10.6: Distribution of Sample Means (4 of 4)

Learning Objectives

- We need to be careful before using the normal model to find probabilities associated with sample means.
  - Note: The logic of inference in this module is familiar. We make a claim about a population mean. We use a random sample to test our claim. We determine whether it is probable that random samples have means as extreme as the actual sample. If this is very unlikely, then we conclude this sample probably could not have come from this population and that the claim about the population mean is probably false. We used logic like this in Modules 7, 8, and 9 in the context of proportions. In this module, we further develop this idea in the context of means.

Let's Summarize

- Many questions regarding quantitative variables require us to say something about the mean of a large population. It is often necessary to compute statistics from a random sample and use them to make an estimate or an inference about the population mean.
- We need to be able to compute the probability that the mean of a random sample falls in a given range. This probability allows us to draw an inference about the population parameter. To compute this probability, we need to understand the distribution of all sample means.
- Let's say we have a quantitative data set from a population with mean μ and standard deviation σ. The model for the theoretical sampling distribution of means of all random samples of size *n* has the following properties:
  - The mean of the sampling distribution of means is μ.
  - The standard deviation of the sampling distribution of means is $\sigma/\sqrt{n}$.
    - Notice that as *n* grows, the standard deviation of the sampling distribution of means shrinks. It means that larger samples give more accurate estimates of population means.
- The central limit theorem states that for large enough sample sizes, the sampling distribution of means is approximately normal, even if the population is not normal.
  - If a variable has a skewed distribution for individuals in the population, a larger sample size is needed to ensure that the sampling distribution has a normal shape.
  - The general rule is that if *n* is more than 30, the sampling distribution of means will be approximately normal. However, if the population is already normal, then any sample size will produce a normal sampling distribution.
- The mechanics of finding a probability associated with a range of sample means usually proceeds as follows.
  - Convert a sample mean $\bar{X}$ into a *z*-score: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.
  - Use technology to find a probability associated with a given range of *z*-scores.