

# Homework 3

*Alla Tamboutseva*

*Deadline: 04 February, 23:59*

## Submission

**Part 1:** answers should be submitted via Google forms: <https://goo.gl/forms/8qePo9VJBSLeC68M2>.

**Part 2:** the pdf-file with answers, code and graphs should be uploaded via Dropbox: <https://www.dropbox.com/request/TuZmOD7qGdZzXUCVTLwu>.

The pdf-file should not be necessarily created in RStudio, it can be a Word/Open Office file converted to pdf. You can create a text file, write answers, copy your R code & outputs and add graphs (export pictures via Export in Plots tab in the right bottom corner in RStudio) and save as pdf.

An alternative way, more convenient, but more demanding: create an Rmd-file with texts and code chunks, then knit it to Word via *Knit* button in RStudio and export to pdf.

## Part 1

### Problem 1

One Catalan polling agency conducted a research dedicated to the Catalan independence. They organized an opinion poll: 1200 respondents were asked whether they approve the independence of Catalonia from Spain. Their analytical report included the following result: 95% confidence interval for the percentage of people voting for the independence of Catalonia is (75%;80%).

A student decided to interpret this result. He wrote:

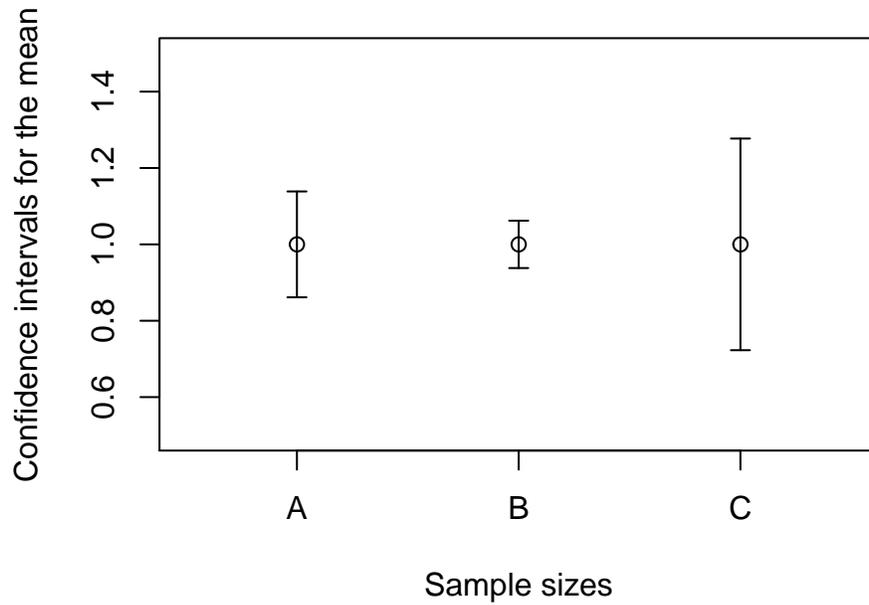
*(1) The percentage of all people supporting the independence of Catalonia lies between 75% and 80%. (2) To be more precise, there is 0.95 probability that the true percentage of people who approve the Catalan independence is between 75% and 80%.*

1.1. This interpretation includes incorrect statements. Find them and explain why they are incorrect. Please, put the number of the wrong statement (1-2) and then write your explanation.

1.2. Provide your own interpretation of the 95% confidence interval described above.

### Problem 2

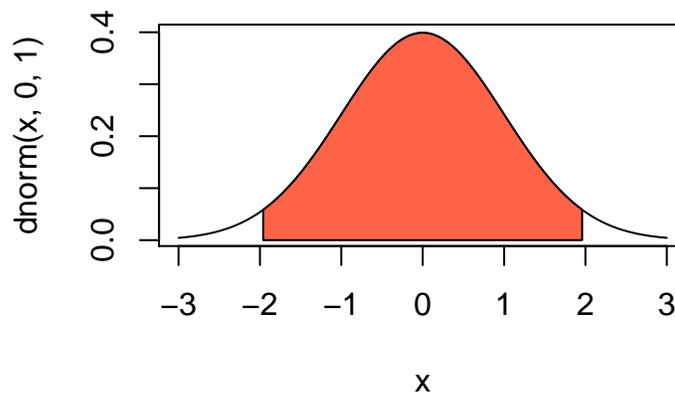
The figures A-C show the 95% confidence intervals for the mean value of the same population. The estimation of the mean was conducted on the samples of different sizes. Match each figure with the sample size  $n$  ( $n = 50$ ,  $n = 200$ ,  $n = 1000$ ) and explain your answer.



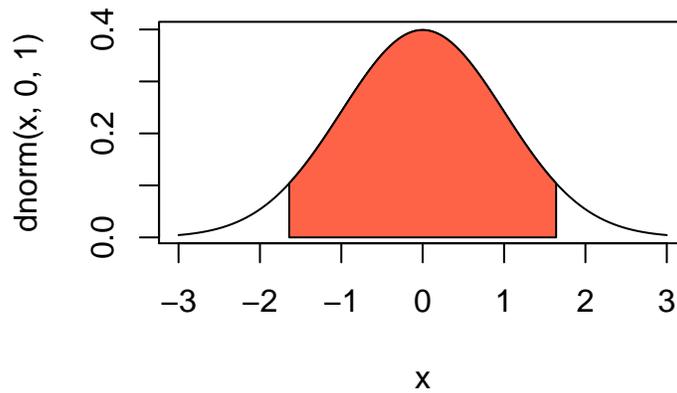
**Problem 3**

The figures A-C show the same normal distribution with the different confidence intervals for the mean marked (shaded area). Decide which figure corresponds to each of the following confidence intervals: 90%, 95%, 99%. Explain your choice.

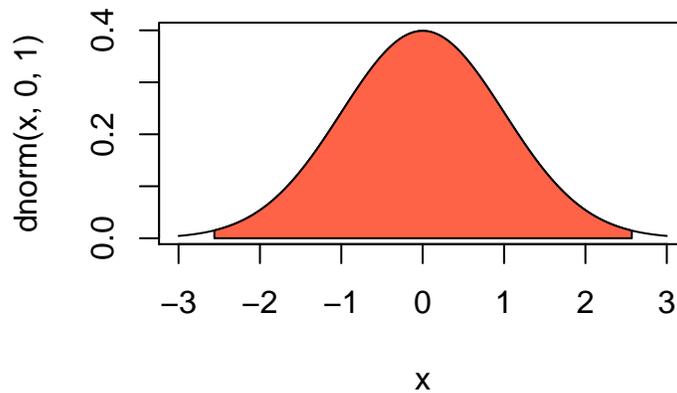
**Figure A**



**Figure B**



**Figure C**



**Problem 4**

One polling agency X organized a survey so as to evaluate citizens' attitude towards the smoking ban policy. If this survey was repeated 100 times by different polling organizations, in 90 cases the sample proportion of people favoring this policy would lie between 0.44 and 0.49, and in 10 cases it would lie outside this range. Express the same idea using statistical concept(s) your know.

**Problem 5**

Someone calculated a 95% confidence interval for the proportion of universities that has a special page dedicated to anti-corruption measures on the official website. Here is the R output (values are rounded):

```
BinomCI(45, 125, conf.level = 0.95)
      est  lwr.ci  upr.ci
[1,] 0.36   0.28   0.45
```

- Find the length of a confidence interval.
- Find the length of the confidence interval if we increase the sample size twice (all else equal)?

## Part 2

### Problem 1

1. Choose one of the following distributions:
  - Student's distribution. It has one parameter in R: `df`. Use `rt(n, df)` to generate a random sample of size `n`.
  - Beta distribution. It has two parameters in R: `shape1` and `shape2`. Use `rbeta(n, shape1, shape2)` to generate a random sample of size `n`.
2. Set a fixed seed so as the grader could reproduce your results, write this line before providing code for this task:

```
set.seed(1234)
```

Generate a random sample of size 10000 from the distribution you have chosen (set any values of parameters you want). It will be your population. Plot a histogram of this population. Does this distribution look like a normal one?

3. Set a fixed seed one more time so that there would be no need to run all your code at once with the previous seed:

```
set.seed(1234)
```

Take 1000 samples of size 900, compute their means and arrange them in a vector. Plot a histogram of this vector. Does this distribution look like a normal one?

4. Compute the expected value and the standard deviation of the distribution obtained in point 3 (use formulas above and the intuition from statistical laws we studied). Compare them with the mean and the standard deviation of a sample obtained in point 3 (calculate sample statistics using R). Do they resemble ones expected from theory? Make your conclusions.

### Problem 2

A social science student conducted a survey so as to decide how many people at his faculty supports right-wing parties. He took a random sample of 80 people and it turned out that 15 students supported right-wing parties. Use R to get a 95% confidence interval for the proportion of right-wing's supporters. Provide your R code and interpret the results you got.

**Note:** make your interpretation as full as you can.

### Problem 3

Imagine that you conduct a research on the parliamentary representation in presidential and parliamentary European states. As a part of your research you need to decide whether the average number of seats held by all opposition parties differs in countries with presidential system and in countries with parliamentary system.

1. Upload the data set using this [link](#). The file contains data on European countries in 1990–2015. It is a subset of the large Database of Political Institutions (DPI). Your variables of interest are `totalseats` (total number of seats), `numopp` (number of seats held by opposition parties) and `system` (0 - presidential, 1 - parliamentary).
2. Create a column that contains the percentage of seats held by opposition parties and call it `opp_perc`.
3. Choose rows that correspond to presidential states and save them to `dat1`. It will be needed for your first sample.

4. Choose rows that correspond to parliamentary states and save them to `dat2`. It will be needed for your second sample.
5. Compute the 95% confidence interval for the mean percentage of seats held by opposition parties in presidential states. Interpret the confidence interval obtained.
6. Compute the 95% confidence interval for the mean percentage of seats held by opposition parties in parliamentary states. Interpret the confidence interval obtained.
7. Based on the results in previous points, can you conclude that the mean percentage of seats held by opposition parties does not differ in parliamentary and presidential states?