Exam: Introduction to R

You have 45 minutes to solve the exam on your own. You may use your laptop computer and any resources on R you find on the internet. If one of your computers breaks unexpectedly, remember that the teacher's computer in the front row has Rstudio installed, too.

Please send your answers before 10:15 to `oliver@kirchkamp.de`

Please send only the commands, *not* the output of your commands!

Thank you for your active participation in the course.

Question 1: Have another look at the dataset `Wages` from `Ecdat`.

1. How would you calculate the relative proportion (in percentages) of female workers. The output of your command might look like this:

   ```
   [1] 11.2605
   ```

2. How do you calculate this proportion separately for workers in the south and in the north? The output of your command might look like this:

   ```
     south        x
   1    no 10.21651
   2   yes 13.81307
   ```

3. How do you calculate this proportion for all workers with an `lwage` larger than 6.5? The output of your command might look like this:

   ```
     south        x
   1    no 4.820416
   2   yes 4.689864
   ```

Question 2:

1. You use the dataset `DoctorAUS` from `Ecdat`. What is the command to run a possion regression that explains the number of consultations with a doctor (`doctorco`) as a function of type of the insurance (`insurance`)?

2. Based on the results of this regression, which insurance category do you expect to have most consultations?

3. You want to compare the insurance category where you have the largest number of consultations with the category where you have the second largest number of consultations. How can you test whether these two coefficients are significantly different? (Have a look at your estimation results to find out what these two coefficients are.)

4. You suspect that the variable `actdays` in your poisson regression has an important influence. How would you include this variable in your regression? How can you see whether the impact is, indeed, significant?

5. Now you suspect that the variable `actdays` has a non-linear impact in the poisson model. You want to include a dummy for each level of this variable. How can you do this?

1

6. What can you say about the impact of the different insurance contracts. Which is now, according to your estimation results, the contract with the largest number of consultations?

Question 3:

1. You again use the dataset `Wages`. Use the command `aggregate` to construct a data frame with three columns: In one column you have `ed`, in the second column you have `sex`, and in the third column you have for each possible combination of `ed` and `sex` the median of `lwage` for this combination. Since there are only 14 levels of `ed` and only two levels of `sex`, your new data frame will have at most 28 rows.

   Which commands do you need to construct this data frame?

2. How many rows does your data frame actually have? Why does it have fewer than 28 rows?

3. Now create from the previous data frame a data frame which you call `d1` and which contains only the observations for females. Create another data frame which you call `d2` which contains only the observations for males. Use `merge` to join these two data frames so that now you have only one row for each level of `ed` and two columns with the wage, one with the wage for female and one with the wage for male.

   The output of your command might look like this:

   ```
      ed lwage.female lwage.male
   1   4           NA   6.063775
   ```

		lwage.female	lwage.male
2	5	NA	6.188260
3	6	NA	6.409275
4	7	5.817105	6.343880
5	8	5.643680	6.551080
6	9	5.393630	6.559615
7	10	5.945420	6.586170
8	11	6.046170	6.620070
9	12	6.233430	6.668230
10	13	6.555145	6.614050
11	14	6.163310	6.841615
12	15	6.700730	6.905750
13	16	6.476970	6.905750
14	17	6.856460	6.907760

Remark: A quicker way to solve this problem would involve `reshape`, but we did not find the time to cover `reshape` in the course.

Question 4:

1. Write a function `f` that takes a matrix as an argument and that returns a matrix of the same size as the argument but where all elements are the reciprocals of the elements in the original matrix. E.g. if x is as follows:

   ```
        [,1] [,2] [,3]
   [1,]    1    2    3
   [2,]    3    4    5
   ```

   then `f(x)` should return the following:

   ```
            [,1] [,2]      [,3]
   [1,] 1.0000000 0.50 0.3333333
   [2,] 0.3333333 0.25 0.2000000
   ```

2. Now modify the previous function so that it checks whether it is really called with a matrix as an argument. In this case it should work as above. In all other cases it should return `NA`.