ADS2 Practical 8 + problem set: Simulation-based comparison of two means

Instructors: GL, DS, WZ, JC

November 9, 2023

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 in the country, 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). So, for example, if raw scores of 10 students were 17, 17.5, 16, 16.4, 18.9, 18.3, 18.6, 20, 15.5, 18.1, their normative scores would be 35, 45, 15, 25, 85, 65, 75, 95, 5, 55 respectively.

Question 1

Let's assume that a 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 5^{th} 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?

Question 5 (homework)

In reality it is difficult to expect that student **raw scores** from different schools and classes will follow the same distribution even if the vast majority of classes didn't have bad luck of missing out one question in five. We expect that due to differences in teaching quality and learning environment (some schools are better, others worse) **the means of each class** follow a normal distribution with mean = 50 and standard deviation = 5 whereas standard deviation is stable across the country (and is equal to 10, except for the unlucky class, for which it's 8). Class sizes also vary as in *question 4*, following a uniform distribution between 5 and 40, with 10000 classes in total.

With these changes generate your new student **raw score** distribution in the country and based on this recalculate a distribution of **normative scores** for the unlucky class (*as you did in question 4*). How do both distributions look and what are their means?

What is the probability that the unlucky class did better than another class in the country that didn't have such bad luck?

How does the average of Leonie's team's normative scores now compare to that of Sheldon's team (as in *question 3*)

Use simulations to obtain these values and inferences (this time *pnorm* may not be of much help, as the overall distribution of raw scores in the country no longer follows the normal distribution with mean = 50 and s.d. = 10 – only that of each class does, albeit with different means that are themselves normally distributed)