STAT 2593

Lecture 029 - Hypothesis and Test Procedures

Dylan Spicker

Hypothesis and Test Procedures

Learning Objectives

1. Understand the framework of null hypothesis significance testing.
2. Be able to identify the null hypothesis and alternative hypothesis.
3. Understand and interpret p-values, and hypothesis testing conclusions.
4. Understand type I and type II errors.

Source: https://www.fredhutch.org/en/news/center-news/2020/02/spinning-science-overhyped-headlines-snarled-statistics-lead-readers-astray.html

Hypothesis Tests

- We have seen both **point estimates** and **interval estimates**.

Hypothesis Tests

- We have seen both **point estimates** and **interval estimates**.
- Often, we will wish to *make a decision* based on what we observe in a sample.

Hypothesis Tests

- We have seen both **point estimates** and **interval estimates**.
- Often, we will wish to *make a decision* based on what we observe in a sample.
- **Hypothesis tests** provide one mechanism for making decisions based on statistical analyses.

Hypothesis Tests

- We have seen both **point estimates** and **interval estimates**.
- Often, we will wish to *make a decision* based on what we observe in a sample.
- **Hypothesis tests** provide one mechanism for making decisions based on statistical analyses.
  - A hypothesis test uses statistical inference to make decisions regarding the **value of a parameter**.

Hypothesis Tests

- We have seen both **point estimates** and **interval estimates**.
- Often, we will wish to *make a decision* based on what we observe in a sample.
- **Hypothesis tests** provide one mechanism for making decisions based on statistical analyses.
  - A hypothesis test uses statistical inference to make decisions regarding the **value of a parameter**.
  - In statistics, a **hypothesis** is a claim about the value of a parameter.

Hypothesis Tests

- We have seen both **point estimates** and **interval estimates**.
- Often, we will wish to *make a decision* based on what we observe in a sample.
- **Hypothesis tests** provide one mechanism for making decisions based on statistical analyses.
  - A hypothesis test uses statistical inference to make decisions regarding the **value of a parameter**.
  - In statistics, a **hypothesis** is a claim about the value of a parameter.
  - **Hypothesis tests** weigh evidence against competing statistical hypotheses to infer which seems more probable, given the observations.

The Hypothesis Testing Framework

- There is generally a **six step procedure** for testing hypotheses:

The Hypothesis Testing Framework

- There is generally a **six step procedure** for testing hypotheses:
    1. State the **null** and **alternative hypothesis**.

The Hypothesis Testing Framework

- There is generally a **six step procedure** for testing hypotheses:

  1. State the **null** and **alternative hypothesis**.
  2. Decide on a **significance level**, α, at which to test.

The Hypothesis Testing Framework

- There is generally a **six step procedure** for testing hypotheses:

  1. State the **null** and **alternative hypothesis**.
  2. Decide on a **significance level**, α, at which to test.
  3. Compute your **test statistic**.

The Hypothesis Testing Framework

- There is generally a **six step procedure** for testing hypotheses:

    1. State the **null** and **alternative hypothesis**.
    2. Decide on a **significance level**, α, at which to test.
    3. Compute your **test statistic**.
    4. Find the **p-value** or consider a **critical value**.

The Hypothesis Testing Framework

- There is generally a **six step procedure** for testing hypotheses:

1. State the **null** and **alternative hypothesis**.
2. Decide on a **significance level**, α, at which to test.
3. Compute your **test statistic**.
4. Find the **p-value** or consider a **critical value**.
5. Make the comparison.

The Hypothesis Testing Framework

- There is generally a **six step procedure** for testing hypotheses:

  1. State the **null** and **alternative hypothesis**.
  2. Decide on a **significance level**, α, at which to test.
  3. Compute your **test statistic**.
  4. Find the **p-value** or consider a **critical value**.
  5. Make the comparison.
  6. **Interpret** the results to draw your conclusions.

The Null and Alternative Hypotheses

- The **null hypothesis** is a claim that is **assumed to be true** without evidence to the contrary.

The Null and Alternative Hypotheses

- The **null hypothesis** is a claim that is **assumed to be true** without evidence to the contrary.
  - It represents the *default* state of the world, and is designed to be conservative.

The Null and Alternative Hypotheses

- The **null hypothesis** is a claim that is **assumed to be true** without evidence to the contrary.
  - It represents the *default* state of the world, and is designed to be conservative.
  - Typically it will take the form of "no effect" ($\theta = \theta_0$) or equivalent in the setting.

The Null and Alternative Hypotheses

- The **null hypothesis** is a claim that is **assumed to be true** without evidence to the contrary.
  - It represents the *default* state of the world, and is designed to be conservative.
  - Typically it will take the form of "no effect" ($\theta = \theta_0$) or equivalent in the setting.
  - We denote this as H_0.

The Null and Alternative Hypotheses

- The **null hypothesis** is a claim that is **assumed to be true** without evidence to the contrary.
  - It represents the *default* state of the world, and is designed to be conservative.
  - Typically it will take the form of "no effect" ($\theta = \theta_0$) or equivalent in the setting.
  - We denote this as H_0.
- The **alternative hypothesis** is the claim that we are interested in seeing if it is **supported by the evidence**.

The Null and Alternative Hypotheses

- The **null hypothesis** is a claim that is **assumed to be true** without evidence to the contrary.
  - It represents the *default* state of the world, and is designed to be conservative.
  - Typically it will take the form of "no effect" ($\theta = \theta_0$) or equivalent in the setting.
  - We denote this as H_0.
- The **alternative hypothesis** is the claim that we are interested in seeing if it is **supported by the evidence**.
  - This is always defined in contrast to the null.

The Null and Alternative Hypotheses

- The **null hypothesis** is a claim that is **assumed to be true** without evidence to the contrary.
  - It represents the *default* state of the world, and is designed to be conservative.
  - Typically it will take the form of "no effect" ($\theta = \theta_0$) or equivalent in the setting.
  - We denote this as H_0.
- The **alternative hypothesis** is the claim that we are interested in seeing if it is **supported by the evidence**.
  - This is always defined in contrast to the null.
  - Typically will take the form of "effect exists" ($\theta \neq \theta_0$) or similar.

The Null and Alternative Hypotheses

- The **null hypothesis** is a claim that is **assumed to be true** without evidence to the contrary.
  - It represents the *default* state of the world, and is designed to be conservative.
  - Typically it will take the form of "no effect" $(\theta = \theta_0)$ or equivalent in the setting.
  - We denote this as H_0.
- The **alternative hypothesis** is the claim that we are interested in seeing if it is **supported by the evidence**.
  - This is always defined in contrast to the null.
  - Typically will take the form of "effect exists" $(\theta \neq \theta_0)$ or similar.
  - We denote this as H_A or H_1.

The Null and Alternative Hypotheses

- The **null hypothesis** is a claim that is **assumed to be true** without evidence to the contrary.
  - It represents the *default* state of the world, and is designed to be conservative.
  - Typically it will take the form of “no effect” ($\theta = \theta_0$) or equivalent in the setting.
  - We denote this as H_0.
- The **alternative hypothesis** is the claim that we are interested in seeing if it is **supported by the evidence**.
  - This is always defined in contrast to the null.
  - Typically will take the form of “effect exists” ($\theta \neq \theta_0$) or similar.
  - We denote this as H_A or H_1.
- Important: your hypotheses **must** be selected **before** looking at the data.

Significance Level

- We have seen significance levels for confidence intervals.

Significance Level

- We have seen significance levels for confidence intervals.
  - The relationship is more than aesthetic: in fact, hypothesis tests are equivalent to confidence intervals.

Significance Level

- We have seen significance levels for confidence intervals.
  - The relationship is more than aesthetic: in fact, hypothesis tests are equivalent to confidence intervals.
  - More on this later.

Significance Level

- We have seen significance levels for confidence intervals.
  - The relationship is more than aesthetic: in fact, hypothesis tests are equivalent to confidence intervals.
  - More on this later.
- The significance level is the threshold of evidence that you need *against* your null hypothesis in order to reject it.

Significance Level

- We have seen significance levels for confidence intervals.
  - The relationship is more than aesthetic: in fact, hypothesis tests are equivalent to confidence intervals.
  - More on this later.
- The significance level is the threshold of evidence that you need *against* your null hypothesis in order to reject it.
  - Lower values of α equate to needing stronger evidence against the null.

Significance Level

- We have seen significance levels for confidence intervals.
  - The relationship is more than aesthetic: in fact, hypothesis tests are equivalent to confidence intervals.
  - More on this later.
- The significance level is the threshold of evidence that you need *against* your null hypothesis in order to reject it.
  - Lower values of α equate to needing stronger evidence against the null.
  - How unlikely do you think that the null hypothesis needs to be in order to act as though it were untrue?

Significance Level

- We have seen significance levels for confidence intervals.
  - The relationship is more than aesthetic: in fact, hypothesis tests are equivalent to confidence intervals.
  - More on this later.
- The significance level is the threshold of evidence that you need *against* your null hypothesis in order to reject it.
  - Lower values of α equate to needing stronger evidence against the null.
  - How unlikely do you think that the null hypothesis needs to be in order to act as though it were untrue?
  - Extraordinary claims require extraordinary evidence.

Significance Level

- We have seen significance levels for confidence intervals.
  - The relationship is more than aesthetic: in fact, hypothesis tests are equivalent to confidence intervals.
  - More on this later.
- The significance level is the threshold of evidence that you need *against* your null hypothesis in order to reject it.
  - Lower values of α equate to needing stronger evidence against the null.
  - How unlikely do you think that the null hypothesis needs to be in order to act as though it were untrue?
  - Extraordinary claims require extraordinary evidence.
  - In practice, people select $\alpha = 0.05$ and move on. Don't do this.

Test Statistic

- A **test statistic** is a value which is computed with a **known distribution** assuming the null is true.

Test Statistic

- A **test statistic** is a value which is computed with a **known distribution** assuming the null is true.
  - The test statistic will be related to the parameter of interest.

Test Statistic

- A **test statistic** is a value which is computed with a **known distribution** assuming the null is true.
  - The test statistic will be related to the parameter of interest.
  - Specifically, it will related to the value of the parameter implied by the null hypothesis.

Test Statistic

- A **test statistic** is a value which is computed with a **known distribution** assuming the null is true.
  - The test statistic will be related to the parameter of interest.
  - Specifically, it will related to the value of the parameter implied by the null hypothesis.
- Test statistics are statistics, and as such, are random variables.

Test Statistic

- A **test statistic** is a value which is computed with a **known distribution** assuming the null is true.
  - The test statistic will be related to the parameter of interest.
  - Specifically, it will related to the value of the parameter implied by the null hypothesis.
- Test statistics are statistics, and as such, are random variables.
  - They have sampling distributions, which can be used to make probability statements.

Test Statistic

- A **test statistic** is a value which is computed with a **known distribution** assuming the null is true.
  - The test statistic will be related to the parameter of interest.
  - Specifically, it will related to the value of the parameter implied by the null hypothesis.
- Test statistics are statistics, and as such, are random variables.
  - They have sampling distributions, which can be used to make probability statements.
  - The premise of hypothesis testing is to compute a test statistic, and use its sampling distribution to make inferences regarding the null hypothesis.

P-Values

- Our **null hypothesis** will be likely be stated in the form of $H_0 : \theta = \theta_0$.

P-Values

- Our **null hypothesis** will be likely be stated in the form of $H_0 : \theta = \theta_0$.
- **If the null hypothesis is true**, then our test statistic T has a known distribution.

P-Values

- Our **null hypothesis** will be likely be stated in the form of $H_0 : \theta = \theta_0$.
- **If the null hypothesis is true**, then our test statistic T has a known distribution.
  - We can ask: What is the probability, *assuming that the null is true* that we would observe a test statistic which is **at least as extreme** as the one we did, by random chance.

P-Values

- Our **null hypothesis** will be likely be stated in the form of $H_0 : \theta = \theta_0$.
- **If the null hypothesis is true**, then our test statistic T has a known distribution.
  - We can ask: What is the probability, *assuming that the null is true* that we would observe a test statistic which is **at least as extreme** as the one we did, by random chance.
  - If this probability is very small, then it is *unlikely* that we would observe the sample we did *if* the null hypothesis were true.

P-Values

- Our **null hypothesis** will be likely be stated in the form of $H_0 : \theta = \theta_0$.
- **If the null hypothesis is true**, then our test statistic T has a known distribution.
  - We can ask: What is the probability, *assuming that the null is true* that we would observe a test statistic which is **at least as extreme** as the one we did, by random chance.
  - If this probability is very small, then it is *unlikely* that we would observe the sample we did *if* the null hypothesis were true.
  - This probability is a **p-value**.

P-Values

- Our **null hypothesis** will be likely be stated in the form of $H_0 : \theta = \theta_0$.
- **If the null hypothesis is true**, then our test statistic T has a known distribution.
  - We can ask: What is the probability, *assuming that the null is true* that we would observe a test statistic which is **at least as extreme** as the one we did, by random chance.
  - If this probability is very small, then it is *unlikely* that we would observe the sample we did *if* the null hypothesis were true.
  - This probability is a **p-value**.
- The p-value is not:

P-Values

- Our **null hypothesis** will be likely be stated in the form of $H_0 : \theta = \theta_0$.
- **If the null hypothesis is true**, then our test statistic T has a known distribution.
  - We can ask: What is the probability, *assuming that the null is true* that we would observe a test statistic which is **at least as extreme** as the one we did, by random chance.
  - If this probability is very small, then it is *unlikely* that we would observe the sample we did *if* the null hypothesis were true.
  - This probability is a **p-value**.
- The p-value is not:
  - The probability that the null hypothesis is false.

P-Values

- Our **null hypothesis** will be likely be stated in the form of $H_0 : \theta = \theta_0$.
- **If the null hypothesis is true**, then our test statistic T has a known distribution.
  - We can ask: What is the probability, *assuming that the null is true* that we would observe a test statistic which is **at least as extreme** as the one we did, by random chance.
  - If this probability is very small, then it is *unlikely* that we would observe the sample we did *if* the null hypothesis were true.
  - This probability is a **p-value**.
- The p-value is not:
  - The probability that the null hypothesis is false.
  - The probability that we make an error in the hypothesis test.

P-Values

- Our **null hypothesis** will be likely be stated in the form of $H_0 : \theta = \theta_0$.
- **If the null hypothesis is true**, then our test statistic T has a known distribution.
  - We can ask: What is the probability, *assuming that the null is true* that we would observe a test statistic which is **at least as extreme** as the one we did, by random chance.
  - If this probability is very small, then it is *unlikely* that we would observe the sample we did *if* the null hypothesis were true.
  - This probability is a **p-value**.
- The p-value is not:
  - The probability that the null hypothesis is false.
  - The probability that we make an error in the hypothesis test.
  - The probability that the alternative is true.

Critical Values and P-Value Comparisons

- We have seen **critical values** for confidence interval construction.

Critical Values and P-Value Comparisons

- We have seen **critical values** for confidence interval construction.
  - The critical value is the value based on the sampling distribution which corresponds to a (one-sided) α p-value.

Critical Values and P-Value Comparisons

- We have seen **critical values** for confidence interval construction.
  - The critical value is the value based on the sampling distribution which corresponds to a (one-sided) α p-value.
  - Values more extreme than critical values happen with probability less than α.

Critical Values and P-Value Comparisons

- We have seen **critical values** for confidence interval construction.
  - The critical value is the value based on the sampling distribution which corresponds to a (one-sided) α p-value.
  - Values more extreme than critical values happen with probability less than α.
  - P-values can be thought of as **observed significance levels**.

Critical Values and P-Value Comparisons

- We have seen **critical values** for confidence interval construction.
  - The critical value is the value based on the sampling distribution which corresponds to a (one-sided) α p-value.
  - Values more extreme than critical values happen with probability less than α.
  - P-values can be thought of as **observed significance levels**.
  - They are the smallest α that we would have drawn our relevant conclusion for.

Critical Values and P-Value Comparisons

- We have seen **critical values** for confidence interval construction.
  - The critical value is the value based on the sampling distribution which corresponds to a (one-sided) α p-value.
  - Values more extreme than critical values happen with probability less than α.
  - P-values can be thought of as **observed significance levels**.
  - They are the smallest α that we would have drawn our relevant conclusion for.
- With either a p-value or a critical value we can **draw comparisons**.

Critical Values and P-Value Comparisons

- We have seen **critical values** for confidence interval construction.
  - The critical value is the value based on the sampling distribution which corresponds to a (one-sided) α p-value.
  - Values more extreme than critical values happen with probability less than α.
  - P-values can be thought of as **observed significance levels**.
  - They are the smallest α that we would have drawn our relevant conclusion for.
- With either a p-value or a critical value we can **draw comparisons**.
  - If the p-value is below our significance level we have *strong enough* evidence *against* the null hypothesis.

Critical Values and P-Value Comparisons

- We have seen **critical values** for confidence interval construction.
  - The critical value is the value based on the sampling distribution which corresponds to a (one-sided) α p-value.
  - Values more extreme than critical values happen with probability less than α.
  - P-values can be thought of as **observed significance levels**.
  - They are the smallest α that we would have drawn our relevant conclusion for.
- With either a p-value or a critical value we can **draw comparisons**.
  - If the p-value is below our significance level we have *strong enough* evidence *against* the null hypothesis.
  - If the test statistic falls beyond the critical value(s) then we have *strong enough* evidence *against* the null hypothesis.

Drawing Conclusions and Interpretation

- If we observe $p < \alpha$ (or equivalent with critical values):

Drawing Conclusions and Interpretation

- If we observe $p < \alpha$ (or equivalent with critical values):
  - We have strong enough evidence against the null hypothesis.

Drawing Conclusions and Interpretation

- If we observe $p < \alpha$ (or equivalent with critical values):
  - We have strong enough evidence against the null hypothesis.
  - We decide to **reject the null hypothesis** in favour of the alternative.

Drawing Conclusions and Interpretation

- If we observe $p < \alpha$ (or equivalent with critical values):
  - We have strong enough evidence against the null hypothesis.
  - We decide to **reject the null hypothesis** in favour of the alternative.
  - In practice, we act as though the null is untrue, and that the alternative is supported.

Drawing Conclusions and Interpretation

- If we observe $p < \alpha$ (or equivalent with critical values):
  - We have strong enough evidence against the null hypothesis.
  - We decide to **reject the null hypothesis** in favour of the alternative.
  - In practice, we act as though the null is untrue, and that the alternative is supported.
- If we observe $p \geq \alpha$ (or equivalent with critical values):

Drawing Conclusions and Interpretation

- If we observe $p < \alpha$ (or equivalent with critical values):
  - We have strong enough evidence against the null hypothesis.
  - We decide to **reject the null hypothesis** in favour of the alternative.
  - In practice, we act as though the null is untrue, and that the alternative is supported.
- If we observe $p \geq \alpha$ (or equivalent with critical values):
  - We do not have strong enough evidence against the null hypothesis.

Drawing Conclusions and Interpretation

- If we observe $p < \alpha$ (or equivalent with critical values):
  - We have strong enough evidence against the null hypothesis.
  - We decide to **reject the null hypothesis** in favour of the alternative.
  - In practice, we act as though the null is untrue, and that the alternative is supported.
- If we observe $p \geq \alpha$ (or equivalent with critical values):
  - We do not have strong enough evidence against the null hypothesis.
  - We **fail to reject** the null hypothesis.

Drawing Conclusions and Interpretation

- If we observe $p < \alpha$ (or equivalent with critical values):
  - We have strong enough evidence against the null hypothesis.
  - We decide to **reject the null hypothesis** in favour of the alternative.
  - In practice, we act as though the null is untrue, and that the alternative is supported.
- If we observe $p \geq \alpha$ (or equivalent with critical values):
  - We do not have strong enough evidence against the null hypothesis.
  - We **fail to reject** the null hypothesis.
  - We **never** accept the null hypothesis (innocent until proven guilty).

Drawing Conclusions and Interpretation

- If we observe $p < \alpha$ (or equivalent with critical values):
  - We have strong enough evidence against the null hypothesis.
  - We decide to **reject the null hypothesis** in favour of the alternative.
  - In practice, we act as though the null is untrue, and that the alternative is supported.
- If we observe $p \geq \alpha$ (or equivalent with critical values):
  - We do not have strong enough evidence against the null hypothesis.
  - We **fail to reject** the null hypothesis.
  - We **never** accept the null hypothesis (innocent until proven guilty).
  - We continue to act as though the null is true.

Speaking of Conclusions

- If we **reject the null hypothesis** we say that the result is **statistically significant**.

Speaking of Conclusions

- If we **reject the null hypothesis** we say that the result is **statistically significant**.
  - Should clarify "statistically significant at an α level of significance."

Speaking of Conclusions

- If we **reject the null hypothesis** we say that the result is **statistically significant**.
  - Should clarify "statistically significant at an α level of significance."
  - A reported p-value indicates how strong the evidence *against* the null hypothesis is.

Speaking of Conclusions

- If we **reject the null hypothesis** we say that the result is **statistically significant**.
  - Should clarify "statistically significant at an α level of significance."
  - A reported p-value indicates how strong the evidence *against* the null hypothesis is.
- If we **fail to reject the null hypothesis** we say that the result is **not statistically significant**.

Speaking of Conclusions

- If we **reject the null hypothesis** we say that the result is **statistically significant**.
  - Should clarify "statistically significant at an α level of significance."
  - A reported p-value indicates how strong the evidence *against* the null hypothesis is.
- If we **fail to reject the null hypothesis** we say that the result is **not statistically significant**.
  - Based on our sample, there is a strong enough chance that the result is due to random chance.

What if we are Wrong?

- We can make a mistake in our conclusion in two possible ways:

What if we are Wrong?

- We can make a mistake in our conclusion in two possible ways:
  - A **false positive**, where we **reject** H_0 even if it were true.

What if we are Wrong?

- We can make a mistake in our conclusion in two possible ways:
  - A **false positive**, where we **reject** H_0 even if it were true.
  - A **false negative**, where we **fail to reject** H_0 when it is false.

What if we are Wrong?

- We can make a mistake in our conclusion in two possible ways:
  - A **false positive**, where we **reject** H_0 even if it were true.
  - A **false negative**, where we **fail to reject** H_0 when it is false.
- We call **false positives** Type I errors.

What if we are Wrong?

- We can make a mistake in our conclusion in two possible ways:
  - A **false positive**, where we **reject** H_0 even if it were true.
  - A **false negative**, where we **fail to reject** H_0 when it is false.
- We call **false positives** Type I errors.
  - The probability of a Type I error is α, the significance level. We call **false negatives** Type II errors.

What if we are Wrong?

- We can make a mistake in our conclusion in two possible ways:
  - A **false positive**, where we **reject** H_0 even if it were true.
  - A **false negative**, where we **fail to reject** H_0 when it is false.
- We call **false positives** Type I errors.
  - The probability of a Type I error is α, the significance level. We call **false negatives** Type II errors.
  - The probability of a Type II error is denoted β. We call $1 - \beta$ the **power** of a test.

What if we are Wrong?

- We can make a mistake in our conclusion in two possible ways:
  - A **false positive**, where we **reject** H_0 even if it were true.
  - A **false negative**, where we **fail to reject** H_0 when it is false.
- We call **false positives** Type I errors.
  - The probability of a Type I error is α, the significance level. We call **false negatives** Type II errors.
  - The probability of a Type II error is denoted β. We call $1-\beta$ the **power** of a test.
- We want α and β to be as small as possible, but they trade-off against one another.

What if we are Wrong?

- We can make a mistake in our conclusion in two possible ways:
  - A **false positive**, where we **reject** H_0 even if it were true.
  - A **false negative**, where we **fail to reject** H_0 when it is false.
- We call **false positives** Type I errors.
  - The probability of a Type I error is α, the significance level. We call **false negatives** Type II errors.
  - The probability of a Type II error is denoted β. We call $1 - \beta$ the **power** of a test.
- We want α and β to be as small as possible, but they trade-off against one another.
  - Typically we set α and then try to minimize β at that level.

Summary

- Hypothesis testing is a framework by which we assess evidence for (or against) statistical hypotheses.
- The null hypothesis, stated in contrast to an alternative, is the default state of the world.
- Test statistics are computed with known sampling distributions to assess hypothesis.
- P-values indicate the chance of observing, by random chance, evidence as strong as what was observed.
- The null hypothesis is rejected or fail to be rejected, never accepted.
- Type I errors and type II errors need to be balanced against one another.