Case

Noelia, a chief psychometrician at Educational Testing Service, is responsible for analyzing SAT scores. She selected a sample data set of 400 students from the population. According to the preliminary analyses, the sample has a mean of 1000 and a standard deviation of 200.

- What is the typical distance that a sample mean of SAT scores deviates from the population mean?

The typical distance that a sample mean deviates from the population mean is the meaning of the standard error. The standard error is calculated by dividing the standard deviation of the sample by the square root of the sample size:

$$s_{\bar{X}} = \frac{s_X}{\sqrt{n}} = \frac{200}{\sqrt{400}} = \frac{200}{20} = 10$$

Therefore, the answer is 10.

- What is the probability that the one sample of 400 students randomly drawn has a mean SAT score of 1010 or higher?

Note that we are interested in the distribution of the sample mean. That is, we would like to know the percentage of the randomly drawn sample mean that is 1010 or higher. Therefore, our focus is to use the sampling distribution of the sample mean.

There are two ways to approach this question. First, we can use the idea of the confidence interval (according to the lecture note). Second, we can use the idea of the normal distribution (do you remember that the sampling distribution of the mean follows the normal distribution?). Either way leads to the same solution, so choose one method that corresponds to your taste!

Let me solve this question in the first way by using the confidence interval. As the lecture slide does (see Class Exercise in Lecture 6), we equate the upper limit of the confidence interval with 1010 (note that we leave a z-score as a letter to solve for that). Therefore, what you need to do is:

$$\bar{X} + z \times \frac{200}{\sqrt{400}} = 1000 + 10z = 1010$$

If you solve the above equation, it becomes:

$$1000 + 10z = 1010$$

$$10z = 10$$

$$z = 1$$

If you read the z-table, the percentile of the z-score of 1 is 0.8413. This means that 84.13% of the sample means fall below the sample mean of 1010. Since we are interested in the probability of obtaining a sample mean that is 1010 or higher, we subtract 0.8413 from 1, which becomes 0.1587. Therefore, 15.87% is the probability that the one sample of 400 students randomly drawn has a mean SAT score of 1010 or higher.

Let me solve this question again in the second way (which is I personally prefer). We know, from the central limit theorem, that the sampling distribution of the mean is always normally distributed regardless of the shape of the population distribution. Therefore, we calculate the z-score of 1010 and read the z-table (You can always calculate the z-score once you know the mean and the standard deviation only when the distribution is normally distributed). Note that the mean of the sampling distribution of the mean is 1000. Also, note that the standard deviation of the sampling distribution of the mean is 10 (Can you see that this 10 is equal to the standard error?) Consequently, if I calculate the z-score of 1010, it becomes:

$$z = \frac{1010 - 1000}{10} = \frac{10}{10} = 1$$

If you read the z-table, the percentile of the z-score of 1 is 0.8413. This means that 84.13% of the sample means fall below the sample mean of 1010. Since we are interested in the probability of obtaining a sample mean that is 1010 or higher, we subtract 0.8413 from 1, which becomes 0.1587.

Therefore, 15.87% is the probability that the one sample of 400 students randomly drawn has a mean SAT score of 1010 or higher.

- If Noelia calculates the 95% confidence interval of the mean, what would be?

The 95% confidence interval of the mean is calculated as:

$$\bar{X} \pm 1.96 \times \frac{s_X}{\sqrt{n}} = 1000 \pm 1.96 \times \frac{200}{\sqrt{400}} = 1000 \pm 19.6 = [980.4, 1019.6]$$

- Say Joel, a statistician in the U.S. Department of Education, visited Noelia and received a sample of 625 students to do some research.
  - Between Noelia and Joel, who do you think will have a sampling distribution of the mean that is less dispersed? Why?

Joel's sample size is bigger than Noelia's sample size. Recall that the standard error depends on the standard deviation and the sample size. The sample size is in the denominator, so a bigger sample size leads to a smaller standard error. Therefore, Joel will have a sampling distribution of the mean that is less dispersed.