Sampling Distributions

6.1 Minimum Variance Unbiased Point Estimators

The Concept of a Sampling Distribution

The main objective of this section is to understand the concept of a sampling distribution of a statistic. A sampling distribution of a statistic is the distribution of all values of the statistic when all possible samples of the same size are taken from the population.

In this section, we will also see that some statistics are better than others for estimating population parameters.

Recall that simple random sampling is a sampling plan that ensures that if a sample of size n is drawn from a population of size N, each sample has a 1/ C_n^N chance of being selected.

The **Sampling Distribution** of a sample statistic calculated from a sample of n measurements is the probability distribution of the statistic. In other words, it is the probability distribution for all of the possible values of the statistic that could result when taking samples of size n.

We need to think of our statistic as a random variable to understand the concept of a sampling distribution. Imagine a very small population consisting of the elements 1, 2 and 3. Below are the possible samples that could be drawn, along with the means of the samples and the *mean of the means*.

Samples for n = 1	$\bar{x}$	Samples for n = 2	$\bar{x}$	Samples for n = 3	$\bar{x}$
1	1	1, 2	1.5	1, 2, 3	2
2	2	1, 3	2		
3	3	2, 3	2.5		
Mean of the $\bar{x}$'s	$\dfrac{\sum \bar{x}}{3} = 2$	Mean of the $\bar{x}$'s	$\dfrac{\sum \bar{x}}{3} = 2$	Mean of the $\bar{x}$'s	$\dfrac{\sum \bar{x}}{1} = 2$

<mark>Example 92:</mark> Let Y = the number of boys in a family with three children, then assuming a 50% chance of having a boy, the probability distribution for the number of boys is:

Y	0	1	2	3
P(Y)	1/8	3/8	3/8	1/8

Find the sampling distribution for the sample mean when we look at two randomly selected families (n = 2). Draw a histogram for the sample mean when taking samples of size two. Compare this to the histogram for Y.

<mark>Solution:</mark>

Samples of n = 2	$\bar{x}$
0, 0	0
0, 1	.5
1, 0	.5
0, 2	1
1, 1	1
2, 0	1
0, 3	1.5
1, 2	1.5
2, 1	1.5
3, 0	1.5
3, 1	2
1, 3	2
2, 2	2
3, 2	2.5
2, 3	2.5
3, 3	3

Sample Means	Probability
0	$\frac{1}{8} \cdot \frac{1}{8} = 0.0156$
.5	$\frac{1}{8} \cdot \frac{3}{8} + \frac{3}{8} \cdot \frac{1}{8} = 0.0938$
1	$\frac{1}{8} \cdot \frac{3}{8} + \frac{3}{8} \cdot \frac{3}{8} + \frac{3}{8} \cdot \frac{1}{8} = 0.2344$
1.5	$\frac{1}{8} \cdot \frac{1}{8} + \frac{3}{8} \cdot \frac{3}{8} + \frac{3}{8} \cdot \frac{3}{8} + \frac{1}{8} \cdot \frac{1}{8} = 0.3125$
2	$\frac{1}{8} \cdot \frac{3}{8} + \frac{3}{8} \cdot \frac{1}{8} + \frac{3}{8} \cdot \frac{3}{8} = 0.2344$
2.5	$\frac{1}{8} \cdot \frac{3}{8} + \frac{3}{8} \cdot \frac{1}{8} = 0.0938$
3	$\frac{1}{8} \cdot \frac{1}{8} = 0.0156$
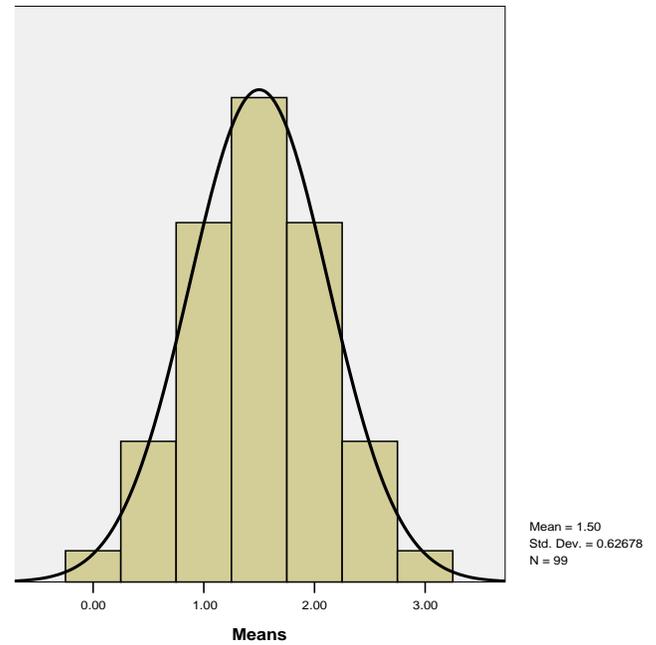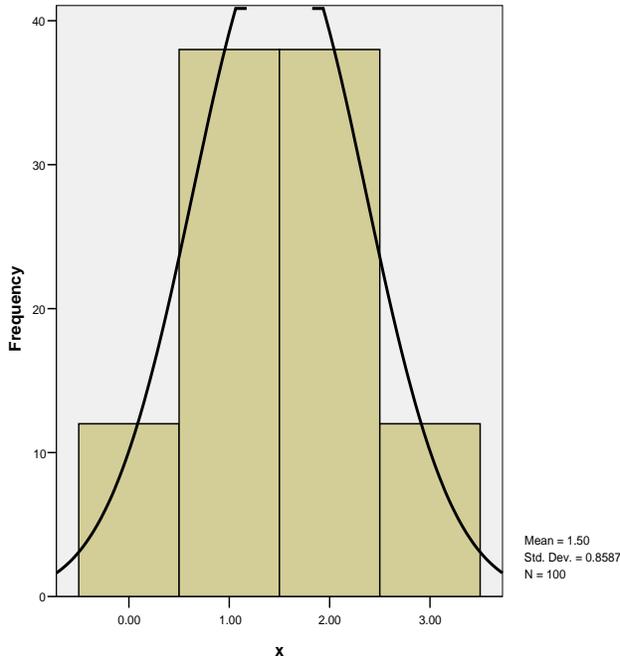
The applet below will allow you to see what the sampling distribution for $\overline{x}$ looks like under different situations:

http://www.ruf.rice.edu/%7elane/stat_sim/sampling_dist/index.html

The value of a statistic, such as the sample mean $\overline{x}$, depends on the particular values included in the sample, and generally varies from sample to sample. This variability of a statistic is called **sampling variability.**

Desirable Properties for Estimators: M.V.U.E.

A **Point Estimator** of a population parameter is a rule or formula that tells us how to use the sample data to calculate a single number that can be used as an estimate of the population parameter.

We could create many different point estimators, so there needs to be some criteria to determine which point estimators should be used.

If the sampling distribution of a sample statistic has a mean equal to the population parameter the statistic is estimating, the statistic is said to be an **unbiased estimator**.

If the mean of the sampling distribution is not equal to the parameter, the statistic is said to be a **biased estimate** of the parameter.
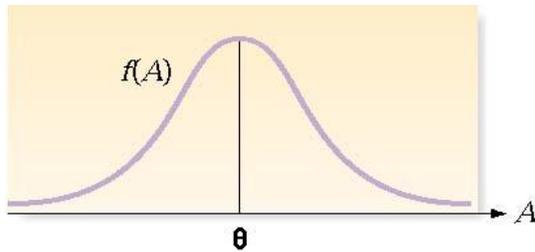


a. Unbiased sample statistic
   for the parameter θ
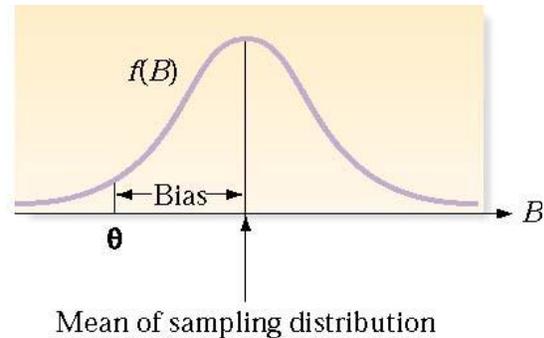
b. Biased sample statistic
   for the parameter θ

Example 92.5 If an estimator L is used to estimate λ and $E(L) = \lambda + 3.1$, is L biased or unbiased as an estimate of λ ?

Another quality we look for in an estimator is **minimum variance**.   This means that among estimators, the estimator with the minimum variance is the estimator that has the smallest standard error for its sampling distribution. We do not want to find the smallest variance among all estimators, we are only interested in the variance of the estimators that are unbiased. Putting these ideas together, we will want our point estimators to be **minimum variance unbiased estimators** (MVUE).

Example 93: Both of the estimators below are unbiased, which one would be a better choice to estimate the parameter, A or B?
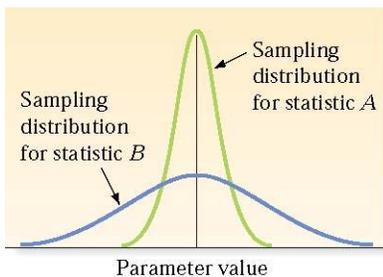
Solution: A is better. Estimator A has the better sampling distribution, since it has less variance than B.

Example 94:

X	0	1	4
P(x)	1/3	1/3	1/3

    a. Find the mean and standard deviation for this distribution.
    b. Find the sampling distribution for the sample mean for a random sample of n = 2 measurements from the distribution.
    c. Show $\bar{x}$ is an unbiased estimator of μ (show E($\bar{x}$) = $\sum \bar{x}p(\bar{x}) = \mu$)

Solution:

    a. $\mu = \dfrac{5}{3}$, $\sigma = 1.6997$

    b.

X	0	0.5	1	2	2.5	4
P(x)	1/9	2/9	1/9	2/9	2/9	1/9

    c. μ = 0/9 + 1/9 + 1/9 + 4/9 + 5/9 + 4/9 = 15/9 = 5/3

6.2 Using the Central Limit Theorem

The Sampling Distribution of $\bar{x}$ and the Central Limit Theorem

The **Central Limit Theorem** states that if random samples of size n are drawn from a non-normal population with a finite mean μ and standard deviation σ, then when n is large, the sampling distribution of the sample mean is approximately normally distributed. Regardless of the sample size the mean and standard deviation for the sample mean are as follows:

$$\mu_{\bar{x}} = E(\bar{x}) = \mu \text{ and } \sigma_{\bar{x}} = \text{the standard error of the mean} = \frac{\sigma}{\sqrt{n}}$$

The approximation gets better as n approaches infinity (i.e. n $\rightarrow \infty$)

Note: for a finite population $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}\left(\sqrt{\dfrac{N-n}{N-1}}\right)$

Question: What happens to the finite correction factor, $\sqrt{\dfrac{N-n}{N-1}}$ when N >> n? (N > > n means N is much bigger than n)

The Central Limit Theorem also applies to the statistic $\displaystyle\sum_{i=1}^{n} x_i$, and it also states that *if the sampled population is normally distributed the distribution of the sample mean will be exactly normal no matter the sample size.*

==Example 95:== How does the standard deviation of X compare to the standard error of the mean?

Solution: The standard deviation of X = σ > $\sigma_{\bar{x}}$ = Standard error of the mean. The standard error of the mean is always smaller.

Please study the figure below:
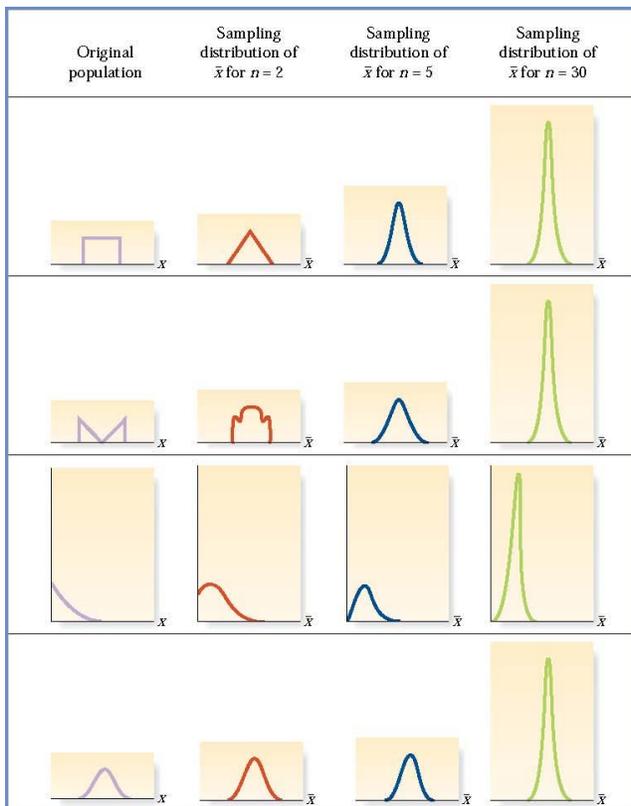
Applying the Central Limit Theorem:

1. For samples of size *n* larger than 30, the distribution of the sample means can be approximated reasonably well by a normal distribution. The approximation gets better as the sample size *n* becomes larger.

2. If the original population is itself normally distributed, then the sample means will be normally distributed for any sample size *n* (not just the values of *n* larger than 30).

Example 96: Researchers conducted a study on the age of sexual debut for different genders and ethnic groups. For the Euro-American males surveyed the average age at the time of first intercourse was 16.61 years old. The standard deviation was 2.33 years old. We are interested in the average age at the time of first intercourse for groups of 33 randomly selected Euro-American males.   Use this information to answer the questions below:

a) What is the mean of the sample means for randomly selected groups of 33 Euro-American males? What is the standard deviation of the sample means for groups of 33 Euro-American males? (i.e. – what is the standard error of the sample means?)
b) Find the probability of a sample of 33 randomly selected Euro-American males having an average age at the time of first intercourse greater than 17.

Example 97: In 2011, the highest starting salaries for recent college graduates were paid to chemical engineering majors. The average starting salary for chemical engineers in 2011 was $66,886, and the standard deviation was $7,500.   The distribution of starting salaries for chemical engineers appears to be normal.   Find the following probabilities:

a) Find the probability that one randomly selected chemical engineer receives a starting salary less than $58,000.
b) Find the probability of 36 randomly selected chemical engineers having an average starting salary less than $58,000.
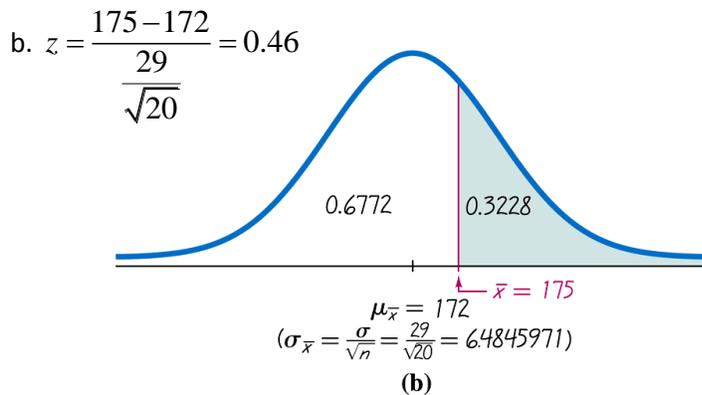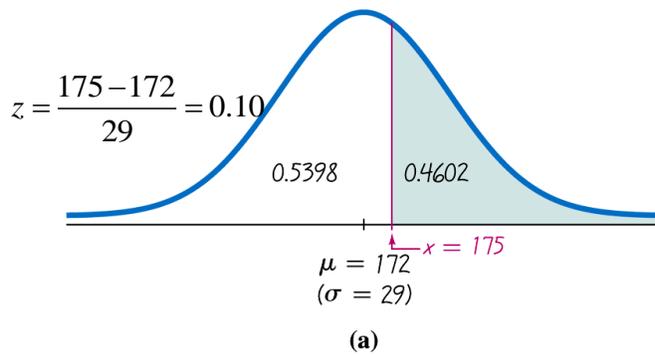
Example 98: Given the population of men has normally distributed weights with a mean of 172 lbs and a standard deviation of 29 lbs:

a) If one man is randomly selected, find the probability that his weight is greater than 175 lbs.

b) If 20 different men are randomly selected, find the probability that their mean weight is greater than 175 lbs (their total weight would then exceed the carrying capacity of an elevator).
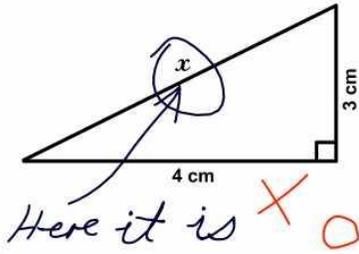
Solution:

a.

$$z = \frac{175 - 172}{29} = 0.10$$

0.5398   0.4602

$\mu = 172$
$(\sigma = 29)$

$x = 175$

(a)

b. $z = \dfrac{175 - 172}{\frac{29}{\sqrt{20}}} = 0.46$

0.6772   0.3228

$\mu_{\bar{x}} = 172$
$(\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{29}{\sqrt{20}} = 6.4845971)$

$\bar{x} = 175$

(b)

Conclusion: It is much easier for an individual to deviate from the mean than it is for a group of 20 to deviate from the mean.

A final exam in Math 160 is normally distributed and has a mean of 73 with a standard deviation of 7.8. If 24 students are randomly selected, find the probability that the mean of their test scores is less than 70.

3. Find x.

3 cm

4 cm

Here it is ✗ ○

A. 0.1006
B. -1.880
C. 0.0301
D. 0.9699

Solution: **C. 0.0301**