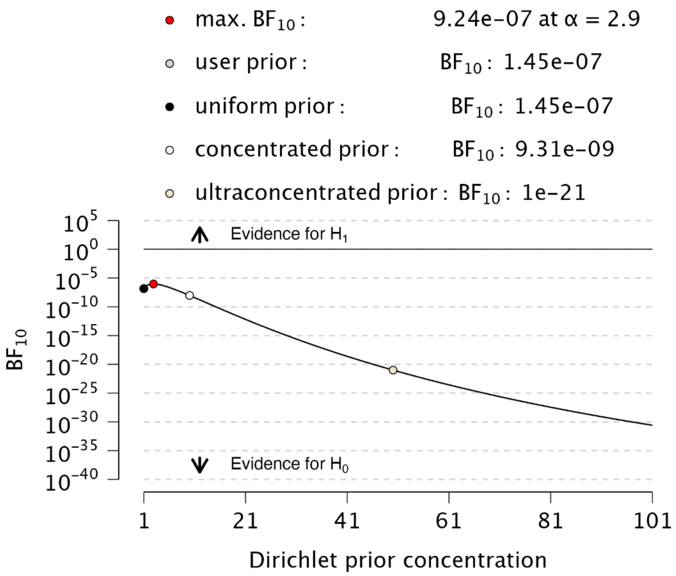
- **Plots: Observed vs. expected:** Check this box to display a figure that illustrates the observed frequencies compared to the expected frequencies.

The figure in the output visualizes the observed relative frequencies compared to the expected ones, with the digit 8 highlighted in red. From this figure, it is immediately clear that the transactions starting with the digit 8 may warrant further inspection.



- **Plots: Bayes factor robustness check:** Check this box to display a figure that shows the Bayes factor under different specifications of the prior concentration parameter.

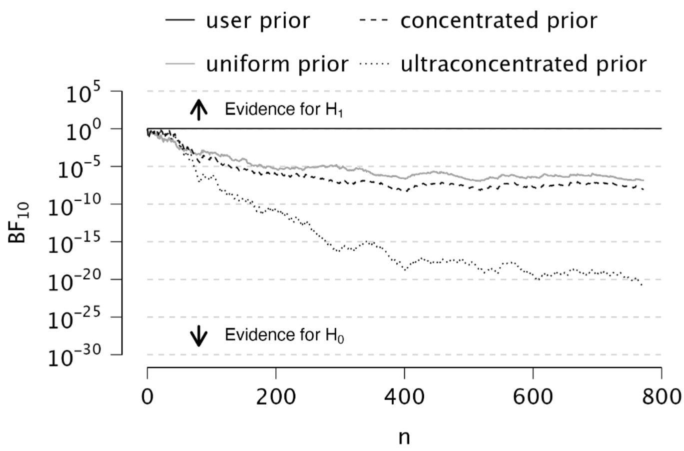
  The figure below is referred to as a robustness check. If the Bayes factor supports a particular hypothesis across all reasonable values of the prior concentration parameter, the result is considered robust regarding the choice of prior distribution. In this instance, the figure demonstrates that the Bayes factor consistently provides evidence in favor of the null hypothesis, regardless of the prior concentration parameter values.



- **Plots: Sequential analysis:** Select this box to display a figure illustrating the Bayes factor as a function of sample size, across various prior specifications.

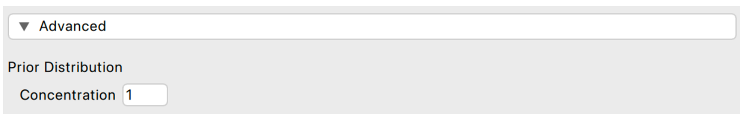
  In the example analysis, the sequential analysis plot demonstrates that the

Bayes factor provides increasing evidence in favor of H_0 as the sample size grows. Additionally, this evidence is more pronounced when using a more concentrated prior distribution.



7.2.4 Advanced

The following advanced settings enable you to customize the statistical computations according to your preferences.



- **Prior distribution: Concentration:** Specify the concentration parameter for the Dirichlet prior distribution. Adjusting this value will alter the Bayes factor in the main output table. A larger concentration parameter indicates a more concentrated prior distribution, suggesting that the population proportions are more similar. When testing against the uniform distribution, this implies a stronger belief in H_0. Conversely, when testing against Benford's law, it indicates a stronger belief in H_1.

Chapter 8

Repeated Values

This chapter is about the ‘Repeated Values’ analysis in the ‘Data Auditing’ section of the module.

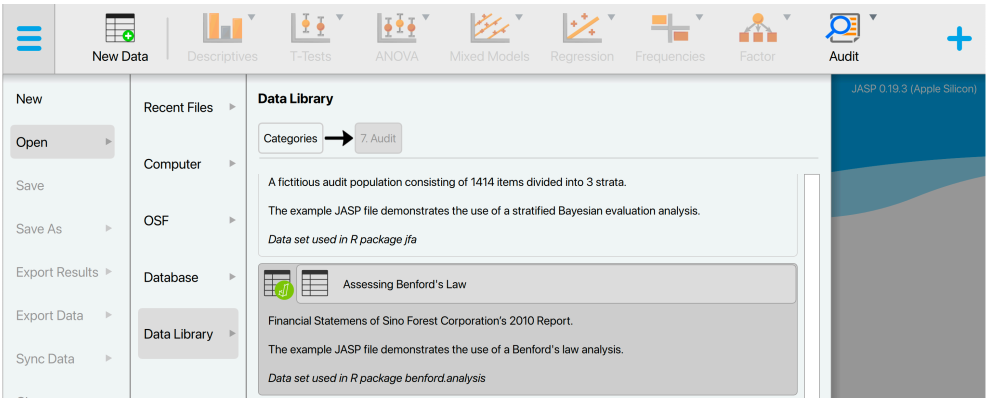
8.1 Purpose of the analysis

The repeated values analysis examines the frequency of value repetitions within a dataset (referred to as “number-bunching”) to statistically determine if the data were likely tampered with (Simohnsohn, 2019). This can be tested statistically. The null hypothesis H_0 posits that the data do not contain an unexpected amount of repeated values, while the alternative hypothesis H_1 suggests they do. Unlike Benford’s law, this approach analyzes the entire number at once, not just the first or last digit.

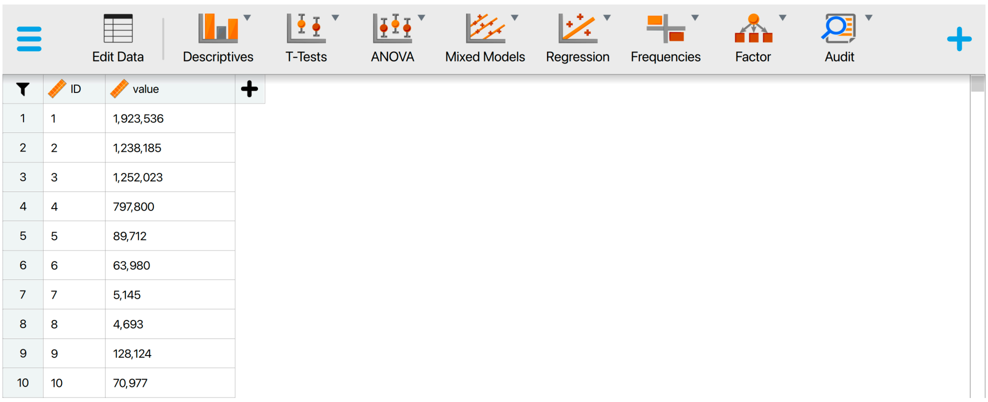
The purpose of the analysis in JASP is to identify whether the data exhibit excessive repeated values. In auditing, this could indicate that certain items or transactions within a population may require further investigation.

8.2 Practical example

Let’s explore an example analysis of repeated values. To follow along, open the ‘Assessing Benford’s Law’ dataset from the Data Library. Navigate to the top-left menu, click ‘Open’, then ‘Data Library’, select ‘7. Audit’, and finally click on the text ‘Assessing Benford’s Law’ (not the green JASP-icon button).

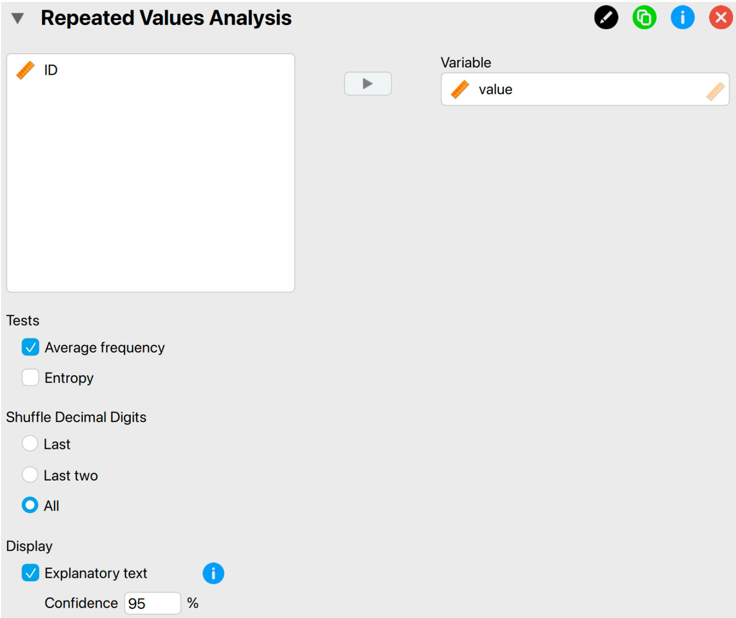


This will open a dataset with 772 rows and two columns: 'ID' and 'value'. The 'ID' column represents the identification number of the items in the population. The 'value' column shows the recorded values of the items.



8.2.1 Main settings

In this example, we will test whether the values in the 'value' column show an excessive amount of repeated values. To test this, we open the 'Repeated Values' analysis from the Audit module. The interface of the repeated values analysis is shown below.



These are the main settings for the analysis:

- **Variable:** Start by entering the variable whose digits should be analyzed for repeated values in the designated box. In this example, the variable is 'value', so we drag this variable to the field on the right.
- **Tests: Average frequency** Check this box to test if the average frequency of values differs from what is expected. In this example, we only examine the average frequency, so we check this box.
- **Tests: Entropy** Check this box to test if the entropy of values differs from what is expected. In this example, we do not check this box as we are only looking at the average frequency.
- **Shuffle decimal digits:** This setting determines which decimal digits are shuffled in the analysis. In this example, we select all decimal digits to be shuffled.
- **Display: Explanatory text:** Finally, select whether to show explanatory text in the output.
- **Display: Confidence:** Set the confidence level used in the explanatory text. In this example, we use a confidence level of 95%.

8.2.2 Main output

The main table in the output below displays the sample size (n), the average frequency of 1.324, and the p-value for the test. This indicates that each unique value in the data occurs, on average, 1.324 times. The p-value is smaller than the significance level of 5%, leading us to reject the null hypothesis and conclude that there is an excessive amount of repeated values in the data.

Table 1. *Repeated Values Test*

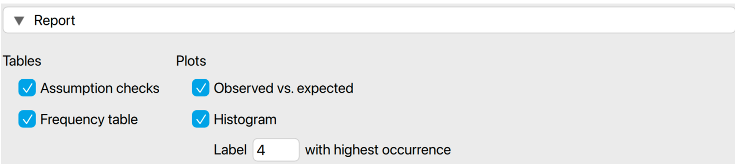
		Frequency	
	n	Average	p
value	772	1.324	0.024

Note. The displayed *p*-value is one-sided and is computed on the basis of 500 samples.

Note that rejecting the null hypothesis does not necessarily indicate fraud. A repeated values analysis should therefore only be used to acquire insight into whether a population might need further investigation.

8.2.3 Report

The following settings enable you to expand the report with additional output, such as tables and figures.



- **Tables: Assumption checks:** To quantify expectations, this test assumes that the integer portions of the numbers are not correlated with their decimal portions. The table below tests this assumption and confirms it holds, as indicated by the non-significant p-value of 0.461.

Table 2. *Assumption Checks*

			n	r	t	df	p
Integer values	–	Decimal values	772	−0.027	−0.738	770	0.461

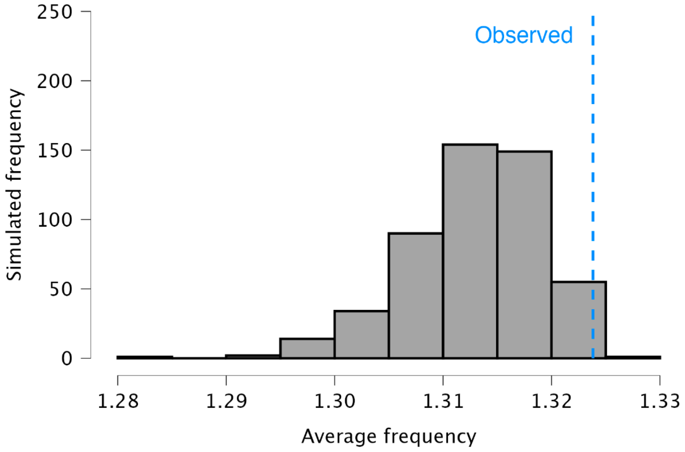
Note. The displayed *p*-value is for a two-sided test against H_0: $r = 0$.

- **Tables: Frequency table:** The frequency table displays the occurrence of each unique value in the data, ordered from highest to lowest frequency. For example, it shows that the value 87,670 appeared five times, representing 0.6% of the total values.

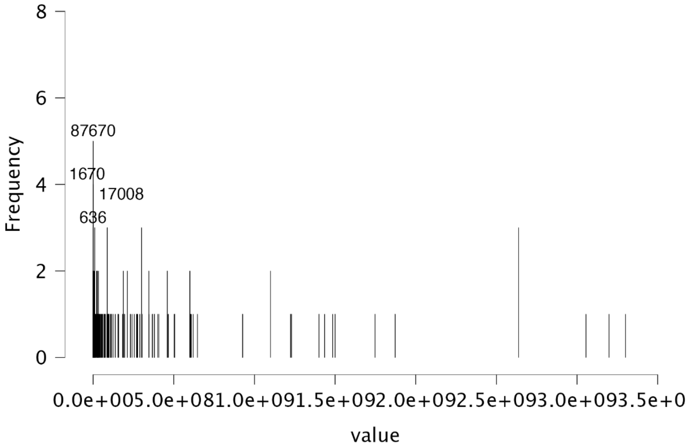
Table 3. *Frequency Table* ▼

Value	Count	Percentage
87,670	5	0.6 %
1,670	4	0.5 %
17,008	4	0.5 %
636	3	0.4 %
1,403	3	0.4 %
8,756	3	0.4 %
50,000	3	0.4 %
150,000	3	0.4 %
1,473,353	3	0.4 %

- **Plots: Observed vs. expected:** Check this box to generate a histogram of the expected distribution of the average frequency or entropy, assuming the decimal portions of the numbers are random and not associated with their integer portions. The observed average frequency will be indicated in the figure.

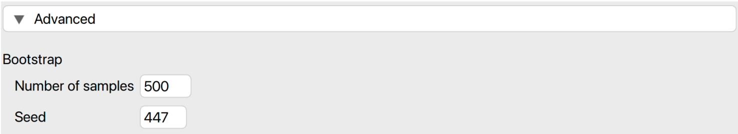


- **Plots: Histogram:** The histogram visualizes the frequency table using bars to represent the values. Similar to the frequency table, the histogram indicates that the most frequently occurring value is 87,670, which appears five times.



8.2.4 Advanced

The following advanced settings enable you to customize the statistical computations according to your preferences.



- **Bootstrap: Number of samples:** This setting specifies the number of bootstratp samples used to compute the expected distribution of the average frequency or the entropy. The default value for this setting is 500.

- **Bootstrap: Seed:** A seed in computing is a starting point for generating random numbers. By setting a seed, you ensure that the results of the analysis can be reproduced across computers, which is useful for sharing your analysis.

Part III

Algorithm Auditing

Chapter 9

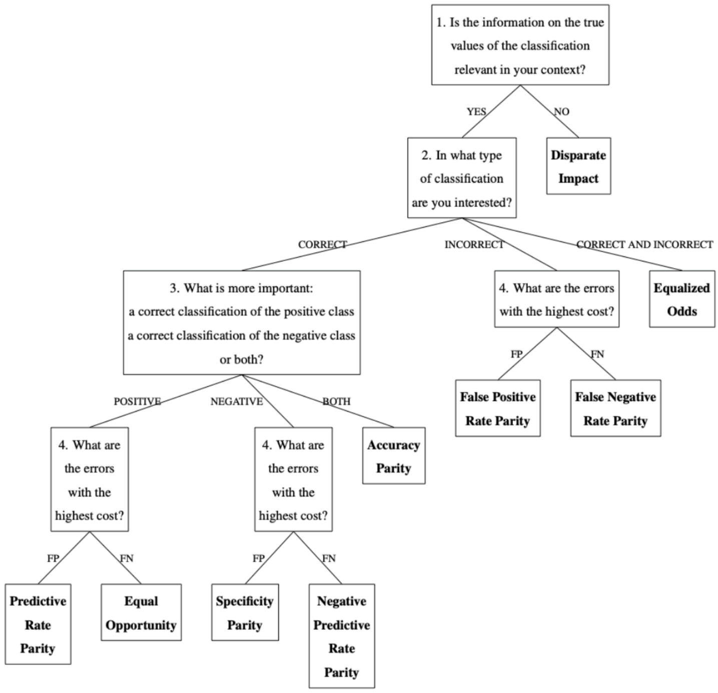
Fairness Workflow

The goal of a statistical fairness audit is to assess whether an algorithm treats unprivileged groups unfairly, based on its outcomes evaluated through a fairness measure. This can be challenging, but the Audit module simplifies the process into two stages: selection and evaluation.

More detailed information about the individual stages in the audit fairness workflow is provided below.

9.1 The two stages in the fairness workflow

In the selection stage, you choose the most suitable fairness measure to apply to the algorithm's outcomes you want to audit. This decision is based on the dataset's characteristics and the audit context. You will answer up to 4 questions, depending on the path taken in the decision-making workflow (see image below).



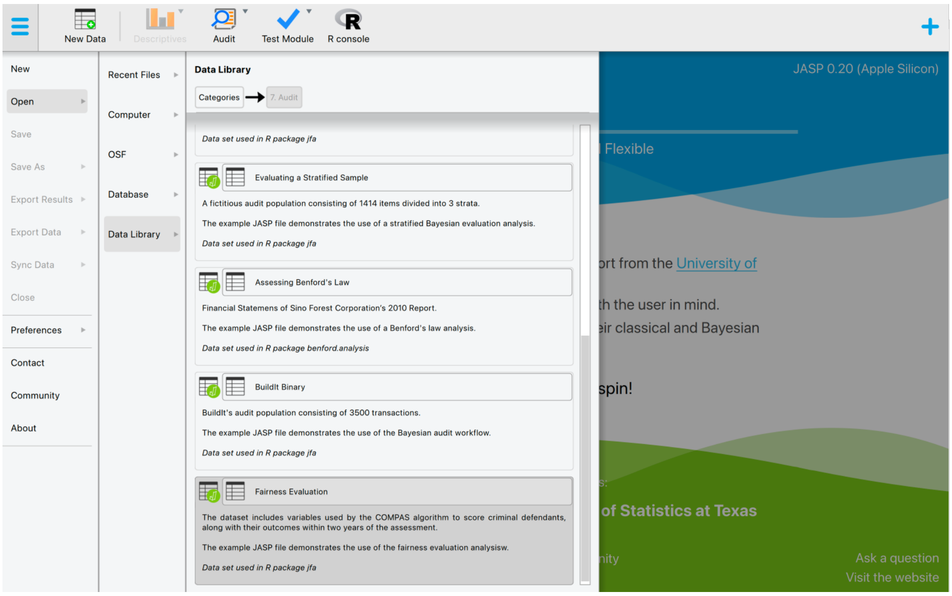
In the evaluation stage, you apply the selected fairness measure to assess the algorithm's outcomes. Statistical techniques calculate the values of the evaluation metric on which the fairness measure is based, and then test whether differences across groups indicate unfairness.

This manual emphasizes the practical application of the fairness workflow in JASP.

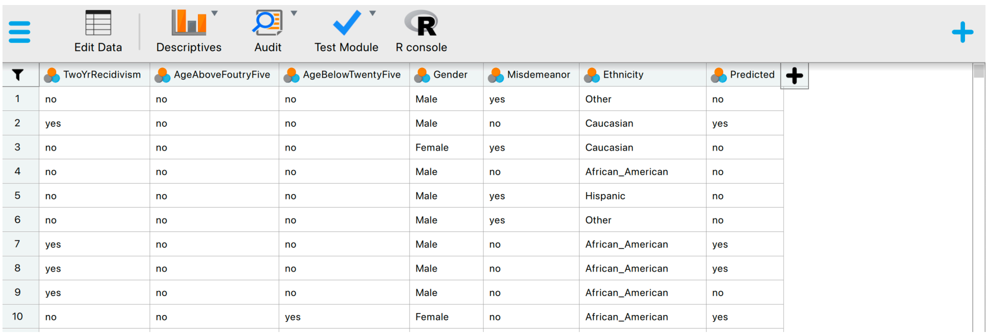
9.2 Practical example

The Audit module in JASP offers two ways to navigate the fairness workflow: the Fairness Workflow analysis, which guides you through all two stages, and an individual analysis for the Evaluation. This chapter uses the fairness workflow analysis to explain the Audit module's core functionality.

Let's explore an example of the audit fairness workflow. To follow along, open the 'Fairness Evaluation' dataset from the Data Library. Navigate to the top-left menu, click 'Open', then 'Data Library', select '7. Audit', and finally click on the text 'Fairness Evaluation' (not the green JASP-icon button).



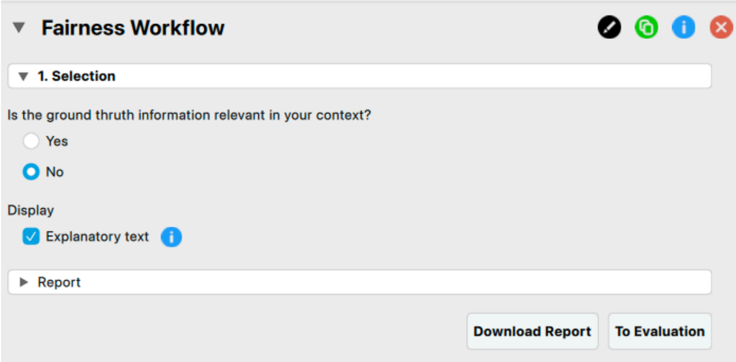
This will open a dataset with 6,172 rows and seven columns. The 'TwoYrRecidivism' column indicates whether the offender committed a crime within two years of being released and, therefore, provides the ground truth information on the offenders' recidivism. The 'Predicted' column indicates whether the COMPAS algorithm classifies the offender as low or high risk of committing another crime after being released. The 'Ethnicity' column indicates the offender's ethnicity (African-American, Asian, Caucasian, Hispanic, Native American, or Other). This column represents the so-called senstitive attribute. In this example, we seek to determine, with 95% confidence, whether the COMPAS algorithm discriminates against offenders based on their ethnicity.



	TwoYrRecidivism	AgeAboveFoutryFive	AgeBelowTwentyFive	Gender	Misdemeanor	Ethnicity	Predicted
1	no	no	no	Male	yes	Other	no
2	yes	no	no	Male	no	Caucasian	yes
3	no	no	no	Female	yes	Caucasian	no
4	no	no	no	Male	no	African_American	no
5	no	no	no	Male	yes	Hispanic	no
6	no	no	no	Male	yes	Other	no
7	yes	no	no	Male	no	African_American	yes
8	yes	no	no	Male	no	African_American	yes
9	yes	no	no	Male	no	African_American	no
10	no	no	yes	Female	no	African_American	yes

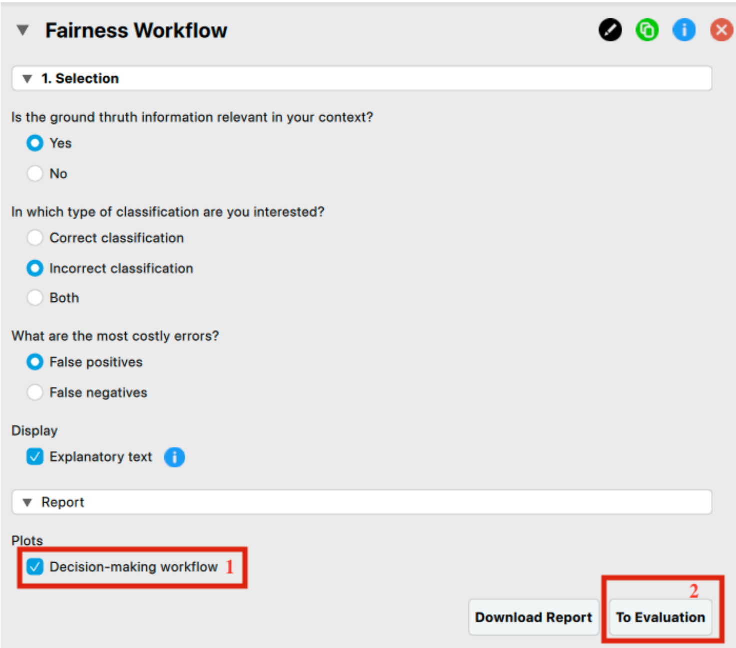
9.2.1 Stage 1: Selection

To start the fairness workflow, click on the Audit module icon and select 'Fairness Workflow'. This will open the following interface. where you need to specify the settings for the analysis.



As you change the default answer to this first question (answer 'no'), the next question appears. This happens for each question in the decision-making workflow.

In this example, the following answers are provided to the input questions, and the following specific options are then selected to generate both a textual and graphical output.



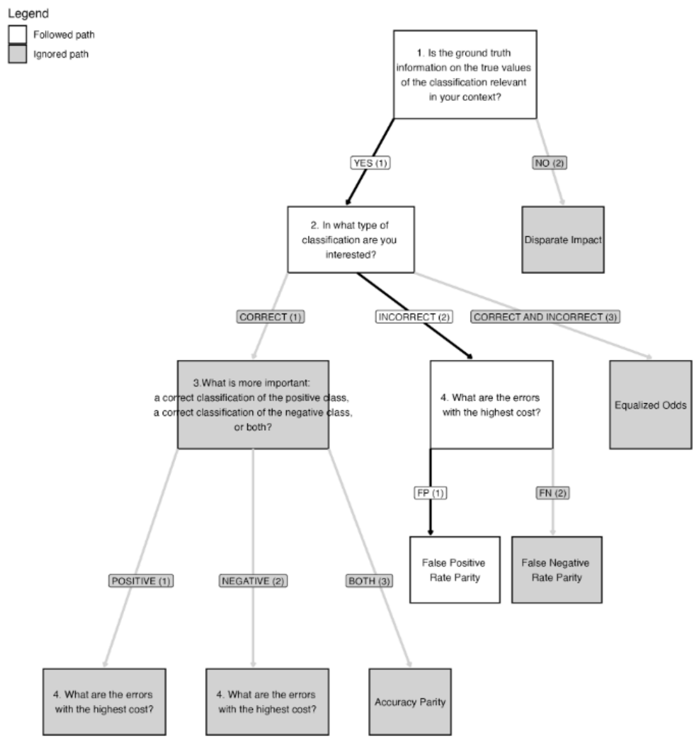
The primary output from the selection stage, shown below, identifies false positive rate parity as the most suitable measure for evaluating the fairness of the algorithm's outcomes.

Selection

The most suitable fairness measure is **false positive rate parity**.

1. **Decision-making workflow:** By selecting this option, you will see a graphical

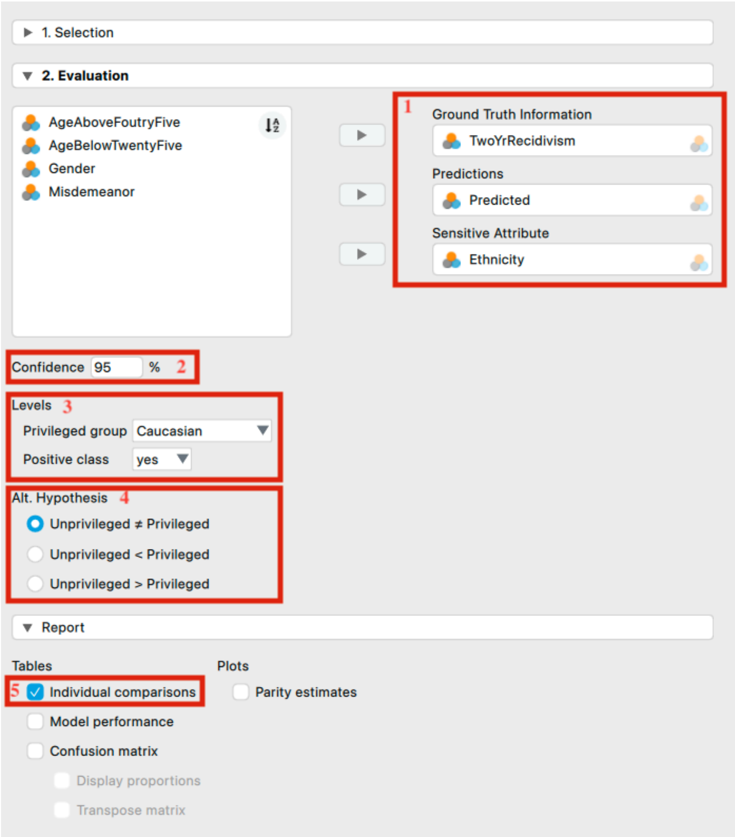
representation of the decision path: the questions you answered are shown in white, while the questions and answers that were skipped (because not relevant to your path) are shown in grey.



2. **Next stage:** Next, progress to the evaluation stage by clicking the 'To Evaluation' button.

9.2.2 Stage 2: Evaluation

In the evaluation stage, you assess the fairness in the algorithm's outcome (the 'Predicted' column in the dataset).



The following settings are required:

1. **Indicate the variables:** First, enter the variable indicating the true recidivism values of the offenders into the 'Ground Truth Information' field. Then, input the variables that hold the predicted recidivism values (the algorithm's outcome to be audited) and the offenders' ethnicity into their respective fields.
2. **Confidence level:** Specify the confidence level for your analysis. This level, which complements the significance level, dictates when to reject the null hypothesis and the amount of work needed to approve the fairness of the algorithm.
3. **Levels:** Select which group, identified by the sensitive attribute's values, is the privileged one (i.e., which group is advantaged by the algorithm's outcome), and which class of the algorithm's predictions is the favorable outcome.
4. **Alt. Hypothesis:** Choose whether to calculate a two-sided test or a one-sided test to evaluate the algorithm's fairness.
5. **Individual comparisons:** In the case of multiple unprivileged groups, it is possible to identify which specific group is being discriminated against if the algorithm is found to be unfair toward at least one group.

For a more detailed explanation of the settings and output in the evaluation stage, see Chapter 10.

The primary output from the evaluation stage, as shown in the first table below, reveals that the false positive rates across the groups—comparing all unprivileged groups to the privileged one—are not equal. Therefore, the algorithm is unfair toward at least one unprivileged group.The second table provides details on which specific group, among all the unprivileged groups, is being discriminated against. This is determined by comparing the algorithm's outcomes for each unprivileged group with those for the privileged group. A group is considered discriminated against if the p-value for its corresponding row is below 0.05.

Table 1. *Omnibus Test – False positive rate parity*

	n	X^2	df	p
Ethnicity	6.172	179,8	5	< 0.001

Note. The null hypothesis specifies that the false positive rate is equal across protected groups.

Table 2. *Comparisons to Privileged (P) Group*

		95% Confidence Interval			95% Confidence Interval		
	False positive rate	Lower	Upper	Parity	Lower	Upper	p[a]
Caucasian (P)	0,222	0,199	0,245	1,000			
African_American	0,415	0,390	0,441	1,874	1,761	1,988	< 0.001
Asian	0,087	0,011	0,280	0,392	0,048	1,265	0,199
Hispanic	0,191	0,149	0,238	0,860	0,672	1,074	0,254
Native_American	0,333	0,043	0,777	1,504	0,195	3,506	0,620
Other	0,151	0,106	0,205	0,680	0,478	0,925	0,020

Note. The null hypothesis specifies that the false positive rate of an unprivileged group is equal to that of the privileged (P) group.

[a] *p*–values are uncorrected for multiple comparisons and are based on Fisher's exact test.

Based on the results of this statistical analysis, the auditor concludes that the algorithm is unfair toward at least one group (Table 1). By comparing each group to the designated privileged group (Caucasian offenders) (Table 2), the auditor further concludes that the algorithm is unfair toward African-American offenders and those classified with 'Other' as their ethnicity.

Chapter 10

Evaluation

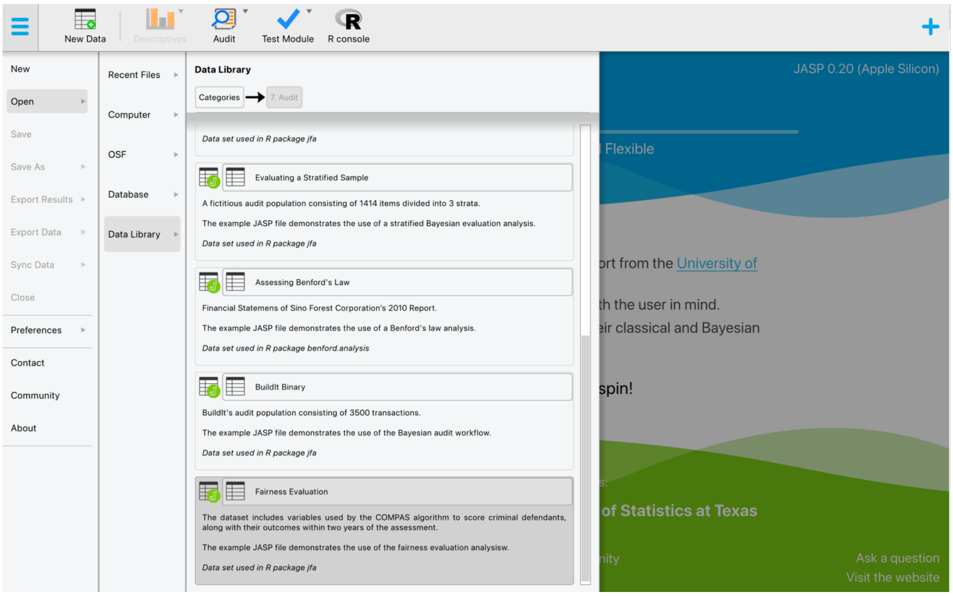
This chapter is about the ‘Evaluation’ analysis in the ‘Algorithm Auditing’ section of the module.

10.1 Purpose of the analysis

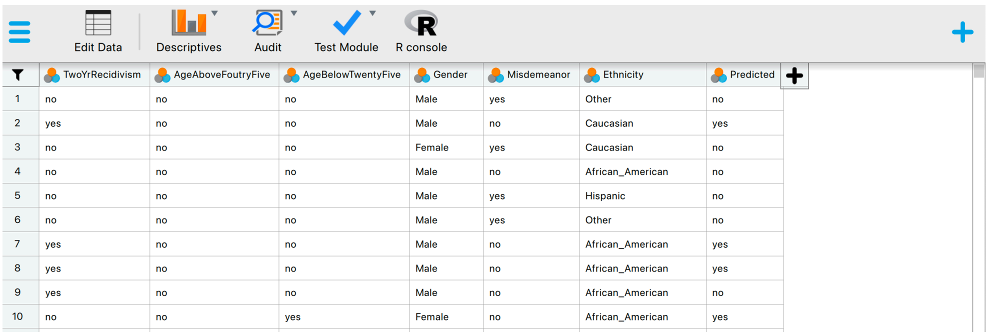
The goal of this procedure is to assess the extent to which an algorithm’s predictions are fair toward protected groups based on a sensitive attribute and to test this fairness with a type I error rate corresponding to a chosen significance level. This allows auditors to determine, with a certain level of assurance, whether the use of the audited algorithm results in discriminatory acts.

10.2 Practical example

Let’s explore an example of an evaluation analysis. To follow along, open the ‘Fairness Evaluation’ dataset from the Data Library. Navigate to the top-left menu, click ‘Open’, then ‘Data Library’, select ‘7. Audit’, and finally click on the text ‘Fairness Evaluation’ (not the green JASP-icon button).



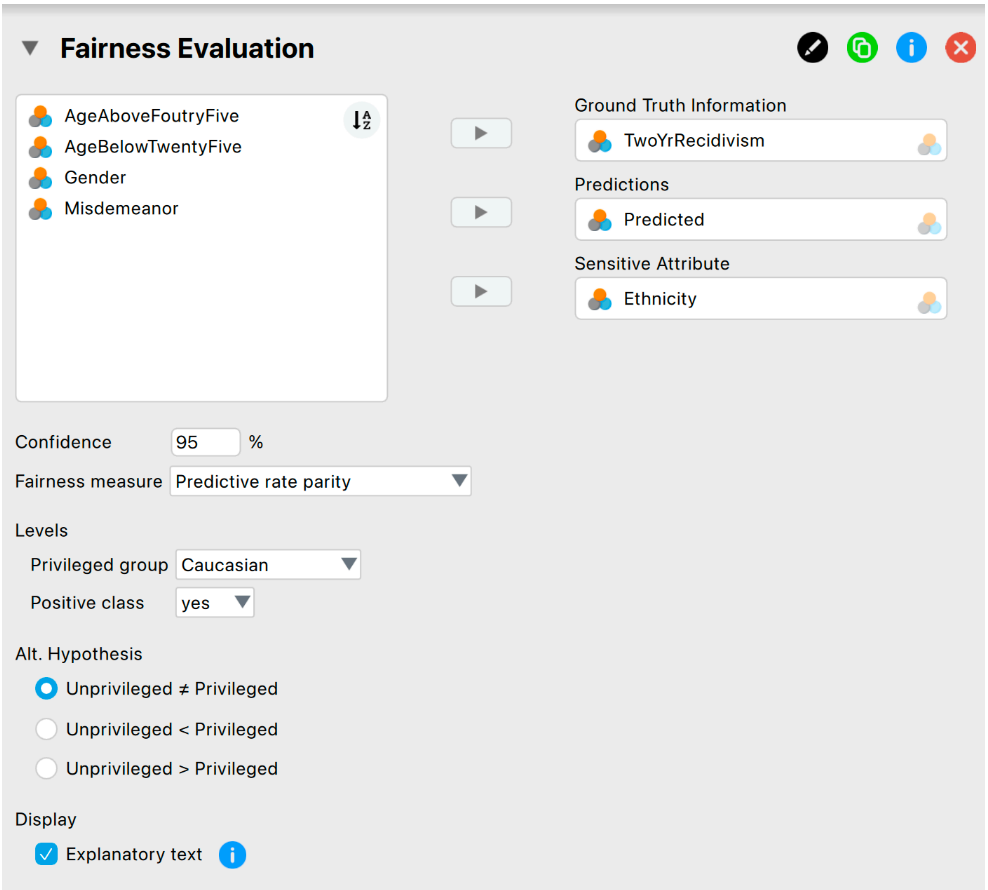
This will open a dataset with 6,172 rows and seven columns. The 'TwoYrRecidivism' column indicates whether the offender committed a crime within two years of being released and, therefore, provides the ground truth information on the offenders' recidivism. The 'Predicted' column indicates whether the COMPAS algorithm classifies the offender as low or high risk of committing another crime after being released. The 'Ethnicity' column indicates the offender's ethnicity (African-American, Asian, Caucasian, Hispanic, Native American, or Other). This column represents the so-called senstitive attribute. In this example, we seek to determine, with 95% confidence, whether the COMPAS algorithm discriminates against offenders based on their ethnicity.



	TwoYrRecidivism	AgeAboveFoutryFive	AgeBelowTwentyFive	Gender	Misdemeanor	Ethnicity	Predicted
1	no	no	no	Male	yes	Other	no
2	yes	no	no	Male	no	Caucasian	yes
3	no	no	no	Female	yes	Caucasian	no
4	no	no	no	Male	no	African_American	no
5	no	no	no	Male	yes	Hispanic	no
6	no	no	no	Male	yes	Other	no
7	yes	no	no	Male	no	African_American	yes
8	yes	no	no	Male	no	African_American	yes
9	yes	no	no	Male	no	African_American	no
10	no	no	yes	Female	no	African_American	yes

10.2.1 Main settings

To evaluate the fairness of the algorithm, we open the 'Evaluation' analysis within the section 'Algorithm Auditing' within the Audit module. The interface for the evaluation analysis is displayed below.



These are the main settings for the analysis:

- **Variables:** Begin by entering the variable that contains the ground truth information into the 'Ground Truth Information' field. Then, input the variables that hold the predictions of the algorithm values and the sensitive attribute into their respective fields.
- **Confidence:** Specify the confidence level for your analysis. This level, which complements the significance level, dictates when to reject the null hypothesis and the amount of work needed to determine the fairness of the algorithm. A higher confidence level requires more audit evidence to conclude that the algorithm is not discriminating against certain social groups. In this example, we use a confidence level of 95%.
- **Fairness measure:** Specify the fairness measure for your analysis. The possible options are: 'Predictive rate parity', 'Negative predictive rate parity', 'False positive rate parity', 'False negative rate parity', 'Equal opportunity, 'Specificity parity', 'Disparate impact', 'Equalized odds', and 'Accuracy parity'. In this example, we use the 'False positive rate parity'. Note that the 'Fairness Workflow' analysis in this module section guides you in choosing a fairness measure.
- **Levels: Privileged group:** Select which group, identified by the sensitive attribute's values, is the privileged one (i.e., which group is advantaged by the

algorithm's outcome). In this example, the privileged group is 'Caucasian'.

- **Levels: Positive class:** Select which class of the algorithm's predictions is the favorable outcome. This refers to the outcome that is beneficial to those who receive it. In this example, the positive class is 'yes'.
- **Alt. Hypothesis:** Choose whether to calculate a two-sided test or a one-sided test to evaluate the algorithm's fairness. In this example, we select the two-sided option 'Unprivileged $\neq$ privileged'.
- **Display: Explanatory text:** Finally, select whether to show explanatory text in the output.

10.2.2 Main output

The main table in the output below shows the omnibus test (Pearson's chi-square test), which compares all the groups identified by the values of the sensitive attribute, based on the direction of the alternative hypothesis. The first column specifies the total number of items in the dataset. The column 'χ^2' column displays the value of the test statistic for the overall test. The 'df' column indicates the degrees of freedom, which is the parameter for the chi-square distribution. The 'p' column shows the p-value of the omnibus test, based on which the null hypothesis is either rejected or not.

Table 1. *Omnibus Test – False positive rate parity* ▼

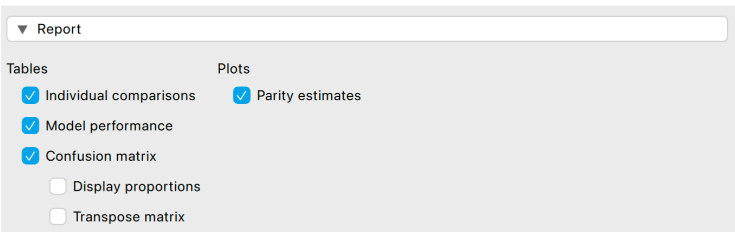
	n	X²	df	p
Ethnicity	6.172	179,8	5	< 0.001

Note. The null hypothesis specifies that the false positive rate is equal across protected groups.

In this example, the offenders' ethnicity is the sensitive attribute. As previously mentioned, the total number of rows is 6,127. The value of the test statistic is 179.8, and the degrees of freedom for the chi-square distribution are 5. The p-value associated with the omnibus test is less than 0.001, which is below the significance level (0.05). Therefore, we reject the hypothesis of equal false positive rates across the groups, concluding that the algorithm is unfair toward at least one of the groups.

10.2.3 Report

The following settings enable you to expand the report with additional output, such as tables and figures.



- **Tables: Individual comparisons:** Check this box to generate a table showing the individual comparisons between the privileged group and each of the different unprivileged groups. This allows you, when the p-value from the overall test leads to rejection of the null hypothesis, to identify which specific group is being discriminated against by the algorithm. To determine which group is being unfairly treated by the algorithm, you can focus on the last column, which shows statistical significance for the individual comparisons. In this example, the first three columns show the false positive rate by offenders' ethnicity, along with their respective 95% confidence intervals. The fourth and sixth columns display the false positive rate by offenders' ethnicity relative to the group of Caucasian offenders, again with corresponding confidence intervals. In the last columm, when the p-value is lower than the significance level (0.05), the false positive rate is significantly different between the unprivileged group and the privileged group, indicating that the unprivileged group is being discriminated against.

Table 2. *Comparisons to Privileged (P) Group*

		95% Confidence Interval			95% Confidence Interval		
	False positive rate	Lower	Upper	Parity	Lower	Upper	p[a]
Caucasian (P)	0,222	0,199	0,245	1,000			
African_American	0,415	0,390	0,441	1,874	1,761	1,988	< 0.001
Asian	0,087	0,011	0,280	0,392	0,048	1,265	0,199
Hispanic	0,191	0,149	0,238	0,860	0,672	1,074	0,254
Native_American	0,333	0,043	0,777	1,504	0,195	3,506	0,620
Other	0,151	0,106	0,205	0,680	0,478	0,925	0,020

Note. The null hypothesis specifies that the false positive rate of an unprivileged group is equal to that of the privileged (P) group.
[a] *p*–values are uncorrected for multiple comparisons and are based on Fisher's exact test.

- **Tables: Model performance:** Check this box to generate a table showing the model performance metric values for each group identified by the sensitive attribute — both privileged and unprivileged — in order to evaluate the algorithm's performance within each group.

Table 3. *Model Performance*

	Support	Accuracy	Precision	Recall	F1 Score
African_American	3.175	0,672	0,665	0,753	0,706
Asian	31	0,742	0,500	0,250	0,333
Caucasian	2.103	0,659	0,577	0,472	0,519
Hispanic	509	0,682	0,591	0,466	0,521
Native_American	11	0,636	0,600	0,600	0,600
Other	343	0,694	0,612	0,419	0,498

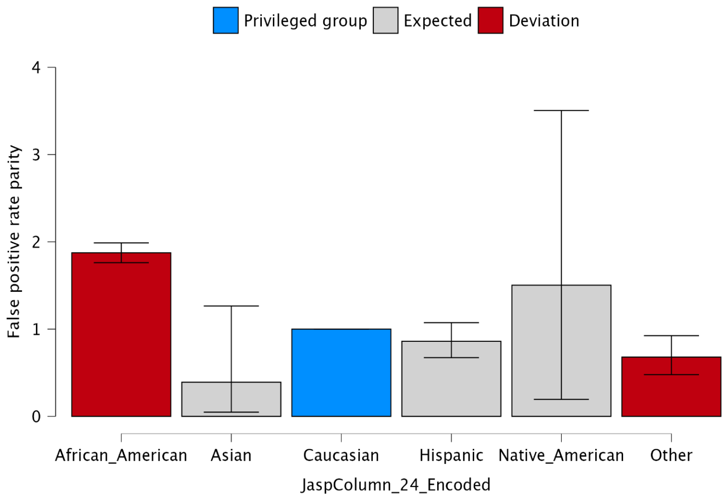
- **Tables: Confusion matrix:** Check this box to generate a table showing the confusion matrix, obtained by comparing the predicted values of the algorithm with the true values for each group identified by the sensitive attribute, along with the value of the model evaluation metric on which the selected fairness measure is based (in this example, the false positive rate). It is also possible to display the numbers as proportions or transpose the matrix.

***Table 4.** Confusion Matrix* ▼

		Predicted		
	Observed	no	yes	False positive rate[a]
African_American	no	885	629	0,415
	yes	411	1.250	
Asian	no	21	2	0,087
	yes	6	2	
Caucasian	no	997	284	0,222
	yes	434	388	
Hispanic	no	259	61	0,191
	yes	101	88	
Native_American	no	4	2	0,333
	yes	2	3	
Other	no	186	33	0,151
	yes	72	52	

Note. True Positives (TP) are cells in which the positive class (*yes*) is correctly predicted, while True Negatives (TN) are cells where the negative class is correctly predicted. False Negatives (FN) are cells in which the positive class is incorrectly predicted, and False Positives (FP) are cells in which the negative class is incorrectly predicted.
[a] False positive rate = FP / (TN + FP)

- **Plots: Parity estimates:** Check this box to generate a figure displaying the ratio between the model evaluation metric on which the selected fairness measure is based (in this example, the false positive rate) for each unprivileged group compared to the privileged group. Blue is used to represent the privileged group. Grey indicates that the ratio does not differ significantly from 1, meaning there is no significant difference in treatment between the unprivileged group and the privileged group. Red indicates that the ratio differs significantly from 1, suggesting a difference in treatment between the unprivileged group and the privileged group.



This is the user manual for **JASP for Audit**, a module within the free and open-source statistical software program **JASP (https://jasp-stats.org)** built to support the statistical aspects of an audit. The module offers a user-friendly graphical interface, simplifying complex statistical auditing procedures to make them as effortless as possible.

Next to the classical frequentist statistical techniques that are standard in audit practice, the module offers state-of-the-art Bayesian techniques that enable auditors to work more efficiently and transparently. Furthermore, JASP for Audit helps the auditor in interpreting, explaining, and reporting the analyses and leaves a transparent audit trail.