# Hypothesis Testing

The two main areas of statistics are **estimation** and **hypothesis testing**

# Hypothesis Testing

The two main areas of statistics are **estimation** and **hypothesis testing**

The objective in estimation is to determine the unknown value of some parameter assiciated with a statistical distribution.

# Hypothesis Testing

The two main areas of statistics are **estimation** and **hypothesis testing**

The objective in estimation is to determine the unknown value of some parameter assiciated with a statistical distribution.

The objective in hypothesis testing is to decide which of two contradictory claims or assertions is likely to be true.

# Hypothesis Testing

The two main areas of statistics are **estimation** and **hypothesis testing**

The objective in estimation is to determine the unknown value of some parameter assiciated with a statistical distribution.

The objective in hypothesis testing is to decide which of two contradictory claims or assertions is likely to be true.

In both estimation and hypothesis testing, a sample from the population under consideration is the basis of all procedures.

# Terminology

Hypothesis testing has strongly established terminology.

# Terminology

Hypothesis testing has strongly established terminology.

The first part of this terminology has to do with the contradictory assertions. These are known as:

- The **null hypothesis** $H_0$
- The **alternative hypothesis** $H_a$ (or in some texts $H_1$)

# Terminology

Hypothesis testing has strongly established terminology.

The first part of this terminology has to do with the contradictory assertions. These are known as:

- The **null hypothesis** $H_0$
- The **alternative hypothesis** $H_a$ (or in some texts $H_1$)

Unfortunately these names are rather nondescript which leads to a certain amount confusion over which of the two claims is the null hypothesis, and which is the alternative.

# Terminology

Hypothesis testing has strongly established terminology.

The first part of this terminology has to do with the contradictory assertions. These are known as:

- The **null hypothesis** $H_0$
- The **alternative hypothesis** $H_a$ (or in some texts $H_1$)

Unfortunately these names are rather nondescript which leads to a certain amount confusion over which of the two claims is the null hypothesis, and which is the alternative.

The null hypothesis is *favored* in the sense that the burden of proof is on the proponents of the alternative hypothesis.

# Terminology

Hypothesis testing has strongly established terminology.

The first part of this terminology has to do with the contradictory assertions. These are known as:

- The **null hypothesis** $H_0$
- The **alternative hypothesis** $H_a$ (or in some texts $H_1$)

Unfortunately these names are rather nondescript which leads to a certain amount confusion over which of the two claims is the null hypothesis, and which is the alternative.

The null hypothesis is *favored* in the sense that the burden of proof is on the proponents of the alternative hypothesis.

Only when presented with strong evidence (based on the sample) that it is false will we **reject the null hypothesis**

# Terminology

The mechanism for rejecting or failing to reject the null hypothesis is a **test procedure** which consists of:

- A **test statistic** derived from the sample
- A **rejection region** which is a set of values for the test statistic for which $H_0$ will be rejected.

# Terminology

The mechanism for rejecting or failing to reject the null hypothesis is a **test procedure** which consists of:

- A **test statistic** derived from the sample
- A **rejection region** which is a set of values for the test statistic for which $H_0$ will be rejected.

If the test statistic falls in the rejection region, we say that we **reject the null hypothesis $H_0$ in favor of the alternative hypothesis $H_a$**.

# Terminology

The mechanism for rejecting or failing to reject the null hypothesis is a **test procedure** which consists of:

- A **test statistic** derived from the sample
- A **rejection region** which is a set of values for the test statistic for which $H_0$ will be rejected.

If the test statistic falls in the rejection region, we say that we **reject the null hypothesis $H_0$ in favor of the alternative hypothesis $H_a$**.

If the test statistic falls outside the rejection region, we say that we **fail to reject the null hypothesis $H_0$**.

# Terminology

Additional terminlolgy has to do with the types of *errors* that are possible.

# Terminology

Additional terminlolgy has to do with the types of *errors* that are possible.

The null and alternative hypotheses are always constructed to be contradictory: one and only one of the two must be true, the other must be false.

# Terminology

Additional terminlolgy has to do with the types of *errors* that are possible.

The null and alternative hypotheses are always constructed to be contradictory: one and only one of the two must be true, the other must be false.

This means there are only two ways to be wrong when we reject or fail to reject $H_0$:

- Reject $H_0$ when it is true (**type I error**)
- Fail to reject $H_0$ when it is false (**type II error**)

# Terminology

Additional terminlolgy has to do with the types of *errors* that are possible.

The null and alternative hypotheses are always constructed to be contradictory: one and only one of the two must be true, the other must be false.

This means there are only two ways to be wrong when we reject or fail to reject $H_0$:

- Reject $H_0$ when it is true (**type I error**)
- Fail to reject $H_0$ when it is false (**type II error**)

- The probability of a **type I error** is denoted by $\alpha$
- The probability of a **type II error** is denoted by $\beta$

# Example

An automobile manufacturer wants to refute criticism that its products have more defects than those of its competitors.

# Example

An automobile manufacturer wants to refute criticism that its products have more defects than those of its competitors.

The null and alternative hypotheses would be stated as:

- $H_0$: Automobiles made by X have about the same defect rate as those of other manufacturers
- $H_a$: Automobiles made by X have a higher defect rate than those of other manufacturers

# Example

- $H_0$: Automobiles made by X have about the same defect rate as those of other manufacturers
- $H_a$: Automobiles made by X have a higher defect rate than those of other manufacturers

A **type I error** in this case would be concluding based on a sample that X had a higher defect rate when in fact its defect rate was not higher than those of its competitors (i.e, reject $H_0$ when it is true).

# Example

- $H_0$: Automobiles made by X have about the same defect rate as those of other manufacturers
- $H_a$: Automobiles made by X have a higher defect rate than those of other manufacturers

A **type I error** in this case would be concluding based on a sample that X had a higher defect rate when in fact its defect rate was not higher than those of its competitors (i.e, reject $H_0$ when it is true).

A **type II error** in this case would be concluding based on a sample that X *did not* have a higher defect rate when in fact its defect rate *was* higher than those of its competitors (i.e., fail to reject $H_0$ when it is false)

# Example

- $H_0$: Automobiles made by X have about the same defect rate as those of other manufacturers
- $H_a$: Automobiles made by X have a higher defect rate than those of other manufacturers

The **test statistic** in this case would be the average number of defects per vehicle for manufacturer X

# Example

- $H_0$: Automobiles made by X have about the same defect rate as those of other manufacturers
- $H_a$: Automobiles made by X have a higher defect rate than those of other manufacturers

The **test statistic** in this case would be the average number of defects per vehicle for manufacturer X

The **rejection region** would be based on the average defect rate for competitors, and would consist of all values above some threshold.

# Example

- $H_0$: Automobiles made by X have about the same defect rate as those of other manufacturers
- $H_a$: Automobiles made by X have a higher defect rate than those of other manufacturers

The **test statistic** in this case would be the average number of defects per vehicle for manufacturer X

The **rejection region** would be based on the average defect rate for competitors, and would consist of all values above some threshold.

The threshold would be set somewhat above the competitor's defect rate (to allow for sampling error). If X's defect rate falls above the threshold, we reject $H_0$ in favor of $H_a$. Otherwise we fail to reject $H_0$.

# Example

Manufacturing defects occur at the rate of 10.17 serious defects per 1,000 vehicles across all manufacturers. A sample of 200 vehicles from manufacturer X had a rate of 10.43 with a sample standard deviation of 2.47.

# Example

Manufacturing defects occur at the rate of 10.17 serious defects per 1,000 vehicles across all manufacturers. A sample of 200 vehicles from manufacturer X had a rate of 10.43 with a sample standard deviation of 2.47.

Test the hypothesis: $H_0$: Automobiles made by X have about the same defect rate as those of other manufacturers

against the alternative $H_a$: Automobiles made by X have a higher defect rate than those of other manufacturers at the level $\alpha = 0.05$.

# Example

Manufacturing defects occur at the rate of 10.17 serious defects per 1,000 vehicles across all manufacturers. A sample of 200 vehicles from manufacturer X had a rate of 10.43 with a sample standard deviation of 2.47.

Test the hypothesis: $H_0$: Automobiles made by X have about the same defect rate as those of other manufacturers

against the alternative $H_a$: Automobiles made by X have a higher defect rate than those of other manufacturers at the level $\alpha = 0.05$.

Use the spreadsheet on the technology link under "hypothesis testing: sigma unknown" with $Mu_0 = 10.17$, $x - bar = 10.43$, $s = 2.47$, $n = 200$, and $\alpha = 0.05$.

# Example

Since we are only concerned with the question of whether X has a *higher* defect rate (as opposed to the question of whether X has a *different* defect rate), we look at the right-tailed test.

# Example

Since we are only concerned with the question of whether X has a *higher* defect rate (as opposed to the question of whether X has a *different* defect rate), we look at the right-tailed test.

In this case we fail to reject $H_0$, but it is somewhat close. Note that the upper limit for the one-sided confidence interval is 10.46, and our sample mean was 10.43.

# Example

Since we are only concerned with the question of whether X has a *higher* defect rate (as opposed to the question of whether X has a *different* defect rate), we look at the right-tailed test.

In this case we fail to reject $H_0$, but it is somewhat close. Note that the upper limit for the one-sided confidence interval is 10.46, and our sample mean was 10.43.

Had the sample mean been 10.47, we would have rejected $H_0$ and concluded that X did have a higher defect rate than its competitors.

# Example

The EPA estimates the fuel efficiency of a certain car at $28.1mpg$. To test the accuracy of this claim, a news agency conducts a test in which $83$ owners of that make and model record their actual mileage for one tank of gas. The text cars average $27.2mpg$ with a standard deviation of $4.5$.

# Example

The EPA estimates the fuel efficiency of a certain car at $28.1mpg$. To test the accuracy of this claim, a news agency conducts a test in which $83$ owners of that make and model record their actual mileage for one tank of gas. The text cars average $27.2mpg$ with a standard deviation of $4.5$.

Test the hypothesis: $H_0$: The fuel efficiency is $28.1mpg$

against the alternative $H_a$: The fuel efficiency is different from $28.1mpg$ at the level $\alpha = 0.05$.

# Example

The EPA estimates the fuel efficiency of a certain car at $28.1mpg$. To test the accuracy of this claim, a news agency conducts a test in which $83$ owners of that make and model record their actual mileage for one tank of gas. The text cars average $27.2mpg$ with a standard deviation of $4.5$.

Test the hypothesis: $H_0$: The fuel efficiency is $28.1mpg$

against the alternative $H_a$: The fuel efficiency is different from $28.1mpg$ at the level $\alpha = 0.05$.

Use the spreadsheet on the technology link under "hypothesis testing: sigma unknown" with $Mu_0 = 28.1$, $x - bar = 27.8$, $s = 4.5$, $n = 83$, and $\alpha = 0.05$.

# Example

A sample of $100$ students in an inner city school district take the SAT math exam. The average score is $489.3$. The superintendent claims that this shows his districts math scores are not significantly lower than the national average of $500$.

# Example

A sample of $100$ students in an inner city school district take the SAT math exam. The average score is $489.3$. The superintendent claims that this shows his districts math scores are not significantly lower than the national average of $500$.

Test the hypothesis: $H_0$: The mean SAT math score in the district is $500$

against the alternative $H_a$: The mean SAT math score is less than $500$ at the level $\alpha = 0.05$.

# Example

A sample of $100$ students in an inner city school district take the SAT math exam. The average score is $489.3$. The superintendent claims that this shows his districts math scores are not significantly lower than the national average of $500$.

Test the hypothesis: $H_0$: The mean SAT math score in the district is $500$

against the alternative $H_a$: The mean SAT math score is less than $500$ at the level $\alpha = 0.05$.

Use the spreadsheet on the technology link under "hypothesis testing: sigma known" with $Mu_0 = 500$, $x - bar = 489.3$, $sigma = 100$, $n = 100$, and $\alpha = 0.05$.

# Testing Proportions

A special case of hypothesis testing occurs when the quantity of interest is the proportion of the population that has a certain characteristic (such as supporting a certain bond issue or not).

# Testing Proportions

A special case of hypothesis testing occurs when the quantity of interest is the proportion of the population that has a certain characteristic (such as supporting a certain bond issue or not).

In this case, a sample of size $n$ is taken and the test statistic is the proportion of people in the sample who have the characteristic,

$$\hat{p} = \frac{X}{n}$$

where $X$ represents the number of people in the sample that have the characteristic.

# Testing Proportions

A special case of hypothesis testing occurs when the quantity of interest is the proportion of the population that has a certain characteristic (such as supporting a certain bond issue or not).

In this case, a sample of size $n$ is taken and the test statistic is the proportion of people in the sample who have the characteristic,

$$\hat{p} = \frac{X}{n}$$

where $X$ represents the number of people in the sample that have the characteristic.

We can think of the underlying population as having a Bernoulli distribution with probability of success $p$.

# Example

A sample of $350$ likely voters indicates that $158$ would support a bill to allow casino gambling in a certain city.

# Example

A sample of $350$ likely voters indicates that $158$ would support a bill to allow casino gambling in a certain city.

Test the hypothesis: $H_0$: The proportion of people in the city support who support the bill is $50\%$

against the alternative $H_a$: The proportion supporting the bill is less than $50\%$ at the level $\alpha = 0.05$.

# Example

A sample of $350$ likely voters indicates that $158$ would support a bill to allow casino gambling in a certain city.

Test the hypothesis: $H_0$: The proportion of people in the city support who support the bill is $50\%$

against the alternative $H_a$: The proportion supporting the bill is less than $50\%$ at the level $\alpha = 0.05$.

Use the spreadsheet on the technology link under "hypothesis testing: proportions" with $p_0 = .5$, $\hat{p} = 158/350$, $n = 350$, and $\alpha = 0.05$.