

Chapter 7: Sampling Distributions

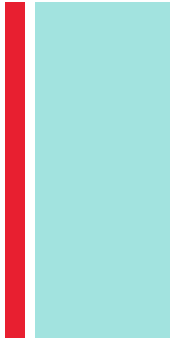
Section 7.3

Sample Means

The Practice of Statistics, 4th edition – For AP*
STARNES, YATES, MOORE

+ Chapter 7

Sampling Distributions

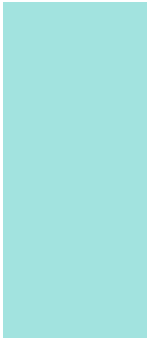


- 7.1 What is a Sampling Distribution?
- 7.2 Sample Proportions
- 7.3 **Sample Means**



Section 7.3

Sample Means



Learning Objectives

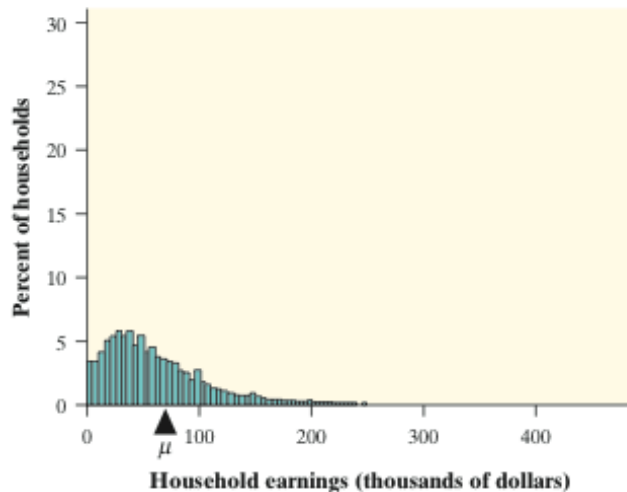
After this section, you should be able to...

- ✓ FIND the mean and standard deviation of the sampling distribution of a sample mean
- ✓ CALCULATE probabilities involving a sample mean when the population distribution is Normal
- ✓ EXPLAIN how the shape of the sampling distribution of sample means is related to the shape of the population distribution
- ✓ APPLY the central limit theorem to help find probabilities involving a sample mean

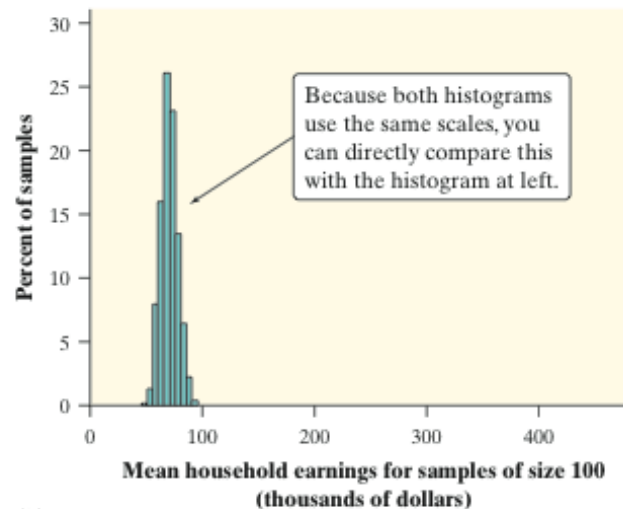
■ Sample Means

Sample proportions arise most often when we are interested in categorical variables. When we record quantitative variables we are interested in other statistics such as the median or mean or standard deviation of the variable. 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?



(a)

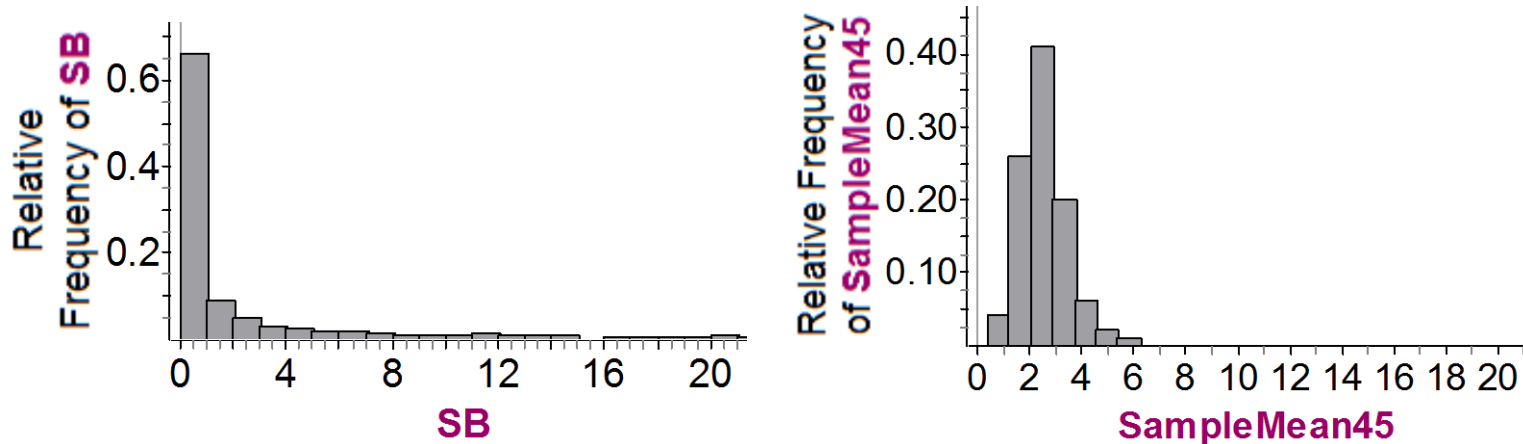


(b)

■ Alternate Example – Stealing bases

The first histogram shows the distribution of stolen bases (SB) for the 1341 Major League Baseball players who had at least 1 plate appearance in the 2009 season. The right tail of the histogram actually extends to 70, but we cut off the scale at 20 to be able to focus on the majority of the observations, which are near 0.

The second is a histogram showing the distribution of the sample mean number of stolen bases for 100 SRSs of size $n = 45$. It is graphed on the same scale to make it easier to see the difference in variability.



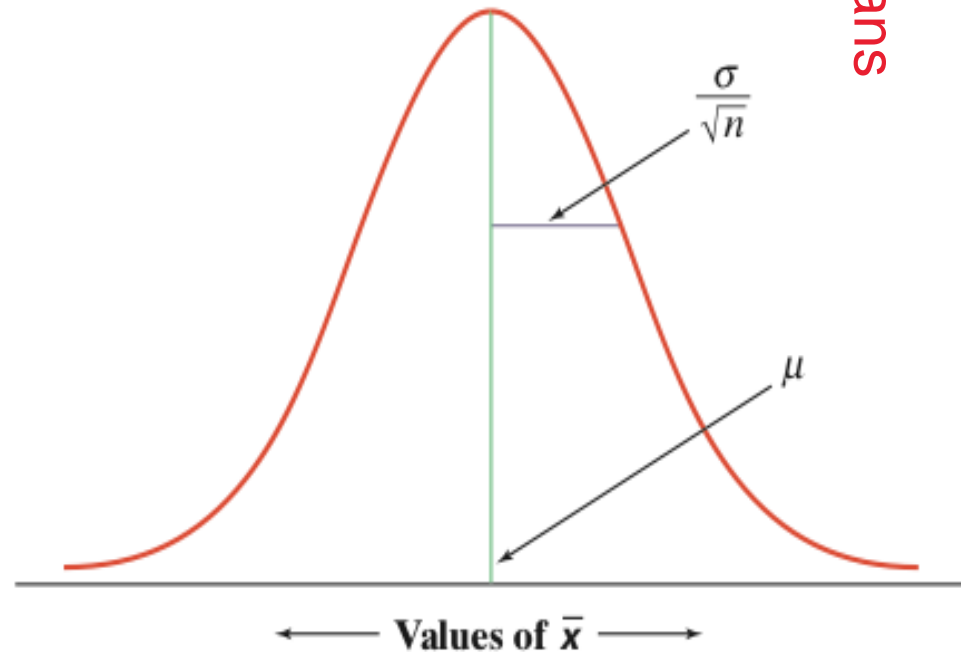
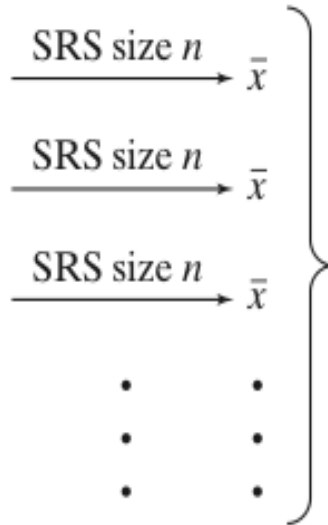
In addition to being much less spread out, the distribution of is also much more symmetric than the population distribution. However, the center of each distribution is about the same.

■ 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.



Population
Mean μ



■ Alternate Example – Moviegoing students

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.

(a) What is the mean of the sampling distribution of \bar{x}

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

(b) What is the standard deviation of the sampling distribution of \bar{x} ?

Check whether the 10% condition is satisfied.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15.8}{\sqrt{100}} = 1.58.$$

The 10% condition is met because there are more than $10(100) = 1000$ high school students.

■ 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 of \bar{x} . *This is true no matter what the sample size is.*

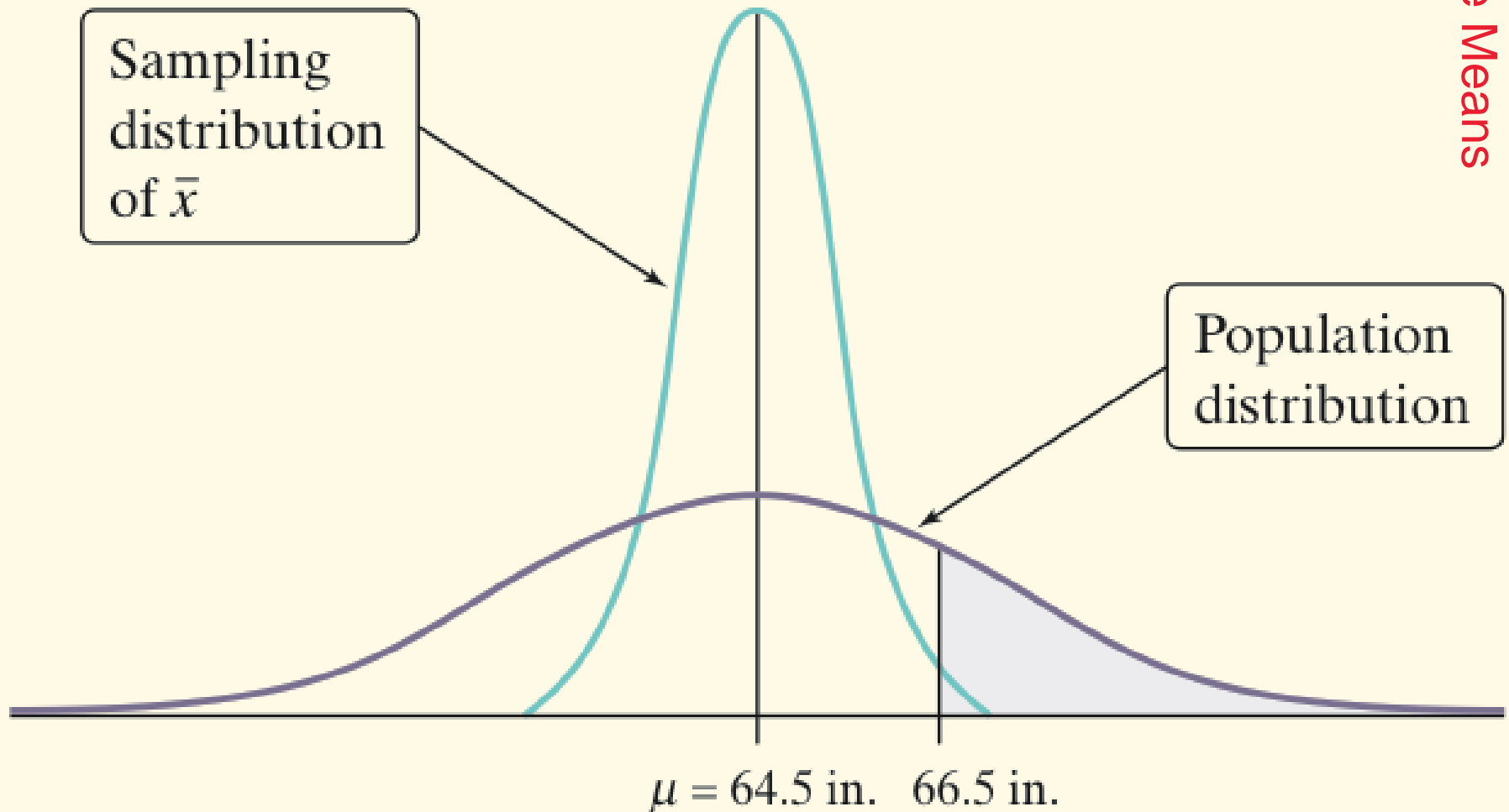
Sampling Distribution of a Sample Mean from a Normal Population

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 σ/\sqrt{n} , provided that the 10% condition is met.

Example: Young Women's Heights

The height of young women follows a Normal distribution with mean $\mu = 64.5$ inches and standard deviation $\sigma = 2.5$ inches.

Sample Means



■ Alternate Example – Buy me some peanuts and sample means

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

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

Since averages are less variable than individual measurements, I would expect the sample mean of 10 jars to be closer, on average, to the true mean of 16.1 ounces. Thus, it is more likely that a single jar would weigh less than 16 ounces than the average of 10 jars to be less than 16 ounces.

(b) Find the probability of each event described above.

Let X = weight of the contents of a randomly selected jar of peanuts. X is $N(16.1, 0.15)$.

$$P(X < 16) = \text{normalcdf}(-100, 16, 16.1, 0.15) = 0.2525.$$

Let \bar{x} = average weight of the contents of a random sample of 10 jars. \bar{x} has the $N(16.1, 0.15/\sqrt{10})$ distribution.

$$P(\bar{x} < 16) = \text{normalcdf}(-100, 16, 16.1, 0.15/\sqrt{10}) = 0.0175.$$

This answer agrees with the answer to part (a) because this probability is much smaller than 0.2525.

■ 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?

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.

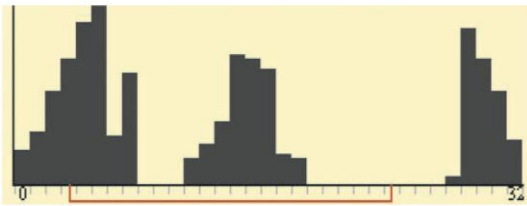
Definition:

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.

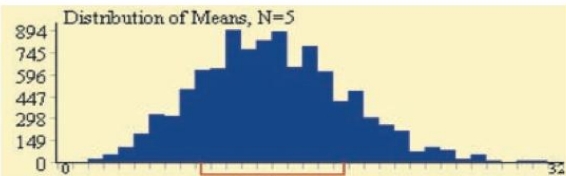
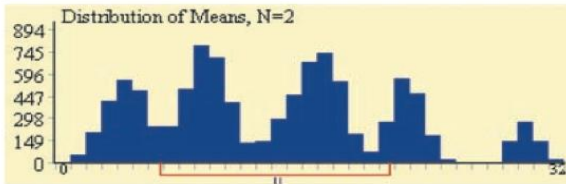
Note: How large a sample size n is needed for the sampling distribution to be close to Normal depends on the shape of the population distribution. More observations are required if the population distribution is far from Normal.

■ The Central Limit Theorem

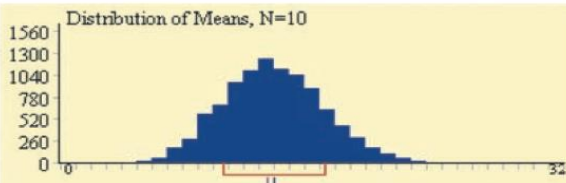
Consider the strange population distribution from the Rice University sampling distribution applet.



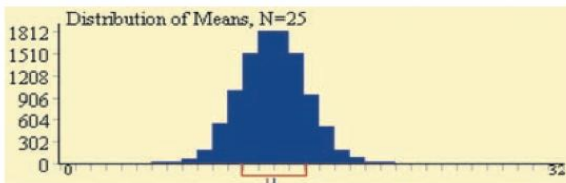
Describe the shape of the sampling distributions as n increases. What do you notice?



(h)



(c)



(d)

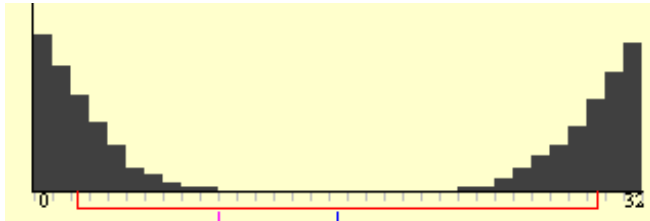
Normal Condition for Sample Means

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.

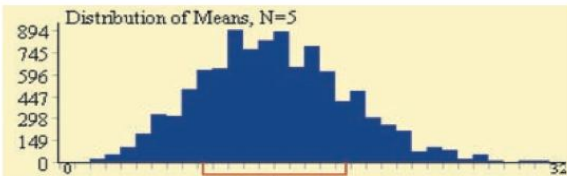
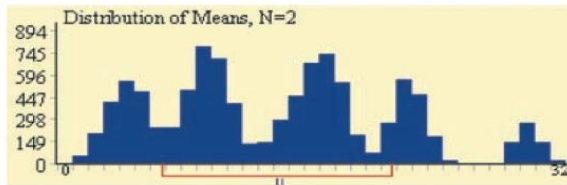
If the population distribution is not Normal the central limit theorem tells us that the sampling distribution of \bar{x} will be approximately Normal in most cases if $n \geq 30$.

■ Alternate Example – Another strange population

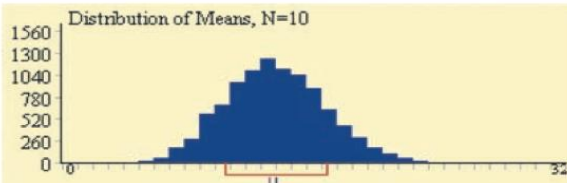
■ Here is another population distribution with a strange shape:



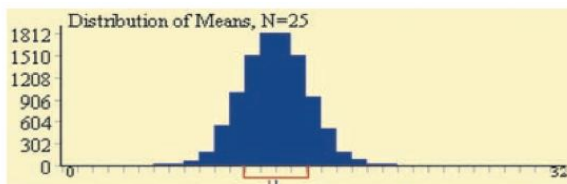
What do you think the sampling distribution of \bar{x} will look like for samples of size 2? What about samples of size 5? Size 25?



(h)



(c)

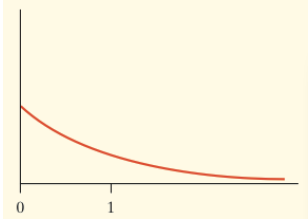


(d)

Here are the results of 10,000 SRSs of each size. The first graph has three peaks, since there are only 4 basic outcomes for a sample: two small values, which gives a small mean, two large values, which gives a large mean, or one of each, which gives a mean in the middle. Since there are two ways to get one of each, the middle pile is roughly twice as big.

Example: Servicing Air Conditioners

Based on service records from the past year, the time (in hours) that a technician requires to complete preventative maintenance on an air conditioner follows the distribution that is strongly right-skewed, and whose most likely outcomes are close to 0. The mean time is $\mu = 1$ hour and the standard deviation is $\sigma = 1$

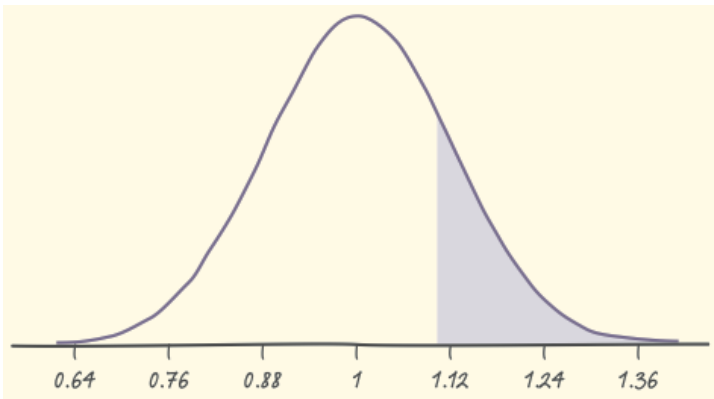


Your company will service an SRS of 70 air conditioners. You have budgeted 1.1 hours per unit. Will this be enough?

Since the 10% condition is met (there are more than $10(70)=700$ air conditioners in the population), the sampling distribution of the mean time spent working on the 70 units has

$$\mu_{\bar{x}} = \mu = 1 \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{70}} = 0.12$$

The sampling distribution of the mean time spent working is approximately $N(1, 0.12)$ since $n = 70 \geq 30$.



We need to find $P(\text{mean time} > 1.1 \text{ hours})$

$$z = \frac{1.1 - 1}{0.12} = 0.83 \quad P(\bar{x} > 1.1) = P(Z > 0.83) \\ = 1 - 0.7967 = 0.2033$$

If you budget 1.1 hours per unit, there is a 20% chance the technicians will not complete the work within the budgeted time.

Alternate Example – Mean texts

- Suppose that the number of texts sent during a typical day by a randomly selected high school student follows a right-skewed distribution with a mean of 15 and a standard deviation of 35. Assuming that students at your school are typical texters, how likely is it that a random sample of 50 students will have sent more than a total of 1000 texts in the last 24 hours?
- *State:* What is the probability that the total number of texts in the last 24 hours is greater than 1000 for a random sample of 50 high school students?

Plan: A total of 1000 texts among 50 students is the same as an average number of texts of $1000/50 = 20$. We want to find $P(\bar{x} > 20)$, where \bar{x} = sample mean number of texts. Since n is large ($50 > 30$), the distribution of \bar{x} is approximately $N(15, 35/\sqrt{50})$.

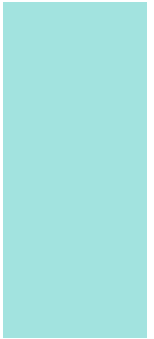
$$\text{Do: } P(\bar{x} > 20) \approx \text{normalcdf}(20, 9999, 15, 35/\sqrt{50}) = 0.1562.$$

Conclude: There is about a 16% chance that a random sample of 50 high school students will send more than 1000 texts in a day.



Section 7.3

Sample Means



Summary

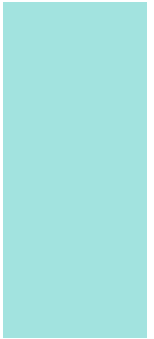
In this section, we learned that...

- ✓ When we want information about the population mean μ for some variable, we often take an SRS and use the sample mean \bar{x} to estimate the unknown parameter μ . The **sampling distribution** of \bar{x} describes how the statistic varies in all possible samples of the same size from the population.
- ✓ The mean of the sampling distribution is μ , so that \bar{x} is an unbiased estimator of μ .
- ✓ The standard deviation of the sampling distribution of \bar{x} is σ/\sqrt{n} for an SRS of size n if the population has standard deviation σ . This formula can be used if the population is at least 10 times as large as the sample (10% condition).



Section 7.3

Sample Means



Summary

In this section, we learned that...

- ✓ Choose an SRS of size n from a population with mean μ and standard deviation σ . If the population is Normal, then so is the sampling distribution of the sample mean \bar{x} . If the population distribution is not Normal, the **central limit theorem (CLT)** states that when n is large, the sampling distribution of \bar{x} is approximately Normal.
- ✓ We can use a Normal distribution to calculate approximate probabilities for events involving \bar{x} whenever the Normal condition is met

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.



Looking Ahead...



In the next Chapter...

We'll learn how to estimate population parameters with confidence, using sample statistics.

We'll learn about

- ✓ **Confidence Intervals**
- ✓ **Estimating Population Proportions**
- ✓ **Estimating Population Means**