Questions Module 2.2

Select the correct order of the names for the following symbols:

- *a.* population standard deviation, sample mean, sample standard deviation, population mean
- *b.* sample standard deviation, sample mean, population mean, population standard deviation
- *c.* population mean, population standard deviation, sample mean, sample standard deviation
- *d.* population standard deviation, sample standard deviation, sample mean, population mean

---

Incorrect.
The correct answer is **c**.

Let's say you are studying a coating process, and that coating thickness is a critical characteristic. The customer has tight specifications for thickness and is complaining that many parts are out of spec. You want to estimate coating thickness and also study potential causes of variation in thickness. Data are automatically collected on thickness and a variety of characteristics. You use data that were collected during the previous day.

1. Why is this considered convenience sampling?
2. What are some potential issues with using these data to estimate coating thickness?
3. What are some potential issues with using these data to study potential causes of variation in coating thickness?

Answers (there are many possible answers):

1. You are using the most readily available data rather than randomly selecting data from the target population. The data from this one day might not represent the population, which is all parts produced with the given thickness specifications.
2. The estimate of thickness from one day might not reflect the average thickness for the population. There might have been a problem with the process on that particular day, or there might be a lot of variability in thickness from one day to another. You can use statistical process control charts to determine whether the average thickness is consistent, or stable, over time. There also might be an issue with the measurement system for thickness. You learn about control charts and measurement system studies in the Quality Methods module.
3. You have the same issue as in #2. Here's another potential issue. You have data on a variety of characteristics, but you might not have data on all of the characteristics that might be causing the problem. When you use only available data, you might not have data on the root cause or causes of the problem. Without systematically thinking about the process and potential root causes of the problem, you might miss an important variable.

If your data are normally distributed, approximately what percent of the observations fall within two standard deviations of the mean?

- *a.* 99.73%
- *b.* 95%

- *c.* 68%
- *d.* 5%

Incorrect.
The correct answer is **b**. Approximately 95.48% of the observations fall within two standard deviations of the mean if the data are normally distributed.