

8.2: Describing...

Firstly, look on the basic characteristics of every character:

Since [SEX](#) and [COLOR](#) are categorical, the output in these columns has no sense, but you may want to convert these columns into “true” categorical data. There are multiple possibilities but the simplest is the conversion into *factor*:

(To retain the original data, we copied it first into new object [data1](#). Please check it now with [summary\(\)](#) yourself.)

[summary\(\)](#) command is applicable not only to the whole data frame but also to individual characters (or variables, or columns):

It is possible to calculate characteristics from [summary\(\)](#) one by one. Maximum and minimum:

... median:

... mean for [WEIGHT](#) and for each character:

and

... and also round the result to one decimal place:

(Again, the output of [colMeans\(\)](#) has no sense for [SEX](#) and [COLOR](#).)

Unfortunately, the commands above (but not [summary\(\)](#)) do not work if the data have missed values ([NA](#)):

To calculate mean without noticing missing data, enter

Another way is to remove rows with [NA](#) from the data with:

Then, [data2.o](#) will be free from missing values.

Sometimes, you need to calculate the sum of all character values:

... or the sum of all values in one row (we will try the second row):

... or the sum of all values for *every* row:

(These summarizing exercises are here for training purposes only.)

For the categorical data, it is sensible to look how many times every value appear in the data file (and that also help to know all values of the character):

Now transform frequencies into percents (100% is the total number of bugs):

One of the most important characteristics of data variability is the *standard deviation*:

Calculate standard deviation for each numerical column (columns 3 and 4):

If you want to do the same for data with a missed value, you need something like:

Calculate also the *coefficient of variation* (CV):

We can calculate any characteristic separately for males and females. Means for insect weights:

How many individuals of each color are among males and females?

(Rows are colors, columns are males and females.)

Now the same in percents:

Finally, calculate mean values of weight separately for every combination of color and sex (i.e., for red males, red females, green males, green females, and so on):

This page titled 8.2: Describing... is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by Alexey Shipunov via source content that was edited to the style and standards of the LibreTexts platform.