9.3: Sampling Means

Sample Means

Learning Objectives

After this section, you should be able to:

- ✓ FIND the mean and standard deviation of the sampling distribution of a sample mean. CHECK the 10% condition before calculating the standard deviation of a sample mean.
- ✓ EXPLAIN how the shape of the sampling distribution of a sample mean is affected by the shape of the population distribution and the sample size.
- ✓ If appropriate, use a Normal distribution to CALCULATE probabilities involving sample means.

Sampling Means

- Sample **proportions** arise most often when we are interested in **categorical** variables. Examples of questions we might want to answer are: “What proportion of US adults have watched Survivor?” or “What percent of the adult population attended church last week?”

- **Quantitative** variables are usually reported as sample **means**—household income, blood pressure, GPA. Because sample means are just averages of observations, they are among the most common statistics.

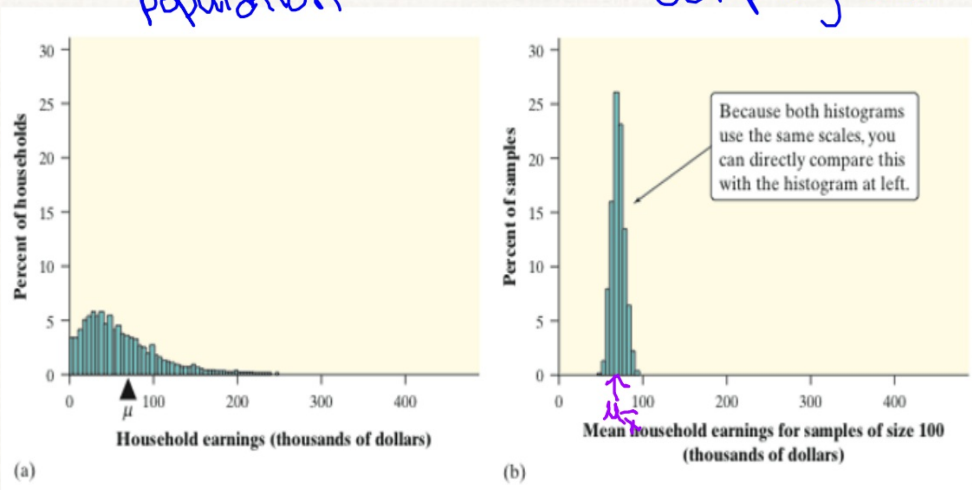
The Sampling Distribution of $\overline{x}$

Consider the mean household earnings for samples of size 100. Compare the population distribution on the left with the sampling distribution on the right. What do you notice about the shape, center, and spread of each?

population

sampling dist of $\overline{x}$

$n = 100$



The Sampling Distribution of $\bar{x}$

When we choose many SRSs from a population, the sampling distribution of the sample mean is centered at the population mean μ and is less spread out than the population distribution. Here are the facts.

Sampling Distribution of a Sample Mean

Suppose that $\bar{x}$ is the mean of an SRS of size n drawn from a large population with mean μ and standard deviation σ. Then:

The **mean** of the sampling distribution of $\bar{x}$ is $\mu_{\bar{x}} = \mu$

The **standard deviation** of the sampling distribution of $\bar{x}$ is

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

st.dev of pop

as long as the *10% condition* is satisfied: $n \leq (1/10)N$.

Note: These facts about the mean and standard deviation of $\bar{x}$ are true *no matter what shape the population distribution has.*

Sampling From a Normal Population

We have described the mean and standard deviation of the sampling distribution of the sample mean $\bar{x}$ but not its shape. That's because the shape of the distribution of $\bar{x}$ depends on the shape of the population distribution.

In one important case, there is a simple relationship between the two distributions. **If the population distribution is Normal, then so is the sampling distribution, regardless of the sample size.**

Sampling Distribution of a Sample Mean from a Normal Population

Suppose that a population is Normally distributed with mean μ and standard deviation σ. Then the sampling distribution of $\bar{x}$ has the Normal distribution with mean μ and standard deviation $\sigma / \sqrt{n}$, provided that the 10% condition is met.

The Central Limit Theorem

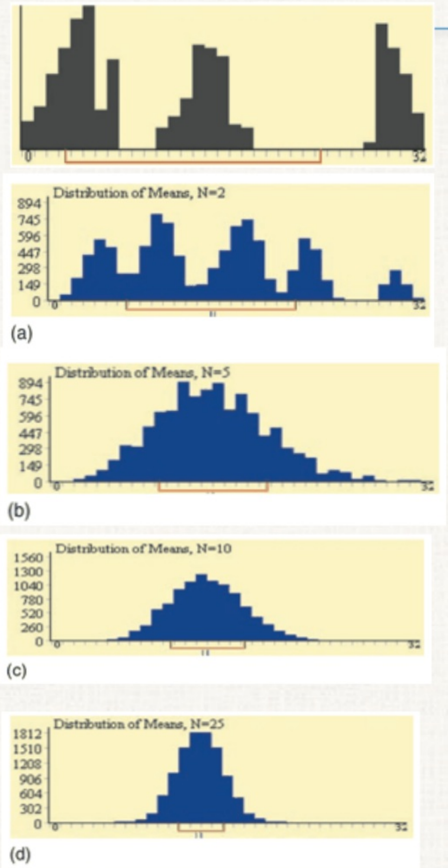
Most population distributions are not Normal. What is the shape of the sampling distribution of sample means when the population distribution isn't Normal?

As we saw in the penny lab, as the sample size increases, the distribution of sample means changes its shape: it looks less like that of the population and more like a Normal distribution!

When the sample is large enough, the distribution of sample means is very close to Normal, *no matter what shape the population distribution has*, as long as the population has a finite standard deviation.

Draw an SRS of size n from any population with mean μ and finite standard deviation σ. The **central limit theorem (CLT)** says that when n is large, the sampling distribution of the sample mean $\bar{x}$ is approximately Normal. $(n \geq 30)$

The Central Limit Theorem



Consider the strange population distribution from the Rice University sampling distribution applet.

Describe the shape of the sampling distributions as n increases. What do you notice?

The Central Limit Theorem

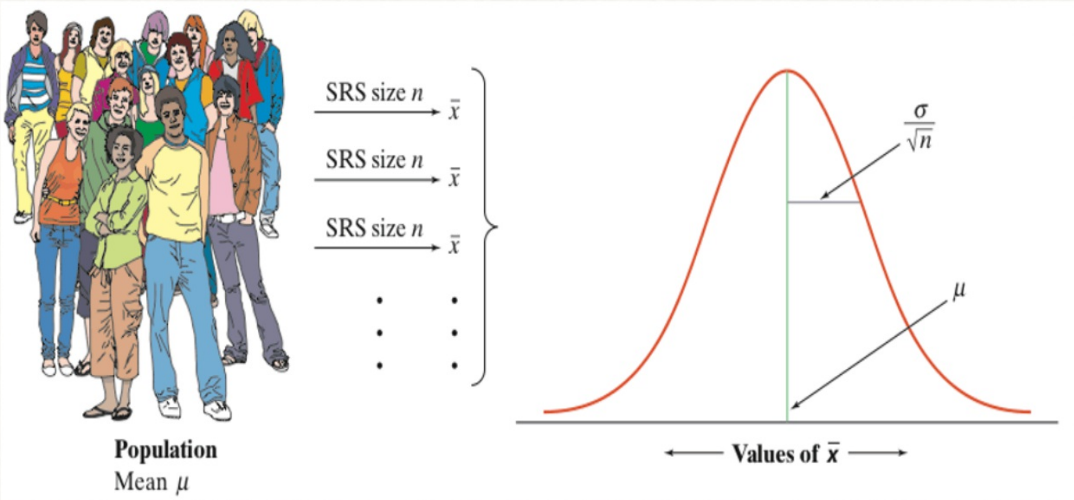
Normal/Large Condition for Sample Means

If the population distribution is Normal, then so is the sampling distribution of $\bar{x}$. This is true no matter what the sample size n is.

If the population distribution is not Normal, the central limit theorem tells us that the sampling distribution of $\bar{x}$ will be approximately Normal in most cases if $n \geq 30$.

The central limit theorem allows us to use Normal probability calculations to answer questions about sample means from many observations even when the population distribution is not Normal.

The Sampling Distribution of $\bar{x}$



Sampling Distribution of a Sample Mean from a Normal Population

- **Example: Suppose the heights of young women are normally distributed with $\mu = 64.5$ inches and $\sigma = 2.5$ inches. What is the probability that the mean height of an SRS of 10 young women is greater than 66.5 inches?**

- Step 1: μ = true mean height of all young women

- Step 2: Find mean, standard deviation, and verify Normality.

$\mu_{\bar{x}} = \mu = 64.5$

The pop. of all young women in the US is more than $10(10) = 100$. (10% rule met)

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{10}} = 0.79$$

The population dist is Normal, so the sampling dist. of $\bar{x}$ is also Normal

Sampling Distribution of a Sample Mean from a Normal Population

- **What is the probability that the mean height of an SRS of 10 young women is greater than 66.5 inches?**
- $n = 10$ $\quad \mu_{\bar{x}} =$ 64.5 $\quad \sigma_{\bar{x}} = \frac{2.5}{\sqrt{10}}$ *or* 0.790569415

- 3) Calculations:

$P(\bar{x} > 66.5) = \text{ncdf}(66.5, 1\times10^{99}, 64.5, \frac{2.5}{\sqrt{10}})$

$= 0.0057$

- 4) Conclusion:

The prob. that the mean height of an SRS of 10 young women is $>$ than 66.5 in is $\sim$ 0.57%

Example 2

- The length of pregnancy from conception to birth varies has a mean of 266 days and a standard deviation of 16 days
  - What is the probability that an SRS of 36 women will have an average pregnancy length of more than 270 days?

Because $n = 36$ is greater than 30, our s.s. is large enough for the CLT to apply. The sample dist. is approx Normal.

Example 3

Why can't you do this problem?????

- The length of pregnancy from conception to birth varies & has a mean of 266 days and a standard deviation of 16 days
  - What is the probability that an SRS of 10 women will have an average pregnancy length of more than 270 days?

$\mu_{\bar{x}} = 266$

10% rule ✓

$\sigma_{\bar{x}} = \frac{16}{\sqrt{10}}$

We don't know the shape of the population dist., and n=10 is too small for the CLT to apply

Sampling Distribution of p-hat

(1) Mean of the Sampling Distribution:

- If p-hat is an unbiased estimator of p (which we assume it is...)
- $\mu_{\hat{p}} = \rho$

(2) Spread/Variability of Sampling Distribution:

- The larger the sample size, the smaller the standard deviation of the sampling distribution (large sample size = less variability)
- **IF** the population is at least 10 times the sample size.....
- $\sigma_{\hat{p}} = \sqrt{\frac{\rho\,(1-\rho\,)}{n}}$

(3) Shape of the Sampling Distribution:

- **IF** n$\rho \geq 10$ and n$(1-\rho) \geq 10$, the overall shape of the distribution is approximately normal.

Sampling Distribution of x-bar

(1) Mean of the Sampling Distribution:

- If x-bar is an unbiased estimator of p (which we assume it is...)
- $\mu_{\bar{x}} = \mu$

(2) Spread/Variability of Sampling Distribution:

- The larger the sample size, the smaller the standard deviation of the sampling distribution (large sample size = less variability)
- **IF** the population is at least 10 times the sample size.....
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

(3) Shape of Sampling Distribution:

- **IF the population is Normally distributed**, then the Sampling distribution is Normally distributed.

OR...

- **IF the sample size is large enough** (depends on shape of population, but generally if **n ≥ 30**) the overall shape of the distribution is approximately normal due to the **central limit theorem**.

Sample Means

Section Summary

In this section, we learned how to…

- ✓ FIND the mean and standard deviation of the sampling distribution of a sample mean. CHECK the 10% condition before calculating the standard deviation of a sample mean.
- ✓ EXPLAIN how the shape of the sampling distribution of a sample mean is affected by the shape of the population distribution and the sample size.
- ✓ If appropriate, use a Normal distribution to CALCULATE probabilities involving sample means.