Re coding variables

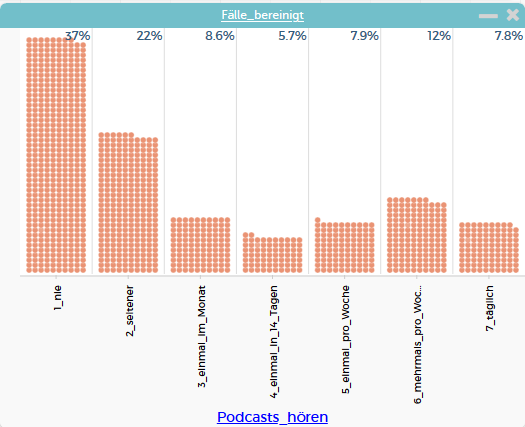
**Link to CODAP:** <https://tinyurl.com/you-pb-50en>

|  |
| --- |
| When **recoding variables,** the aim is to combine variables into groups and recode them in the data set in order to be able to make statements more quickly.  If you combine the values of a variable into two groups, you get a **binary variable**.  Different groupings of characteristics allow **different interpretations**! |

The recoding is worked out using the following question:

# How many respondents in the sample listen to podcasts frequently and how many respondents do so infrequently?

If you look at the variable podcast\_listening, you can see that there is a large proportion of people who stated that they never or rarely listen to podcasts.

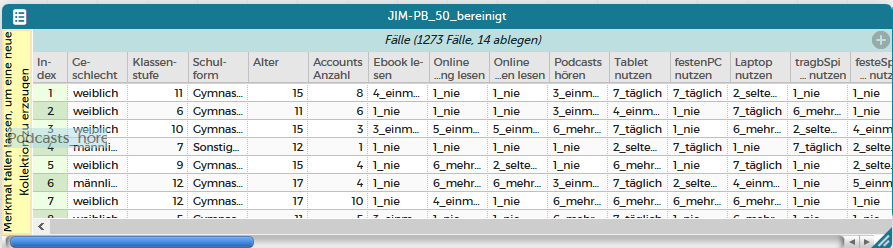


As data scientists, we now influence the evaluation by determining or modeling which characteristics mean that a person **rarely** or **frequently** listens to podcasts. Here we specify that

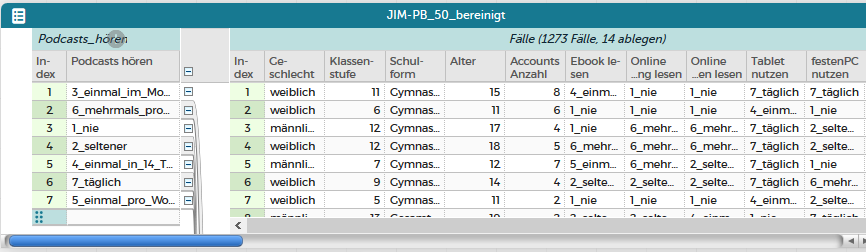
* **rarely** stands for *1\_never*, *2\_seldom*, *3\_once\_a\_month* and *4\_once\_in\_14\_days*,
* **often** stands for *5\_once\_per\_week*, *6\_several\_times\_per\_week*, *7\_daily*.

We now encode this in the data.

First, we drag and drop the variable "Podcast\_listen" to the far left of the table until a yellow field appears.

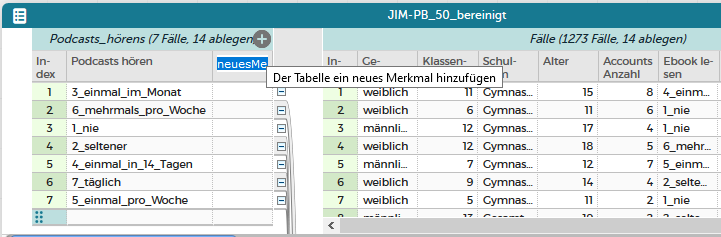


The variable Podcasts\_listening is now at a higher level in the table and the seven different values are displayed. We can see that variable 7 has different values between "1\_never" and "7\_daily".

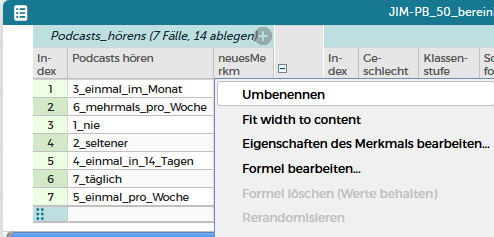


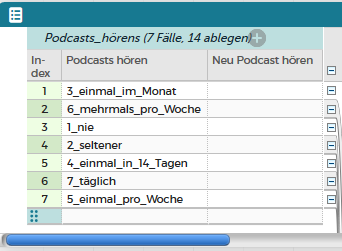
We now want to recode (or summarize) these in a new variable so that the new variable only contains the two characteristics **frequent** and **rare**.

To do this, we create a new variable by clicking on the plus on the left.

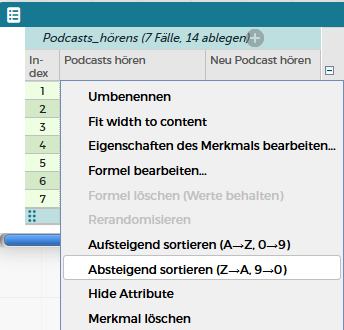


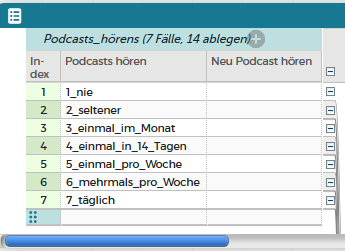
We now rename this new variable appropriately. This is done by left-clicking on "new\_feature" and then "rename". We select "New\_Listen\_to\_Podcasts" as the new variable name so that we know what the variable refers to.





The characteristics must now be recoded appropriately. To do this, we first sort the old characteristics by clicking on the characteristic name and selecting "Sort in ascending order".



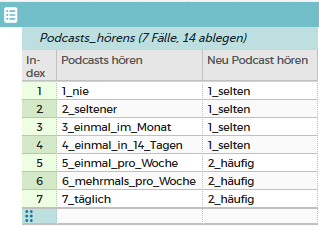


Now the values of the new variable "New\_listen\_to\_podcasts" can be summarized appropriately.

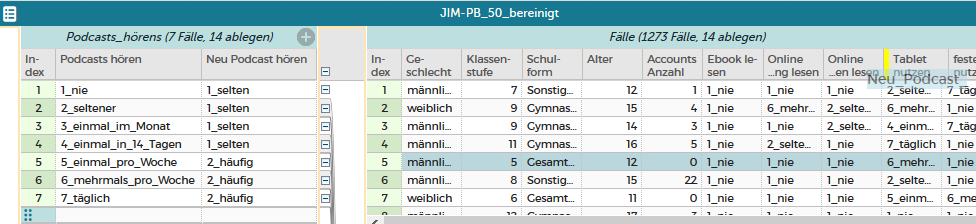
1\_never, 2\_seldom, 3\_once\_a\_month, 4\_once\_in\_14\_days 1\_seldom

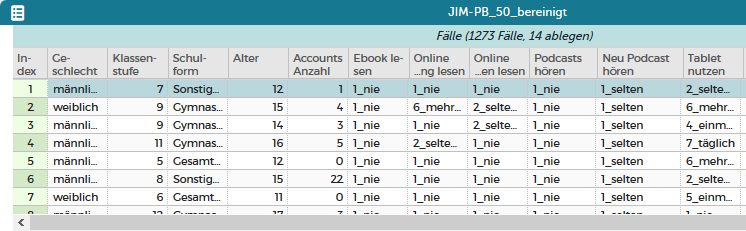
5\_once\_per\_week, 6\_several\_times\_per\_week, 7\_daily 2\_frequently

The corresponding characteristics are written into the table by hand.

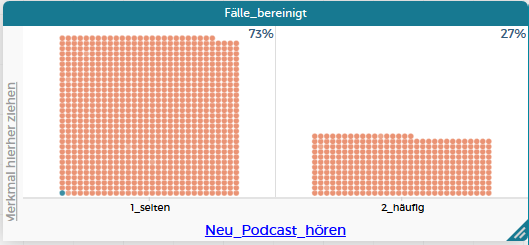


The two variables must then be dragged and dropped back into the right-hand part of the table.





Evaluation is now easier.

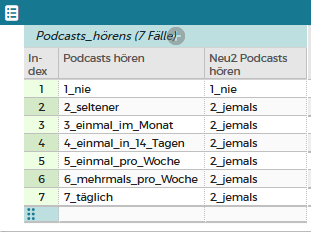
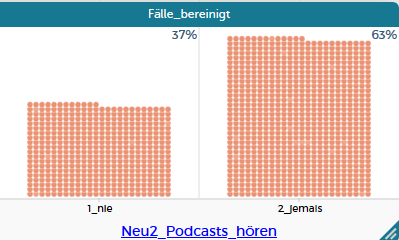


73% of respondents rarely listen to podcasts and 27% listen to podcasts frequently.

# Alternative modeling

When recoding, the data is modeled. This can be done in different ways. Alternatively, a distinction could have been made between respondents who **never** listen to podcasts as opposed to all others who stated that they **ever** listen to podcasts.

The modeling and thus recoding would then look like this and the evaluation:

The interpretation of this is as follows: 37% of respondents never listen to podcasts, whereas 63% of respondents state that they listen to podcasts at all.