Central Limit Theorem for Averages

Suppose that we are taking samples of size n items from a large population with mean μ and standard deviation σ. Each sample taken from the population has its own average $\overline{X}$.

- **The sample averages $\overline{X}$ follow a probability distribution of their own.**
- **The average of the sample averages is the population average:** $\mu_{\overline{X}} = \mu$
- **The standard deviation of the sample averages equals the population standard deviation divided by the square root of the sample size**

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{\textit{sample size}}}$$

- **The shape of the distribution of the sample averages $\overline{X}$ is normally distributed IF the sample size is large enough OR IF the original population is normally distributed**
- **The larger the sample size, the closer the shape of the distribution of sample averages becomes to the normal distribution.**

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Amazingly, this means that even if we don't know the distribution of individuals in the original population, as the sample size grows large we can assume that the sample average follows a normal distribution.

We can find probabilities for sample averages using the normal distribution,

- **even if the original population is not normally distributed.**
- **even if we don't know the shape of the distribution of the original population.**

<u>How large does the sample size need to be in order to use the Central Limit Theorem?</u>

The value of n to be a "large enough" sample size depends on the shape of the original distribution of the individuals in the population

- **If the individuals in the original population follow a normal distribution, then the sample averages will have a normal distribution, no matter how small or large the sample size is.**
- If the individuals in the original population (X) do not follow a normal distribution, then the sample averages $\overline{X}$ become more normally distributed as the sample size grows larger. In this case the sample averages $\overline{X}$ do not follow the same distribution as the original population.
- **The more skewed the original distribution of individual values, the larger the sample size needed.**
- **If the original distribution is symmetric, the sample size needed can be smaller.**
- Some statisticians use the rule of thumb that $n \geq 30$ is the minimum sample size to use the Central Limit Theorem; you will see this in many other textbooks. But in reality there is not a universal minimum sample size that works for all distributions; the sample size needed depends on the shape of the original distribution.
- In your homework in the textbook in chapter 7, assume the sample size is large enough for the Central Limit Theorem to be used to find probabilities for $\overline{X}$.

Central Limit Theorem Practice Problems

Class examples selected from those below; some but not all problems done in class.

1. A biologist finds that the lengths of adult fish in a species of fish he is studying follow a normal distribution with a mean of 10 inches and a standard deviation of 2 inches.

a. Find the probability that an individual adult fish is between 9.5 and 10.5 inches long.
b. Find the probability that for a sample of 16 adult fish, the average length is between 9.5 and 10.5 inches
c. Find the probability that for a sample of 25 adult fish, the average length is between 9.5 and 10.5 inches.
d. Sketch the graphs of the probability distributions for a, b, and c on the same axes showing how the shape of the distribution changes as the sample size changes.

2. a. Explain what happens to the standard deviation of $\bar{X}$ as the sample size increases.
b. Explain in words how this shows up in the graph of the sample averages $\bar{x}$ as the sample size increases.

3. Suppose that the length of time that a student waits for help in the Student Services Building during the first week of class follows an exponential distribution with a mean of 5 minutes.

a. Find the probability that an individual student waits between 4 and 6 minutes.
b. Find the probability that for a sample of 36 students the average wait is between 4 and 6 minutes.
c. Find the probability that for a sample of 64 students the average wait is between 4 and 6 minutes.
d. Sketch the graphs of the probability distributions for a, b, and c on the same axes showing how the shape of the distribution changes as the sample size changes.

4. The ages of students riding school busses in a large city are uniformly distributed between 6 and 16 years old.

a. Find the probability that in individual student is between 10 and 12 years old.
b. Find the probability that for a randomly selected sample of 36 students who ride school busses in this city, the average age is between 10 and 12 years old.
c. Find the probability that for a randomly selected sample of 64 students who ride school busses in this city, the average age is between 10 and 12 years old.
d. Sketch the graphs of the probability distributions for a, b, and c on the same axes showing how the shape of the distribution changes as the sample size changes.

5. What happens to the shape of the distribution when you look at sample averages instead of individuals?

6. Power plants and industrial processes use water from sources such as rivers to regulate temperature. The water is taken from the river, run through cooling pipes to cool the power or production process, and then is released (clean) back into the river. The temperature of the released water must be monitored closely. Fish and plants living in the river are very sensitive to the water temperature; seemingly small temperature differences can affect survival. Suppose that water used to cool an industrial process is released into a river; the temperature of the released water follows an unknown distribution that is skewed to the right with an average temperature of 14.1°C with a standard deviation of 2.5°C.

a. Explain why you can't find the probability that the water released into the river is more than 15°C
b. Find the probability that for a sample of 42 days, the average temperature of released water is more than 15°C.
c. For samples of 60 days, would the shape of the probability distribution for sample averages be skewed to the right, like the original distribution of temperatures on individual days? Explain why or why not.

A. Question 1 revisited: Refer to parts a, b

a. Find the 80^{th} percentile of individual adult fish lengths. Interpret the 80^{th} percentile in the context of the problem.
b. Find the 80^{th} percentile of average fish lengths for samples of 16 adult fish

If we were to take repeated samples of 16 fish, 80% of all possible samples of 16 fish would have average lengths of less than ____ inches.

B. Question 4 revisited: Refer to parts a, b

a. Find the third quartile of ages of individual students riding the school bus and write the interpretation.
b. Find the third quartile of average ages of samples of 36 students who ride school busses in this city.

If we were to take repeated samples of 36 students riding school busses, 75% of all possible samples would have average ages less than _____ years