The following example will help us understand <u>The Sampling Distribution of the Mean</u>

Review:

- The population is the entire collection of all individuals or objects of interest
- The sample is the portion of the population that is selected for study.
- Inferential Statistics is the process of using sample information to draw inferences or conclusions about the population.

Consider a population of 5 commuters who are all neighbors. Each commuter was asked how many miles he/she commutes to work each day.

C1	C2	C3	C4	C5
50 miles	84 miles	38 miles	120 miles	48 miles

Find the **population mean** of the set of data: $\mu =$

Find the **population standard deviation**: $\sigma =$

Look at a sample of two commuters ($n = 2$), and find the mean $\bar{x}$ to estimate μ.
Like: C1 = 50 and C2 = 84, then C1 and C3, then C1 and C4, then C1 and C5… then C4 and C5.
How many should we have?

List all possible samples of two commuters and calculate the mean, $\bar{x}$, for each sample.

Commuters	**Data Values**	**Mean,** $\bar{x}$

The data set of all the sample means in column 3 is called a **sampling distribution of the means**.
The **sampling error** is the difference between the value of a sample mean, $\bar{x}$, and the population mean, μ.

Sampling error of the mean = $= \bar{x} - \mu$

For the sample mean 67, calculate the sample error:

For the sample mean 102, calculate the sample error:

- Looking back at column 3 of our table, this data set of all the sample means is called a **sampling distribution of the means**.

Calculate the mean of this data set – we are calculating the mean of the sampling distribution of the means:

The mean of all the sample means of a sampling distribution has special notation:
"mu sub x-bar" or $\mu_{\overline{X}}$

Calculate the standard deviation of this data set – we are calculating the standard deviation of the sampling distribution of the means:

The standard deviation of the all the sample means of a sampling distribution has special notation:
"sigma sub x-bar" or $\sigma_{\overline{X}}$

The standard deviation is a measure of how spread out the sample means are from μ.

The **mean of the sampling distribution of the means** is: $\boxed{\mu_{\overline{X}} =}$

What is the **population mean** which we calculated before? $\boxed{\mu =}$

So, the mean of the sampling distribution of the means **is equal** to the mean of the population from which the samples were selected:

$$\mu_{\overline{X}} = \mu$$

The **standard deviation of the sampling distribution of the means** is: $\boxed{\sigma_{\overline{X}} =}$

What is the **population standard deviation** which we calculated before? $\boxed{\sigma =}$

The standard deviation of the sampling distribution of the mean, also known as the **standard error of the mean,** will **always be smaller** than the population standard deviation and the **formula** is:

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

- The larger the standard error of the mean is, the more dispersed the samples means are from the population mean.
- The smaller the standard error, the closer the sample means are to the population mean.

Finite Correction Factor: For a finite population (having a limit), the formula for the standard deviation of the sampling distribution of the mean, σ_{X} is: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$

8.1 – *The Sampling Distribution of the Mean*

- Populations are generally either quite large, have an unlimited number of data values (that is, infinite), or partly unobtainable. Since a population can rarely be studied completely, a sample randomly selected from the population serves as a convenient and economical procedure to estimate the characteristics of a population.
- Sample information is used to estimate the mean of a population. When the population mean is being estimated using sample data, the mean of the sample will probably not be equal to the mean of the population. In fact, if a few samples were randomly selected from the same population, it is very likely that none of these sample means would be exactly equal to the mean of the population.
- Although there is only one value for the true mean of a population, there are many different values for the mean when different samples are selected from the population. Therefore, the estimates of a true population mean will vary from sample to sample due to the data values randomly selected within each sample, that is, by chance alone. These variations in the estimates of the population mean from sample to sample are due to chance and are called **sampling errors**.

Sampling error is the difference between the value of a sample statistic, such as the sample mean, $\bar{x}$ and the value of the corresponding population parameter, such as the population mean, μ. Thus, the sampling error for the mean is:

sampling error = sample mean − population mean

assuming the sample is random and there are no nonsample errors.

In symbols, **sampling error of the mean** = $\bar{x} - \mu$.

- Sampling error is inevitable because we are using a chosen number of data values which are randomly selected, by chance, from the population. In practice, we will select only one sample from the population to estimate the mean of the population.
- A population can have only one mean. Yet, depending upon which sample is selected from a population, the mean of a sample can vary from sample to sample as different samples of the same size are randomly selected from the same population. Thus, the sample mean is a **random variable** because it is dependent upon the particular data values which are randomly selected from the population.

Example 8.1 pg. 423 – Suppose the population of seven college students on the Student Government Association (SGA) have the following ages: 23, 19, 20, 21, 18, 19, 25

a. Compute the population mean age of all the students on the SGA.

b. If a random sample of 3 students was selected from this population having ages: 19, 21, and 18, then compute the sample mean age for this sample.

c. Determine the sampling error if this sample (from part b) was used to estimate the population mean age. *Explain or interpret the meaning of this sampling error.*

*Review parts d, e, and f of this example too!

The **sampling distribution of the mean** is a probability distribution which lists the sample means from all possible samples of the same sample size selected from the same population along with the probability associated with each sample mean.

Notation for the <u>Mean of the Sampling Distribution of the Mean</u>

The **mean of the sampling distribution of the mean** is denoted by $\mu_{\bar{x}}$, read mu sub x bar. Thus, $\mu_{\bar{x}}$ = mean of all the sample means of the sampling distribution.

Notation for the <u>Standard Deviation of the Sampling Distribution of the Mean</u>

The **standard deviation of the sampling distribution of the mean** is denoted by $\sigma_{\bar{x}}$, read sigma sub x bar. Thus, $\sigma_{\bar{x}}$ = standard deviation of all the sample means of the sampling distribution.

8.2 – *The Mean and Standard Deviation of the Sampling Distribution of the Mean*

Mean of the Sampling Distribution of the Mean, $\mu_{\bar{x}}$

The mean of the sample means of all possible samples of size *n* is called the mean of the sampling distribution of the mean, denoted by $\mu_{\bar{x}}$. It is equal to the mean of the population from which the samples were selected. In symbols, this is expressed as:

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

Standard Deviation of the Sampling Distribution of the Mean or Standard Error of the Mean, denoted by $\sigma_{\bar{x}}$

The standard error of the mean is the standard deviation of the sample means of all possible samples of size *n* of the sampling distribution, denoted by $\sigma_{\bar{x}}$. The standard error of the mean is equal to the standard deviation of the population, σ, divided by the square root of the sample size *n*. That is:

$$\textbf{standard error of the mean} = \frac{\text{population standard deviation}}{\sqrt{\text{sample size}}} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Interpretation of the Standard Error of the Mean

- The standard deviation of the sampling distribution of the mean is referred to as the standard error of the mean because it is a measure of how much a sample mean is likely to deviate from the population mean, that is, a measure of the average sampling error.
- If the standard error of the mean, $\sigma_{\bar{x}}$ **,** is a small number, then the sampling distribution of the mean has relatively little dispersion and the sample means will be relatively close to the population mean. On the other hand, if the standard error of the mean, $\sigma_{\bar{x}}$ **,** is a large number, then the sampling distribution of the mean has a relatively large dispersion and the sample means will be relatively far from the population mean.

Example 8.3 pg. 430 – According to a study of TV viewing habits, the average number of hours a teenager watches MTV per week is $\mu = 17.9$ hours with a standard deviation of $\sigma = 3.8$ hours. If a sample of 64 teenagers is randomly selected from the population, then determine the mean and standard error of the mean of the sampling distribution of the mean.

Example 8.4 pg. 430-431 – The registrar at a large University states that the mean grade point average of all the students is $\mu = 2.95$ with a population standard deviation of $\sigma = 0.20$.

a. Determine the mean and standard error of the sampling distribution if the sampling distribution of the mean consists of all possible sample means from samples of size 25.

b. Determine the mean and standard error of the sampling distribution if the sampling distribution of the mean consists of all possible sample means from samples of size 100.

c. What effect did increasing the sample size have on the mean and standard error of the sampling distribution?

d. In which sampling distribution of the mean (n = 25 or n = 100) would you have a better chance of selecting a sample mean which is closer to the population mean grade point average?

Review Example 8.5 on pg. 432

8.4 – *The Shape of the Sampling Distribution of the Mean*

Sampling from a <u>Normal Population</u>

THEOREM 8.1 - The Shape of the Sampling Distribution when Sampling from a Normal Population
If the population being sampled is a normal distribution, then the sampling distribution of the mean is a normal distribution regardless of the sample size, *n*.

Characteristics of the Sampling Distribution of the Mean When Sampling from a Normal Population Whose Mean is μ and Standard Deviation is σ
If all possible samples of size *n* are selected from a normal population, then the sampling distribution of the mean has the following three characteristics:

1. The sampling distribution of the mean is a normal distribution, regardless of sample size, *n*.
2. The mean of the sampling distribution of the mean, $\mu_{\bar{x}}$**,** is equal to the mean of the population, μ: $\mu_{\bar{x}} = \mu$.
3. The standard error of the sampling distribution of the mean, $\sigma_{\bar{x}}$**,** is equal to the standard deviation of the population, σ, divided by the square root of the sample size, *n*: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

Example 8.7 pg. 438 – At a large New England college, the grade point average (GPA) of all attending students is normally distributed with a mean of $\mu = 2.95$ and a population standard deviation of $\sigma = 0.32$. A sample is randomly selected from this population and its sample mean, $\bar{x}$, is calculated. Determine the mean, $\mu_{\bar{x}}$, the standard error, $\sigma_{\bar{x}}$ identify the shape of the sampling distribution of the mean of the samples of size:

a. n = 9

b. n = 49

c. n = 100

Sampling from a <u>Non-Normal Population</u>

THEOREM 8.2 – The Central Limit Theorem
For any population, the sampling distribution of the mean approaches a normal distribution as the sample size n becomes large. This is true regardless of the shape of the population being randomly sampled.

General Rule for Applying the Central Limit Theorem: The n Greater than 30 Rule
For most applications, a sample size n greater than 30 is considered large enough to apply the Central Limit Theorem. Thus, the sampling distribution of the mean can be reasonably approximated by a normal distribution whenever the sample size n is greater than 30.

THEOREM 8.3 Characteristics of the Sampling Distribution of the Mean when Sampling from a Non-Normal Population

If, the following three conditions are satisfied:

- given any infinite population with mean, μ, and standard deviation, σ, and
- all possible random samples of size *n* are selected from the population to form a sampling distribution of the mean, and
- the sample size, *n*, is large (**greater than 30**).

then:

1. the sampling distribution of the mean is **approximately normal**.
2. the **mean of the sampling distribution of the mean** is equal to the mean of the population. This is expressed as:
   $$\mu_{\bar{x}} = \mu$$
3. the **standard error of the sampling distribution of the mean** is equal to the standard deviation of the population divided by the square root of the sample size. This is written as:
   $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Example 8.8 pg. 442 – GRAM-HAM Bell, a telephone company, states that the average length of time of long-distance telephone calls is $\mu = 21.3$ minutes with a standard deviation of $\sigma = 3.8$ minutes. Determine the mean and the standard error of the sampling distribution of the mean and describe the shape of the sampling distribution of the mean when the sample size is:

a. n = 36

b. n = 100

c. Compare the sampling distribution of the mean for the sample size of n = 36 and n = 100.

d. If you had to estimate the mean of the population by either randomly selecting a sample of size n = 36 or of n = 100 from the population, then which sample size would give you a better chance of obtaining a smaller sampling error? Explain.

USING THE CENTRAL LIMIT THEOREM FOR VARIOUS SHAPES AND SAMPLE SIZES:

POPULATIONS

SHAPE OF THE POPULATION	SKEWED TO RIGHT	UNIFORM	SKEWED TO LEFT	UNKNOWN	BELL-SHAPED

SAMPLING DISTRIBUTION OF THE MEAN ***$n \leq 30$***

SHAPE					

SAMPLING DISTRIBUTION OF THE MEAN n> 30

SHAPE					

SAMPLING DISTRIBUTION OF THE MEAN

FOR THE FOLLOWING USE $\mu = 8$ AND $\sigma = 2$

SAMPLE SIZE n	MEAN $\mu_{\bar{x}} = \mu$	STANDARD DEVIATION $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.	SHAPE OF THE SAMPLING DISTRIBUTION OF THE MEAN
4			
16			
25			
49			
100			
400			

Note: **As the size of the sample increases, the mean _______________________ and the standard deviation______________.**

As the sample size increases, sample means become more clustered about the mean and the more confidence you have in the accuracy of the sample means as an estimate of the population mean.

8.5 – *Calculating Probabilities Using the Sampling Distribution of the Mean*

Now in this chapter we are working with the sampling distribution of the mean so the data values are sample means, $\bar{x}$, thus, the z score formula becomes:

$$z \text{ score of a sampling mean} = \frac{\text{sample mean} - \text{mean of the sampling distribution}}{\text{standard error of the sampling distribution}}$$

or

$$z \text{ of } \bar{x} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

OR we can Use the TI-83/84 normalcdf function:

$\boxed{2^{nd}}$ $\boxed{\text{DISTR:2}}$ (lower sample mean value , higher sample mean value , $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$) $\boxed{\text{ENTER}}$

Remember: $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Example 8.9 pg. 444-445 **–** At a large public state college in Virginia, the mean Verbal SAT score of all attending students was $\mu = 600$ with a population standard deviation of $\sigma = 65$. If a random sample of 100 students is selected from a population of students to determine:

a. The probability that the mean Verbal SAT score of the selected sample will be less than 615.

b. The probability that the mean Verbal SAT score of the selected sample will be within 10 points of the population mean.

Example 8.10 pg. 447 – The population mean weight of newborn babies for a western suburb is 7.4 lbs. with a standard deviation of 0.8 lbs. What is the probability that a sample of 64 newborns selected at random will have a mean weight greater than 7.5 lbs.?

Example 8.11 pg. 449 – The population of the ages of all U. S. college students is skewed to the right with a mean age of 27.4 years and a standard deviation of 5.8 years. Determine the probability that a random sample of 49 students selected from the population will have a sample mean age within one year of the population mean age?