Exam BST02

March 1st 2019 14:15-17:00

This exam consists of three questions of increasing difficulty. You need two data files: AEGrades.csv and patinfo.sav, that you can download from Canvas. Each question is worth the same number of points and within each question all subquestions are worth the same number of points.

The file patinfo.sav contains the patient characteristics of the patients that participated in a clinical trial. The file AEGrades.csv contains information on the System Organ Class (SOC) and grade of the adverse events (AE) the patients had. Note that a patient can either have no events, a single event or also multiple adverse events.

If you are finished with the exam. Submit your *code* and the *output* it generates using Canvas.

Good luck!

Question 1

Import the file patinfo.sav to R. Write an R script to answer the following questions:

(a) How many patients do we have in each arm?
(b) What are the mean and the standard deviation of age of the patients in the sample as a whole?
(c) What are the mean age and heart rate (HR) for men and for women?
(d) Perform a t-test to compare the age between men and women.

Question 2

We want to investigate the smooth relation between age and heart rate.

(a) Write a function with the name `f` that takes two parameters: `x` and `data`. This functions should calculate the median heart rate (HR variable) for all patients in `data` for which x-5 <= age < x+5.

(b) Now make a scatter plot with age on the x-axis and `f(age, dat)` on the y-axis where `dat` is the imported data from `patinfo.sav`. Use the labels "Age" for the x-axis and "Median heart rate" for the y axis.

Question 3

Now also import the file `AEGrades.csv`. This file contains all adverse events for the patients (note some patients did not get an adverse event). For the following questions we want to count each patient only once (even if there were multiple adverse events), i.e. we are counting patients with adverse events not the adverse events themselves. For a patient with multiple adverse events we only count the one with the highest grade. So if a patient had three adverse events: one cardiac disorder of grade 1, one cardiac disorder of grade 2 and a nervous system disorder of grade 3 we would count this as 1 event of grade 2 when looking at the cardiac disorder system organ class and as grade 3 when looking at all adverse events.

Keeping in mind that patients should only be counted once when focusing either on one System Organ Class (e,g, cardiac disorder) or all System Organ Classes:

(a) Create a barchart for each of the two arms (hint 1: for this you will have to merge this data set with `patinfo.sav`) with the information in `AEGrades.csv`. Here we use adverse events of all System Organ Classes together.
Hint: The data that you plot should look like the following table:

```
##                           Arm
## No. Patients with AE gade Control Active
##                         1       8      9
##                         2       6      3
##                         3       2      3
##                         4       3      1
```

However you should extract these numbers from the data set not from the table above (so only use it if you cannot continue otherwise!)

(b) Create tables for the adverse events grades and their frequencies in the two arms for each System Organ Class separately. Hint: So for 'Blood and lymphatic system disorders' the results should be:

```
##         Arm
## AE gade Control Active
##       1       6      8
##       2       2      0
##       3       0      1
##       4       2      0
```

(c) Write a function with the name `logFit` and parameter `y`. The function should perform logistic regression were the outcome per patient is 1 if the patient has any adverse event of grade y or higher and 0 if the patient has no adverse events of grade y or higher. The arm of the trial should be the predictor (hint: for this you will have to merge the `AEGrades.csv` data set with `patinfo.sav` again). As output the function should give the coefficient of the log odds ratio, the standard error and the p-value. Call the function for all AE grades in a for loop (for y from 1 until 4) and present the results in a `data.frame` where rows indicate the model (grade) and columns the statistic. The row names should be "grade 1", "grade 2", "grade 3", "grade 4" and the row names should be "log OR", "SE", "p-value". (Note that here we combine the adverse events of all SOCs again.)