Statistical Significance II

Hypothesis Testing

Notation:

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

What we test:

$$H_0(\text{Null Hypothesis})$$

Hypothesis Testing

Notation:

What we expect:

$$H_a (\text{Alternative Hypothesis})$$

What we test:

$$H_0 (\text{Null Hypothesis})$$

Examples:

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

What we test:

$$H_0(\text{Null Hypothesis})$$

Examples:

- "Higher poverty rates increase incarceration rates"

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

What we test:

$$H_0(\text{Null Hypothesis})$$

Examples:

- "Higher poverty rates increase incarceration rates"
- "Higher poverty rates change incarceration rates"

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

What we test:

$$H_0(\text{Null Hypothesis})$$

Examples:

- "Higher poverty rates increase incarceration rates"
- "Higher poverty rates change incarceration rates"
- "Higher poverty rates decrease incarceration rates"

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

What we test:

$$H_0(\text{Null Hypothesis})$$

Examples:

- "Higher poverty rates increase incarceration rates"
- "Higher poverty rates change incarceration rates"
- "Higher poverty rates decrease incarceration rates"

Corresponding nulls:

Hypothesis Testing

Notation:

What we expect:

$$H_a (\text{Alternative Hypothesis})$$

What we test:

$$H_0 (\text{Null Hypothesis})$$

Examples:

- "Higher poverty rates increase incarceration rates"
- "Higher poverty rates change incarceration rates"
- "Higher poverty rates decrease incarceration rates"

Corresponding nulls:

- "Higher poverty rates have no effect on or decrease incarceration rates"

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

What we test:

$$H_0(\text{Null Hypothesis})$$

Examples:

- "Higher poverty rates increase incarceration rates"
- "Higher poverty rates change incarceration rates"
- "Higher poverty rates decrease incarceration rates"

Corresponding nulls:

- "Higher poverty rates have no effect on or decrease incarceration rates"
- "Higher poverty rates have no effect on incarceration rates"

Hypothesis Testing

Notation:

What we expect:

$$H_a(\text{Alternative Hypothesis})$$

What we test:

$$H_0(\text{Null Hypothesis})$$

Examples:

- "Higher poverty rates increase incarceration rates"
- "Higher poverty rates change incarceration rates"
- "Higher poverty rates decrease incarceration rates"

Corresponding nulls:

- "Higher poverty rates have no effect on or decrease incarceration rates"
- "Higher poverty rates have no effect on incarceration rates"
- "Higher poverty rates have no effect on or increase incarceration rates"

What do we do with a null hypothesis?

We can never say:

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.
2. The null hypothesis is false.

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.
2. The null hypothesis is false.

Instead, we can say:

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.
2. The null hypothesis is false.

Instead, we can say:

1. There is OR is not significant evidence supporting our hypothesis.

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.
2. The null hypothesis is false.

Instead, we can say:

1. There is OR is not significant evidence supporting our hypothesis.
2. We can reject OR fail to reject the null hypothesis

What do we do with a null hypothesis?

We can never say:

1. Our hypothesis is true/correct.
2. The null hypothesis is false.

Instead, we can say:

1. There is OR is not significant evidence supporting our hypothesis.
2. We can reject OR fail to reject the null hypothesis

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

We will run a statistical test designed to tell us how likely it is our results are due to random chance

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

We will run a statistical test designed to tell us how likely it is our results are due to random chance

A variety of tests:

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

We will run a statistical test designed to tell us how likely it is our results are due to random chance

A variety of tests:

- t-test

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

We will run a statistical test designed to tell us how likely it is our results are due to random chance

A variety of tests:

- t-test
- f-test

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

We will run a statistical test designed to tell us how likely it is our results are due to random chance

A variety of tests:

- t-test
- f-test
- χ^2 test

The next step: standard of evidence

How do we know what to make of our hypothesis and the null hypothesis?

We will run a statistical test designed to tell us how likely it is our results are due to random chance

A variety of tests:

- t-test
- f-test
- χ^2 test

All give us a value we must interpret

What do we do with this value?

We impose a threshold.

What do we do with this value?

We impose a threshold.

If the result of the statistical test meets the established threshold:

What do we do with this value?

We impose a threshold.

If the result of the statistical test meets the established threshold:

- then we have significance

What do we do with this value?

We impose a threshold.

If the result of the statistical test meets the established threshold:

- then we have significance

If the result of the statistical test does not meet the established threshold:

What do we do with this value?

We impose a threshold.

If the result of the statistical test meets the established threshold:

- then we have significance

If the result of the statistical test does not meet the established threshold:

- then we do not have significance

Where does this threshold come from?

Sir Ronald A. Fisher

Where does this threshold come from?

Sir Ronald A. Fisher

Where does this threshold come from?

- Originally a “rough guide” for the strength of evidence against the null hypothesis.

Where does this threshold come from?

- Originally a “rough guide” for the strength of evidence against the null hypothesis.
- Honestly, very arbitrary.

Where does this threshold come from?

- Originally a “rough guide” for the strength of evidence against the null hypothesis.
- Honestly, very arbitrary.
  - We use a 5% chance of a false positive.

Where does this threshold come from?

- Originally a “rough guide” for the strength of evidence against the null hypothesis.
- Honestly, very arbitrary.
  - We use a 5% chance of a false positive.
  - Physicist use a .0005% threshold

Where does this threshold come from?

- Originally a “rough guide” for the strength of evidence against the null hypothesis.
- Honestly, very arbitrary.
  - We use a 5% chance of a false positive.
  - Physicist use a .0005% threshold

We use a lower threshold because of sample size limitations and cost concerns.

What do we do with the results from a test?

- Three “signs” of significance

What do we do with the results from a test?

- Three "signs" of significance
- All are equivalent at the 5% threshold

What do we do with the results from a test?

- Three "signs" of significance
- All are equivalent at the 5% threshold

1. T-statistic $>$ 1.96 or t-statistic $<$ -1.96

What do we do with the results from a test?

- Three “signs” of significance
- All are equivalent at the 5% threshold

1. T-statistic $>$ 1.96 or t-statistic $<$ -1.96
2. P-value $<$.05

What do we do with the results from a test?

- Three "signs" of significance
- All are equivalent at the 5% threshold

1. T-statistic $>$ 1.96 or t-statistic $<$ -1.96
2. P-value $<$.05
3. 95% confidence intervals do *not* include 0

What do we do with the results from a test?

- Three "signs" of significance
- All are equivalent at the 5% threshold

1. T-statistic $>$ 1.96 or t-statistic $<$ -1.96
2. P-value $<$.05
3. 95% confidence intervals do *not* include 0

These ideas are all connected

- A p-value is computed from a t-statistic

These ideas are all connected

- A p-value is computed from a t-statistic
- 95% confidence intervals are computed from (indirectly) t-statistics

These ideas are all connected

- A p-value is computed from a t-statistic
- 95% confidence intervals are computed from (indirectly) t-statistics

What do they mean?

The t-statistic

A statistic to evaluate if a mean or a difference in means when hypothesis testing

The t-statistic

A statistic to evaluate if a mean or a difference in means when hypothesis testing

- Provides information on if what we observed matches what we expected

The t-statistic

A statistic to evaluate if a mean or a difference in means when hypothesis testing

- Provides information on if what we observed matches what we expected

Significant: If the t-statistic is large (positive or negative), it means the difference between the groups is unlikely to have happened by random chance.

The t-statistic

A statistic to evaluate if a mean or a difference in means when hypothesis testing

- Provides information on if what we observed matches what we expected

Significant: If the t-statistic is large (positive or negative), it means the difference between the groups is unlikely to have happened by random chance.

Not significant: If the t-statistic is small, it means the difference could easily be due to random variation.

P-values

A p-value is a number that helps determine the significance of results when running a hypothesis test.

- More formally, the probability of observing our results if the null hypothesis were true.

Significant: A low p-value ($<=.05$) indicates we should reject the null hypothesis—there might be a real effect or difference.

P-values

A p-value is a number that helps determine the significance of results when running a hypothesis test.

- More formally, the probability of observing our results if the null hypothesis were true.

Significant: A low p-value ($<=$ .05) indicates we should reject the null hypothesis—there might be a real effect or difference.

Not significant: A high p-value ($>$.05) suggests that there's not enough evidence to reject the null hypothesis, and any observed difference might be due to random chance.

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

- With a sample we can be wrong!

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

- With a sample we can be wrong!
- There will be error.

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

- With a sample we can be wrong!
- There will be error.
- But how close did we get?

Confidence intervals are more complex

Assume we can not measure attitudes or features of everyone and that we will rely on a sample.

- With a sample we can be wrong!
- There will be error.
- But how close did we get?

A 95% confidence interval gives a range of values, that should, 95% of the time, contain the true value of what we are estimating.

How to think about a confidence interval

- Assume we can repeat a survey 100 times.

How to think about a confidence interval

- Assume we can repeat a survey 100 times.
- Assume we compute a confidence interval for each survey

How to think about a confidence interval

- Assume we can repeat a survey 100 times.
- Assume we compute a confidence interval for each survey
- 95% of those intervals would contain the true population mean.

How do we compute a confidence interval

Two components:

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

The confidence interval is computed by adding and subtracting the measure of error (and a standard value).

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

The confidence interval is computed by adding and subtracting the measure of error (and a standard value).

So, let estimate heights of sixth grade students. We get a a mean of 50in with a standard error of 5in would have the following CI:

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

The confidence interval is computed by adding and subtracting the measure of error (and a standard value).

So, let estimate heights of sixth grade students. We get a a mean of 50in with a standard error of 5in would have the following CI:

$$\text{Upper CI} = 50 + 1.96 * 5 = 59.8$$

$$\text{Lower CI} = 50 - 1.96 * 5 = 40.2$$

How do we compute a confidence interval

Two components:

1. A mean, coefficient, or mean difference
2. A measure of error

The confidence interval is computed by adding and subtracting the measure of error (and a standard value).

So, let estimate heights of sixth grade students. We get a a mean of 50in with a standard error of 5in would have the following CI:

$$\text{Upper CI} = 50 + 1.96 * 5 = 59.8$$

$$\text{Lower CI} = 50 - 1.96 * 5 = 40.2$$

We would write this as "95% CI [40.2, 59.8]"

Interpretation:

- We can be 95% confident that the true average height of students is between 40.2in and 59.8in.

Interpretation:

- We can be 95% confident that the true average height of students is between 40.2in and 59.8in.
- This is **not absolute certainty.**

Interpretation:

- We can be 95% confident that the true average height of students is between 40.2in and 59.8in.
- This is **not absolute certainty.**

A 95% confidence interval is a way to estimate the true value of a population parameter, providing a range that likely includes the true value based on your sample data.

Example 1

Testing the impact of a college degree on voting behavior.

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout.

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Difference: 20% increase in voter turnout.

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Difference: 20% increase in voter turnout. p-value: 0.04; t-statistic: 5.23; 95% Confidence Interval: [17.23, 22.14]

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Difference: 20% increase in voter turnout. p-value: 0.04; t-statistic: 5.23; 95% Confidence Interval: [17.23, 22.14]

Conclusion:

Example 1

Testing the impact of a college degree on voting behavior.

Hypotheses:

H_a: Holding a college degree increase voter turnout. H_0: Holding a college degree has no effect on voter turnout.

Results:

Difference: 20% increase in voter turnout. p-value: 0.04; t-statistic: 5.23; 95% Confidence Interval: [17.23, 22.14]

Conclusion:

A college degree increases voting by 20% (95% CI [17.23, 22.14]). This is a statistically significant effect.

Example 2

Scenario:

Testing the relationship between anger and support for violence

Example 2

Scenario:

Testing the relationship between anger and support for violence

Hypotheses:

Example 2

Scenario:

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence.

Example 2

Scenario:

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

Example 2

Scenario:

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

Control Group Support: 15%. Angry Group Support: 72%. Difference: 57%

Example 2

Scenario:

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

Control Group Support: 15%. Angry Group Support: 72%. Difference: 57%P-Value < 0.001; t-statistic: 19.23; 95% Confidence Interval: [53.7, 60.3]

Example 2

Scenario:

Testing the relationship between anger and support for violence

Hypotheses:

H_a: Anger decreases support for violence. H_0: Anger has no effect on support for violence.

Control Group Support: 15%. Angry Group Support: 72%. Difference: 57%P-Value $<$ 0.001; t-statistic: 19.23; 95% Confidence Interval: [53.7, 60.3]

Conclusion: Anger increases support for violence 15% (95% CI [53.7, 60.3]). This is a statistically significant effect.

Example 3

Scenario:

Testing the relationship between SAT scores and college graduation rates.

Example 3

Scenario:

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

Example 3

Scenario:

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates.

Example 3

Scenario:

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

Example 3

Scenario:

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

LOW SAT Group: 70%; High SAT Group: 73%; Difference: 3%

Example 3

Scenario:

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

LOW SAT Group: 70%; High SAT Group: 73%; Difference: 3% p-value: 0.08; t-statistic: 1.45; 95% Confidence Interval: [-1.2, 7.2]

Example 3

Scenario:

Testing the relationship between SAT scores and college graduation rates.

Hypotheses:

H_a: SAT scores increase college graduation rates. H_0: SAT scores have no effect on college graduation rates.

LOW SAT Group: 70%; High SAT Group: 73%; Difference: 3% p-value: 0.08; t-statistic: 1.45; 95% Confidence Interval: [-1.2, 7.2]

Conclusion: SAT scores are not significantly related to college graduation rates, with a difference in graduation rates of 3% (95% CI [-1.2, 7.2]). This is a statistically significant effect.