ADS2 Practical 8: Simulation-based comparison of two means

Facilitators: Gedi Lukšys, Chaochen Wang

Mon, November 4, 2019, 14-16

Learning objectives:

- Understand the logic behind hypothesis tests using simulations
- Learn how to compare mean of a sample from a distribution to a value
- Learn how to compare means of two samples from two distributions
- Appreciate the advantages of simulation-based approach

The task:

Imagine that a country runs an exam system using normative scoring, where each student first gets a raw score based on performance in the exam, and later student's normative score is computed as corresponding to the percentile of all students, hence it can be effectively approximated using a uniform distribution between 0 and 100 (for the sake of simplicity let's assume all scores are not rounded).

Question 1

Let's assume that our class has 26 students whose score distribution follows the same one as in the whole country. We want to know the probability that the mean of their normative scores is lower than 40. How can we find this out using simulation?

For uniform distribution you can use function *runif* (use *?* to figure out its parameters)
Once you get your result (probability of the mean being below 40) try to explain it.
Is it sensitive to number of repetitions in your simulation?

You can plot the distribution of means.
Does it look similar to the uniform distribution? If not, why not?
How would it look if the class only had 5 students?

Question 2

Now imagine that one class of 26 students had a careless administrator who didn't notice the 5th exam question and only printed the first four. Let's **assume** that as a result of this, the normative scores of students from this class followed a uniform distribution between 0 and 80 (with the mean of 40). We now want to know what is the probability that this class did better than another class of 26 that didn't have such bad luck. How can we do this using simulation?

Is the result the same as in the previous question? If not, why not? You can also plot both distributions of means (ideally on one figure).

Question 3

As this exam had a clear scoring system and students could take a copy of their papers, a student Leonie found that she got a 64. Based on statistics from previous years of students taking a similar exam, she has found out that their raw scores follow a normal distribution with a mean of 50 and standard deviation of 10 (for which you can use function *rnorm*).

What would be the expected normative score Leonie likely got? Why?

Leonie and her three friends got the following raw scores: 64, 63, 62, 59. Her friend Sheldon is a very bright student from another class. He and his friends got the following raw marks: 70, 63, 61, 56. As a result, Sheldon is boastful that their average is higher. However, having a good understanding of data science Leonie thinks that her team will have the last laugh once the normative scores are out. Is she right? Why or why not? Use simulations and plots to support your argument.

Question 4

Now remember the unlucky class for which one exam question was not printed. What is wrong with the assumption in bold in question 2? Let's find and plot a distribution of normative (percentile-based) scores for this class if 1 out of 5 answers are missing, hence their raw scores are on average 20% lower but follow the same distribution (normal with mean = 40 and standard deviation = 8).

How does this distribution look? Why is it such? What is its mean value?

What is the probability that this class got a higher mean normative score than another class of 26 that didn't have bad luck of missing one problem?

Does the school's principal have a valid reason to worry that the unlucky class would be the worst in the country, assuming it has about 10000 classes of the same size?

Use simulations to answer these questions.

Now assume the number of students per class actually varies, following a uniform distribution between 5 and 40 (with 10000 classes in total). Would that alleviate the principal's worries of having a class with the worst normative score average in the country? Why or why not?