Chapter 6: Introduction to hypothesis testing

Ott & Longnecker Sections: 5.1, 5.4, 5.6

Duzhe Wang

the Department of Statistics, UW-Madison

https://dzwang91.github.io/stat371/

What do we study

Key concepts: Hypothesis, null hypothesis, alternative hypothesis, test statistic, rejection region, Type I error, Type II error, power, p-value, significance level

Hypothesis testing

Philosophy: **dis**prove(reject) a claim by **contradiction**

Hypothesis testing

Philosophy: **dis**prove(reject) a claim by **contradiction**

Analogy: “dependent love” story

Hypothesis testing

Philosophy: **dis**prove(reject) a claim by **contradiction**

Analogy: “dependent love” story

Question: Do you love me? (think about when you will have such a question?)

Hypothesis testing

Philosophy: **dis**prove(reject) a claim by **contradiction**

Analogy: "dependent love" story

Question: Do you love me? (think about when you will have such a question?)
Claim: You love me.

Hypothesis testing

Philosophy: **dis**prove(reject) a claim by **contradiction**

Analogy: “dependent love” story

Question: Do you love me? (think about when you will have such a question?)
Claim: You love me.
Reasoning: If you loved me, you would take the trash out every week and put your socks away.

Hypothesis testing

Philosophy: **dis**prove(reject) a claim by **contradiction**

Analogy: “dependent love” story

Question: Do you love me? (think about when you will have such a question?)
Claim: You love me.
Reasoning: If you loved me, you would take the trash out every week and put your socks away.
Data: Some weeks you don’t take the trash out or leave your socks where they fall.

Hypothesis testing

Philosophy: **dis**prove(reject) a claim by **contradiction**

Analogy: “dependent love” story

Question: Do you love me? (think about when you will have such a question?)
Claim: You love me.
Reasoning: If you loved me, you would take the trash out every week and put your socks away.
Data: Some weeks you don’t take the trash out or leave your socks where they fall.
Conclusion: I don’t believe you love me(disprove the claim).

Ingredients of a hypothesis test

- In hypothesis testing, there are two competing hypotheses: **null hypothesis H_0 and alternative hypothesis H_A**.

$$H_0 = \text{" you love me"},\ H_A = \text{" you don't love me"}.$$

Ingredients of a hypothesis test

- In hypothesis testing, there are two competing hypotheses: **null hypothesis H_0 and alternative hypothesis H_A**.

  $$H_0 = \text{" you love me"},\ H_A = \text{" you don't love me"}.$$

  In statistical hypothesis testing, the hypotheses are complementary: that is, they do not overlap, and between them cover all possibilities.

Ingredients of a hypothesis test

- Data $X_1, ..., X_n$ are gathered, choose a **test statistic** $T_n = T_n(X_1, ..., X_n)$. The test statistic is an RV. Based on data, we can calculate the realization of the test statistic.

Ingredients of a hypothesis test

- Data $X_1, ..., X_n$ are gathered, choose a **test statistic** $T_n = T_n(X_1, ..., X_n)$. The test statistic is an RV. Based on data, we can calculate the realization of the test statistic.
- We specify a set of values of the test statistic such that, if it realizes to one of these values, we reject H_0. This region is called the **rejection region**. The rejection region consists of values that comprise evidence **against** H_0.

Ingredients of a hypothesis test

- Data $X_1, ..., X_n$ are gathered, choose a **test statistic** $T_n = T_n(X_1, ..., X_n)$. The test statistic is an RV. Based on data, we can calculate the realization of the test statistic.
- We specify a set of values of the test statistic such that, if it realizes to one of these values, we reject H_0. This region is called the **rejection region**. The rejection region consists of values that comprise evidence **against** H_0.
  If the test statistic falls outside of the rejection region, there is insufficient evidence against the null, and we say we **fail to reject the null or retain the null**. Note that we generally don't formulate our result in terms of the alternative hypothesis.

Type I and Type II error

Type I Error vs Type II Error

		Decision (based on sample)	
		Reject H_0	Not Reject H_0
Truth (for population studied)	H_0 True	Type I Error α	😄
	H_0 False	😄	Type II Error β

- α=P(reject H_0 | H_0 is true). Smaller α is better.
- β=P(not reject H_0 | H_0 is false). Smaller β is better.
- Power$=1-\beta$=P(reject H_0 | H_0 is false). Larger power is better.

Type I and Type II error

relationship between α and β: We desire a small α and a small β. Unfortunately, in general, for a given fixed sample size, if we adjust our rejection region to decrease α, β will go up, and vice versa. The only way to decrease both α and β simultaneously is to increase the sample size.

Hypothesis testing

Our goal: we need to choose the **test statistic** and the **rejection region** so that the test has good statistical properties.

Hypothesis testing

Our goal: we need to choose the **test statistic** and the **rejection region** so that the test has good statistical properties.

- **What are good statistical properties?**

Hypothesis testing

Our goal: we need to choose the **test statistic** and the **rejection region** so that the test has good statistical properties.

- **What are good statistical properties?**
  small errors.

Hypothesis testing

Our goal: we need to choose the **test statistic** and the **rejection region** so that the test has good statistical properties.

- **What are good statistical properties?**
  small errors.
- **What is a good test statistic?**

Hypothesis testing

Our goal: we need to choose the **test statistic** and the **rejection region** so that the test has good statistical properties.

- **What are good statistical properties?**
  small errors.
- **What is a good test statistic?**
  It's good if we know its distribution.

Hypothesis testing

Our goal: we need to choose the **test statistic** and the **rejection region** so that the test has good statistical properties.

- **What are good statistical properties?**
  small errors.
- **What is a good test statistic?**
  It's good if we know its distribution.
- **How to construct the rejection region?**

Hypothesis testing

Our goal: we need to choose the **test statistic** and the **rejection region** so that the test has good statistical properties.

- **What are good statistical properties?**
  small errors.
- **What is a good test statistic?**
  It's good if we know its distribution.
- **How to construct the rejection region?**
  Use the distribution to determine a rejection region that **limits the type I error at significance level** α.

P-value

- The p-value is defined to be the probability of a test statistic realizing to a value that is as or more extreme than the one actually observed **when the null hypothesis is true**.

P-value

- The p-value is defined to be the probability of a test statistic realizing to a value that is as or more extreme than the one actually observed **when the null hypothesis is true**.
- Smaller p-values indicate relatively more evidence against the null hypothesis.

P-value

- The p-value is defined to be the probability of a test statistic realizing to a value that is as or more extreme than the one actually observed **when the null hypothesis is true**.
- Smaller p-values indicate relatively more evidence against the null hypothesis.
- **If the p-value is smaller than the given significance level α, we would reject the null, otherwise we would not reject the null.**

P-value

- The p-value is defined to be the probability of a test statistic realizing to a value that is as or more extreme than the one actually observed **when the null hypothesis is true**.
- Smaller p-values indicate relatively more evidence against the null hypothesis.
- **If the p-value is smaller than the given significance level α, we would reject the null, otherwise we would not reject the null.**
- In most situations, reporting the p-values so that it may be used as the degree of evidence against the null is better than only stating the reject or not-reject decision.

What's the next?

We'll give examples of some specific tests based on samples from one population in the next lecture.