Rapport package team

Descriptive statistics

2011-04-26 20:25 CET

## Description

This template will return descriptive statistics of a numerical or frequency table of a categorical variable.

### *gender* ("Gender")

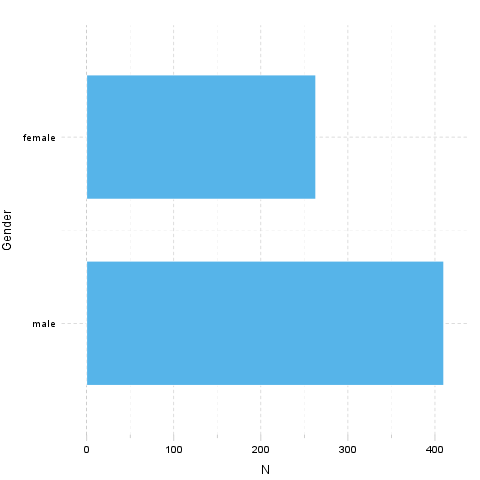
The dataset has *709* observations with *673* valid values (missing: *36*).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| gender | N | % | Cumul. N | Cumul. % |
| male | 410 | 60.92 | 410 | 60.92 |
| female | 263 | 39.08 | 673 | 100 |
| Total | 673 | 100 | 673 | 100 |

Frequency table: Gender

The most frequent value is *male*.

#### Charts

[](plots/Descriptives-1-hires.png)

## Description

This template will return descriptive statistics of a numerical or frequency table of a categorical variable.

### *age* ("Age")

The dataset has *709* observations with *677* valid values (missing: *32*).

#### Base statistics

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | mean | sd | var |
| Age | 24.57 | 6.849 | 46.91 |

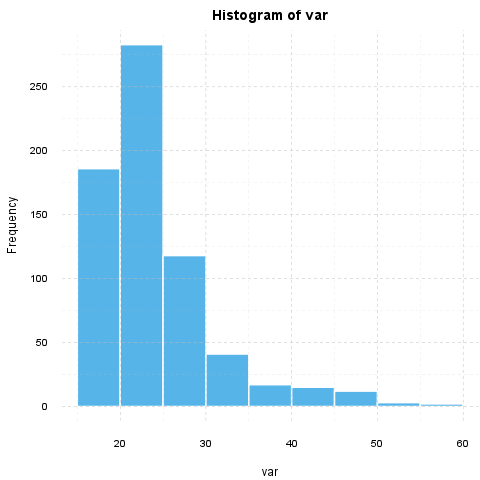
Descriptives: Age

The [standard deviation](http://en.wikipedia.org/wiki/Standard_deviation) equals to *6.849* (variance: *46.91*), which shows the unstandardized degree of [homogenity](http://en.wikipedia.org/wiki/Homogeneity_(statistics)): how much variation exists from the average. The [expected value](http://en.wikipedia.org/wiki/Mean) is around *24.57*, somewhere between *24.06* and *25.09* with the standard error of *0.2632*.

The highest value found in the dataset is *58*, which is exactly *3.625* times higher than the minimum (*16*). The difference between the two is described by the [range](http://en.wikipedia.org/wiki/Range_(statistics)): *42*.

#### Chart

A [histogram](http://en.wikipedia.org/wiki/Histogram) visually shows the [distribution](http://en.wikipedia.org/wiki/Probability_distribution) of the dataset based on artificially allocated [frequencies](http://en.wikipedia.org/wiki/Statistical_frequency). Each bar represents a theoretical interval of the data, where the height shows the count or density.

[](plots/Descriptives-2-hires.png)

If we *suppose* that *Age* is not near to the [normal distribution](http://en.wikipedia.org/wiki/Normal_distribution) (see for example [skewness](http://en.wikipedia.org/wiki/Skewness): *1.925*, [kurtosis](http://en.wikipedia.org/wiki/Kurtosis): *4.463*), checking the median (*23*) might be a better option instead of the mean. The [interquartile range](http://en.wikipedia.org/wiki/Interquartile_range) (*6*) measures the statistics dispersion of the variable (similar to standard deviation) based on median.

## Description

This template will return descriptive statistics of a numerical or frequency table of a categorical variable.

### *hp*

The dataset has *32* observations with *32* valid values (missing: *0*).

#### Base statistics

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | mean | sd | var |
| hp | 146.7 | 68.56 | 4701 |

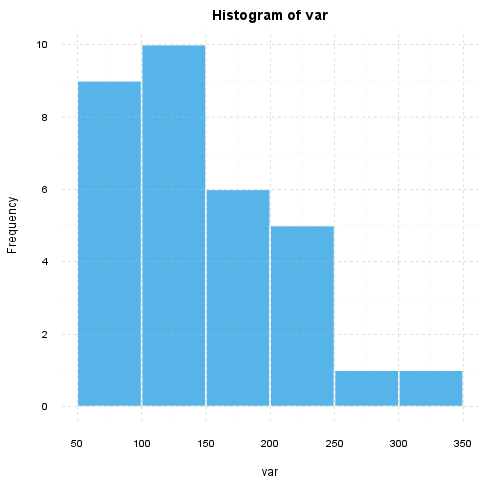
Descriptives: hp

The [standard deviation](http://en.wikipedia.org/wiki/Standard_deviation) equals to *68.56* (variance: *4701*), which shows the unstandardized degree of [homogenity](http://en.wikipedia.org/wiki/Homogeneity_(statistics)): how much variation exists from the average. The [expected value](http://en.wikipedia.org/wiki/Mean) is around *146.7*, somewhere between *122.9* and *170.4* with the standard error of *12.12*.

The highest value found in the dataset is *335*, which is exactly *6.442* times higher than the minimum (*52*). The difference between the two is described by the [range](http://en.wikipedia.org/wiki/Range_(statistics)): *283*.

#### Chart

A [histogram](http://en.wikipedia.org/wiki/Histogram) visually shows the [distribution](http://en.wikipedia.org/wiki/Probability_distribution) of the dataset based on artificially allocated [frequencies](http://en.wikipedia.org/wiki/Statistical_frequency). Each bar represents a theoretical interval of the data, where the height shows the count or density.

[](plots/Descriptives-3-hires.png)

If we *suppose* that *hp* is not near to the [normal distribution](http://en.wikipedia.org/wiki/Normal_distribution) (see for example [skewness](http://en.wikipedia.org/wiki/Skewness): *0.726*, [kurtosis](http://en.wikipedia.org/wiki/Kurtosis): *-0.1356*), checking the median (*123*) might be a better option instead of the mean. The [interquartile range](http://en.wikipedia.org/wiki/Interquartile_range) (*83.5*) measures the statistics dispersion of the variable (similar to standard deviation) based on median.

This report was generated with [R](http://www.r-project.org/) (3.0.1) and [rapport](http://rapport-package.info/) (0.51) in *1.105* sec on x86\_64-unknown-linux-gnu platform.

