7.1 The Central Limit Theorem

The following theorem is considered to be the most important theorem in statistics.

The Central Limit Theorem	
As the sample size n increases without limit, the shape of the distribution of the sample means taken from a population with mean μ and standard deviation σ will approach a normal distribution. This distribution will have a mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.	
When applying the Central Limit Theorem, we use the following formula for z-values:	$z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$

For reference, here are the formulas for the mean, variance, and standard deviation of a population and of a sample

	Population	**Sample**
Mean	$\mu = \dfrac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \dfrac{\sum X}{N}$	$\bar{X} = \dfrac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \dfrac{\sum X}{n}$
Variance	$\sigma^2 = \dfrac{\Sigma(X-\mu)^2}{N}$	$s^2 = \dfrac{\Sigma(X-\mu)^2}{n-1}$
Standard Deviation	$\sigma = \sqrt{\dfrac{\Sigma(X-\mu)^2}{N}}$	$s = \sqrt{\dfrac{\Sigma(X-\mu)^2}{n-1}}$

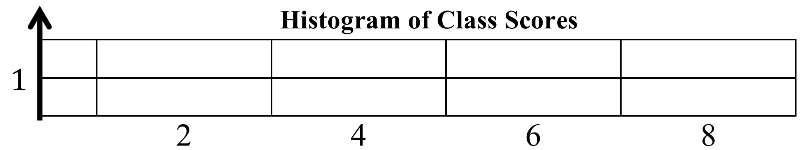
The Central Limit theorem refers to the distribution of the sample means of a given population. The following is an example of how to find the sample means of a small population.

Example 1: Consider a tiny class of 4 students with the following scores on an 8-point quiz: 2, 4, 6, 8

Find the mean and the standard deviation of this population:

Score X	$(X-\mu)^2$
2	
4	
6	
8	
$\Sigma X =$	$\Sigma(X-\mu)^2 =$
$\mu =$	$\sigma^2 =$ $\sigma =$

Construct a histogram of the data. Notice that this is a uniform distribution.

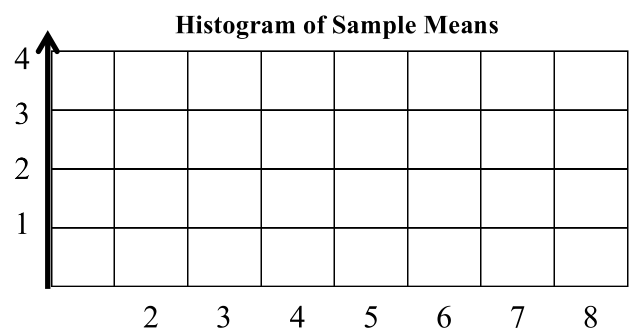


List all of the samples of size 2 that can be formed from the data set 2, 4, 6, 8, the means for each, the mean of the means, and the standard deviation of the means. Then complete a frequency distribution and histogram for the sample means.

Sample	$\bar{X}$	$(\bar{X} - \mu_{\bar{X}})^2$
{2, 2}	2	
{2, 4}	3	
{2, 6}		
{2, 8}		
	$\Sigma\bar{X} =$	$\Sigma\left(\bar{X} - \mu_{\bar{X}}\right)^2 =$
	$\mu_{\bar{X}} =$	$\sigma_{\bar{X}}^2 =$ $\sigma_{\bar{X}} =$

Frequency Distribution of Sample Means

$\bar{X}$	f



Notice that

$\mu =$ $\qquad = \mu_{\bar{X}}$

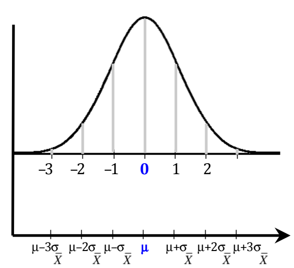
and

$\dfrac{\sigma}{\sqrt{n}} =$ $\qquad =$ $\qquad = \sigma_{\bar{X}}$

Example 2:

In 1998, homeowners in a Sacramento neighborhood became concerned by a high prevalence of leukemia among the residents. Five cases occurred in a neighborhood of 800 homes, approximately 2,000 residents, over a period of 4 years. The homeowners demanded an investigation. It was discovered that propane had been stored in an underground cavern from which the propane was seeping into the groundwater.

If the prevalence of leukemia is 10.3 in 100,000 US residents in a 4 year period with a standard deviation of 0.8 in 100,000, did the homeowners have cause for alarm? We can answer this question with the aid of the Central Limit Theorem.

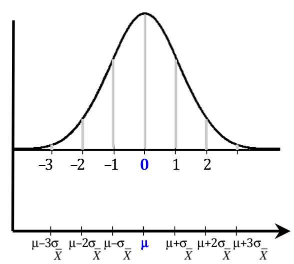
Step 1	Label the X-axis with μ, label the z-axis with z_L, shade the appropriate region	**Step 3**
Identify the variable values $\bar{X}_L =$ $\sigma =$ $\mu =$ $n =$		Use the z-table or statistics calculator to find the associated area under the normal curve
Step 2		**Step 4**
Find $z_L = \dfrac{\bar{X}_L - \mu}{\frac{\sigma}{\sqrt{n}}}$		Find the desired probability $P(\bar{X} >$ $)$ $= P(z >$ $)$ $=$

A well-known similar case involved residents of Hinkley, CA and environmental activist, Erin Brockovich. Groundwater was contaminated by PG&E at the Hinkley compressor station from which wastewater was pumped into unlined ponds and, from there, percolated into the groundwater. A high incident of cancer emerged, Brockovich became aware of the increase, investigated, sued PG&E, and a settlement of $333 million in damages in 1996. From 1996 to 2008, another 196 cases of cancer were reported in a population of 1915 residents, yet the CA Cancer Registry considers this to be "unremarkable." The cancer incident rate for Imperial Valley is 307 per 100,000. Now you can do the math to see just how "unremarkable" this number truly is.

Visit http://www.cancer-rates.info/ca/ for an interactive map of cancer incident rates per county of California.

Example 3:

It is recommended by the American Academy of Pediatrics (AAP) that children between the ages of 2 and 5 should engage in no more than 10 hours per week of screen time, involving television, computers, and video games. The AAP reported that American children in this age group spend an average of 25 hours per week engaged in screen time. Assume the variable is normally distributed and the standard deviation is 3 hours. If 20 children between the ages of 2 and 5 are randomly selected, find the probability that the mean of the number of hours they watch television will be greater than 23 hours.

<table>
<tr>
<td>Step 1
Identify the variable values
$\overline{X}_U$ = σ =
μ = n =</td>
<td rowspan="2">Label the X-axis with μ, label the z-axis with z_L, shade the appropriate region
</td>
<td>Step 3
Use the z-table or statistics calculator to find the associated area under the normal curve</td>
</tr>
<tr>
<td>Step 2
Find $z_U = \dfrac{\overline{X}_U - \mu}{\frac{\sigma}{\sqrt{n}}}$</td>
<td>Step 4
Find the desired probability
$P(\overline{X} >$ $)$
$= P(z >$ $)$
$=$</td>
</tr>
</table>

Example 4:

The average number of pounds of meat a person consumes per year is 218.4 pounds. Assume that the standard deviation is 25 pounds and the distribution is approximately normal.

(a) Find the probability that a person selected at random consumes less than 224 pounds per year.

Step 1

Identify the variable values

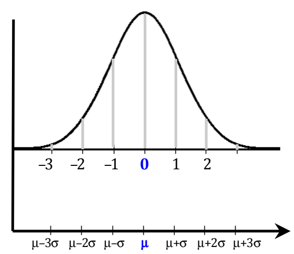
$X_U =$ $\sigma =$

$\mu =$

Step 2

Find $z_U = \dfrac{X_U - \mu}{\sigma}$

Label the X-axis with μ, label the z-axis with z_U, shade the appropriate region



Step 3

Use the z-table or statistics calculator to find the associated area under the normal curve

Step 4

Find the desired probability

$P(X <$ $)$

$= P(z <$ $)$

$=$

(b) If a sample of 40 individuals is selected, find the probability that the mean of the sample will be less than 224 pounds per year.

Step 1

Identify the variable values

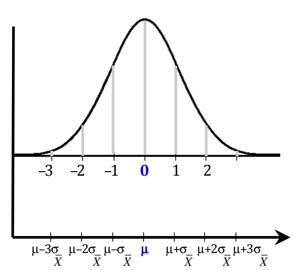
$\bar{X}_U =$ $\sigma =$

$\mu =$ $n =$

Step 2

Find $z_U = \dfrac{\bar{X}_U - \mu}{\frac{\sigma}{\sqrt{n}}}$

Label the X-axis with μ, label the z-axis with z_U, shade the appropriate region



Step 3

Use the z-table or statistics calculator to find the associated area under the normal curve

Step 4

Find the desired probability

$P(\bar{X} <$ $)$

$= P(z <$ $)$

$=$