BRM1_IBA_2021: Week 2

team 117, Ashley Verdies, Merle, Demi Maric

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

```
df<-read.csv("GRC_PHL.csv")
str(df)
```

```
## 'data.frame':    2400 obs. of  39 variables:
##  $ Country : chr  "GRC" "GRC" "GRC" "GRC" ...
##  $ Q27     : int  1 2 1 2 3 1 1 1 2 NA ...
##  $ Q28     : int  3 3 3 3 3 2 3 3 4 3 ...
##  $ Q29     : int  4 4 4 4 4 4 NA 4 1 3 ...
##  $ Q30     : int  4 4 4 4 4 3 4 4 4 4 ...
##  $ Q31     : int  4 3 3 4 3 2 3 4 4 NA ...
##  $ Q32     : int  1 3 2 4 1 1 1 2 NA 2 ...
##  $ Q33     : int  2 5 2 5 5 2 5 5 5 5 ...
##  $ Q34     : int  5 2 1 1 1 1 5 1 5 1 ...
##  $ Q35     : int  1 4 3 5 1 1 3 NA 5 5 ...
##  $ Q36     : int  5 5 5 5 3 5 NA 5 1 5 ...
##  $ Q37     : int  1 2 1 3 5 1 1 1 1 5 ...
##  $ Q38     : int  1 2 1 1 3 1 1 1 NA 1 ...
##  $ Q39     : int  4 4 1 5 5 1 4 2 2 5 ...
##  $ Q40     : int  3 1 1 2 1 2 2 1 3 2 ...
```

```
##  $ Q41     : int  2 4 1 4 5 2 3 2 3 2 ...
##  $ Q46     : int  3 1 3 3 2 3 3 4 2 3 ...
##  $ Q47     : int  1 2 3 2 1 2 1 3 1 2 ...
##  $ Q48     : int  6 4 4 6 9 6 3 5 6 7 ...
##  $ Q49     : int  5 7 5 5 10 5 4 3 7 7 ...
##  $ Q50     : int  2 3 1 7 5 4 1 3 3 7 ...
##  $ Q51     : int  2 4 3 4 4 4 1 2 4 4 ...
##  $ Q260    : int  2 1 2 2 2 2 2 1 2 1 ...
##  $ Q261    : int  1984 1935 1939 1954 1968 1940 1989 1934 1989 1956 ...
##  $ Q262    : int  33 82 78 63 49 77 28 83 28 61 ...
##  $ Q273    : int  1 1 5 3 3 1 1 5 1 1 ...
##  $ Q274    : int  2 2 0 1 2 2 2 0 2 0 ...
##  $ Q275    : int  3 4 1 3 6 3 1 1 3 4 ...
##  $ Q275A   : int  300004 300005 300002 300004 300008 300004 300002 300002 300004 300005 ...
##  $ Q275R   : int  2 3 1 2 4 2 1 1 2 3 ...
##  $ Q276    : int  3 3 1 NA 6 6 3 1 3 3 ...
##  $ Q276A   : int  300004 300004 300002 NA 300008 300008 300004 300002 300004 300004 ...
##  $ Q276R   : int  2 2 1 NA 4 4 2 1 2 2 ...
##  $ Q286    : int  4 2 3 2 2 4 4 3 3 2 ...
##  $ Q287    : int  5 4 5 3 2 4 5 5 4 3 ...
##  $ Q288    : int  1 4 3 6 6 4 2 3 2 5 ...
##  $ Q288R   : int  1 2 1 2 2 2 1 1 1 2 ...
##  $ Q289    : int  3 3 3 3 3 3 3 3 3 3 ...
##  $ W_WEIGHT: num  0.991 0.707 1.108 0.707 0.939 ...
```

1. t-test

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

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

```
##
##  One Sample t-test
##
## data:  df$Q50
## t = 111.97, df = 2391, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  5.570847 5.769454
## sample estimates:
## mean of x
##  5.670151
```

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.

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

```
##
##  Pearson's Chi-squared test
##
## data:  df$Q31 and df$Q276R
## X-squared = 56.932, df = 9, p-value = 5.207e-09
```

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.

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

```
##            Q28        Q29       Q30        Q31         Q32        Q33
## Q28 1.00000000 0.24626511 0.1402132 0.25149977  0.05732369 0.20138827
## Q29 0.24626511 1.00000000 0.3636181 0.62353359  0.03318167 0.32532147
## Q30 0.14021322 0.36361809 1.0000000 0.42935697  0.11339243 0.26453654
## Q31 0.25149977 0.62353359 0.4293570 1.00000000  0.08816039 0.32943031
## Q32 0.05732369 0.03318167 0.1133924 0.08816039  1.00000000 0.08595674
## Q33 0.20138827 0.32532147 0.2645365 0.32943031  0.08595674 1.00000000
## Q35 0.29802895 0.19287645 0.1956279 0.26915751 -0.05775870 0.29522549
##            Q35
## Q28  0.2980290
## Q29  0.1928765
## Q30  0.1956279
## Q31  0.2691575
## Q32 -0.0577587
## Q33  0.2952255
## Q35  1.0000000
```

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

```
##           Q28        Q29       Q30       Q31        Q32        Q33        Q35
## Q28 1.0000000 0.17243394 0.1542285 0.1593451 0.15096249 0.14525836 0.10307829
## Q29 0.1724339 1.00000000 0.3763962 0.3672903 0.08108371 0.25843832 0.09649637
## Q30 0.1542285 0.37639622 1.0000000 0.4794227 0.13893919 0.19088408 0.16035313
## Q31 0.1593451 0.36729034 0.4794227 1.0000000 0.17519142 0.18181777 0.12013301
## Q32 0.1509625 0.08108371 0.1389392 0.1751914 1.00000000 0.07870677 0.06668170
## Q33 0.1452584 0.25843832 0.1908841 0.1818178 0.07870677 1.00000000 0.06751470
## Q35 0.1030783 0.09649637 0.1603531 0.1201330 0.06668170 0.06751470 1.00000000
```

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)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   1.000   2.000   3.000   2.558   3.000   4.000      21
```

```
summary(df$Q29)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   1.000   2.000   3.000   2.618   3.000   4.000      45
```

```
summary(df$Q30)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   1.000   2.000   3.000   2.932   4.000   4.000      20
```

```
summary(df$Q31)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   1.000   2.000   3.000   2.736   3.000   4.000      46
```

```
summary(df$Q32)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   1.000   2.000   2.000   2.093   3.000   4.000      25
```

```
summary(df$Q33)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##    1.00    2.00    2.00    2.68    4.00    5.00       7
```

```
summary(df$Q35)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   1.000   2.000   3.000   3.235   4.000   5.000      16
```

```
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")
```

```
## [1] "Q28" "Q29" "Q30" "Q31" "Q32" "Q33" "Q35"
```

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

```
## [1] 1.166667
```

```
# 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)
```

```
##       Q28       Q29       Q30       Q31       Q32       Q33       Q35
## 0.6495240 0.7475650 0.7192136 0.6531654 0.5942064 1.5333503 1.0737125
```

```
sum(diag(varGRC))
```

```
## [1] 5.970737
```

```
# 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.

```r
# 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!

```r
# 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`.

```r
# 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.

```r
# 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?

```r
# 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`)

```r
# write your code here
```