TOPIC-9
TEST OF HYPOTHESIS
LARGE SAMPLE
Introduction to Hypothesis Testing
INTRODUCTION

• Suppose that a pharmaceutical company is concerned that the mean potency \( \mu \) of an antibiotic meet the minimum government potency standards. They need to decide between two possibilities:

1. The mean potency \( \mu \) does not exceed the mean allowable potency.
2. The mean potency \( \mu \) exceeds the mean allowable potency.

• This is an example of a test of hypothesis.
INTRODUCTION

- Hypothesis simply as a statement about one or more populations.

- Hypothesis of interest here are those that are concerned with one or more parameters of the population(s) about which we are making the statement.

- An advertising executive may hypothesize that a certain type of newspaper advertisement attracts a larger proportion of readers than some other type of ad.

- A quality control engineer may hypothesize that the variance of the measurements generated by some process is equal to some specific value.

- A marketing analyst may hypothesize that the mean family income in a certain area is some specific value.
INTRODUCTION

• Similar to a courtroom trial. In trying a person for a crime, the jury needs to decide between one of two possibilities:
  • The person is guilty.
  • The person is innocent.
• To begin with, the person is assumed innocent.
• The prosecutor presents evidence, trying to convince the jury to reject the original assumption of innocence, and conclude that the person is guilty.
Assume the population mean age is 50. (Null Hypothesis)

Is $\bar{X} = 20 \approx \mu = 50$?

No, not likely!

REJECT

Null Hypothesis

The Sample Mean Is 20

Sample

Population
Parts of a Statistical Test

1. The hypothesis
   a) The null hypothesis, $H_0$: hypothesis that is tested
      • Assumed to be true until we can prove otherwise.
   b) The alternative hypothesis, $H_a$:
      • Will be accepted as true if we can disprove $H_0$

<table>
<thead>
<tr>
<th>Court trial:</th>
<th>Pharmaceuticals:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$: innocent</td>
<td>$H_0$: $\mu$ does not exceed allowed amount</td>
</tr>
<tr>
<td>$H_a$: guilty</td>
<td>$H_a$: $\mu$ exceeds allowed amount</td>
</tr>
</tbody>
</table>
PARTS OF A STATISTICAL TEST (CONT’D)

2. **The test statistic or its \( p \)-value:**
   - A single statistic calculated from the sample which will allow us to reject or not reject \( H_0 \), and
   - A probability, calculated from the test statistic that measures whether the test statistic is likely or unlikely, assuming \( H_0 \) is true.
   - Smallest Value of \( \alpha \) so that the \( H_0 \) Can Be Rejected

3. **The rejection region:**
   - A rule that tells us for which values of the test statistic, or for which \( p \)-values, the null hypothesis should be rejected.

4. **Conclusion:**
   - Either “**Reject** \( H_0 \)” or “**Fail to reject** \( H_0 \)”, along with a statement about the reliability of your conclusion.
How do you decide when to reject $H_0$?

- Depends on the significance level, $\alpha$, the maximum tolerable risk you want to have of making a mistake, if you decide to reject $H_0$. Usually, the significance level is $\alpha = .01$ or $\alpha = .05$.
- Used to **Rejection Rule**

1. **Based on $t$ or $z$ statistics**
   
   $|t| > t_{\text{table}}$
   
   $|z| > z_{\text{table}}$

   Reject $H_0$

2. **Based on $p$-value**

   $p < \alpha$

   Reject $H_0$
**Parts of a Statistical Test (cont’d)**

About Conclusion

- If we **reject the null** hypothesis, we conclude that there is enough evidence to infer that the alternative hypothesis is true.
- If we **fail to reject the null** hypothesis, we conclude that there is not enough statistical evidence to infer that the alternative hypothesis is true. **This does not mean that we have proven that the null hypothesis is true!**
### Result Possibilities

<table>
<thead>
<tr>
<th>Decision</th>
<th>Actual Situation</th>
<th>H₀ is true</th>
<th>H₁ is true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail to Reject H₀</td>
<td>1 − α</td>
<td>β</td>
<td>Type II Error</td>
</tr>
<tr>
<td>Reject H₀</td>
<td>α</td>
<td>1 − β</td>
<td>Power</td>
</tr>
</tbody>
</table>

**Define:**

\[
\alpha = P(\text{Type I error}) = P(\text{reject } H₀ \text{ when } H₀ \text{ is true})
\]

\[
\beta = P(\text{Type II error}) = P(\text{accept } H₀ \text{ when } H₁ \text{ is true})
\]
**Two Types of Errors**

We want to keep the probabilities of error as small as possible.

- **The value of \( \alpha \) is the significance level, and is controlled by the experimenter.**
- **The value of \( \beta \) is difficult to calculate.**

Rather than “accepting \( H_0 \)” as true without being able to provide a measure of goodness, we choose to “not reject” \( H_0 \).

We write: There is insufficient evidence to reject \( H_0 \).
\( \alpha \) & \( \beta \) Have an Inverse Relationship

Reduce probability of one error and the other one goes up.
EXAMPLE

- The mayor of a small city claims that the average income in his city is $35,000 with a standard deviation of $5000. We take a sample of 64 families, and find that their average income is $30,000. Is his claim correct?

1. We want to test the hypothesis:

   \( \mathcal{H}_0 : \mu = 35,000 \) (mayor is correct) versus

   \( \mathcal{H}_a : \mu \neq 35,000 \) (mayor is wrong)

   Start by assuming that \( \mathcal{H}_0 \) is true and \( \mu = 35,000 \).
2. The best estimate of the population mean \( m \) is the sample mean, \$30,000:

- From the Central Limit Theorem the sample mean has an approximate normal distribution with mean \( \mu = 35,000 \) and standard error \( SE = 5000/8 = 625 \).
- The sample mean, \$30,000 lies \( z = (30,000 - 35,000)/625 = -8 \) standard deviations below the mean.
- The probability of observing a sample mean this far from \( \mu = 35,000 \) (assuming \( H_0 \) is true) is nearly zero.
EXAMPLE (CONT’D)

3. From the Empirical Rule, values more than three standard deviations away from the mean are considered **extremely unlikely**. Such a value would be extremely unlikely to occur if indeed $H_0$ is true, and would give reason to reject $H_0$.

Since the observed sample mean, $30,000 is so unlikely, we choose to **reject $H_0$: $\mu = 35,000** and conclude that the mayor’s claim is incorrect.

4. The probability that $m = 35,000$ and that we have observed such a small sample mean just by chance is **nearly zero**.
Take a random sample of size $n \geq 30$ from a population with mean $\mu$ and standard deviation $\sigma$.

We assume that either

1. $\sigma$ is known or
2. $s \approx \sigma$ since $n$ is large

The hypothesis to be tested is

- $H_0: \mu = \mu_0$ versus $H_a: \mu \neq \mu_0$
Assume to begin with that $H_0$ is true. The sample mean $\bar{x}$ is our best estimate of $\mu$, and we use it in a standardized form as the test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \approx \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

since $\bar{x}$ has an approximate normal distribution with mean $\mu_0$ and standard deviation $\sigma / \sqrt{n}$. 
**Test Statistic**

- If $H_0$ is true the value of $\bar{x}$ should be close to $\mu_0$, and $z$ will be close to 0. If $H_0$ is false, $z$ will be much larger or smaller than $\mu_0$, and $z$ will be much larger or smaller than 0, indicating that we should reject $H_0$. 
L I K E L Y  O R  U N L I K E L Y ?

• Once you’ve calculated the observed value of the test statistic, calculate its \( p \)-value:

\[ \text{\textbf{p-value:}} \text{ The probability of observing, just by chance, a test statistic as extreme or even more extreme than what we’ve actually observed. If } H_0 \text{ is rejected this is the actual probability that we have made an incorrect decision.} \]

- If this probability is very small, less than some \textbf{pre-assigned significance level, } \( \alpha \), \( H_0 \) can be rejected.


**Parts of a Statistical Test**

1. **Null hypothesis**: a contradiction of the alternative hypothesis

2. **Alternative hypothesis**: the hypothesis the researcher wants to support.

3. **Test statistic and its p-value**: sample evidence calculated from sample data.

4. **Rejection region**—critical values and significance levels: values that separate rejection and nonrejection of the null hypothesis

5. **Conclusion**: Reject or do not reject the null hypothesis, stating the practical significance of your conclusion.
Part - I
Large Sample Test of Hypothesis:

One Population mean
and
one population proportion
The daily yield for a chemical plant has averaged 880 tons for several years. The quality control manager wants to know if this average has changed. She randomly selects 50 days and records an average yield of 871 tons with a standard deviation of 21 tons.

\[ H_0 : \mu_0 = 880 \]
\[ H_a : \mu_0 \neq 880 \]

Test statistic:
\[
z \approx \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{871 - 880}{21 / \sqrt{50}} = -3.03
\]
What is the probability that this test statistic or something even more extreme (far from what is expected if $H_0$ is true) could have happened *just by chance*?

**Example**

$p$-value : $P(z > 3.03) + P(z < -3.03)$

$= 2P(z < -3.03) = 2(.0012) = .0024$

This is an unlikely occurrence, which happens about 2 times in 1000, assuming $\mu = 880!$
EXAMPLE

To make our decision clear, we choose a significance level, say \( \alpha = .01 \).

If the \( p \)-value is less than \( \alpha \), \( H_0 \) is rejected as false. You report that the results are statistically significant at level \( \alpha \).

If the \( p \)-value is greater than \( \alpha \), \( H_0 \) is not rejected. You report that the results are not significant at level \( \alpha \).

Since our \( p \)-value = .0024 is less than, we reject \( H_0 \) and conclude that the average yield has changed.
USING A REJECTION REGION

If $\alpha = .01$, what would be the **critical value** that marks the “dividing line” between “not rejecting” and “rejecting” $H_0$?

- If $p$-value $< \alpha$, $H_0$ is rejected.
- If $p$-value $> \alpha$, $H_0$ is not rejected.

The dividing line occurs when $p$-value $= \alpha$. This is called the **critical value** of the test statistic.

- Test statistic $> \text{critical value}$ implies $p$-value $< \alpha$, $H_0$ is rejected.
- Test statistic $< \text{critical value}$ implies $p$-value $> \alpha$, $H_0$ is not rejected.
EXAMPLE

What is the critical value of $z$ that cuts off exactly $\alpha/2 = .01/2 = .005$ in the tail of the $z$ distribution?

Rejection Region: Reject $H_0$ if $z > 2.58$ or $z < -2.58$. If the test statistic falls in the rejection region, its $p$-value will be less than $\alpha = .01$.

For our example, $z = -3.03$ falls in the rejection region and $H_0$ is rejected at the 1% significance level.
ONE TAILED TESTS

- Sometimes we are interested in detecting a specific directional difference in the value of m.
- The alternative hypothesis to be tested is one tailed:
  - $H_a: \mu > \mu_0$ or $H_a: \mu < \mu_0$
- Rejection regions and $p$-values are calculated using only one tail of the sampling distribution.
Example

A homeowner randomly samples 64 homes similar to her own and finds that the average selling price is $252,000 with a standard deviation of $15,000. Is this sufficient evidence to conclude that the average selling price is greater than $250,000? Use $\alpha = .01$.

$H_0 : \mu_0 = 250,000$
$H_a : \mu_0 > 250,000$

Test statistic:

$$z \approx \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{252,000 - 250,000}{15,000 / \sqrt{64}} = 1.07$$
Critical Value Approach

What is the critical value of $z$ that cuts off exactly $\alpha = .01$ in the right-tail of the $z$ distribution?

For our example, $z = 1.07$ does not fall in the rejection region and $H_0$ is not rejected. There is not enough evidence to indicate that $\mu$ is greater than $250,000$.

**Rejection Region:** Reject $H_0$ if $z > 2.33$. If the test statistic falls in the rejection region, its $p$-value will be less than $\alpha = .01$. 

**P-Value Approach**

- The probability that our sample results or something even more unlikely would have occurred *just by chance*, when \( \mu = 250,000 \).

\[
p\text{-value : } P(z > 1.07) = 1 - .8577 = .1423
\]

Since the \( p \)-value is greater than \( \alpha = .01 \), \( H_0 \) is not rejected. There is insufficient evidence to indicate that \( m \) is greater than $250,000.
STATISTICAL SIGNIFICANCE

- The critical value approach and the $p$-value approach produce identical results.
- The $p$-value approach is often preferred because
  - Computer printouts usually calculate $p$-values
  - You can evaluate the test results at any significance level you choose.
- What should you do if you are the experimenter and no one gives you a significance level to use?
**Statistical Significance**

- If the *p*-value is less than *.01*, reject $H_0$. The results are highly significant.
- If the *p*-value is between *.01* and *.05*, reject $H_0$. The results are statistically significant.
- If the *p*-value is between *.05* and *.10*, do not reject $H_0$. But, the results are tending towards significance.
- If the *p*-value is greater than *.10*, do not reject $H_0$. The results are not statistically significant.
TESTING A BINOMIAL PROPORTION: ONE PROPORTION

A random sample of size $n$ from a binomial population to test

$H_0 : p = p_0$ versus

$H_a :$ one of three alternatives

Test statistic: $z \approx \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}}$

with rejection regions and/or $p$-values based on the standard normal $z$ distribution.
EXAMPLE

Regardless of age, about 20% of American adults participate in fitness activities at least twice a week. A random sample of 100 adults over 40 years old found 15 who exercised at least twice a week. Is this evidence of a decline in participation after age 40? Use $\alpha = .05$.

$H_0 : p = .2$
$H_a : p < .2$

Test statistic :

$$z \approx \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}} = \frac{.15 - .2}{\sqrt{\frac{.2(.8)}{100}}} = -1.25$$
CRITICAL VALUE APPROACH
What is the critical value of \( z \) that cuts off exactly \( \alpha = .05 \) in the left-tail of the \( z \) distribution?

For our example, \( z = -1.25 \) does not fall in the rejection region and \( H_0 \) is not rejected. There is not enough evidence to indicate that \( p \) is less than .2 for people over 40.

**Rejection Region:** Reject \( H_0 \) if \( z < -1.645 \). If the test statistic falls in the rejection region, its \( p \)-value will be less than \( \alpha = .05 \).
Part-II
Large Sample Test of Hypothesis:

The Different between

Two Population Means
And
Two Population Proportion
TESTING THE DIFFERENCE BETWEEN TWO MEANS

A random sample of size $n_1$ drawn from population 1 with mean $\mu_1$ and variance $\sigma_1^2$.

A random sample of size $n_2$ drawn from population 2 with mean $\mu_2$ and variance $\sigma_2^2$.

• The hypothesis of interest involves the difference, $\mu_1 - \mu_2$, in the form:
  • $H_0$: $\mu_1 - \mu_2 = \Delta_0$ versus $H_a$: one of three
    where $\Delta_0$ is some hypothesized difference, usually 0.
THE SAMPLING DISTRIBUTION OF $\bar{x}_1 - \bar{x}_2$

• Applying the laws of expected value and variance we have:

\[
E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2
\]

\[
V(\bar{x}_1 - \bar{x}_2) = V(\bar{x}_1) + V(\bar{x}_2) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}
\]

• We can define:

\[
Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}
\]
**TESTING THE DIFFERENCE BETWEEN TWO MEANS**

\[ H_0 : \mu_1 - \mu_2 = \Delta_0 \]

\[ H_a : \text{one of three alternatives} \]

Test statistic: 
\[
z \approx \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\]

with rejection regions and/or \(p\)-values based on the standard normal \(z\) distribution.
**EXAMPLE**

<table>
<thead>
<tr>
<th>Avg Daily Intakes</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Sample mean</td>
<td>756</td>
<td>762</td>
</tr>
<tr>
<td>Sample Std Dev</td>
<td>35</td>
<td>30</td>
</tr>
</tbody>
</table>

- Is there a difference in the average daily intakes of dairy products for men versus women? Use $a = .05$.

$$H_0 : \mu_1 - \mu_2 = 0 \text{ (same)} \quad H_a : \mu_1 - \mu_2 \neq 0 \text{ (different)}$$

Test statistic:

$$z \approx \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{756 - 762 - 0}{\sqrt{\frac{35^2}{50} + \frac{30^2}{50}}} = -0.92$$
**P-Value Approach**

- The probability of observing values of $z$ that are as far away from $z = 0$ as we have, *just by chance*, if indeed $\mu_1 - \mu_2 = 0$.

$$p\text{-value} : P(z > 0.92) + P(z < -0.92)$$

$$= 2(0.1788) = 0.3576$$

Since the $p$-value is greater than $\alpha = 0.05$, $H_0$ is not rejected. There is insufficient evidence to indicate that men and women have different average daily intakes.
THE SAMPLING DISTRIBUTION OF $\hat{p}_1 - \hat{p}_2$

- Proportions observed in independent random samples are independent. Thus, we can add their variances.
- The Mean is $\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$

$$\sigma^2_{\hat{p}_1 - \hat{p}_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$$

- Thus, the Standard deviation is $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$

- So, the $\hat{p}_1 - \hat{p}_2$ is approximately normally distributed for large $n_1$ and $n_2$
**Testing the Difference between Two Proportions**

\[ H_0 : p_1 - p_2 = 0 \] versus

\[ H_a : \text{one of three alternatives} \]

Test statistic: \( z \approx \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \)

with \( \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \) to estimate the common value of \( p \)

and rejection regions or \( p \)-values

based on the standard normal \( z \) distribution.
**Example**

<table>
<thead>
<tr>
<th>Youth Soccer</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>80</td>
<td>70</td>
</tr>
<tr>
<td>Played soccer</td>
<td>65</td>
<td>39</td>
</tr>
</tbody>
</table>

- Compare the proportion of male and female college students who said that they had played on a soccer team during their K-12 years using a test of hypothesis.

\[
\begin{align*}
H_0 &: p_1 - p_2 = 0 \text{ (same)} \\
H_a &: p_1 - p_2 \neq 0 \text{ (different)} \\
\text{Calculate} &: \hat{p}_1 = \frac{65}{80} = .81 \quad \hat{p}_2 = \frac{39}{70} = .56 \\
\hat{p} &= \frac{x_1 + x_2}{n_1 + n_2} = \frac{104}{150} = .69
\end{align*}
\]
EXAMPLE

<table>
<thead>
<tr>
<th>Youth Soccer</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>80</td>
<td>70</td>
</tr>
<tr>
<td>Played soccer</td>
<td>65</td>
<td>39</td>
</tr>
</tbody>
</table>

Test statistic:

\[ z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \sqrt{.69(.31)\left(\frac{1}{80} + \frac{1}{70}\right)} = 3.30 \]

\[ p\text{-value}: P(z > 3.30) + P(z < -3.30) = 2(.0005) = .001 \]

Since the \( p \)-value is less than \( \alpha = .01 \), \( H_0 \) is rejected. The results are highly significant. There is evidence to indicate that the rates of participation are different for boys and girls.
**Key Concepts**

I. Parts of a Statistical Test

1. **Null hypothesis**: a contradiction of the alternative hypothesis

2. **Alternative hypothesis**: the hypothesis the researcher wants to support.

3. **Test statistic and its p-value**: sample evidence calculated from sample data.

4. **Rejection region**—critical values and significance levels: values that separate rejection and nonrejection of the null hypothesis

5. **Conclusion**: Reject or do not reject the null hypothesis, stating the practical significance of your conclusion.
KEY CONCEPTS

II. Errors and Statistical Significance

1. The **significance level** $\alpha$ is the probability if rejecting $H_0$ when it is in fact true.

2. The **$p$-value** is the probability of observing a test statistic as extreme as or more than the one observed; also, the smallest value of $\alpha$ for which $H_0$ can be rejected.

3. When the **$p$-value** is less than the **significance level** $\alpha$, the null hypothesis is rejected. This happens when the test statistic exceeds the **critical value**.

4. In a **Type II error**, $\beta$ is the probability of accepting $H_0$ when it is in fact false. The **power of the test** is $(1 - \beta)$, the probability of rejecting $H_0$ when it is false.
III. Large-Sample Test Statistics Using the $z$ Distribution

To test one of the four population parameters when the sample sizes are large, use the following test statistics:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$</td>
</tr>
<tr>
<td>$\hat{p} - p_0$</td>
<td>$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}}$</td>
</tr>
<tr>
<td>$\mu_1 - \mu_2$</td>
<td>$z = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$</td>
</tr>
<tr>
<td>$p_1 - p_2$</td>
<td>$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$</td>
</tr>
</tbody>
</table>