Introduction to Business Statistics
QMIS 220
Chapter 7
The Normal distribution

- A large number of real world phenomena are either exactly or approximately normally distributed and that leads to make the normal distribution to be the most important and widely used one.

- The normal distribution or Gaussian distribution is given by a bell-shaped curve.

- A continuous RV $x$ that has a normal distribution is said to be a normal RV with a mean $\mu$ and a standard deviation $\sigma$ or simply $x \sim N(\mu, \sigma)$.

- The mean $\mu$ and the standard deviation $\sigma$ are the parameters of the normal distribution.

- Given the values of these two parameters, we can find the area under the normal curve for any interval.
The Normal distribution

- The normal probability distribution, when plotted, gives a bell-shaped curve such that
  - The total area under the curve is 1.0.
  - The curve is symmetric around the mean.
  - The two tails of the curve extended indefinitely.
The Normal distribution

- There are a family of normal distribution.
- Each different set of values of $\mu$ and $\sigma$ gives different normal curve.
- The value of $\mu$ determines the center of the curve on the horizontal axis and the value of $\sigma$ gives the spread of the normal curve.
The Normal distribution

- Like all other distributions, the normal probability distribution can be expressed by a mathematical function

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left[ \frac{x-\mu}{\sigma} \right]^2} \]

in which the probability that \( x \) falls between \( a \) and \( b \) is the integral of the above function from \( a \) to \( b \), i.e.

\[ P(a < x < b) = \int_a^b \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left[ \frac{x-\mu}{\sigma} \right]^2} dx \]
The Normal distribution

- But as you know, we will not use the formula to find the probability. Instead, we will use a table.

Table IV  Standard Normal Distribution Table

The entries in this table give the cumulative area under the standard normal curve to the left of \( z \) with the values of \( z \) equal to 0 or negative.

<table>
<thead>
<tr>
<th>( z )</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>−3.4</td>
<td>.0003</td>
<td>.0003</td>
<td>.0003</td>
<td>.0003</td>
<td>.0003</td>
<td>.0003</td>
<td>.0003</td>
<td>.0003</td>
<td>.0003</td>
<td>.0002</td>
</tr>
<tr>
<td>−3.3</td>
<td>.0005</td>
<td>.0005</td>
<td>.0005</td>
<td>.0004</td>
<td>.0004</td>
<td>.0004</td>
<td>.0004</td>
<td>.0004</td>
<td>.0004</td>
<td>.0003</td>
</tr>
<tr>
<td>−3.2</td>
<td>.0007</td>
<td>.0007</td>
<td>.0006</td>
<td>.0006</td>
<td>.0006</td>
<td>.0006</td>
<td>.0006</td>
<td>.0005</td>
<td>.0005</td>
<td>.0005</td>
</tr>
<tr>
<td>−3.1</td>
<td>.0010</td>
<td>.0009</td>
<td>.0009</td>
<td>.0008</td>
<td>.0008</td>
<td>.0008</td>
<td>.0008</td>
<td>.0008</td>
<td>.0007</td>
<td>.0007</td>
</tr>
<tr>
<td>−3.0</td>
<td>.0013</td>
<td>.0013</td>
<td>.0013</td>
<td>.0012</td>
<td>.0012</td>
<td>.0011</td>
<td>.0011</td>
<td>.0011</td>
<td>.0010</td>
<td>.0010</td>
</tr>
<tr>
<td>−2.9</td>
<td>.0019</td>
<td>.0018</td>
<td>.0018</td>
<td>.0017</td>
<td>.0016</td>
<td>.0016</td>
<td>.0015</td>
<td>.0015</td>
<td>.0014</td>
<td>.0014</td>
</tr>
<tr>
<td>−2.8</td>
<td>.0026</td>
<td>.0025</td>
<td>.0024</td>
<td>.0023</td>
<td>.0023</td>
<td>.0023</td>
<td>.0022</td>
<td>.0021</td>
<td>.0020</td>
<td>.0020</td>
</tr>
<tr>
<td>−2.7</td>
<td>.0035</td>
<td>.0034</td>
<td>.0033</td>
<td>.0032</td>
<td>.0031</td>
<td>.0030</td>
<td>.0029</td>
<td>.0028</td>
<td>.0027</td>
<td>.0026</td>
</tr>
<tr>
<td>−2.6</td>
<td>.0047</td>
<td>.0045</td>
<td>.0044</td>
<td>.0043</td>
<td>.0041</td>
<td>.0040</td>
<td>.0039</td>
<td>.0038</td>
<td>.0037</td>
<td>.0036</td>
</tr>
<tr>
<td>−2.5</td>
<td>.0062</td>
<td>.0060</td>
<td>.0059</td>
<td>.0057</td>
<td>.0055</td>
<td>.0054</td>
<td>.0053</td>
<td>.0051</td>
<td>.0049</td>
<td>.0048</td>
</tr>
<tr>
<td>−2.4</td>
<td>.0080</td>
<td>.0078</td>
<td>.0076</td>
<td>.0075</td>
<td>.0073</td>
<td>.0071</td>
<td>.0069</td>
<td>.0068</td>
<td>.0066</td>
<td>.0064</td>
</tr>
<tr>
<td>−2.3</td>
<td>.0104</td>
<td>.0102</td>
<td>.0099</td>
<td>.0095</td>
<td>.0092</td>
<td>.0090</td>
<td>.0089</td>
<td>.0087</td>
<td>.0085</td>
<td>.0084</td>
</tr>
<tr>
<td>−2.2</td>
<td>.0136</td>
<td>.0134</td>
<td>.0132</td>
<td>.0129</td>
<td>.0125</td>
<td>.0122</td>
<td>.0119</td>
<td>.0116</td>
<td>.0113</td>
<td>.0110</td>
</tr>
<tr>
<td>−2.1</td>
<td>.0174</td>
<td>.0172</td>
<td>.0170</td>
<td>.0166</td>
<td>.0162</td>
<td>.0158</td>
<td>.0154</td>
<td>.0150</td>
<td>.0146</td>
<td>.0143</td>
</tr>
<tr>
<td>−2.0</td>
<td>.0219</td>
<td>.0217</td>
<td>.0215</td>
<td>.0212</td>
<td>.0207</td>
<td>.0202</td>
<td>.0197</td>
<td>.0192</td>
<td>.0188</td>
<td>.0183</td>
</tr>
<tr>
<td>−1.9</td>
<td>.0268</td>
<td>.0266</td>
<td>.0263</td>
<td>.0260</td>
<td>.0256</td>
<td>.0250</td>
<td>.0244</td>
<td>.0239</td>
<td>.0233</td>
<td>.0233</td>
</tr>
</tbody>
</table>
The Normal distribution

- The standard normal distribution is a special case of the normal distribution where the $\mu$ is zero and the $\sigma$ is 1.
- The RV that possesses the standard normal distribution is denoted by $z$ and it is called $z$ values or $z$ scores.
The Normal distribution

- Since $\mu$ is zero and the $\sigma$ is 1 for the standard normal, a specific value of $z$ gives the distance between the mean and the point represented by $z$ in terms of the standard deviation.

- The $z$ values to the right side of the mean are positive and those on the left are negative $BUT$ the area under the curve is always positive.

- For a value of $z = 2$, we are 2 standard deviations from the mean (to the right).

- Similarly, for $z = -2$, we are 2 standard deviations from the mean (to the left)
The Normal distribution
The Normal distribution

**Empirical rule**

![Standard normal (z) distribution](image)

*Figure 5–6. Standard normal (z) distribution.*
The Normal distribution

- The table can be used only to find the areas under the standard normal curve.

- However, in real-world applications, most of continuous RVs that are normally distributed come with mean and standard deviation different from 0 and 1, respectively.

- What shall we do? Is there any way to bring \( \mu \) to zero and \( \sigma \) to 1?

- Yes, it can be done by subtracting \( \mu \) from \( x \) and dividing the result by \( \sigma \) (standardizing)
The Normal distribution

- **Standardizing \( x \)**

For a normal RV \( x \) with mean \( \mu \) and standard deviation \( \sigma \). The standardized RV \( z \) can be found using the following formula 

\[
z = \frac{x - \mu}{\sigma}
\]

- **To find the area between two values of \( x \) for a normal distribution**
  - Convert both values of \( x \) to their respective \( z \) values
  - Find the area under the standard normal curve between those two values.

- **We reverse the procedure of finding the area under the normal curve for a specific value of \( z \) or \( x \) to finding a specific value of \( z \) or \( x \) for a known area under the normal curve.**
Example: Find a point \( z \) such that the area under the standard normal curve between 0 and \( z \) is .4251 and the value of \( z \) is positive.
The Normal distribution

**Finding an x value for a normal distribution:**

- To find an x value when an area under a normal distribution curve is given, we do the following

1. Find the z value corresponding to that x value from the standard normal curve.

2. Transform the z value to x by substituting the values of μ, σ, and z in the following formula

\[ x = \mu + \sigma z \]
Population and sampling distribution

Random Sampling:

- The term *random sampling* refers to a sampling procedure where every member in the population has a chance of being selected.

- The objective of the sampling procedure is to ensure that the final sample is representative of the population from which it was taken.

- A biased sample is a sample that doesn’t represent the intended population and can lead to distorted findings.
Population and sampling distribution

Simple Random Sampling (SRS):

- A simple random sample is a sample in which every member of the population has an equal chance of being selected.

- Unfortunately, this is easier said than done!

- Using a random numbers table we can start randomly selecting our elements.
Population and sampling distribution

If we want to conduct a survey in a mall using SRS:

- Depending on the population size, we can choose the 1st digits from each random number.
- Suppose the were 1000 shoppers in the mall, from which we were drawing a sample size of 100.
- We would list these shoppers with the assigned number: 24, 19, 46, 17, 22, 93, 52, 62….and so forth.
Systematic Sampling:

- One way to avoid a personal bias when selecting people at random is to use *symmetric sampling*.
- This technique results in selecting every $k^{th}$ member of the population to be in the sample.
- $K \approx \frac{N}{n}$, where $N$ is the population size and $n$ is the sample size.
- Much easier and cheaper to conduct.
- Very dangerous if there is a pattern in the population.
Population and sampling distribution

Cluster Sampling:

- If the population can be divided into groups, or clusters, then simple random sample can be selected from these clusters to form the final sample.

- These clusters could be classes in a school or stores in a mall.

- Once the cluster is selected, all the elements inside that cluster must be included in the sample.
Population and sampling distribution

Stratified Sampling:

- In *stratified sampling*, we divide the population into mutually exclusive groups, or strata, and randomly sample from each of these groups.
- Using our mall example, our strata could be defined as male and female shoppers.
- We can be sure that our final sample contains a number of male and female shoppers with the same proportion in the population.
- Other examples of stratus: age, income, occupation.
Population and sampling distribution

**Sampling Error:**

- So far, we have stressed the benefits of drawing a sample from a population.
- However, in statistics, as in life, there's no such thing as a free lunch.
- By sampling, we expose ourselves to errors that can lead to inaccurate conclusions about the population.
- The type of error that a statistician is most concerned about is called *sampling error*, which occurs when the sample measurement is different from the population measurement.
Population and sampling distribution

- Because the entire population is rarely measured, the sampling error cannot be directly calculated.

- However, with inferential statistics, we'll be able to assign probabilities to certain amounts of sampling error later.

- Sampling errors occur because we might have the unfortunate luck of selecting a sample that is not a perfect match to the entire population.

- Sampling errors are expected and usually are a small price to pay to avoid measuring an entire population.

- One way to reduce the sampling error of a statistical study is to increase the size of the sample.
Population and sampling distribution

- In general, the larger the sample size, the smaller the sampling error.
- If you increase the sample size until it reaches the size of the population, then the sampling error will be reduced to 0.
- But in doing so, we lose the benefits of sampling.
Population and sampling distribution

**Examples of poor samplings techniques:**

- The technique of sampling has been widely used, both properly and improperly, in the area of politics.
- One of the most famous mistakes with sampling occurred during the 1936 presidential race where the *Literary Digest* predicted Alf Landon to win the election over Franklin D. Roosevelt.
- Sure somebody had an egg on his face after this Election Day.
Population and sampling distribution

Population distribution:

- The population distribution is the probability distribution derived from the information on all elements of a population.

Example:

- Suppose there are only five students in a MBA class and the midterm scores of these five students are.

<table>
<thead>
<tr>
<th>Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>70</td>
</tr>
<tr>
<td>B</td>
<td>78</td>
</tr>
<tr>
<td>C</td>
<td>80</td>
</tr>
<tr>
<td>D</td>
<td>80</td>
</tr>
<tr>
<td>E</td>
<td>95</td>
</tr>
</tbody>
</table>

QMIS 220, CH 7 by M. Zainal
Population and sampling distribution

- The population frequency and relative frequency distributions table of the scores is:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$f$</th>
<th>$R.f.$</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>1</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>78</td>
<td>1</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>80</td>
<td>2</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>95</td>
<td>1</td>
<td>0.20</td>
<td>0.20</td>
</tr>
</tbody>
</table>

$N = 5 \quad \text{Sum} = 1.00 \quad \Sigma P(x) = 1.00$

- The values of the mean and the standard deviation calculated for the probability distribution above give $\mu = 80.60$ and $\sigma = 8.09$ (How & Why?)

- Are these values constant?
Population and sampling distribution

**Sampling distribution:**

- Suppose we want to draw a sample (without replacement) of three students from that class and see how much is their mean and standard deviation.

- In this case, we will have 10 different samples.
  
  ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE

- Can you tell how?

- Each sample will have different mean and standard deviation depending on the elements included in the sample.
### Population and sampling distribution

<table>
<thead>
<tr>
<th>Sample</th>
<th>Scores</th>
<th>$\bar{x}$</th>
<th>Frequency and relative frequency distributions</th>
<th>Sampling distribution of $x$ for $n = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\bar{x}$</td>
<td>$f$</td>
</tr>
<tr>
<td>ABC</td>
<td>70, 78, 80</td>
<td>76.00</td>
<td>76.00</td>
<td>2</td>
</tr>
<tr>
<td>ABD</td>
<td>70, 78, 80</td>
<td>76.00</td>
<td>76.00</td>
<td>1</td>
</tr>
<tr>
<td>ABE</td>
<td>70, 78, 95</td>
<td>81.00</td>
<td>79.33</td>
<td>1</td>
</tr>
<tr>
<td>ACD</td>
<td>70, 80, 80</td>
<td>76.67</td>
<td>81.00</td>
<td>1</td>
</tr>
<tr>
<td>ACE</td>
<td>70, 80, 95</td>
<td>81.67</td>
<td>84.33</td>
<td>2</td>
</tr>
<tr>
<td>ADE</td>
<td>70, 80, 95</td>
<td>81.67</td>
<td>84.33</td>
<td>2</td>
</tr>
<tr>
<td>BCD</td>
<td>78, 80, 80</td>
<td>79.33</td>
<td>85.00</td>
<td>1</td>
</tr>
<tr>
<td>BCE</td>
<td>78, 80, 95</td>
<td>84.33</td>
<td>85.00</td>
<td>1</td>
</tr>
<tr>
<td>BDE</td>
<td>78, 80, 95</td>
<td>84.33</td>
<td>85.00</td>
<td>1</td>
</tr>
<tr>
<td>CDE</td>
<td>80, 80, 95</td>
<td>85.00</td>
<td>85.00</td>
<td>1</td>
</tr>
</tbody>
</table>

QMIS 220, CH 7 by M. Zainal
Sampling and nonsampling errors

**Sampling errors:**
- The difference between the sample mean (statistic) and the population mean (parameter) is called *sampling error*.
- This difference is only due to the chance of including some elements and excluding others in the random sample only.

\[
\text{Sampling error} = \bar{x} - \mu
\]

**Nonsampling errors:**
- The error that occurs in the collection, recording, and tabulation of data are called *nonsampling errors*

\[
\text{Nonsampling error} = \text{Incorrect } \bar{x} - \text{ correct } \bar{x}
\]
Example:

Reconsider the population of the five scores. Suppose one sample of three scores is selected, and the sample is 70, 80, and 95.

1. Find the sampling error

2. Suppose we mistakenly record the second score as 82 instead of 80. Find the nonsampling error
Mean and standard deviation of $\bar{x}$

- If we calculate the mean and the standard deviation of all possible samples (with the same size $n$) mean $\mu_{\bar{x}}$ and standard deviations, we obtain the mean and the standard deviation $\sigma_{\bar{x}}$ of $\bar{x}$.

- The mean of the sampling distribution of $\bar{x}$ is always equal to the mean of the population. Thus,

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

- The sample mean is an estimator of the population mean and if they are equal, the statistic is said to be say it is unbiased estimator.
Mean and standard deviation of $\bar{x}$

- The standard deviation $\sigma_{\bar{x}}$, of $\bar{x}$ is not equal to the standard deviation, $\sigma$, of the population distribution (unless $n = 1$)

- To find the standard deviation of $\bar{x}$ we use:

$$\sigma_{\bar{x}} = \begin{cases} \frac{\sigma}{\sqrt{n}} & \text{if } \frac{n}{N} \leq 0.05 \\ \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} & \text{Otherwise} \end{cases}$$
Mean and standard deviation of $\bar{x}$

**Example:** The mean wage per hour for all 5000 employees who work at a large company is $17.50 and the standard deviation is $2.9. Let $\bar{x}$ be the wage per hour for a random sample of certain employees selected from this company. Find the mean and the standard deviation of $\bar{x}$ for a sample size of

a) 30  b) 75  c) 200
Mean and standard deviation of $\bar{x}$

**Example:** The living spaces of all homes in a city have a mean of 2300 square feet and a standard deviation of 450 square feet. Let $\bar{x}$ be the mean living space for a random sample of 25 homes selected from this city. Find the mean and the standard deviation of the sampling distribution $\bar{x}$.
Shape of the sampling distribution of $\bar{x}$

- The shape of the sampling distribution of $\bar{x}$ depends on whether the population from which samples are drawn has a normal distribution or not.

**Sampling from a normally distributed population**

- If the sample is drawn from a normally distributed population with a mean $\mu$ and a standard deviation $\sigma$, then the sampling distribution of $\bar{x}$ will be normally distributed with the following mean and standard deviation, irrespective of the sample size:

$$
\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
$$
Shape of the sampling distribution of $\bar{x}$

2. Sampling distribution of $\bar{x}$ for $n = 5$.
3. Sampling distribution of $\bar{x}$ for $n = 16$.
4. Sampling distribution of $\bar{x}$ for $n = 30$.
5. Sampling distribution of $\bar{x}$ for $n = 100$. 

Normal distribution
Shape of the sampling distribution of $\overline{x}$

**Sampling from a population that is not normally distributed**

- Central limit theorem (CLT): For a large sample size, the sampling distribution of $\overline{X}$ is approximately normal, irrespective of the shape of the population distribution.

- The mean and standard deviation of $\overline{X}$ are

$$
\mu_{\overline{x}} = \mu \quad \text{and} \quad \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{for } n \geq 30
$$
Shape of the sampling distribution of $\bar{x}$

(a) Several sampled populations

(b) Corresponding populations of all possible sample means for different sample sizes
**Shape of the sampling distribution of $\bar{x}$**

**Example:** In a recent SAT, the mean score for all examinees was 1020. Assume that the distribution of SAT scores of all examinees is normal of 1020 and a standard deviation of 135. Let $\bar{x}$ be the mean SAT score of a random sample of certain examinees. Calculate the mean and the standard deviation of $\bar{x}$ and describe the shape of its sampling distribution when the sample size is

a) 16  

b) 50  

c) 1000
Example: The weight of all people living in a town has a distribution that is skewed to the right with a mean of 133 pounds and a standard deviation of 24 pounds. Let $\bar{x}$ be the mean weight of a random sample of 45 persons selected from this town. Find the mean and standard deviation of $\bar{x}$ and comment on the shape of its sampling distribution.
Application of the sampling distribution of $\bar{x}$

Based on the CLT, we can make the following statements:

$$P(\mu - 1\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 1\sigma_{\bar{x}}) = .6826$$

$$P(\mu - 2\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 2\sigma_{\bar{x}}) = .9544$$

$$P(\mu - 3\sigma_{\bar{x}} \leq \bar{x} \leq \mu + 3\sigma_{\bar{x}}) = .9974$$
Application of the sampling distribution of $\bar{x}$

**Example:** Assume that the weights of all packages of a certain brand of cookies are normally distributed with a mean of 32 ounces and a standard deviation of .3 ounce. Find the probability that the mean weight, $\bar{x}$, of a random sample of 20 packages of this brand of cookies will be between 31.8 and 31.9 ounces.
Application of the sampling distribution of $\bar{x}$

**Example**: The time that college students spend studying per week have a distribution that is skewed to the right with a mean 8.4 hours and a standard deviation of 2.7 hours. Find the probability that the mean time spent studying per week for a random sample of 45 students would be

a) Within 1 hour from the mean

b) Between 8 and 9 hours

c) Less than 8 hours
Population and sample proportion

What if we are dealing with qualitative variable?

Population proportion, denoted by $p$, is obtained by taking the ratio of the number of elements in a population with a specific characteristic to the total number of elements in the population. It is calculated as

$$p = \frac{X}{N}$$

Sample proportion, denoted by $\hat{p}$ (pronounced by p hat) gives a similar ratio for a sample and it is given by

$$\hat{p} = \frac{x}{n}$$
Population and sample proportion

Example: Suppose a total of 20,000 students are registered in KU for this semester and 14,500 of them own a laptop. A sample of 120 students is selected and 78 of them found to own a laptop. Find

a) the proportion of students who own a laptop in the population

b) the proportion of students who own a laptop in the sample

c) the sampling error
Population and sample proportion

**Sampling distribution of \( \hat{p} \).**

- Like in the sample mean \( \bar{x} \), the sample proportion \( \hat{p} \) is a random variable that possesses a probability distribution which is called its sampling distribution.

- Remember that the probability distribution gives the various values that a RV may assume and their probabilities.

- The value of \( \hat{p} \) for a particular sample depends on what elements of the population are included in that sample.
Population and sample proportion

Mean, standard deviation, and shape of the sampling distribution of $\hat{p}$

Example: The following table gives the names of the MBA students class along with their opinion of whether or not they like Statistics.

<table>
<thead>
<tr>
<th>Name</th>
<th>opinion</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>yes</td>
</tr>
<tr>
<td>B</td>
<td>no</td>
</tr>
<tr>
<td>C</td>
<td>no</td>
</tr>
<tr>
<td>D</td>
<td>yes</td>
</tr>
<tr>
<td>E</td>
<td>yes</td>
</tr>
</tbody>
</table>
Population and sample proportion

Mean, standard deviation, and shape of the sampling distribution of $\hat{p}$

- The population proportion is $p = 3/5 = .60$
- The following table lists the 10 possible samples and the proportion of students who like statistics for each of those samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$\hat{p}$</th>
<th>Sample</th>
<th>$\hat{p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>1/3 = .33</td>
<td>ADE</td>
<td>3/3 = 1.00</td>
</tr>
<tr>
<td>ABD</td>
<td>2/3 = .67</td>
<td>BCD</td>
<td>1/3 = .33</td>
</tr>
<tr>
<td>ABE</td>
<td>2/3 = .67</td>
<td>BCE</td>
<td>1/3 = .33</td>
</tr>
<tr>
<td>ACD</td>
<td>2/3 = .67</td>
<td>BDE</td>
<td>2/3 = .67</td>
</tr>
<tr>
<td>ACE</td>
<td>2/3 = .67</td>
<td>CDE</td>
<td>2/3 = .67</td>
</tr>
</tbody>
</table>
## Population and sample proportion

### Mean, standard deviation, and shape of the sampling distribution of \( \hat{p} \)

<table>
<thead>
<tr>
<th>( \hat{p} )</th>
<th>( f )</th>
<th>( R.f )</th>
<th>( \hat{p} )</th>
<th>( P(\hat{p}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.33</td>
<td>3</td>
<td>3/10 = .30</td>
<td>.33</td>
<td>0.30</td>
</tr>
<tr>
<td>.67</td>
<td>6</td>
<td>6/10 = .60</td>
<td>.67</td>
<td>0.60</td>
</tr>
<tr>
<td>1.00</td>
<td>1</td>
<td>1/10 = .10</td>
<td>1.00</td>
<td>0.10</td>
</tr>
<tr>
<td>( \Sigma f = 10 )</td>
<td>( \Sigma P(x) = 1.00 )</td>
<td>( \Sigma f = 10 )</td>
<td>( \Sigma P(x) = 1.00 )</td>
<td></td>
</tr>
</tbody>
</table>
Population and sample proportion

**Mean, standard deviation, and shape of the sampling distribution of \( \hat{p} \)**

- The mean of \( \hat{p} \) is always equal to the population proportion. That is
  \[
  \mu_{\hat{p}} = p
  \]
  In which it is said to be unbiased estimator of the population proportion.

- The standard deviation of \( \hat{p} \) is given by
  \[
  \sigma_{\hat{p}} = \begin{cases} \sqrt{\frac{pq}{n}} & \text{if } \frac{n}{N} \leq 0.05 \\ \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}} & \text{Otherwise} \end{cases}
  \]
Population and sample proportion

**Mean, standard deviation, and shape of the sampling distribution of \( \hat{p} \)**

- According to the CLT, the shape of \( \hat{p} \) and the sampling distribution of \( \hat{p} \) will be approximately normal if

  \[ np > 5 \quad \text{and} \quad nq > 5 \]

**Example:** The National survey of Student Engagement shows about 87% of freshmen and seniors rate their college experience as “good” or “excellent”. Let \( \hat{p} \) be the proportion of freshmen and seniors in a random sample of 900 who hold this view. Find the mean and the standard deviation of \( \hat{p} \) and describe the shape of its sampling distribution.
Applications of the sampling distribution of $\hat{p}$

Example: According to a 2002 University of Michigan survey, only about one third of Americans expected the next five years to bring continuous good times. Assume that 33% of the current population of all Americans hold this opinion. Let $\hat{p}$ be the proportion in a random sample of 800 Americans who will hold this opinion. Find the probability that the value of $\hat{p}$ is between .35 and .37.
Population and sample proportion

Applications of the sampling distribution of $\hat{p}$

Example: Ahmad Ali, who was running for the parliament last summer, thinks that he is favored by 53% of all eligible voters in his district. Assume that what he think is true. What is the probability that in a random sample of 400 voters attended one of his campaign speeches, less than 49% will favor him.