Ch. 8 – Hypothesis Testing

Statistical Hypotheses

A **hypothesis** is an assertion about a distribution or its parameters.

Statistical Hypotheses

A **hypothesis** is an assertion about a distribution or its parameters. For example,

- Given a coin, one hypothesis is that each toss has probability $p = .5$ of coming up heads.

Statistical Hypotheses

A **hypothesis** is an assertion about a distribution or its parameters. For example,

- Given a coin, one hypothesis is that each toss has probability $p = .5$ of coming up heads. Another hypothesis would be that $p \neq .5$.

Statistical Hypotheses

A **hypothesis** is an assertion about a distribution or its parameters. For example,

- Given a coin, one hypothesis is that each toss has probability $p = .5$ of coming up heads. Another hypothesis would be that $p \neq .5$.
- Given a certain type of candy bar, labeled as having a mass of 60 grams, one hypothesis is that the mean mass is as labeled, $\mu = 60$.

Statistical Hypotheses

A **hypothesis** is an assertion about a distribution or its parameters. For example,

- Given a coin, one hypothesis is that each toss has probability $p = .5$ of coming up heads. Another hypothesis would be that $p \neq .5$.
- Given a certain type of candy bar, labeled as having a mass of 60 grams, one hypothesis is that the mean mass is as labeled, $\mu = 60$. Another hypothesis would be that the mean mass is smaller than labeled: $\mu < 60$.

In a hypothesis-testing problem, we consider two contradictory hypotheses H_0 and H_a.

Hypothesis Testing

In a hypothesis-testing problem, we consider two contradictory hypotheses H_0 and H_a.

- H_0, the **null hypothesis**, is the hypothesis which we initially presume to be true.

Hypothesis Testing

In a hypothesis-testing problem, we consider two contradictory hypotheses H_0 and H_a.

- H_0, the **null hypothesis**, is the hypothesis which we initially presume to be true.
- The other hypothesis H_a is called the **alternative hypothesis**.

Hypothesis Testing

In a hypothesis-testing problem, we consider two contradictory hypotheses H_0 and H_a.

- H_0, the **null hypothesis**, is the hypothesis which we initially presume to be true.
- The other hypothesis H_a is called the **alternative hypothesis**.
- The hypothesis H_0 is rejected only if the sample evidence strongly contradicts it. Otherwise we continue to believe that H_0 is plausible.

Hypothesis Testing

In a hypothesis-testing problem, we consider two contradictory hypotheses H_0 and H_a.

- H_0, the **null hypothesis**, is the hypothesis which we initially presume to be true.
- The other hypothesis H_a is called the **alternative hypothesis**.
- The hypothesis H_0 is rejected only if the sample evidence strongly contradicts it. Otherwise we continue to believe that H_0 is plausible.
- The two possible outcomes of the analysis are that we **reject** the null hypothesis H_0, or we **do not reject** the null hypothesis.

Example

Suppose someone claims that a coin is unfair, that it gives heads more than half the time.

Example

Suppose someone claims that a coin is unfair, that it gives heads more than half the time. The null hypothesis would be $H_0 : p = .5$, that the coin is fair.

Example

Suppose someone claims that a coin is unfair, that it gives heads more than half the time. The null hypothesis would be $H_0 : p = .5$, that the coin is fair. The alternative hypothesis is $H_a : p > .5$.

Example

Suppose someone claims that a coin is unfair, that it gives heads more than half the time. The null hypothesis would be $H_0 : p = .5$, that the coin is fair. The alternative hypothesis is $H_a : p > .5$.

To test the claim, we could toss the coin for several trials and reject H_0 if the number of heads we obtain is larger than a certain amount.

Example

Suppose someone claims that a coin is unfair, that it gives heads more than half the time. The null hypothesis would be $H_0 : p = .5$, that the coin is fair. The alternative hypothesis is $H_a : p > .5$.

To test the claim, we could toss the coin for several trials and reject H_0 if the number of heads we obtain is larger than a certain amount.

For example, one procedure would be to toss the coin 10 times and reject the null hypothesis if we obtain 8 heads or more.

Test Procedures

- In general, to test a statistical hypothesis, we select a **test statistic** that can be calculated from a random sample.

Test Procedures

- In general, to test a statistical hypothesis, we select a **test statistic** that can be calculated from a random sample.
- Out of the possible values of the test statistic, we select a subset of values which are unlikely to occur if the null hypothesis H_0 is true; this subset is called the **rejection region**.

Test Procedures

- In general, to test a statistical hypothesis, we select a **test statistic** that can be calculated from a random sample.
- Out of the possible values of the test statistic, we select a subset of values which are unlikely to occur if the null hypothesis H_0 is true; this subset is called the **rejection region**.
- We then obtain data from a random sample, calculate the test statistic for the data, and reject H_0 if the test statistic is in the rejection region.

Test Procedures

- In general, to test a statistical hypothesis, we select a **test statistic** that can be calculated from a random sample.
- Out of the possible values of the test statistic, we select a subset of values which are unlikely to occur if the null hypothesis H_0 is true; this subset is called the **rejection region**.
- We then obtain data from a random sample, calculate the test statistic for the data, and reject H_0 if the test statistic is in the rejection region.

For the example of testing for an unfair coin, the test statistic was the number of heads X out of 10 tosses, and the rejection region was $\{8, 9, 10\}$.

Errors in Hypothesis Testing

For essentially any test procedure, there is a chance that the test will give a misleading conclusion:

Errors in Hypothesis Testing

For essentially any test procedure, there is a chance that the test will give a misleading conclusion:

- When the null hypothesis H_0 is true but is rejected, this is called a **Type I error**.

Errors in Hypothesis Testing

For essentially any test procedure, there is a chance that the test will give a misleading conclusion:

- When the null hypothesis H_0 is true but is rejected, this is called a **Type I error**.
- When the null hypothesis H_0 is false but is not rejected, this is called a **Type II error**.

Errors in Hypothesis Testing

For essentially any test procedure, there is a chance that the test will give a misleading conclusion:

- When the null hypothesis H_0 is true but is rejected, this is called a **Type I error**.
- When the null hypothesis H_0 is false but is not rejected, this is called a **Type II error**.

We cannot eliminate the possibility of these errors. However, we can quantify their probability of occuring:

Errors in Hypothesis Testing

For essentially any test procedure, there is a chance that the test will give a misleading conclusion:

- When the null hypothesis H_0 is true but is rejected, this is called a **Type I error**.
- When the null hypothesis H_0 is false but is not rejected, this is called a **Type II error**.

We cannot eliminate the possibility of these errors. However, we can quantify their probability of occuring:

- The probability of a Type I error is denoted by α.

Errors in Hypothesis Testing

For essentially any test procedure, there is a chance that the test will give a misleading conclusion:

- When the null hypothesis H_0 is true but is rejected, this is called a **Type I error**.
- When the null hypothesis H_0 is false but is not rejected, this is called a **Type II error**.

We cannot eliminate the possibility of these errors. However, we can quantify their probability of occuring:

- The probability of a Type I error is denoted by α.
- The probability of a Type II error is denoted by β.

Errors in Hypothesis Testing

For essentially any test procedure, there is a chance that the test will give a misleading conclusion:

- When the null hypothesis H_0 is true but is rejected, this is called a **Type I error**.
- When the null hypothesis H_0 is false but is not rejected, this is called a **Type II error**.

We cannot eliminate the possibility of these errors. However, we can quantify their probability of occuring:

- The probability of a Type I error is denoted by α.
- The probability of a Type II error is denoted by β.

Typically, the choice of rejection region involves a tradeoff between the two types of errors.

Errors in Hypothesis Testing

For essentially any test procedure, there is a chance that the test will give a misleading conclusion:

- When the null hypothesis H_0 is true but is rejected, this is called a **Type I error**.
- When the null hypothesis H_0 is false but is not rejected, this is called a **Type II error**.

We cannot eliminate the possibility of these errors. However, we can quantify their probability of occuring:

- The probability of a Type I error is denoted by α.
- The probability of a Type II error is denoted by β.

Typically, the choice of rejection region involves a tradeoff between the two types of errors. But by using larger samples, both error probabilities may be reduced.

Example

Suppose we test a coin by tossing it 10 times and rejecting it if we get 8 or more heads. If in reality the coin is fair, what is the probability α of a Type I error? If the coin is unfair with probability $p = .75$ of being heads, what is the probability β of a Type II error?

Example

Suppose we test a coin by tossing it 10 times and rejecting it if we get 8 or more heads. If in reality the coin is fair, what is the probability α of a Type I error? If the coin is unfair with probability $p = .75$ of being heads, what is the probability β of a Type II error?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 8:

Example

Suppose we test a coin by tossing it 10 times and rejecting it if we get 8 or more heads. If in reality the coin is fair, what is the probability α of a Type I error? If the coin is unfair with probability $p = .75$ of being heads, what is the probability β of a Type II error?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 8:

$$\alpha = P(X \geq 8) = \sum_{x=8}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .055$$

Example

Suppose we test a coin by tossing it 10 times and rejecting it if we get 8 or more heads. If in reality the coin is fair, what is the probability α of a Type I error? If the coin is unfair with probability $p = .75$ of being heads, what is the probability β of a Type II error?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 8:

$$\alpha = P(X \geq 8) = \sum_{x=8}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .055$$

To find the Type II error in the case $p = .75$, we calculate the probability that the number of heads X is less than 8:

Example

Suppose we test a coin by tossing it 10 times and rejecting it if we get 8 or more heads. If in reality the coin is fair, what is the probability α of a Type I error? If the coin is unfair with probability $p = .75$ of being heads, what is the probability β of a Type II error?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 8:

$$\alpha = P(X \geq 8) = \sum_{x=8}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .055$$

To find the Type II error in the case $p = .75$, we calculate the probability that the number of heads X is less than 8:

$$\beta = P(X < 8) = \sum_{x=0}^{7} \binom{10}{x} (.75)^x (.25)^{10-x} = .474$$

Example

In the previous example, how do the error probabilities change if we instead reject the coin if we get 7 or more heads out of 10?

Example

In the previous example, how do the error probabilities change if we instead reject the coin if we get 7 or more heads out of 10?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 7:

Example

In the previous example, how do the error probabilities change if we instead reject the coin if we get 7 or more heads out of 10?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 7:

$$\alpha = P(X \geq 7) = \sum_{x=7}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .171$$

Example

In the previous example, how do the error probabilities change if we instead reject the coin if we get 7 or more heads out of 10?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 7:

$$\alpha = P(X \geq 7) = \sum_{x=7}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .171$$

To find the Type II error in the case $p = .75$, we calculate the probability that the number of heads X is less than 7:

Example

In the previous example, how do the error probabilities change if we instead reject the coin if we get 7 or more heads out of 10?

To find the Type I error, we assume $p = .5$ and calculate the probability that the number of heads X is at least 7:

$$\alpha = P(X \geq 7) = \sum_{x=7}^{10} \binom{10}{x} (.5)^x (1 - .5)^{10-x} = .171$$

To find the Type II error in the case $p = .75$, we calculate the probability that the number of heads X is less than 7:

$$\beta = P(X < 7) = \sum_{x=0}^{6} \binom{10}{x} (.75)^x (.25)^{10-x} = .224$$

By enlarging the rejection region, we increased α but decreased β.

Problem

A type of candy bar is labeled 60 grams. We decide to test the label's accuracy using a random sample $X_1, \ldots, X_5$ by rejecting the null hypothesis $H_0 : \mu = 60$ if $\overline{X} < 59$. If the masses are normally distributed with $\sigma = 0.8$, what is the Type I error probability α?

Problem

A type of candy bar is labeled 60 grams. We decide to test the label's accuracy using a random sample $X_1, \ldots, X_5$ by rejecting the null hypothesis $H_0 : \mu = 60$ if $\overline{X} < 59$. If the masses are normally distributed with $\sigma = 0.8$, what is the Type I error probability α?

If H_0 is true, then $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 60}{0.8/\sqrt{5}}$ is a standard normal random variable,

Problem

A type of candy bar is labeled 60 grams. We decide to test the label's accuracy using a random sample $X_1, \dots, X_5$ by rejecting the null hypothesis $H_0 : \mu = 60$ if $\overline{X} < 59$. If the masses are normally distributed with $\sigma = 0.8$, what is the Type I error probability α?

If H_0 is true, then $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-60}{0.8/\sqrt{5}}$ is a standard normal random variable, so

$$\begin{aligned}
\alpha &= P(\overline{X} < 59) \\
&= P\left(\frac{\overline{X}-60}{0.8/\sqrt{5}} < \frac{59-60}{0.8/\sqrt{5}}\right) \\
&= P(Z < -2.80) \\
&= .0026
\end{aligned}$$

Problem

A type of candy bar is labeled 60 grams. We decide to test the label's accuracy using a random sample $X_1, \ldots, X_5$ by rejecting the null hypothesis $H_0 : \mu = 60$ if $\overline{X} < 59$. If the actual mean mass is $\mu = 58.5$, and $\sigma = 0.8$, what is the Type II error probability β?

Problem

A type of candy bar is labeled 60 grams. We decide to test the label's accuracy using a random sample $X_1, \ldots, X_5$ by rejecting the null hypothesis $H_0 : \mu = 60$ if $\overline{X} < 59$. If the actual mean mass is $\mu = 58.5$, and $\sigma = 0.8$, what is the Type II error probability β?

In this case, $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 58.5}{0.8/\sqrt{5}}$ is a standard normal random variable,

Problem

A type of candy bar is labeled 60 grams. We decide to test the label's accuracy using a random sample $X_1, \ldots, X_5$ by rejecting the null hypothesis $H_0 : \mu = 60$ if $\overline{X} < 59$. If the actual mean mass is $\mu = 58.5$, and $\sigma = 0.8$, what is the Type II error probability β?

In this case, $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-58.5}{0.8/\sqrt{5}}$ is a standard normal random variable, so

$$\begin{aligned}\beta &= P(\overline{X} \geq 59) \\ &= P\left(\frac{\overline{X}-58.5}{0.8/\sqrt{5}} \geq \frac{59-58.5}{0.8/\sqrt{5}}\right) \\ &= P(Z \geq 1.40) \\ &= .0808\end{aligned}$$

Problem: Two-tailed Test

A machine is specified to drill holes with diameter 4 mm. We test the hypothesis $H_0 : \mu = 4$, using a random sample $X_1, \ldots, X_{30}$. We reject H_0 if $\overline{X} > 4.1$ or $\overline{X} < 3.9$. If the diameters are normally distributed with $\sigma = .2$, find the Type I error probability α.

Problem: Two-tailed Test

A machine is specified to drill holes with diameter 4 mm. We test the hypothesis $H_0 : \mu = 4$, using a random sample $X_1, \ldots, X_{30}$. We reject H_0 if $\overline{X} > 4.1$ or $\overline{X} < 3.9$. If the diameters are normally distributed with $\sigma = .2$, find the Type I error probability α.

If H_0 is true, then $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{\overline{X} - 4}{0.2/\sqrt{30}}$ is a standard normal random variable,

Problem: Two-tailed Test

A machine is specified to drill holes with diameter 4 mm. We test the hypothesis $H_0 : \mu = 4$, using a random sample $X_1, \ldots, X_{30}$. We reject H_0 if $\overline{X} > 4.1$ or $\overline{X} < 3.9$. If the diameters are normally distributed with $\sigma = .2$, find the Type I error probability α.

If H_0 is true, then $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\overline{X}-4}{0.2/\sqrt{30}}$ is a standard normal random variable, so

$$\begin{aligned}\alpha &= P(\overline{X} < 3.9) + P(\overline{X} > 4.1) \\ &= 2P(\overline{X} < 3.9) \\ &= 2P\left(\frac{\overline{X} - 4}{0.2/\sqrt{30}} < \frac{3.9 - 4}{0.2/\sqrt{30}}\right) \\ &= 2P(Z < -2.74) \\ &= .0062\end{aligned}$$

Significance Level

As we have seen, for a fixed sample size, selecting a rejection region for a test involves a tradeoff between the Type I error probabilities α and the Type II error probability β.

Significance Level

As we have seen, for a fixed sample size, selecting a rejection region for a test involves a tradeoff between the Type I error probabilities α and the Type II error probability β.

A common practice is to design the test to achieve a specified small value of α, such as $\alpha = .1, .05$, or $.01$. The choice for α is called the **significance level**.

z Test for Mean of Normal Distribution with Known σ

Given a random sample $X_1, \ldots, X_n$ from a normal distribution with known standard deviation σ, the z **test** for the null hypothesis $H_0 : \mu = \mu_0$, based on the test statistic $Z = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$, is given by the following rejection region, depending on whether a one-tailed or two-tailed test is desired:

Alternative hypothesis		Rejection region
(Upper-tailed test)	$H_a : \mu > \mu_0$	$Z \geq z_\alpha$
(Lower-tailed test)	$H_a : \mu < \mu_0$	$Z \leq -z_\alpha$
(Two-tailed test)	$H_a : \mu \neq \mu_0$	$\lvert Z \rvert \geq z_{\alpha/2}$

Here α is the significance level (Type I error probability), and z_α is a critical value from the standard normal distribution.

Example

A machine is specified to drill holes with diameter 4 mm. We wish to test the null hypothesis $H_0 : \mu = 4$ against the alternative $H_a : \mu \neq 4$. If the diameters are normally distributed with $\sigma = .2$, and we observe $\overline{X} = 3.87$ in a sample of size 10, do we reject the null hypothesis at the significance level $\alpha = .05$?

Example

A machine is specified to drill holes with diameter 4 mm. We wish to test the null hypothesis $H_0 : \mu = 4$ against the alternative $H_a : \mu \neq 4$. If the diameters are normally distributed with $\sigma = .2$, and we observe $\overline{X} = 3.87$ in a sample of size 10, do we reject the null hypothesis at the significance level $\alpha = .05$?

The test statistic is

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.87 - 4}{.2/\sqrt{10}} = -2.06$$

Example

A machine is specified to drill holes with diameter 4 mm. We wish to test the null hypothesis $H_0 : \mu = 4$ against the alternative $H_a : \mu \neq 4$. If the diameters are normally distributed with $\sigma = .2$, and we observe $\overline{X} = 3.87$ in a sample of size 10, do we reject the null hypothesis at the significance level $\alpha = .05$?

The test statistic is

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.87 - 4}{.2/\sqrt{10}} = -2.06$$

The rejection region is $\{|Z| > z_{\alpha/2}\}$ where $z_{\alpha/2} = z_{.025} = 1.96$.

Example

A machine is specified to drill holes with diameter 4 mm. We wish to test the null hypothesis $H_0 : \mu = 4$ against the alternative $H_a : \mu \neq 4$. If the diameters are normally distributed with $\sigma = .2$, and we observe $\overline{X} = 3.87$ in a sample of size 10, do we reject the null hypothesis at the significance level $\alpha = .05$?

The test statistic is

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.87 - 4}{.2/\sqrt{10}} = -2.06$$

The rejection region is $\{|Z| > z_{\alpha/2}\}$ where $z_{\alpha/2} = z_{.025} = 1.96$.

Since $|Z| = 2.06 > 1.96$, the test statistic is in the rejection region, so we reject the null hypothesis.

Large-Sample z Test for Mean with Unknown σ

Given a random sample $X_1, \ldots, X_n$ from a distribution with unknown standard deviation, the z **test** for the null hypothesis $H_0 : \mu = \mu_0$, based on the test statistic $Z = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$, is given by the following rejection region:

Alternative hypothesis		Rejection region
(Upper-tailed test)	$H_a : \mu > \mu_0$	$Z \geq z_\alpha$
(Lower-tailed test)	$H_a : \mu < \mu_0$	$Z \leq -z_\alpha$
(Two-tailed test)	$H_a : \mu \neq \mu_0$	$\lvert Z\rvert \geq z_{\alpha/2}$

Here α is the nominal significance level. If n is large, then under H_0, Z is approximately standard normal by the Central Limit Theorem, so the true significance level is approximately α.

Note: Here we don't need to assume that the distribution of $X_1, \ldots, X_n$ is normal.

Example

A machine is specified to drill holes with diameter 4 mm. We wish to test the null hypothesis $H_0 : \mu = 4$ against the alternative $H_a : \mu \neq 4$. If we observe $\overline{X} = 3.97$ and $S = .21$ in a sample of size 100, do we reject the null hypothesis at the significance level $\alpha = .05$?

Example

A machine is specified to drill holes with diameter 4 mm. We wish to test the null hypothesis $H_0 : \mu = 4$ against the alternative $H_a : \mu \neq 4$. If we observe $\overline{X} = 3.97$ and $S = .21$ in a sample of size 100, do we reject the null hypothesis at the significance level $\alpha = .05$?

The test statistic is

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{3.97 - 4}{.21/\sqrt{100}} = -1.43$$

Example

A machine is specified to drill holes with diameter 4 mm. We wish to test the null hypothesis $H_0 : \mu = 4$ against the alternative $H_a : \mu \neq 4$. If we observe $\overline{X} = 3.97$ and $S = .21$ in a sample of size 100, do we reject the null hypothesis at the significance level $\alpha = .05$?

The test statistic is

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{3.97 - 4}{.21/\sqrt{100}} = -1.43$$

The rejection region is $\{|Z| > z_{\alpha/2}\}$ where $z_{\alpha/2} = z_{.025} = 1.96$.

Example

A machine is specified to drill holes with diameter 4 mm. We wish to test the null hypothesis $H_0 : \mu = 4$ against the alternative $H_a : \mu \neq 4$. If we observe $\overline{X} = 3.97$ and $S = .21$ in a sample of size 100, do we reject the null hypothesis at the significance level $\alpha = .05$?

The test statistic is

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{3.97 - 4}{.21/\sqrt{100}} = -1.43$$

The rejection region is $\{|Z| > z_{\alpha/2}\}$ where $z_{\alpha/2} = z_{.025} = 1.96$.

Since $|Z| = 1.43 < 1.96$, the test statistic is not in the rejection region, so we do not reject the null hypothesis. In other words, based on the data, it is plausible that the mean is $\mu = 4$ as specified.

t Test for Mean of Normal Distribution with Unknown σ

Given a random sample $X_1, \ldots, X_n$ from a normal distribution with unknown standard deviation, the t **test** for the null hypothesis $H_0 : \mu = \mu_0$, based on the test statistic $T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$, is given by the following rejection region:

Alternative hypothesis		Rejection region
(Upper-tailed test)	$H_a : \mu > \mu_0$	$T \geq t_{\alpha,n-1}$
(Lower-tailed test)	$H_a : \mu < \mu_0$	$T \leq -t_{\alpha,n-1}$
(Two-tailed test)	$H_a : \mu \neq \mu_0$	$\lvert T \rvert \geq t_{\alpha/2,n-1}$

Here α is the significance level (Type I error probability), and $t_{\alpha,n-1}$ is a critical value from the t distribution with $n-1$ degrees of freedom.

Example

A type of candy bar is labeled 60 grams. Someone suggests that the candy bars weigh less than specified. To test this, we gather a random sample of size 5 and observe $\overline{X} = 58.8$ and $S = 0.9$. Do we reject the null hypothesis $H_0 : \mu = 60$ at the significance level $\alpha = .01$?

Example

A type of candy bar is labeled 60 grams. Someone suggests that the candy bars weigh less than specified. To test this, we gather a random sample of size 5 and observe $\overline{X} = 58.8$ and $S = 0.9$. Do we reject the null hypothesis $H_0 : \mu = 60$ at the significance level $\alpha = .01$?

We use the t test with the one-tailed alternative $H_a : \mu < 60$. The test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{58.8 - 60}{0.9/\sqrt{5}} = -2.98$$

Example

A type of candy bar is labeled 60 grams. Someone suggests that the candy bars weigh less than specified. To test this, we gather a random sample of size 5 and observe $\overline{X} = 58.8$ and $S = 0.9$. Do we reject the null hypothesis $H_0 : \mu = 60$ at the significance level $\alpha = .01$?

We use the t test with the one-tailed alternative $H_a : \mu < 60$. The test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{58.8 - 60}{0.9/\sqrt{5}} = -2.98$$

The critical value is $t_{\alpha,n-1} = t_{.01,4} = 3.747$. The rejection region is $\{T < -3.747\}$, whereas in our sample $T = -2.98 > -3.747$, so we do not reject the null hypothesis.

Example

A type of candy bar is labeled 60 grams. Someone suggests that the candy bars weigh less than specified. To test this, we gather a random sample of size 5 and observe $\overline{X} = 58.8$ and $S = 0.9$. Do we reject the null hypothesis $H_0 : \mu = 60$ at the significance level $\alpha = .01$?

We use the t test with the one-tailed alternative $H_a : \mu < 60$. The test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{58.8 - 60}{0.9/\sqrt{5}} = -2.98$$

The critical value is $t_{\alpha,n-1} = t_{.01,4} = 3.747$. The rejection region is $\{T < -3.747\}$, whereas in our sample $T = -2.98 > -3.747$, so we do not reject the null hypothesis.

In other words, the data does *not* allow us to conclude that the average weight of the candy bars is less than specified.

Large-Sample z Test for Proportion

Given a random sample from a Bernoulli distribution with unknown parameter p, the z **test** for the null hypothesis $H_0 : p = p_0$ based on the test statistic $Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$ is given by the following rejection region:

Alternative hypothesis		Rejection region
(Upper-tailed test)	$H_a : p > p_0$	$Z \geq z_\alpha$
(Lower-tailed test)	$H_a : p < p_0$	$Z \leq -z_\alpha$
(Two-tailed test)	$H_a : p \neq p_0$	$\lvert Z \rvert \geq z_{\alpha/2}$

Here α is the nominal significance level, and z_α is a critical value from the standard normal distribution.

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, do we reject the null hypothesis at the $\alpha = .01$ significance level?

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, do we reject the null hypothesis at the $\alpha = .01$ significance level?

We will use a large-sample z test for the null hypothesis $H_0 : p = \frac{1}{2}$ against the alternative $H_0 : p \neq \frac{1}{2}$.

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, do we reject the null hypothesis at the $\alpha = .01$ significance level?

We will use a large-sample z test for the null hypothesis $H_0 : p = \frac{1}{2}$ against the alternative $H_0 : p \neq \frac{1}{2}$. The test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{\frac{54}{80} - \frac{1}{2}}{\sqrt{\frac{1}{2}(1 - \frac{1}{2})/80}} = 3.13$$

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, do we reject the null hypothesis at the $\alpha = .01$ significance level?

We will use a large-sample z test for the null hypothesis $H_0 : p = \frac{1}{2}$ against the alternative $H_0 : p \neq \frac{1}{2}$. The test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{\frac{54}{80} - \frac{1}{2}}{\sqrt{\frac{1}{2}(1-\frac{1}{2})/80}} = 3.13$$

The rejection region is $\{|Z| > z_{\alpha/2}\}$ where $z_{\alpha/2} = z_{.005} = 2.58$.

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, do we reject the null hypothesis at the $\alpha = .01$ significance level?

We will use a large-sample z test for the null hypothesis $H_0 : p = \frac{1}{2}$ against the alternative $H_0 : p \neq \frac{1}{2}$. The test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{\frac{54}{80} - \frac{1}{2}}{\sqrt{\frac{1}{2}(1 - \frac{1}{2})/80}} = 3.13$$

The rejection region is $\{|Z| > z_{\alpha/2}\}$ where $z_{\alpha/2} = z_{.005} = 2.58$. Since $|Z| = 3.13 > 2.58$, we reject the null hypothesis.

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, do we reject the null hypothesis at the $\alpha = .01$ significance level?

We will use a large-sample z test for the null hypothesis $H_0 : p = \frac{1}{2}$ against the alternative $H_0 : p \neq \frac{1}{2}$. The test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{\frac{54}{80} - \frac{1}{2}}{\sqrt{\frac{1}{2}(1 - \frac{1}{2})/80}} = 3.13$$

The rejection region is $\{|Z| > z_{\alpha/2}\}$ where $z_{\alpha/2} = z_{.005} = 2.58$. Since $|Z| = 3.13 > 2.58$, we reject the null hypothesis.

In other words, the test provides strong evidence that the coin indeed gives heads more often than tails when spun.

P-Values

There is a major drawback of the hypothesis-testing procedures considered so far: The tests each only have two outcomes – reject or do not reject – even if the test statistic is near the border of the rejection region.

P-Values

There is a major drawback of the hypothesis-testing procedures considered so far: The tests each only have two outcomes – reject or do not reject – even if the test statistic is near the border of the rejection region.

One way to remedy this is to report the so-called P-value:

The **P-value** is the smallest significance level α for which the test would reject the null hypothesis.

P-Values

There is a major drawback of the hypothesis-testing procedures considered so far: The tests each only have two outcomes – reject or do not reject – even if the test statistic is near the border of the rejection region.

One way to remedy this is to report the so-called P-value:

The **P-value** is the smallest significance level α for which the test would reject the null hypothesis.

For example, in the candy bar example, at the $\alpha = .01$ we failed to reject the null hypothesis; however, if we had used a less strict significance level, $\alpha = .05$, then the test would have rejected.

P-Values

There is a major drawback of the hypothesis-testing procedures considered so far: The tests each only have two outcomes – reject or do not reject – even if the test statistic is near the border of the rejection region.

One way to remedy this is to report the so-called P-value:

The **P-value** is the smallest significance level α for which the test would reject the null hypothesis.

For example, in the candy bar example, at the $\alpha = .01$ we failed to reject the null hypothesis; however, if we had used a less strict significance level, $\alpha = .05$, then the test would have rejected. The P-value would provide a more nuanced summary of the test result by indicating precisely at what significance level the test changes from rejecting to failing to reject.

P-Values

Loosely speaking, another way to describe the P-value is as follows:

The **P-value** is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as extreme as the value actually observed.

P-Values

Loosely speaking, another way to describe the P-value is as follows:

The **P-value** is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as extreme as the value actually observed.

Here are some key points:

- The P-value is a probability.

P-Values

Loosely speaking, another way to describe the P-value is as follows:

The **P-value** is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as extreme as the value actually observed.

Here are some key points:

- The P-value is a probability.
- This probability is calculated assuming that the null hypothesis is true.

P-Values

Loosely speaking, another way to describe the P-value is as follows:

The **P-value** is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as extreme as the value actually observed.

Here are some key points:

- The P-value is a probability.
- This probability is calculated assuming that the null hypothesis is true.
- Beware: The P-value is not the probability that H_0 is true.

P-Values

Loosely speaking, another way to describe the P-value is as follows:

The **P-value** is the probability, calculated assuming that the null hypothesis is true, of obtaining a value of the test statistic at least as extreme as the value actually observed.

Here are some key points:

- The P-value is a probability.
- This probability is calculated assuming that the null hypothesis is true.
- Beware: The P-value is not the probability that H_0 is true.
- The interpretation of "as extreme as" depends on the alternative hypothesis:
  - For an upper-tailed alternative, it means "as large as".
  - For a lower-tailed alternative, it means "as small as".
  - For a two-tailed alternative, it means "as large in absolute value as".

P-Values for z tests and t test

With all of the z test procedures, the P-value may be calculated as follows, where z is the observed value of the test statistic Z (assumed to have a standard normal distribution):

Alternative hypothesis		P-value
(Upper-tailed test)	$H_a : \mu > \mu_0$	$P(Z \geq z)$
(Lower-tailed test)	$H_a : \mu < \mu_0$	$P(Z \leq z)$
(Two-tailed test)	$H_a : \mu \neq \mu_0$	$P(\lvert Z\rvert \geq \lvert z\rvert)$

Similarly, for the t test, the P-value may be calculated as follows, where t is the observed value of the test statistic T (assumed to have a t distribution with $n - 1$ degrees of freedom):

Alternative hypothesis		P-value
(Upper-tailed test)	$H_a : \mu > \mu_0$	$P(T \geq t)$
(Lower-tailed test)	$H_a : \mu < \mu_0$	$P(T \leq t)$
(Two-tailed test)	$H_a : \mu \neq \mu_0$	$P(\lvert T\rvert \geq \lvert t\rvert)$

Example

A type of candy bar is labeled 60 grams. Someone suggests that the candy bars weigh less than specified. To test this, we gather a random sample of size 5 and observe $\overline{X} = 58.8$ and $S = 0.9$. What is the P-value for the test?

Example

A type of candy bar is labeled 60 grams. Someone suggests that the candy bars weigh less than specified. To test this, we gather a random sample of size 5 and observe $\overline{X} = 58.8$ and $S = 0.9$. What is the P-value for the test?

We use the t test with the one-tailed alternative $H_a : \mu < 60$. As before, the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{58.9 - 60}{0.9/\sqrt{5}} = -2.98$$

Example

A type of candy bar is labeled 60 grams. Someone suggests that the candy bars weigh less than specified. To test this, we gather a random sample of size 5 and observe $\overline{X} = 58.8$ and $S = 0.9$. What is the P-value for the test?

We use the t test with the one-tailed alternative $H_a : \mu < 60$. As before, the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{58.9 - 60}{0.9/\sqrt{5}} = -2.98$$

We use a table to find the probability of observing a value for T at least this extreme:

$$P = P(T \leq -2.98)$$

Example

A type of candy bar is labeled 60 grams. Someone suggests that the candy bars weigh less than specified. To test this, we gather a random sample of size 5 and observe $\overline{X} = 58.8$ and $S = 0.9$. What is the P-value for the test?

We use the t test with the one-tailed alternative $H_a : \mu < 60$. As before, the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{58.9 - 60}{0.9/\sqrt{5}} = -2.98$$

We use a table to find the probability of observing a value for T at least this extreme:

$$P = P(T \leq -2.98) \approx .020$$

Example

A type of candy bar is labeled 60 grams. Someone suggests that the candy bars weigh less than specified. To test this, we gather a random sample of size 5 and observe $\overline{X} = 58.8$ and $S = 0.9$. What is the P-value for the test?

We use the t test with the one-tailed alternative $H_a : \mu < 60$. As before, the test statistic is

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{58.9 - 60}{0.9/\sqrt{5}} = -2.98$$

We use a table to find the probability of observing a value for T at least this extreme:

$$P = P(T \leq -2.98) \approx .020$$

So $P = .020$ is the P-value for the test.

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, what is the P-value of the test?

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, what is the P-value of the test?

Again, we are using a large-sample z test for the null hypothesis $H_0 : p = \frac{1}{2}$ against the alternative $H_0 : p \neq \frac{1}{2}$.

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, what is the P-value of the test?

Again, we are using a large-sample z test for the null hypothesis $H_0 : p = \frac{1}{2}$ against the alternative $H_0 : p \neq \frac{1}{2}$. We already calculated the test statistic:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{\frac{54}{80} - \frac{1}{2}}{\sqrt{\frac{1}{2}(1 - \frac{1}{2})/80}} = 3.13$$

Example

We are given a coin which someone suggests may give outcomes with unequal proportions when we spin it on a table. We test this by spinning the coin 80 times. If we observe 54 heads, what is the P-value of the test?

Again, we are using a large-sample z test for the null hypothesis $H_0 : p = \frac{1}{2}$ against the alternative $H_0 : p \neq \frac{1}{2}$. We already calculated the test statistic:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{\frac{54}{80} - \frac{1}{2}}{\sqrt{\frac{1}{2}(1 - \frac{1}{2})/80}} = 3.13$$

The P-value is the probability that we would observe a value of Z this extreme (i.e., a value of Z with $|Z| \geq 3.13$):

$$P = P(|Z| \geq 3.13) = 2\Phi(-3.13) \approx 2(.0009) = .0018$$

Summary

$\alpha =$ probability of Type I error, that H_0 is true but is rejected

$\beta =$ probability of a Type II error, that H_0 is false but is not rejected

$P =$ P-value $=$ smallest α for which the test would reject H_0

Test	Null Hypothesis	Test Statistic
z test	$H_0 : \mu = \mu_0$	$Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$
t test	$H_0 : \mu = \mu_0$	$T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$
z test for a proportion	$H_0 : p = p_0$	$Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$

Alternative hypothesis		P-value for z test
(Upper-tailed test)	$H_a : \mu > \mu_0$	$P(Z \geq z)$
(Lower-tailed test)	$H_a : \mu < \mu_0$	$P(Z \leq z)$
(Two-tailed test)	$H_a : \mu \neq \mu_0$	$P(\lvert Z\rvert \geq \lvert z\rvert)$

Given a fixed significance level α, we reject H_0 if and only if $P \leq \alpha$.