Lesson 7.3 ~ Sample Means

OBJECTIVES

- ✓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.

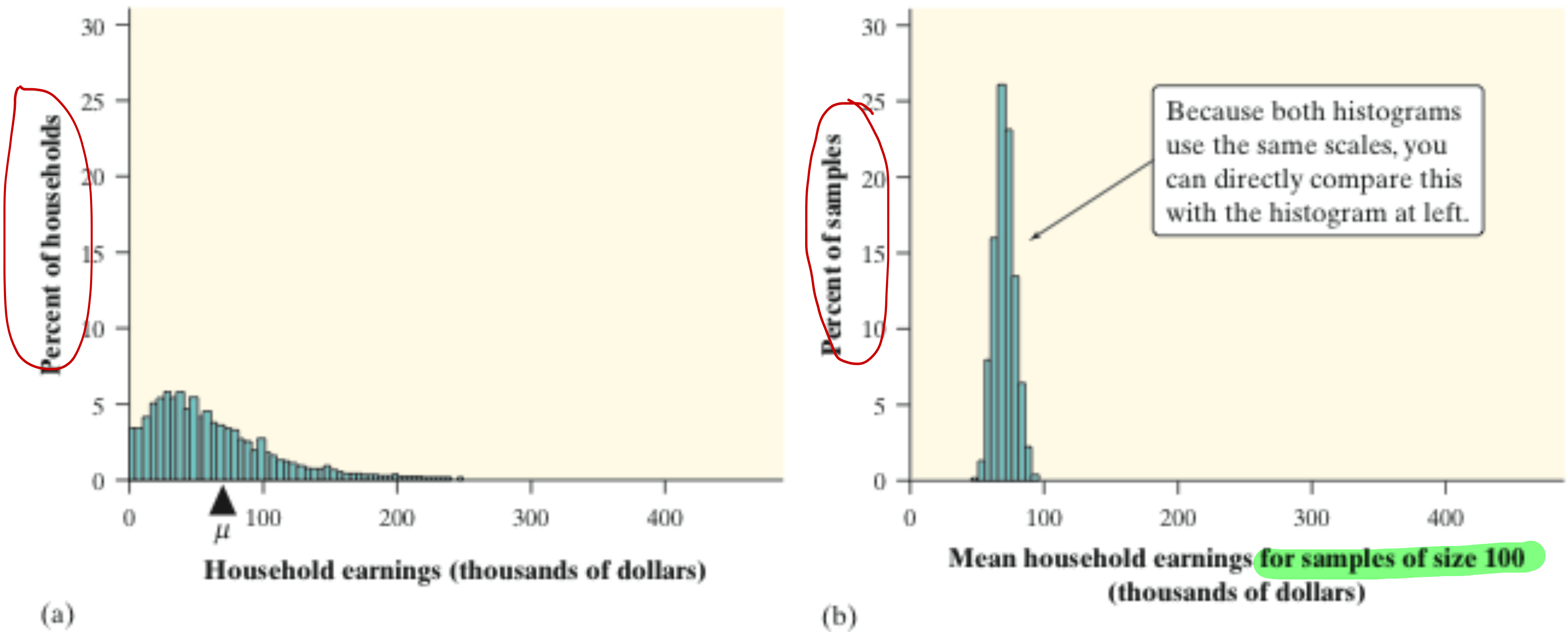
What's the difference between the sampling distributions for sample proportions and sample means?

We use sample **proportions** when we're talking about **categorical variables**. For example: "What proportion of the candies are orange?" or "What percent of the population owns a smart phone?"

We use sample **means** when we're examining **quantitative data**. For example: "What is the average household income in the city?" or "What is the average test score?"

When we record quantitative variables we are interested in other statistics such as the median or mean or standard deviation. Sample means are among the most common statistics.

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?



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

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

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

☆ This is the same rule as we use with sample proportion distributions.

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

The behavior of a sampling distribution of sample means is similar to that of the distribution of sampling proportions.

- The behavior of $\bar{x}$ is an unbiased estimator of the population mean μ .
- The values of $\bar{x}$ are less spread out for larger samples. The standard deviation decreases at the rate of $\sqrt{n}$, so you would need to take a sample four times as large to cut the standard deviation in half.
- We can only use the formula for standard deviation if the 10% condition is followed (the population is 10 times larger than the sample).

Example: Suppose that the number of movies viewed in the last year by high school students has an average of 19.3 with a standard deviation of 15.8. Suppose we take an SRS of 100 high school students and calculate the mean number of movies viewed by the members of the sample.

$\mu = 19.3, \ \sigma = 15.8, \ n = 100$

a) What is the mean of the sampling distribution? Explain.

$$\mu_{\bar{x}} = \mu = \boxed{19.3}$$

b) What is the standard deviation of the sampling distribution? Check that the 10% condition is met.

$100 \leq \frac{1}{10}(\text{HS students}) \checkmark$

$$\sigma_{\bar{x}} = \frac{15.8}{\sqrt{100}} = \frac{15.8}{10} = \boxed{1.58}$$

We said before that the sampling distribution of sample means can be any shape. An interesting fact however, is that the sampling distribution will take on the same shape as the population distribution *no matter what the sample size is*. This is really important! What this means is…

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.

Example: At the P. Nutty Peanut Company, dry-roasted, shelled peanuts are placed in jars by a machine. The distribution of weights in the jars is approximately normal, with a mean of ~~[illegible]~~ 16.1 ounces and a standard deviation of 0.15 ounces.

Without doing any calculations, explain which outcome is more likely: randomly selecting a single jar and finding that the contents weigh less than 16 ounces, or randomly selecting 10 jars and finding that the average contents weigh less than 16 ounces.

The sample of 10 jars is less likely to have the average contents weigh less than 16 ounces because the larger the sample, the smaller the variation.

Example: At the P. Nutty Peanut Company, dry-roasted, shelled peanuts are placed in jars by a machine. The distribution of weights in the jars is approximately normal, with a mean of ~~16.0~~ 16.1 ounces and a standard deviation of 0.15 ounces.

a) Find the probability that a randomly selected jar contains less than 16 ounces of peanuts. $P(x<16) = P\left(z < \frac{16-16.1}{0.15}\right) = P(z < -.67) = \boxed{.2514}$

There is a .2514 probability that a randomly selected jar has less than 16 oz.

b) Find the probability that 10 randomly selected jars contain less than 16 ounces of peanuts on average.

$\mu_{\bar{x}} = 16.1 \qquad \sigma_{\bar{x}} = \frac{.15}{\sqrt{10}}$

$10 \le 1/10$ (# of jars) ✓

Normal ✓ (given)

$P(\bar{x}<16) = P\left(z < \frac{16-16.1}{.15/\sqrt{10}}\right) = P(z < -2.11) = \boxed{.0174}$

There is only a .0174 probability that our sample of 10 jars will have an average fill weight of less than 16 ounces.

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

It is a remarkable fact that 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.

Meaning we're given a value for σ.

The Central Limit Theorem (CLT)

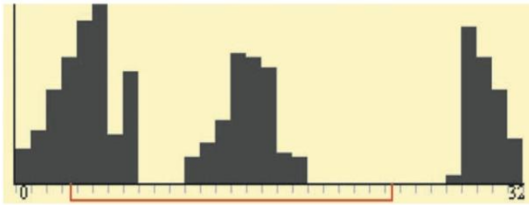
If you draw an SRS of size n from any population with the mean $\bar{x}$ and finite standard deviation σ, the CENTRAL LIMIT THEOREM says that when n is large, the sampling distribution of the sample mean μ is approximately Normal.

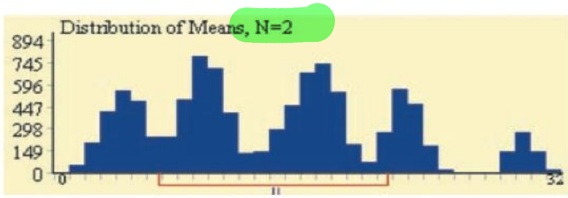
This works the same way that the Large Counts condition allows us to approximate normalcy with sample proportions.

Large Counts → proportions, CLT → means

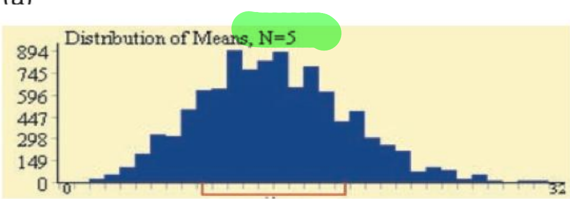
How large does the sample have to be? It depends on how far from normal the population distribution is. The farther from normal the population is, the bigger the sample size must be.

Please note! The CLT is only about the SHAPE of the distribution. While it's true that the larger the sample size is the less variability we will have, the CLT is not related to that, so you can't say that we have less variability because of the CLT.

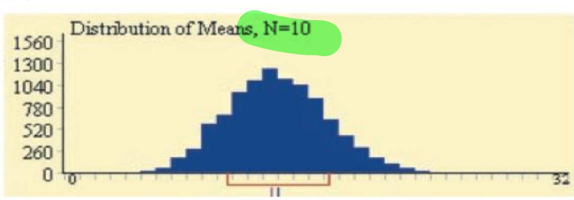




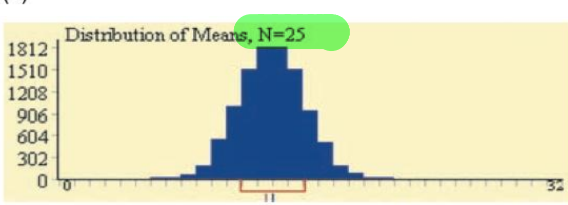
(a)



(b)



(c)



(d)

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

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

As the previous example illustrates, even when the population distribution is very non-Normal, the sampling distribution of the sample mean often looks approximately Normal with sample sizes as small as n = 25

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 $n \geq 30$.

This is a rule of thumb value. Some books say $n \geq 40$.

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.

Example: Suppose that the number of texts sent during a typical day by students at a large high school follows a right-skewed distribution with a mean of 45 and a standard deviation of 35.

How likely is it that a random sample of 50 students will have sent more than a total of 2500 texts in the last day?

$\bar{x} = \frac{2500}{50} = 50$

$\mu_{\bar{x}} = 45$

$\sigma_{\bar{x}} = 35/\sqrt{50}$

$50 \le \frac{1}{10}\left(\begin{array}{l}\text{all texts in}\\ \text{a day}\end{array}\right) \checkmark$

$n \ge 30$, so Normal by the CLT ✓

$P(\bar{x} > 50) = P\left(z > \frac{50-45}{35/\sqrt{50}}\right) = P(z > 1.01)$

$= 1 - .8438 = \boxed{.1562}$

There is a 15.62% probability that the sample of 50 students will send more than 2500 texts in a given day.

Homework:

Page 461: #49-53 odds, 54, 55, 57-59 all, 61, 63, 65-72 all