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



Philosophy: disprove(reject) a claim by contradiction



Philosophy: disprove(reject) a claim by contradiction

Analogy: "dependent love" story



Philosophy: disprove(reject) a claim by contradiction

Analogy: "dependent love" story

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



Philosophy: disprove(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.



Philosophy: disprove(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.



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.



Philosophy: disprove(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).



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

 H_0 =" you love me", H_A =" you don't love me".



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

 H_0 =" you love me", H_A =" 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.



• 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.



- 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 H₀. This region is called the
 rejection region. The rejection region consists of values that
 comprise evidence against H₀.

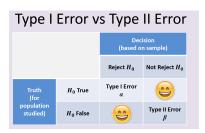


- Data $X_1,, X_n$ are gathered, choose a **test statistic** $T_n = T_n(X_1, ..., X_n)$. The test statistic is an RV. Based on data, we can calculate the realization of the test statistic.
- We specify a set of values of the test statistic such that, if it realizes to one of these values, we reject H₀. This region is called the rejection region. The rejection region consists of values that comprise evidence against H₀.

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





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





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?



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



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



- 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?



- 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 .



 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.



- 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.



- 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.



- 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.