

Frequency Tables

The **FREQUENCY TABLES** commands construct tables of frequency counts and percentages for discrete or continuous variables. A frequency table is a tabular representation of data that can be used to summarize a frequency distribution of a categorical, nominal or ordinal variable. Each row consists of five values: *value* (the value of input variable for discrete data or the class interval for continuous data), *count* (for discrete data: the number of times the value occurs within the data set; for continuous data: the number of observations that fall into the bin), *cumulative count*, *percent*, *cumulative percent*.

How To

- ✓ Run: **STATISTICS->BASIC STATISTICS->FREQUENCY TABLES (DISCRETE DATA)...** or **FREQUENCY TABLES (CONTINUOUS DATA)...**
- ✓ Select one or multiple input **VARIABLES**.
- ✓ Optionally, select a frequency variable. **FREQUENCY VARIABLE** specifies the number of observations that each row represents. When omitted, each row represents a single observation.
- ✓ Optionally, select a break (layer) variable. **BREAK (LAYER) VARIABLE** distinct values will cause separate tables to be generated.
- ✓ *Only complete rows are included (values for an input variable and for the frequency and layer variables, if any, are not missing).*
- ✓ Use the **PLOT HISTOGRAM** option to add a histogram for every frequency table.

CONTINUOUS DATA ONLY:

- ✓ Enter the **NUMBER OF class INTERVALS** or leave the **NUMBER OF INTERVALS** option (**ADVANCED OPTIONS**) equal to zero to create a set of evenly distributed bins between the variable's minimum and maximum values, the number of bins is defined as $k = \lceil \log_2 N \rceil + 1$, and N is the total number of observations for a variable (Sturges, 1926). Please note, that the actual number of intervals may differ due to using *neat bins* (or "round" intervals, i.e. intervals with a width whose last digit in scientific notation is 1, 2 or 5). The number of intervals is calculated separately for each input variable.
- ✓ To use an exact number and location of class intervals (same for all input variables) – specify **both** the lower bound (**START VALUE**) and the width of each interval (**INTERVAL WIDTH**). If there is an observation that is less than the start value or greater than the upper bound, a bin "Up to" or "More than" is added to the table.

Results

Report includes tables with frequency counts for each variable (for each level of the break variable, if any).

Discrete variables

X_i – i^{th} observation of the input variable.

COUNT - the number of observations for each unique value of the input variable (X_i).

CUMULATIVE COUNT - the number of observations with a value less than or equal to the X_i .

PERCENT – percentage of X_i compared to the count of all observation.

CUMULATIVE PERCENT - percentage of the observations with a value less than or equal to the X_i compared to the count of all observation.

Continuous variables

x_i TO x_{i+1} – interval (bin range).

COUNT - the number of observations falling within bin range.

CUMULATIVE COUNT - the number of observations with the value less than *or equal* to the right boundary of the bin (for left-closed bins – *strictly less* than the right boundary of the bin).

PERCENT – percentage of observations compared to the count of all observation.

CUMULATIVE PERCENT - percentage of the observations with a value less than or equal to the right boundary of the range compared to the count of all observation.

Example

The grades awarded for the assignment set for the class of 22 students are as follows:

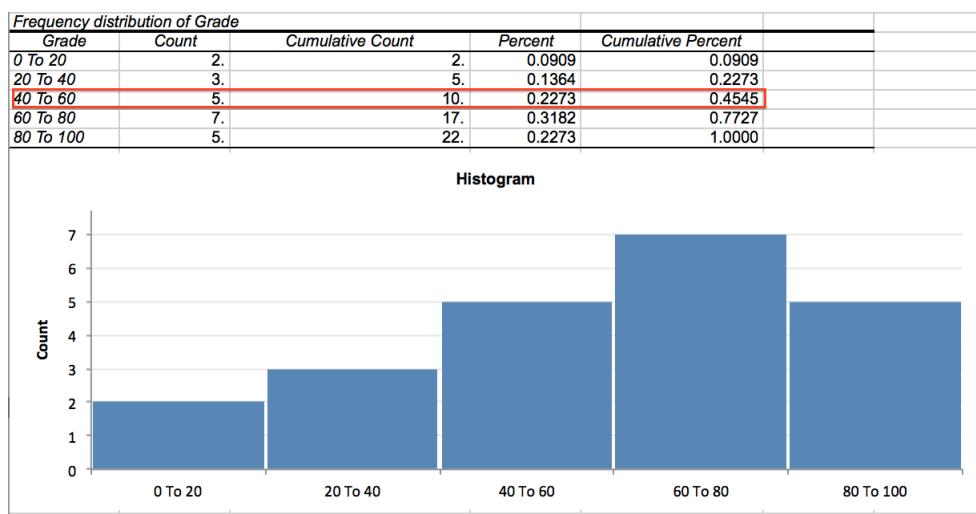
21	95	43	49	30	46	90	46	77	65	85
35	80	53	79	93	93	65	17	19	70	70

EXAMPLE: [FREQ TABLES dataset, “GRADE” variable]

To construct the frequency table: run the **STATISTICS->BASIC STATISTICS-> FREQUENCY TABLES (CONTINUOUS DATA)** command and select the **GRADE** column as the input variable. Then group the data into five class intervals from 1 to 100: set the **INTERVAL WIDTH** to 20, set the **NUMBER OF INTERVALS** to 5 and set the **START VALUE** to 0.

Number of intervals [Leave 0 for auto]	5
* Start value [leave empty for auto]	0
* Interval width [leave empty for auto]	20
* Hide empty bins	<input type="checkbox"/>
Bin intervals	Right-closed (left open) <input type="button" value="▼"/>

The report produced includes the frequency table and the histogram for the Grade variable.



From the third row we can see that 22.7% of students have grades between 40 and 60 and approximately 45% of students were scored less than or equal to 60 (right-closed intervals).

References

- Sturges, H. A. (1926). The choice of a class interval. *Journal of the American Statistical Association*, 21, 65–66.
- Velleman, P. F., & Hoaglin, D. C. (1981). Applications, basics, and computing of exploratory data analysis. Boston, Mass: Duxbury Press.