Exploring two-dimensional distributions (categorical variables) in CODAP, Part 1

**Link to CODAP:** <https://tinyurl.com/you-pb-50en>

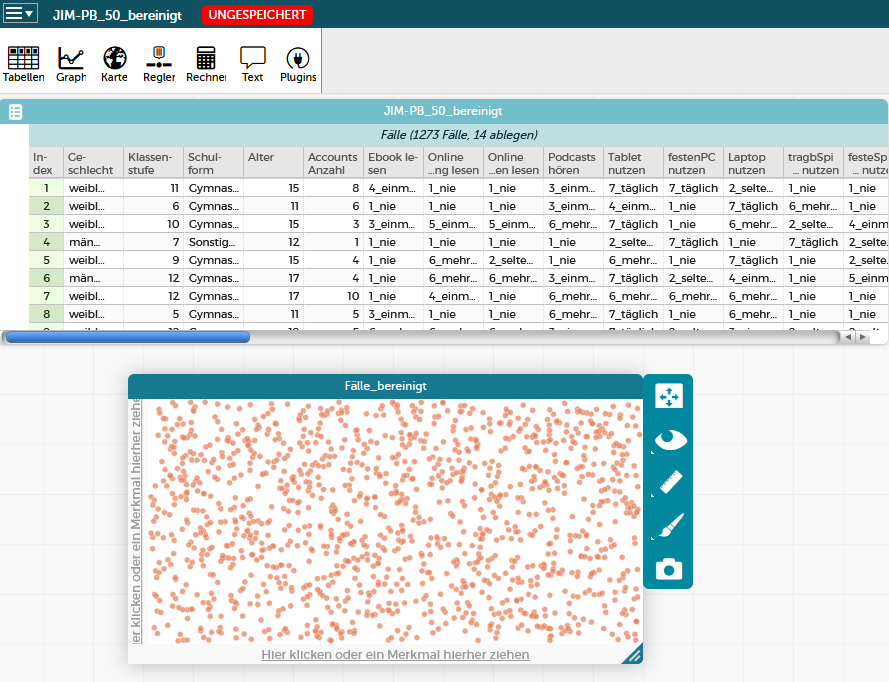
Percentages are very important when comparing two categorical distributions. In this guide, you will learn what row percentages, column percentages and cell percentages are and what you can say with which percentages.

This is done using the following question as an example.

# Podcasts: More for boys or for girls in this sample?

**It's about the two variables: Podcasts\_listening vs. gender**

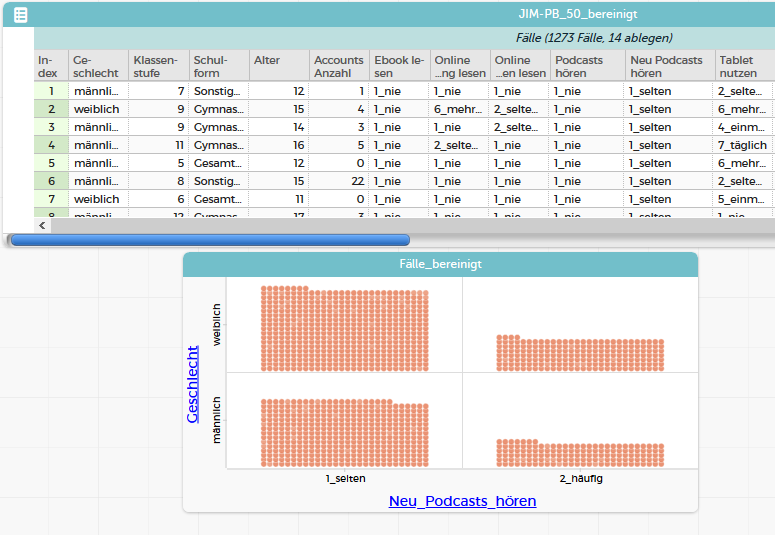
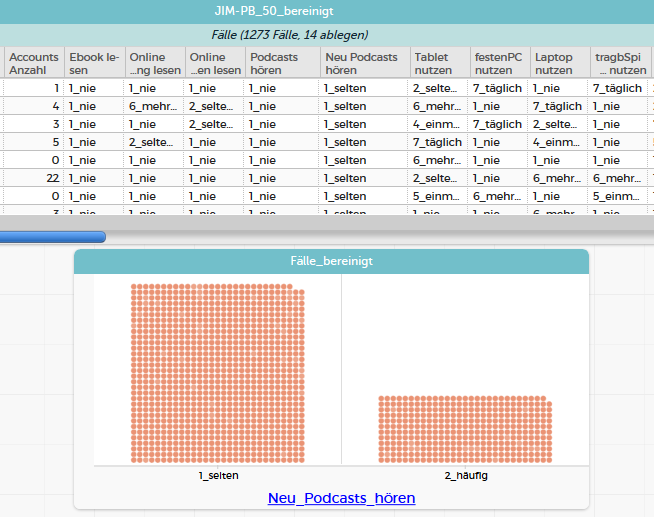
First, we drag a graph into the workspace:



We can then use drag & drop to select the variables that are relevant to our investigation and drag them onto the axes of the graph ("Click here or drag a variable here").

ATTENTION: The variable Podcast\_listen should have been recoded to a binary variable beforehand! See **instructions\_1**!

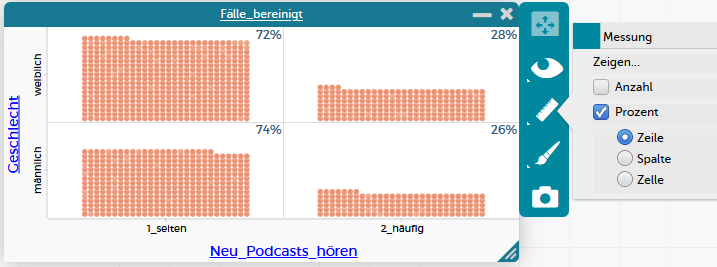
First, let's drag the variable "New\_listen\_to\_podcasts" onto the x-axis and the variable "Gender" onto the y-axis.



In the workbar, CODAP offers various evaluation options (e.g. percentages) to investigate the question.

There are different evaluation options with different percentage types. These are presented below. Each percentage type allows its own interpretations.

## Line percentages

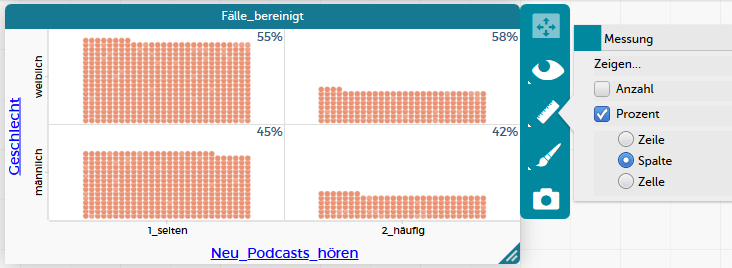


Line percentages are interpreted line by line. The reference group here is therefore the characteristics of the variable gender: female and male.

It makes sense to use line percentages to answer the question. It can be seen that in this sample, girls (top row) tend to listen to podcasts more often than boys. 28% of girls stated that they listen to podcasts frequently. This figure is only 26% for boys. However, the difference is actually very small.

If, on the other hand, column percentages are used, the perspective in the interpretation changes.

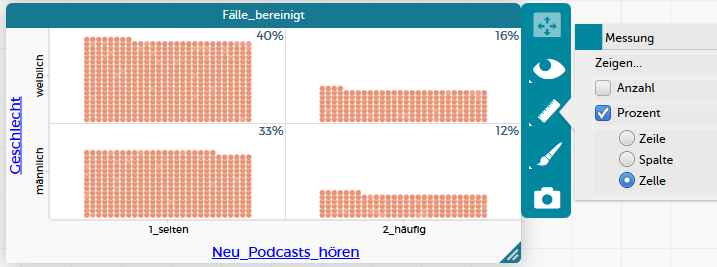
### Column percentages



Now the columns are in focus, an interpretation now takes place column by column.

The percentages now refer to the values of the variable New\_listen\_to\_podcasts. For example: Of those in this sample who state that they frequently listen to podcasts (column on the right), 58% are female and 42% are male. There are therefore clear differences between the genders among the "frequent listeners" of podcasts.

### Cell percentages



This can now be interpreted as follows:

Top left: 40% of all respondents in this sample are female AND state that they rarely listen to podcasts. This contrasts with, for example, of all respondents who are female AND listen to podcasts frequently. Also related to this are, for example, of all respondents who rarely listen to podcasts AND are male.

|  |
| --- |
| Summary percentages  **Row percentages** allow statements to be made in relation to a specific row, and thus to a specific characteristic of the variable that lies on the y-axis.  **Column percentages** can be used to make statements regarding the values of the variable on the x-axis.  **Cell percentages** allow statements to be made about all respondents. |