BRM1_IBA_2021: Week 2

Fill in all student names and team number

11/3/2020

before you start, install your working environment

- check your working directory
- setwd() to change working directory, ?setwd to learn more
- start with an empty environment to prevent mistakes:
  - remove all objects from the environment
  - read in the raw data file

```
getwd()
```

```
## [1] "/Users/Demi/github/BRM117"
```

```
rm(list=ls()) # remove all objects
```

Assignment week 2

Continue with the data file of week 1

```
# read in data file
# you can copy the code of last week
```

1. *t*-test

Test whether financial satisfaction `Q50` is lower than 6, and report the *t*-value and *p*-value.

```
# t.test(df$Q50)
```

How would you label this variable measuring satisfaction? **Likert, semantic difference scale, paired comparison, rank order, continuous rating scale, or Stapel.**

2. Chisquare X^2 test

Analyze dependency between emancipation `Q31` and education recoded `Q276R`. Report the X^2 (=chisquare value), the p-value and interpret whether they are dependent or not.

```
# chisq.test(df$Q31, df$Q276R)
```

How would you label this variable measuring emancipation? **Likert, semantic difference scale, paired comparison, rank order, continuous rating scale, or Stapel.**

3. Validity

Explore all seven items that measure female emancipation (`Q28`,`Q29`,`Q30`,`Q31`, `Q32`, `Q33`, and `Q35`) per country. Use pairwise deletion. Inspect the correlation matrix, and argue whether or not there is a high convergent validity for each country.

Note: Country 1 is the country that comes first alphabetically.

```
# MatrixGRC<-cor(df[df$Country=="GRC",c("Q28","Q29","Q30","Q31","Q32","Q33","Q35")],use="pairwise.compl
# MatrixGRC
# MatrixPHL<-cor(df[df$Country=="PHL",c("Q28","Q29","Q30","Q31","Q32","Q33","Q35")],use="pairwise.compl
# MatrixPHL
```

Does each correlation depend on the same observations?

4. Reliability

Recoding

Recode the seven emancipation items so that a high number reflects a positive attitude towards female emancipation. Use indexing and what you've learned in the last assignment. Do NOT use `recode` from library `car`! Attach the recoded variables to your data with the addition `R` to their original variable names. Try to use a `for` loop, and mind the `NA`s. Do not use listwise deletion! Check your results with cross tabulation.

```
# summary(df$Q28)
# summary(df$Q29)
# summary(df$Q30)
# summary(df$Q31)
# summary(df$Q32)
# summary(df$Q33)
# summary(df$Q35)
# emancipation<-c("Q28","Q29","Q30","Q31","Q32","Q33","Q35")
```

Cronbach's Alpha

There are multiple formulas to calculate the Cronbach's alpha. One is defined as follows:

$$alpha = \frac{k}{k-1}\left(1 - \frac{\sum_i^k \sigma_{y_i}^2}{\sigma_x^2}\right)$$

where k is the number of items, σ_x^2 is the variance of the observed total test scores, and $\sigma_{y_i}^2$ is the variance of the ith item.

```
# first calculate number of items and save to object k
# c("Q28","Q29","Q30","Q31","Q32","Q33","Q35")

# k<-7
# k/(k-1)


# then calculate the variance for all items per item using var()
# to make it easier, we use listwise deletion (use="complete.obs")
# use diag() to get to the variances
# sum the variances using sum()
# varGRC<-var(df[,c("Q28","Q29","Q30","Q31","Q32","Q33","Q35")], use="complete.obs")
# diag(varGRC)
# sum(diag(varGRC))

# lastly, calculate the variance of total test score
# which is the sum of all covariances and variances

# fill in formula
```

The Spearman-Brown Prophecy Formula is easier to calculate, and referred to as the standardized Cronbach's alpha. The standardized Cronbach's alpha is defined as follows:

$$alpha_{st} = \frac{k * \overline{r}}{1 + (k-1) * \overline{r}}$$

Calculate the standardized Cronbach's alpha in the same way as is done with the Cronbach's alpha above but now use the correlation matrix (see point 3) instead of the variance-covariance matrix. It helps if you save the correlation matrix in an R object, and then retrieve the correlations with `lower.tri` using indexing. Do this separately per country, and use pairwise deletion.

```
# write your code here
```

You can check whether you did it correctly, using the `psych` package. Install the package first. I also had to download Rtools. Select the .exe file suitable for your operating system. Make sure you do this separately for each country!

```
# check with alpha() from psych package

# write your code here
```

5. Calculate mean scores

Calculate for each person in your data a mean score across the seven items mentioned above. Use `apply`, mind the missings. Attach this variable as a new variable labeled `emanci` to your data file. Report the absolute mean differences between these countries in `emanci` using `abs`.

```
# write your code here
```

6. Scatter plot

Plot the relationship of satisfaction financial situation (x-axis) and emancipation (y-axis). Make sure you use the mean scores calculated under 5). Create the plot separately for each country.

```
# write your code here
```

7. Create age dummy

Create a dummy variable equal to 1 if millennial (age 15 – 34) and 0 otherwise. To clarify: people aged 34 are labeled 1, and people aged 35 are labeled 0. Use the age variable `Q262`. Make sure missing cases in the original variable are also missing in the dummy variable. What percentage of the population is millennial in country 1?

```
# write your code here
```

Inspect the average value and standard deviation of emancipation item `Q31` per this dummy variable. Use `tapply`. Is the mean difference significantly different from zero (using a 95% confidence interval)? Use `t.test` on the whole dataset including both countries. *Note*: make sure the decimal is recognized as such by Excel (check worksheet named `OUTPUT`)

```
# write your code here
```