CH. 7: SAMPLING DISTRIBUTIONS

Ch. 7-3: Sample Means
7-3 Sample Means

Example: In a certain high school east of Woodland Hills, the distribution of ACT scores for Seniors is approximately normal with an average ACT score of 16 with a standard deviation of 5. Suppose that Mr. Brinkhus goes to Taft to select a random sample of 5 students. He will find these 5 students and record each of their ACT scores and then find the mean of the sample.

What is the population? Seniors at Taft High School

What are the parameters? $\mu = 16 \quad \sigma = 5$

What is the sample? 5 students

Simulate taking a sample of 5 students. What is the sample mean? $\bar{x} =$

http://onlinestatbook.com/stat_sim/sampling_dist/
Now simulate taking many many many many samples of 5 students. Draw the distribution below.

**Population Distribution**

\[
\bar{x} =
\]

**Sample Data Distribution**

\[
\bar{x} =
\]

**Approximate Sampling Distribution**

\[
\bar{x} =
\]
What if the population distribution was NOT normally distributed? Would the sampling distribution still be approximately normal? **Depends on the sample size.**

What happens to the sampling distribution if we increase the sample size? **Less variability**

**More normal**

**Central Limit Theorem (CLT):** No matter the shape of a population distribution, as sample size increases, the sampling distribution of $\bar{x}$ turns more Normal.

- If the population distribution is Normal, so is the sampling distribution of $\bar{x}$.
- If $n \geq 30$, the CLT tells us that the sampling distribution of $\bar{x}$ will be approximately Normal in most cases.

Summarize (circle the correct answer):

1. Averages (means) are **more/less** variable than individual values.
2. Averages are **more/less normal** than the individual values.
3. Increasing the sample size increases/decreases the variability of the sampling distribution.
In Section 7-2 we looked at proportions. We were in search of a proportion of the population with a certain characteristic.

The actual proportion of the entire population is called a parameter, denoted \( p \).

The estimate of the proportion based on a sample is called a statistic, denoted \( \hat{p} \).

Example: Proportion of orange Reese’s Pieces.

In Section 7-3, we will look at means. We are in search of the average value (mean) and standard deviation of some characteristic within the entire population.

The actual mean of the entire population is called a parameter, denoted \( \mu \).

The estimate of the mean based on a sample is called a statistic, denoted \( \bar{x} \).

Example: Mean ACT scores for Taft Seniors.
Mean and Standard Deviation of a Sample Mean

Suppose that \( \bar{x} \) is the mean of an SRS of size \( n \) drawn from a large population with mean \( \mu \) and standard deviation \( \sigma \). If the population distribution is approximately normal, then the sampling distribution is approximately **normal** with:

\[
\mu_{\bar{x}}
\]

The mean of the sampling distribution of \( \bar{x} \) is **\( \mu \)**.

\[
\sigma_{\bar{x}}
\]

The standard deviation of the sampling distribution of \( \bar{x} \) is **\( \sigma / \sqrt{n} \)**.

Look back at the formula for standard deviation. It tells us that the standard deviation **decreases** as the size of the SRS increases. Specifically, if we want to cut the standard deviation in half, we must **multiply sample size by 4**.

\[
\frac{\sigma}{\sqrt{4n}} = \frac{\sigma}{2\sqrt{n}} = \frac{1}{2} \cdot \frac{\sigma}{\sqrt{n}}
\]

Remember the **10% condition**. It says we can only use the above formula for standard deviation if

\[
n \leq \frac{1}{10}N
\]

If true, then use...
The height of all ECRCHS seniors approximately follows a normal distribution with $N(68.8, 3.5)$ (where height is in inches).

1. Pick an ECRCHS senior at random. What is the probability that their height is greater than 70 inches?

There is about a 37% chance that a randomly selected senior is taller than 70 inches.

$$z = \frac{x - \mu}{\sigma} = \frac{70 - 68.8}{3.5} = 0.34$$

$$\text{Area} = 1 - 0.6331 = 0.3669$$

$$\text{normalcdf}(70, 999999, 68.8, 3.5) = 0.3658$$
2. Pick 10 ECRCHS seniors and find the average of their heights. What is the probability that the average of their heights is greater than 70 inches?

Recall: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$$z = \frac{x - \mu}{\sigma} = \frac{70 - 68.8}{1.107} = 1.02$$

Area = $1 - 0.8599 = 0.1401$

$\text{normalcdf}(70, 99999, 68.8, \frac{3.5}{\sqrt{10}}) = 0.1391$

There is about a 14% chance that the average height of an SRS of 10 seniors would be greater than 70 inches.
3. Pick 50 ECRCHS seniors and find the average of their heights. What is the probability that the average of their heights is greater than 70 inches?

How would the sampling distribution for this problem compare with the shape of the previous?
Since \( n \) is larger, the distribution would be skinner (less variability).

**Sampling Distribution of \( \bar{x} \)**

\[
N \left( 68.8, \frac{3.5}{\sqrt{50}} \right)
\]

\[
z = \frac{\bar{x} - \mu}{\sigma} = \frac{70 - 68.8}{0.495} = 2.42
\]

\[
\text{Area} = 1 - 0.9922 = 0.0078
\]

\[
\text{normalcdf}(70, 99999, 68.8, \frac{3.5}{\sqrt{50}}) = 0.0077
\]

There is less than a 1% chance that the average height of an SRS of 50 seniors would be greater than 70 inches.