

What are the descriptive statistics for one numeric variable in terms of frequencies?

Authored by
stats writer

June 24, 2024

RECOMMENDED CITATION

stats writer (2024). *What are the descriptive statistics for one numeric variable in terms of frequencies?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=149529>

Descriptive statistics for one numeric variable in terms of frequencies refer to the analysis and summary of a single numerical variable by examining its distribution and frequency of occurrence. This includes measures such as mean, median, mode, standard deviation, and range, which provide information on the central tendency and variability of the data. By utilizing these statistics, one can gain a better understanding of the data and make informed decisions or draw conclusions. This approach is commonly used in various fields, including finance, economics, and social sciences, to describe and analyze numerical data.

Introduction

When summarizing quantitative (continuous/interval/ratio) variables, we are typically interested in questions like:

What is the "center" of the data? (Mean, median) How spread out is the data? (Standard deviation/variance) What are the extremes of the data? (Minimum, maximum; Outliers) What is the "shape" of the distribution? Is it symmetric or asymmetric? Are the values mostly clustered about the mean, or are there many values in the "tails" of the distribution? (Skewness, kurtosis)

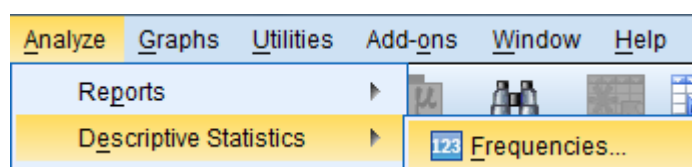
In SPSS, the Frequencies procedure is typically used on categorical variables, but it also has special settings that can be applied for continuous numeric variables. In particular, the Frequencies procedure can compute percentiles that are not otherwise included in the Descriptives, Compare Means, or Explore procedures. In all, the Frequencies procedure can compute the following statistics for one or more continuous variables:

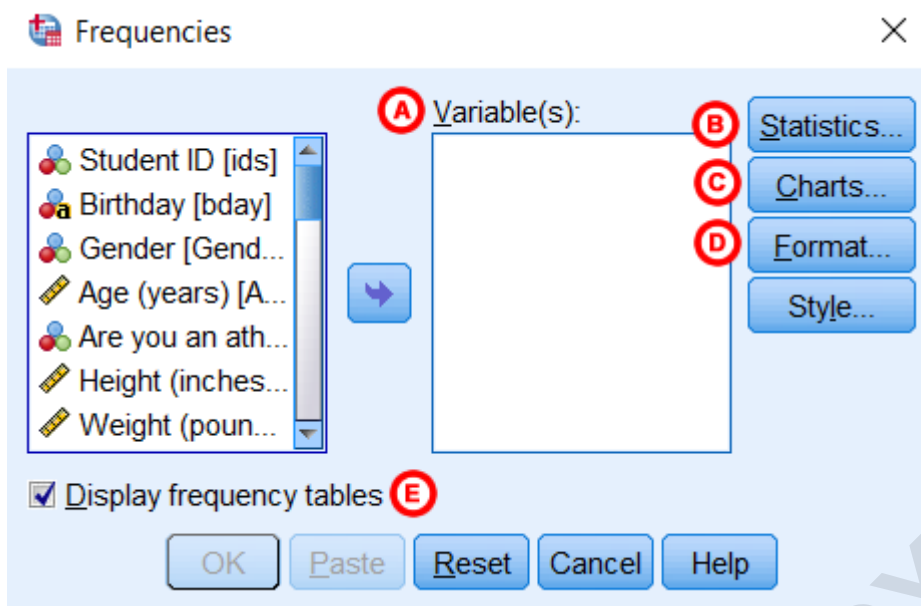
N valid responses N missing responses Mean Standard deviation Variance Sum Minimum Maximum Range Skewness Kurtosis Median Quartiles (25th, 50th, 75th percentiles) Percentiles Mode

The Frequencies procedure can also produce histograms with or without a normal distribution overlaid on the graph.

Using the Frequencies Procedure with Scale Variables

To call the Frequencies procedure, click **Analyze > Descriptive Statistics > Frequencies**.





A Variable(s): The variables to analyze with the Frequencies procedure. To include a variable for analysis, double-click on its name to move it to the Variables box. You can add several variables to this box to obtain statistics for each variable.

B Statistics: Opens the Frequencies: Statistics window, which contains various descriptive statistics, most of which are suitable for continuous numeric variables.

Frequencies: Statistics

Percentile Values

☐ Quartiles

☐ Cut points for: 10 equal groups

☐ Percentile(s):

Add

Change

Remove

Central Tendency

☐ Mean

☐ Median

☐ Mode

☐ Sum

☐ Values are group midpoints

Dispersion

☐ Std. deviation ☐ Minimum

☐ Variance ☐ Maximum

☐ Range ☐ S.E. mean

Characterize Posterior Di...

☐ Skewness

☐ Kurtosis

Continue Cancel Help

Most of the statistics in the **Central Tendency**, **Dispersion**, and **Distribution** groups are valid for continuous variables; the only exception is the **Mode**, which very rarely has a useful interpretation for situations involving continuous variables. Most of these statistics are identical to the ones that can be obtained with Descriptives, Compare Means, or Explore, so they will not be covered again here. One noticeable exception to this is the **Percentile Values** group, which is unique to the Frequencies procedure:

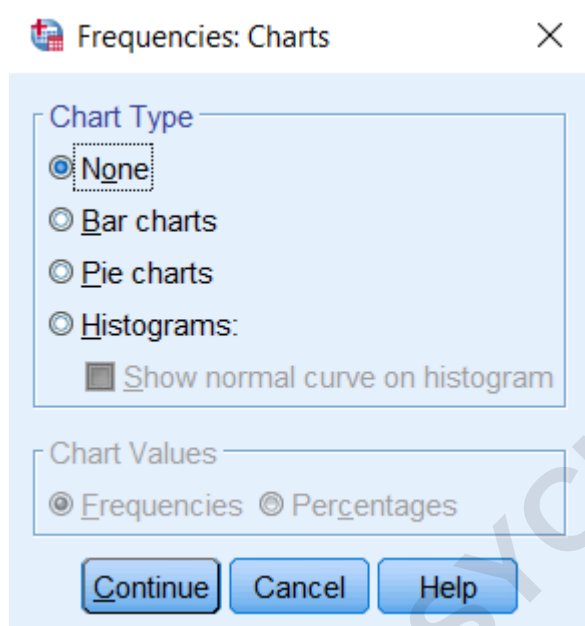
The **Quartiles** option produces the first, second, and third quartiles (i.e., the 25th, 50th, and 75th percentiles, respectively). The **Cut points for n equal groups** option will divide the dataset into n equally sized groups and report the percentiles. For example, if the user specifies $n=5$, then the output will report the 20th, 40th, 60th, and 80th percentiles. Or, if the user specifies $n=10$, then the output will report the 10th, 20th, 30th, ..., 90th percentiles. The **Percentiles** option allows the user to specify the exact percentiles to report. The percentiles should be entered as whole numbers.

You can select more than one option in the Percentile Values group. If your selections request overlapping information, that information will not be printed twice.

Note: The **Values are group midpoints** check box should only be selected when your data values represent the midpoint of a range. For example, this would be the case if you had coded anyone

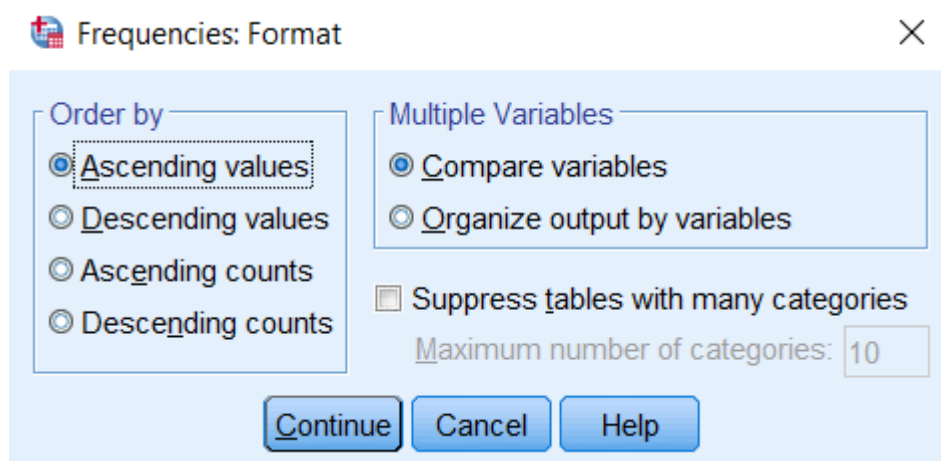
between the ages of 30 and 39 as 35 (source: [IBM SPSS Statistics Information Center](#)). This situation is more often associated with ordinal categorical variables.

CCharts: Opens the Frequencies: Charts window, which contains various graphical options. Options include bar charts, pie charts, and histograms. **Histograms** are the only appropriate option for continuous variables; bar charts and pie charts should *never* be used with continuous variables. If requesting a histogram, the optional **Show normal curve on histogram** option will overlay a normal curve on top of your histogram, which can be useful when assessing the normality of a variable.



Note that the options in the **Chart Values** area apply only to bar charts. These buttons will be greyed out if the radio button for Histograms is selected.

DFormat: Opens the Frequencies: Format window, which contains options for how to sort and organize the table output.



The **Order by** options are not relevant to continuous variables, but the **Multiple Variables** options allow for customization of output when two or more continuous variables are specified.

Compare variables places the descriptive statistics for the numeric variables side-by-side. **Organize output by variables** creates separate summary tables for each numeric variable.

Display frequency tables: When checked, frequency tables will be printed. (This box is checked by default.) If this check box is not checked, no frequency tables will be produced, and the only output will come from supplementary options from **Statistics** or **Charts**. **You will want to uncheck this box if using the Frequencies procedure on a continuous numeric variable.** (If this box is left checked, a frequency table will be produced where each unique number is treated as its own category. This could lead to a table with 100+ categories, depending on the number of observations in your dataset.)

Example: Comparing Percentiles for More Than Two Variables

Problem Statement

For variables with skewed distributions, it is often more useful to look at percentiles than it is to look at means. This is because means are more susceptible to outliers: a single strongly outlying value can "pull" the mean up or down from where it would be otherwise. By comparison, percentiles (including the median) are relatively robust to outliers - that is, percentiles generally do not change much when outliers are present compared to when there aren't outliers present.

When reporting placement or achievement test scores, it's often more useful (and more descriptive) to report the percentiles than it is to report the means. For example, we may want to know the 80th percentile: the score that 80% of students scored below. The sample dataset has placement test scores (out of 100 points) for four subject areas: English, Reading, Math, and

Writing. Let's use the Frequencies procedure to obtain the *quintiles* (i.e., the 20th, 40th, 60th, and 80th percentiles) of the scores.

Running the Procedure

Using the Frequencies Dialog Window

Open the Frequencies window (**Analyze > Descriptive Statistics > Frequencies**). Highlight the four test score variables (click variable English, then hold down Shift and click variable Writing) in the left-hand column. Then click the arrow button to move them to the Variables box. Click **Statistics**. Click the **Cut points** check box, and specify **5** equal groups. Click **Continue** when finished. Click **Format**. In the Multiple Variables area, make sure that **Compare variables** is selected. Then click **Continue**. Uncheck the box for **Display frequency tables**. When finished, click **OK**.

Using Syntax

```
FREQUENCIES VARIABLES=English Reading Math Writing
/FORMAT=NOTABLE
/NTILES=5
/ORDER=ANALYSIS.
```

Output

There is only one box, **Statistics**, that will print to the Output window. This box will contain the number of valid and missing values for each variable, as well as any additional statistics we requested (in this case, the quintiles).

Statistics					
		English	Reading	Math	Writing
N	Valid	409	425	422	404
	Missing	26	10	13	31
Percentiles	20	77.1	75.8	59.3	75.0
	40	81.4	80.6	63.7	77.7
	60	84.5	84.1	67.6	80.9
	80	88.2	88.4	72.1	84.5

Note that, by default, SPSS will determine how many decimal places to use for the percentiles based on the variable's number of decimal places. For this screenshot, we have shortened the output to one decimal place for readability.

The "Compare Groups" option we selected told SPSS to put the results for all four variables in a single table, side-by-side. From this, we can quickly make several observations about the data:

Writing had the most missing test scores (31), while Reading had the fewest (10). The Math test had the lowest scores in general. The bottom 20% of students scored below 59.3; the top 20% scored above 72.1. Contrast with the English test scores, where the bottom 20% scored below 77.1, and the top 20% scored above 88.2.

Example: Histogram with Normal Overlay

Problem Statement

Suppose we are interested in getting a rough estimate of whether or not a variable is normally distributed.

In this example, we will demonstrate what a histogram with a normal overlay looks like using the variable English from the sample dataset. This variable represents the subjects' score (out of 100 points) on an English placement test.

Running the Procedure

Using the Frequencies Dialog Window

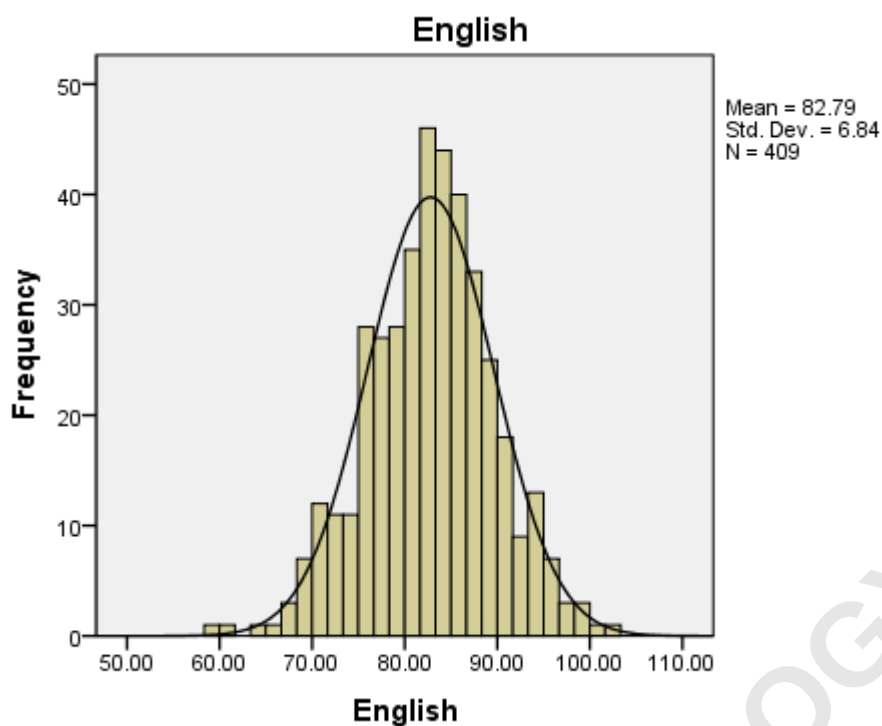
Open the Frequencies window (**Analyze > Descriptive Statistics > Frequencies**). Double-click on variable English to move it to the Variables box. Click **Charts** to open the Frequencies: Charts window. Click **Histogram**, and check the box for **Show normal curve on histogram**. Then click **Continue**. Uncheck the box for **Display frequency tables**. When finished, click **OK**.

Using Syntax

```
FREQUENCIES VARIABLES=English  
/FORMAT=NOTABLE  
/HISTOGRAM NORMAL  
/ORDER=ANALYSIS.
```

Output

You should see a histogram that looks like this:



If the data were perfectly normally distributed, we would expect to see the bars' height match up with normal overlay. While this graph alone is not enough to decide if the English scores are normally distributed, it does allow us to see that the data is reasonably symmetrically distributed about the mean, and there do not appear to be many huge deviations from the normal curve.

How might this curve look with a variable that was non-normal? Try re-running this example using variable Weight. Pay close attention to the tails. A normally distributed variable should have near-equal numbers of observations in the right and the left tails. Also look for any bars of the histogram that are much taller or much shorter than the normal overlay.