Homework 1

Alla Tambovtseva

Deadline: 18 November, 23:59

Submission

Part 1: answers should be submitted via Google forms:

https://goo.gl/forms/UFm6va6d9zZZJQ1J2.

Part 2: the `.R` file with the code should be uploaded via Dropbox:

https://www.dropbox.com/request/UmAaPoe65QXYIyMZ5ada.

The first line in the file should contain a comment with your name like this:

```
### Alla Tambovtseva
```

Do not forget to add comments stating the number of the problem you solve, especially if you skip some problems or change their order in your file.

Part 1

Read the pp.15-23 from *OpenIntro Statistics* and answer the following questions. To submit your answers, fill in the Google form.

1. What is anecdotal evidence? Provide your own examples of anecdotal evidence (2-3 examples) and explain why these cases cannot be used to represent the population correctly.
2. Provide the definition of an explanatory and a response variable.
3. Formulate the difference between a prospective study and a retrospective study.
4. Formulate advantages and disadvantages of a) stratified sampling; b) cluster sampling.

Part 2

Problem 1 (2 points)

Calculate using R:

- $log_2(25)$
- $\pi^2/2$
- $\sqrt{17 \cdot 26}$
- $3^5/2^6 \cdot e^{-3}$

Problem 2 (4 points)

1. Create the variable `x` and assign it the value “23,7”.
2. Make sure `x` has the character type. Check it using the R code.
3. Convert `x` to a floating point number, to a numeric type (you might need a replacement operation first).

Problem 3 (2 points)

Two events A and B are called independent if the following condition holds true: $P(A \cap B) = P(A) \cdot P(B)$. The probabilities of A, B and $A \cap B$ are defined:

```
pA <- 0.8 # P(A)
pB <- 0.6 # P(B)
pAB <- 0.55 # P(A and B), intersection
```

Write one line of code that helps to check whether events A and B are independent. Note that your code should work correctly even if we change the values of `pA`, `pB`, `pAB`.

Problem 4 (6 points)

The number of valid votes at several electoral precincts within a particular district is saved in the vector `valid`, the number of invalid votes - in the vector `invalid`. The total number of registered voters is stored in the vector `voters`.

```
valid <- c(25, 8, 12, 18, 25, 32, 10, 17, 22, 17)
invalid <- c(1, 0, 4, 3, 9, 2, 0, 5, 1, 0)
voters <- c(50, 75, 62, 54, 98, 55, 72, 80, 44, 48)
```

1. Create the vector `turnout` with the turnout at these electoral precincts (turnout is computed as the sum of valid and invalid votes).
2. Create the vector `turnout_perc` with the turnout rate (in %). The turnout rate is calculated as a share of all votes (both valid and invalid) of the total number of registered voters multipled by 100.
3. Round the elements of a vector `turnout_perc` to the first decimal place and save the results as the vector `tround`.

Problem 5 (6 points)

Study the extended lecture on vectors in R and solve the problem 5.

There is the vector `age` that contains respondents' age:

```
age <- c(19, 18, 23, 25, 20, 32, 30, 22, 36, 45)
```

1. Create the vector `age2` that consists of respondents' age squared.
2. How many elements are in the vector `age2`? Write the code that helps you to answer this question automatically.
3. Choose elements of `age2` that are greater that 400. Save these elements to the vector `g`.
4. Choose elements of `age2` that are greater or equal 484, but less than 900 at the same time. Save them to the vector `m`.
5. Print the *indices* of elements of `age2` that are divisible by 5 (can be divided by 5 with no remainder).