Hypothesis Testing

or: How I Learned to Stop Worrying and Love Inferential Statistics

Objectives

1. State the null and alternative hypothesis
2. Understand errors and interpret *p*-value
3. Perform a hypothesis test of the population mean with known population standard deviation

What is Hypothesis Testing?

Hypothesis testing is the process of testing a claim about a (population) parameter by obtaining a sample and analyzing the results against that parameter.

What is Hypothesis Testing?

Hypothesis testing is the process of testing a claim about a (population) parameter by obtaining a sample and analyzing the results against that parameter.

Null Hypothesis

The **null hypothesis** is the original claim about the parameter, denoted H_0, and usually is stated in terms of equality.

What is Hypothesis Testing?

Hypothesis testing is the process of testing a claim about a (population) parameter by obtaining a sample and analyzing the results against that parameter.

Null Hypothesis

The **null hypothesis** is the original claim about the parameter, denoted H_0, and usually is stated in terms of equality.

Alternative Hypothesis

The **alternative hypothesis**, denoted H_A, is the new claim that is made against the null hypothesis.

Types of Hypothesis Testing

- Left-tailed

Types of Hypothesis Testing

- Left-tailed
  - H_A: parameter $< H_0$

Types of Hypothesis Testing

- Left-tailed
  - H_A: parameter $< H_0$
- Right-tailed

Types of Hypothesis Testing

- Left-tailed
  - H_A: parameter $< H_0$
- Right-tailed
  - H_A: parameter $> H_0$

Types of Hypothesis Testing

- Left-tailed
  - H_A: parameter $< H_0$
- Right-tailed
  - H_A: parameter $> H_0$
- Two-tailed

Types of Hypothesis Testing

- Left-tailed
  - H_A: parameter $< H_0$
- Right-tailed
  - H_A: parameter $> H_0$
- Two-tailed
  - H_A: parameter $\neq H_0$

Example 1

Determine the null and alternative hypotheses for each.

(a) A new car's mpg is listed as 33. You want to know if the mpg is not 33.

Example 1

Determine the null and alternative hypotheses for each.

(a) A new car's mpg is listed as 33. You want to know if the mpg is not 33.

$H_0: \mu = 33$ mpg

Example 1

Determine the null and alternative hypotheses for each.

(a) A new car's mpg is listed as 33. You want to know if the mpg is not 33.

$H_0 :\ \mu = 33$ mpg

$H_A :\ \mu \neq 33$mpg

Example 1

(b) A hospital says the mean wait time for patients to see a doctor is 3.5 minutes. You want to know if the mean wait time is more than 3.5 minutes.

Example 1

(b) A hospital says the mean wait time for patients to see a doctor is 3.5 minutes. You want to know if the mean wait time is more than 3.5 minutes.

$H_0: \mu = 3.5$ minutes

Example 1

(b) A hospital says the mean wait time for patients to see a doctor is 3.5 minutes. You want to know if the mean wait time is more than 3.5 minutes.

$H_0: \mu = 3.5$ minutes

$H_A: \mu > 3.5$ minutes

Example 1

(c) A company claims their program will increase your grade in statistics class by 10%. You think it might not be that much.

Example 1

(c) A company claims their program will increase your grade in statistics class by 10%. You think it might not be that much.

$H_0 : \mu = 0.10$

Example 1

(c) A company claims their program will increase your grade in statistics class by 10%. You think it might not be that much.

$H_0: \mu = 0.10$

$H_A: \mu < 0.10$

Objectives

1. State the null and alternative hypothesis
2. Understand errors and interpret p-value
3. Perform a hypothesis test of the population mean with known population standard deviation

What Do Our Sample Results Mean?

Our sample results will allow us to do one of two things:

What Do Our Sample Results Mean?

Our sample results will allow us to do one of two things:

- Reject the null hypothesis

What Do Our Sample Results Mean?

Our sample results will allow us to do one of two things:

- Reject the null hypothesis
- Fail to reject the null hypothesis

Rejecting the Null Hypothesis

If our sample statistics give us reason to believe that the null hypothesis may not in fact be true, then we will state our rejection of it.

Rejecting the Null Hypothesis

If our sample statistics give us reason to believe that the null hypothesis may not in fact be true, then we will state our rejection of it.

> *... we have sufficient evidence to reject [the null hypothesis].*

Rejecting the Null Hypothesis

If our sample statistics give us reason to believe that the null hypothesis may not in fact be true, then we will state our rejection of it.

> *... we have sufficient evidence to reject [the null hypothesis].*

Rejecting the null hypothesis is like a jury declaring a defendant guilty.

Rejecting the Null Hypothesis

If our sample statistics give us reason to believe that the null hypothesis may not in fact be true, then we will state our rejection of it.

> *... we have sufficient evidence to reject [the null hypothesis].*

Rejecting the null hypothesis is like a jury declaring a defendant guilty.

Note: There is still a chance that the defendant is innocent, but the evidence is strong enough to bring a guilty verdict.

Failing to Reject the Null Hypothesis

However, our sample statistics might not give us reason to believe the null hypothesis is false.

Failing to Reject the Null Hypothesis

However, our sample statistics might not give us reason to believe the null hypothesis is false.

> *... we do not have sufficient evidence to reject [the null hypothesis].*

Failing to Reject the Null Hypothesis

However, our sample statistics might not give us reason to believe the null hypothesis is false.

> *... we do not have sufficient evidence to reject [the null hypothesis].*

Failing to reject the null hypothesis is like a jury declaring a defendant not guilty.

Failing to Reject the Null Hypothesis

However, our sample statistics might not give us reason to believe the null hypothesis is false.

> *... we do not have sufficient evidence to reject [the null hypothesis].*

Failing to reject the null hypothesis is like a jury declaring a defendant not guilty.

Note: A declaration of not guilty <u>is not the same as</u> a declaration of innocence. There just is not sufficient evidence to declare guilt, and the defendant *could still actually be guilty*.

Errors

Just like in courts of law, sometimes innocent people are declared guilty and guilty people are declared not guilty. The same can happen in statistics.

Errors

Just like in courts of law, sometimes innocent people are declared guilty and guilty people are declared not guilty. The same can happen in statistics.

Type I Error

A **Type I error** in statistics occurs when we reject a null hypothesis that is actually true.

Errors

Just like in courts of law, sometimes innocent people are declared guilty and guilty people are declared not guilty. The same can happen in statistics.

Type I Error

A **Type I error** in statistics occurs when we reject a null hypothesis that is actually true.

The probability of making a Type I error is α, which we saw earlier with our confidence intervals in the form $1 - \alpha$.

Errors

Just like in courts of law, sometimes innocent people are declared guilty and guilty people are declared not guilty. The same can happen in statistics.

Type I Error

A **Type I error** in statistics occurs when we reject a null hypothesis that is actually true.

The probability of making a Type I error is α, which we saw earlier with our confidence intervals in the form $1 - \alpha$.

In hypothesis testing, α is called the **level of significance**.

Errors

Type II Error

A **Type II error** in statistics occurs when we fail to reject a null hypothesis that is actually false.

Errors

Type II Error

A **Type II error** in statistics occurs when we fail to reject a null hypothesis that is actually false.

The probability of making a Type II error is β.

Errors

Type II Error

A **Type II error** in statistics occurs when we fail to reject a null hypothesis that is actually false.

The probability of making a Type II error is β.

Note: As we decrease the chances of making a Type I error, we increase the chances of making a Type II error.

Errors

Type II Error

A **Type II error** in statistics occurs when we fail to reject a null hypothesis that is actually false.

The probability of making a Type II error is β.

Note: As we decrease the chances of making a Type I error, we increase the chances of making a Type II error.

If we make it harder to put an innocent person in jail, we make it tougher to return a guilty verdict. This will also have the affect of increasing the number of actual guilty defendants who are let go.

Errors

Type II Error

A **Type II error** in statistics occurs when we fail to reject a null hypothesis that is actually false.

The probability of making a Type II error is β.

Note: As we decrease the chances of making a Type I error, we increase the chances of making a Type II error.

If we make it harder to put an innocent person in jail, we make it tougher to return a guilty verdict. This will also have the affect of increasing the number of actual guilty defendants who are let go.

The **power** of a test is given as $1 - P(\beta)$

Errors Summary

H_0	Reject H_0	Fail to reject H_0
H_0 True	Type I error	Correct decision
H_0 False	Correct decision	Type II error

Defendant	Declare Guilty	Declare Not Guilty
Actually Innocent	Type I error	Correct decision
Actually Guilty	Correct decision	Type II error

p-Value

Recall that **statistically significant** means that something is not likely to occur by chance.

p-Value

Recall that **statistically significant** means that something is not likely to occur by chance.

p-Value

The **p-value** is the probability of obtaining a sample as extreme as the one obtained **under the assumption that the null hypothesis is true.**

p-Value

Recall that **statistically significant** means that something is not likely to occur by chance.

p-Value

The **p-value** is the probability of obtaining a sample as extreme as the one obtained **under the assumption that the null hypothesis is true.**

If our *p*-value is less than a given acceptable value (α), then our sample was not likely to occur by chance *assuming the null hypothesis is true*,

p-Value

Recall that **statistically significant** means that something is not likely to occur by chance.

p-Value

The **p-value** is the probability of obtaining a sample as extreme as the one obtained **under the assumption that the null hypothesis is true.**

If our *p*-value is less than a given acceptable value (α), then our sample was not likely to occur by chance *assuming the null hypothesis is true*, so we have sufficient evidence to reject the null hypothesis.

Example 2

(a) A new car's mpg is listed as 33. You want to know if the mpg is not 33, so you perform a hypothesis test at the 5% level of significance. Your sample shows a mean mpg of 37, which has a probability of 2.2% of happening.

Based on your findings, should you reject or fail to reject the null hypothesis that $\mu = 33$?

Example 2

(a) A new car's mpg is listed as 33. You want to know if the mpg is not 33, so you perform a hypothesis test at the 5% level of significance. Your sample shows a mean mpg of 37, which has a probability of 2.2% of happening.

Based on your findings, should you reject or fail to reject the null hypothesis that $\mu = 33$?

If the null hypothesis is true, then the probability we would obtain a sample mean as extreme (or more) than 37 mpg is 2.2%.

Example 2

(a) A new car's mpg is listed as 33. You want to know if the mpg is not 33, so you perform a hypothesis test at the 5% level of significance. Your sample shows a mean mpg of 37, which has a probability of 2.2% of happening.

Based on your findings, should you reject or fail to reject the null hypothesis that $\mu = 33$?

If the null hypothesis is true, then the probability we would obtain a sample mean as extreme (or more) than 37 mpg is 2.2%.

Since our *p*-value is 2.2%, which is less than our significance level of 5%, we will reject the null hypothesis.

Example 2a

Final answer:

At the 5% significance level, we have sufficient evidence to reject the claim that the mean mpg is 33, and conclude that our evidence shows that the mean mpg differs from 33.

Example 2

(b) A hospital says the mean wait time for patients to see a doctor is 3.5 minutes. You want to know if the mean wait time is more than 3.5 minutes, so you perform a hypothesis test at the 10% level of significance. Your sample shoes a mean wait time of 3.8 minutes, which has a probability of 10.5%.

Based on your findings, should you reject or fail to reject the null hypothesis that $\mu = 3.5$?

Example 2

(b) A hospital says the mean wait time for patients to see a doctor is 3.5 minutes. You want to know if the mean wait time is more than 3.5 minutes, so you perform a hypothesis test at the 10% level of significance. Your sample shoes a mean wait time of 3.8 minutes, which has a probability of 10.5%.

Based on your findings, should you reject or fail to reject the null hypothesis that $\mu = 3.5$?

If the null hypothesis is true, then the probability we would obtain a sample mean as extreme (or more) than 3.5 minutes is 10.5%.

Example 2

(b) A hospital says the mean wait time for patients to see a doctor is 3.5 minutes. You want to know if the mean wait time is more than 3.5 minutes, so you perform a hypothesis test at the 10% level of significance. Your sample shoes a mean wait time of 3.8 minutes, which has a probability of 10.5%.

Based on your findings, should you reject or fail to reject the null hypothesis that $\mu = 3.5$?

If the null hypothesis is true, then the probability we would obtain a sample mean as extreme (or more) than 3.5 minutes is 10.5%.

Since our *p*-value is 10.5%, which is greater than our significance level of 10%, we will fail to reject the null hypothesis.

Example 2b

Final answer:

At the 10% significance level, we do not have sufficient evidence to reject the claim that the mean wait time is 3.5 minutes, and thus there is not enough evidence to conclude that the mean wait time is more than 3.5 minutes.

Example 2

(c) A company claims their program will increase your grade in statistics class by 10%. You think it might not be that much, so you perform a hypothesis test at the 1% significance level. You obtain a sample with a grade increase of 7%, which has a probability of 0.8% of occurring.

Based on your findings, should you reject or fail to reject the null hypothesis?

Example 2

(c) A company claims their program will increase your grade in statistics class by 10%. You think it might not be that much, so you perform a hypothesis test at the 1% significance level. You obtain a sample with a grade increase of 7%, which has a probability of 0.8% of occurring.

Based on your findings, should you reject or fail to reject the null hypothesis?

If the null hypothesis is true, then the probability we would obtain a sample mean increase as extreme (or more) than 7% is 0.8%.

Example 2

(c) A company claims their program will increase your grade in statistics class by 10%. You think it might not be that much, so you perform a hypothesis test at the 1% significance level. You obtain a sample with a grade increase of 7%, which has a probability of 0.8% of occurring.

Based on your findings, should you reject or fail to reject the null hypothesis?

If the null hypothesis is true, then the probability we would obtain a sample mean increase as extreme (or more) than 7% is 0.8%.

Since our p-value of 0.8% is less than our significance level of 1%, we reject the null hypothesis.

Example 2c

Final answer:

At the 1% significance level, we have sufficient evidence to reject the claim that the mean grade increase is 10%, and conclude that our evidence shows that the mean grade increase is less than 10%.

Objectives

1. State the null and alternative hypothesis
2. Understand errors and interpret p-value
3. Perform a hypothesis test of the population mean with known population standard deviation

Steps to Performing a Hypothesis Test

1. Check the assumptions

Steps to Performing a Hypothesis Test

1. Check the assumptions
2. State the null and alternative hypotheses

Steps to Performing a Hypothesis Test

1. Check the assumptions
2. State the null and alternative hypotheses
3. Determine the significance level
    - *Note*: Once you determine the significance level, you can not change it after you obtain the results of your sample. Doing so is unethical.

Steps to Performing a Hypothesis Test

1. Check the assumptions
2. State the null and alternative hypotheses
3. Determine the significance level
   - *Note*: Once you determine the significance level, you can not change it after you obtain the results of your sample. Doing so is unethical.
4. Calculate the p-value, test statistic, or confidence interval

Steps to Performing a Hypothesis Test

1. Check the assumptions
2. State the null and alternative hypotheses
3. Determine the significance level
   - *Note*: Once you determine the significance level, you can not change it after you obtain the results of your sample. Doing so is unethical.
4. Calculate the p-value, test statistic, or confidence interval
5. State your conclusion

Hypothesis Testing Methods

We will examine 3 methods of hypothesis testing:

Hypothesis Testing Methods

We will examine 3 methods of hypothesis testing:

- Test statistic and critical value

Hypothesis Testing Methods

We will examine 3 methods of hypothesis testing:

- Test statistic and critical value
- p-value

Hypothesis Testing Methods

We will examine 3 methods of hypothesis testing:

- Test statistic and critical value
- p-value
- Confidence interval

Hypothesis Test for Population Mean

We will look at some examples of hypothesis testing of the population mean when the population standard deviation is known.

Hypothesis Test for Population Mean

We will look at some examples of hypothesis testing of the population mean when the population standard deviation is known.

We will also assume that a simple random sample (or $n \geq 30$) is obtained and the population is normally distributed.

Example 3

A pharmaceutical company claims that the mean time for relief from its headache medicine is 3 minutes. A sample of 50 pills is obtained and the sample's mean is 3.1 minutes. Assuming $\sigma = 0.35$ minutes, test the claim that the mean remedy time is different than 3 minutes at the $\alpha = 0.05$ significance level.

Example 3

A pharmaceutical company claims that the mean time for relief from its headache medicine is 3 minutes. A sample of 50 pills is obtained and the sample's mean is 3.1 minutes. Assuming $\sigma = 0.35$ minutes, test the claim that the mean remedy time is different than 3 minutes at the $\alpha = 0.05$ significance level.

$H_0 : \mu = 3$ minutes

$H_A : \mu \neq 3$ minutes

Example 3

A pharmaceutical company claims that the mean time for relief from its headache medicine is 3 minutes. A sample of 50 pills is obtained and the sample's mean is 3.1 minutes. Assuming $\sigma = 0.35$ minutes, test the claim that the mean remedy time is different than 3 minutes at the $\alpha = 0.05$ significance level.

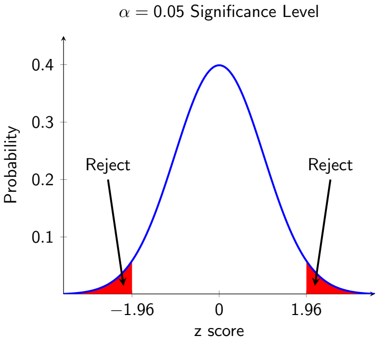
$H_0: \mu = 3$ minutes

$H_A: \mu \neq 3$ minutes

We will first examine the test statistic and critical value approach.

Example 3 – Test Statistic and Critical Value

In this approach, we will compute the z test statistic and see if it falls in the acceptance or rejection region:



Example 3 – Test Statistic and Critical Value

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$$

Example 3 – Test Statistic and Critical Value

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{3.1 - 3}{0.35/\sqrt{50}}$$

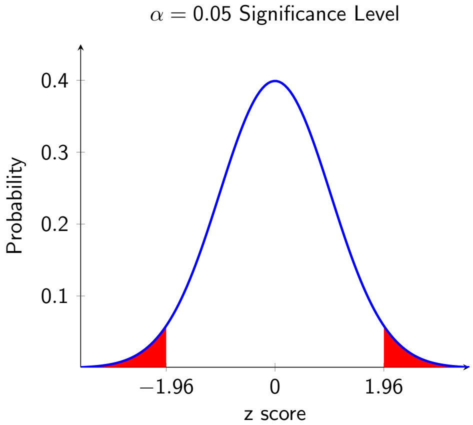
Example 3 – Test Statistic and Critical Value

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{3.1 - 3}{0.35/\sqrt{50}}$$

$$z \approx 2.02$$

Example 3 – Test Statistic and Critical Value



Example 3 – Test Statistic and Critical Value

Since our test statistic of $z = 2.02$ is in the rejection region, we reject the null hypothesis.

Example 3 – Test Statistic and Critical Value

Since our test statistic of $z = 2.02$ is in the rejection region, we reject the null hypothesis.

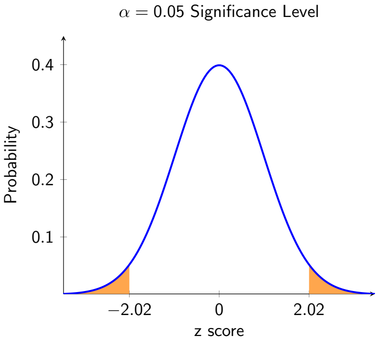
> *At the 5% level of significance, we have sufficient evidence to reject the null hypothesis that the mean remedy time is 3 minutes and conclude that there is evidence to prove the mean remedy time is different.*

Example 3 – p-value

We can obtain the p-value for our sample by computing the total area associated with our z test statistic.

Example 3 – *p*-value

We can obtain the *p*-value for our sample by computing the total area associated with our *z* test statistic.



The *p*-value is approximately 0.0434

Example 3 – *p*-value

Our p-value of 0.0434, or 4.34%, means that assuming the mean remedy time of 3 minutes is true, we would get results as extreme (or more) as ours about 4.34% of the time.

Example 3 – *p*-value

Our *p*-value of 0.0434, or 4.34%, means that assuming the mean remedy time of 3 minutes is true, we would get results as extreme (or more) as ours about 4.34% of the time.

Since this value is less than the $\alpha = 0.05$ significance level, our results are statistically significant, and we thus reject the null hypothesis.

Example 3 – *p*-value

Our *p*-value of 0.0434, or 4.34%, means that assuming the mean remedy time of 3 minutes is true, we would get results as extreme (or more) as ours about 4.34% of the time.

Since this value is less than the $\alpha = 0.05$ significance level, our results are statistically significant, and we thus reject the null hypothesis.

> *At the 5% level of significance, we have sufficient evidence to reject the null hypothesis that the mean remedy time is 3 minutes and conclude that there is evidence to prove the mean remedy time is different.*

Example 3 – Confidence Interval

Recall that confidence intervals for a population mean are in the form

$$\overline{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Example 3 – Confidence Interval

Recall that confidence intervals for a population mean are in the form

$$\overline{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Example 3's $\alpha = 0.05$ significance level means that if we took many samples, we would expect about 95% of them to contain the population mean.

Example 3 – Confidence Interval

Recall that confidence intervals for a population mean are in the form

$$\overline{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Example 3's $\alpha = 0.05$ significance level means that if we took many samples, we would expect about 95% of them to contain the population mean.

As such, if our confidence interval contains the (alleged) population mean of 3 minutes, then we will not reject the null hypothesis.

Example 3 – Confidence Interval

Our sample's 95% confidence interval is

$$3.1 \pm 1.96\left(0.35/\sqrt{50}\right)$$

Example 3 – Confidence Interval

Our sample's 95% confidence interval is

$$3.1 \pm 1.96 \left(0.35/\sqrt{50}\right)$$

3.003 to 3.197

Example 3 – Confidence Interval

Our sample's 95% confidence interval is

$$3.1 \pm 1.96\left(0.35/\sqrt{50}\right)$$

3.003 to 3.197

Since our confidence interval does not contain the population mean of 3, we reject the null hypothesis.

Example 3 – Confidence Interval

Our sample's 95% confidence interval is

$$3.1 \pm 1.96 \left(0.35/\sqrt{50}\right)$$

3.003 to 3.197

Since our confidence interval does not contain the population mean of 3, we reject the null hypothesis.

> *At the 5% level of significance, we have sufficient evidence to reject the null hypothesis that the mean remedy time is 3 minutes and conclude that there is evidence to prove the mean remedy time is different.*

Example 4