Testing, Hypotheses and Errors

Evan Misshula

2017-03-16

Outline

Learning Objectives

Objectives for session 3

- Examine the process of hypothesis testing
- Evaluate research and null hypothesis
- Determine if one- or two-tailed tests are appropriate
- Understand obtained values, significance, and critical regions
- Distinguish between Type I and Type II error

Introduction

Testing

- Researchers compare groups of people
- Hypothesis testing: develop hypothesis and then research to find out if it is true
- One-tailed and two-tailed tests, types of error, and statistical power

Inferential process

Conducting a study

- In order to understand the role and logic of hypothesis testing, we have to know the basic procedures for conducting a study
- Almost all studies begin with a question about a population
- A typical research question in evaluative studies is to see if a treatment has an affect on individuals in the population
- Since it is impossible to study an entire population, we always (randomly) select a sample with certain numbers of people or particular size (e.g., n=140) for study

Hypothesis Testing

Definition (Two types of hypotheses)

- Null hypothesis H_0
- Research hypothesis H_1

Hypothesis Testing

Definition (Two types of hypotheses)

- Null hypothesis H_0
- Research hypothesis H_1
- Typically state goals of research
  - Derived from theory or primary question/research questions

Hypothesis Testing

Definition (Two types of hypotheses)

- Null hypothesis H_0
- Research hypothesis H_1

- Typically state goals of research
  - Derived from theory or primary question/research questions

- Place conceptual language in format for empirical examination
  - The alternative, or research hypothesis, symbolized by H_1, is a statistical hypothesis that states a specific difference between a parameter and a specific value or states that there is a difference between two parameters.

Null and Research Hypotheses

Null hypotheses (H0)

Null and Research Hypotheses

Null hypotheses (H0)

- Research can be statistically validated or refuted
  - The null hypothesis, symbolized by H_0, is a statistical hypothesis that states that there is no difference between a parameter and a specific value or that there is no difference between two parameters.

Null and Research Hypotheses

Null hypotheses (H0)

- Research can be statistically validated or refuted
  - The null hypothesis, symbolized by H_0, is a statistical hypothesis that states that there is no difference between a parameter and a specific value or that there is no difference between two parameters.
- Facilitate hypothesis testing
  - Important to inferential analyses since true parameters of population are unknown

Key concept

- The key concept is *standard error of the mean* or SEM
- variation of the average of $N(\mu, \sigma^2)$:

$$SEM = \frac{\sigma}{n}$$

Key concept

- The key concept is *standard error of the mean* or SEM
- variation of the average of $N(\mu, \sigma^2)$:

$$SEM = \frac{\sigma}{n}$$

- This means that the average has a much narrower range than the sample

One Sample T-test

- For any member it has a 95% chance of falling between 2 std deviations

$$[\mu - 2 * \sigma, \mu + 2 * sigma]$$

$$t = \frac{\overline{x} - \mu_0}{SEM}$$

Null Hypothesis

- The null hypothesis is a statement about a value of a population parameter
- Must contain the condition of equality ($=$)
- We test the null hypothesis directly
- The conclusions of the test are either "reject H0" or "fail to reject H0"
- Cannot absolutely accept research hypothesis but can reject null hypotheses

Research/Alternative Hypothesis

Must be true if H0 is false

Contains the conditions of inequality

- (i.e.: == , < , or >)
- It is the opposite of the null hypothesis H_0
- The result we obtained is not due to mere chance.

Hypotheses:

Conceptually clear and specific

- Clearly spell out the research goals
- Conceptual framework defines operationalization

Hypotheses:

Conceptually clear and specific

- Clearly spell out the research goals
- Conceptual framework defines operationalization

- Empirically based with no moral judgments
- Free from emotional bias

Hypotheses:

Conceptually clear and specific

- Clearly spell out the research goals
- Conceptual framework defines operationalization

- Empirically based with no moral judgments
- Free from emotional bias

- Related to a body of theory
- Should refute, qualify, or support existing theories

Two-Tailed versus One-Tailed Hypothesis

- A two tailed hypothesis does not state the direction in which we expect the groups to differ.
- A one tailed hypothesis specifies a directional relationship between groups.

Two-Tailed versus One-Tailed Hypothesis

- A two tailed hypothesis does not state the direction in which we expect the groups to differ.
- A one tailed hypothesis specifies a directional relationship between groups.

- For example, if we compare income with years of schooling we might hypothesize that years of schooling tend to increase income. That is a one-tailed hypothesis because it specifies that the relationship must be positive.

Two tailed hypothesis

Two tailed hypothesis

- On the other hand, if we were looking at people's heights with their income, we might have no good reason for expecting that the relationship would be positive (income increasing with height) or negative (income decreasing with height). We might just want to find out if there were any relationship at all, and that's a two-tailed hypothesis.

One-Tailed Hypothesis

With directional or one-tailed test

- the region of rejection is entirely within one tail of the distribution.
- "t.s." is our critical test statistic

Two-Tailed Hypothesis

Some hypotheses predict only

- that one value will be different from another, without additionally predicting which will be higher.

Two-Tailed Hypothesis

Some hypotheses predict only

- that one value will be different from another, without additionally predicting which will be higher.
- This hypothesis is non-directional or two-tailed because an extreme test statistic in either tail of the distribution (positive or negative) will lead to the rejection of the null hypothesis of no difference.

Which are the following, one-tailed or two-tailed?

Which are the following, one-tailed or two-tailed?

- Children receiving supportive therapy will get worse over time (at post-treatment).

Which are the following, one-tailed or two-tailed?

- Children receiving supportive therapy will get worse over time (at post-treatment).
- A government official claims that the dropout rate for local schools is 25%.

Which are the following, one-tailed or two-tailed?

- Children receiving supportive therapy will get worse over time (at post-treatment).
- A government official claims that the dropout rate for local schools is 25%.
- We believe it takes more than 30 minutes to find a parking spot in NYC.

Which are the following, one-tailed or two-tailed?

- Children receiving supportive therapy will get worse over time (at post-treatment).
- A government official claims that the dropout rate for local schools is 25%.
- We believe it takes more than 30 minutes to find a parking spot in NYC.
- The average weekly sales at a grocery store will be $10,500.00

Example 1

Example (A medical researcher)

- is interested in finding out whether a new medication will have any undesirable side effects. The researcher is particularly concerned with the pulse rate of the patients who take the medication, and whether it will be different from the mean pulse rate of 82 beats per minute.

Example 1

Example (A medical researcher)

- is interested in finding out whether a new medication will have any undesirable side effects. The researcher is particularly concerned with the pulse rate of the patients who take the medication, and whether it will be different from the mean pulse rate of 82 beats per minute.
- $H_0 : \mu == 82,\ H_1 : \mu ! = 82$

Example 1

Example (A medical researcher)

- is interested in finding out whether a new medication will have any undesirable side effects. The researcher is particularly concerned with the pulse rate of the patients who take the medication, and whether it will be different from the mean pulse rate of 82 beats per minute.
- $H_0 : \mu == 82,\ H_1 : \mu! = 82$
- This is a two-tailed test – the null hypothesis is false if the pulse rate is higher or lower than 82.

EXAMPLE 2

Example (A new prisoner re-entry program)

- being tried in New York facilities asserts that it will increase the time the released prisoners remain free of new charges, beyond the current average of 125 days. The hypotheses are:
- $H_0 : \mu == 125$, $H_1 : \mu >= 125$

EXAMPLE 2

Example (A new prisoner re-entry program)

- being tried in New York facilities asserts that it will increase the time the released prisoners remain free of new charges, beyond the current average of 125 days. The hypotheses are:
- $H_0 : \mu == 125$, $H_1 : \mu >= 125$
- This is a one-tailed test (only valid above 125).

Example 3

Example (A new drug treatment program)

- that helps people stop using IV drugs promises to lower the new AIDS infection rate in a small city from its current rate of 20 cases per 10,000 people each year. The hypotheses about annual AIDS rate will be:
- $H_0 : \mu == 20, \; H_1 : \mu < 20$

Example 3

Example (A new drug treatment program)

- that helps people stop using IV drugs promises to lower the new AIDS infection rate in a small city from its current rate of 20 cases per 10,000 people each year. The hypotheses about annual AIDS rate will be:
- $H_0 : \mu == 20$, $H_1 : \mu < 20$
- This is a one-tailed test (only valid below 20).

Steps in Hypothesis Testing

Step 1: Developing research hypothesis

Steps in Hypothesis Testing

Step 1: Developing research hypothesis

- Research hypotheses are very similar to research questions or propositions

Steps in Hypothesis Testing

Step 1: Developing research hypothesis

- Research hypotheses are very similar to research questions or propositions
- Assumption that the difference is so large that it cannot be attributed to chance

Steps in Hypothesis Testing

Step 1: Developing research hypothesis

- Research hypotheses are very similar to research questions or propositions
- Assumption that the difference is so large that it cannot be attributed to chance
- Must test empirically

Steps in Hypothesis Testing (2)

Step 2: Developing null hypothesis

- Reverse of the research hypotheses
- States difference is so small that it could have occurred by either chance or sampling error

Steps in Hypothesis Testing (2)

Step 2: Developing null hypothesis

- Reverse of the research hypotheses
- States difference is so small that it could have occurred by either chance or sampling error
- Goal is to reject null hypothesis

Steps in Hypothesis Testing (3)

Step 3: Drawing samples

- Samples taken from populations studied
- Samples must represent target population

Steps in Hypothesis Testing (3)

Step 3: Drawing samples

- Samples taken from populations studied
- Samples must represent target population
- Preferably the theoretical population

Steps in Hypothesis Testing (4)

Step 4: Selecting the test

- Two-tailed test
- Without specifying a direction of the difference

Steps in Hypothesis Testing (4)

Step 4: Selecting the test

- Two-tailed test
- Without specifying a direction of the difference

- One-tailed test
- Directional difference

Steps in Hypothesis Testing (4)

Step 4: Selecting the test

- Two-tailed test
- Without specifying a direction of the difference

- One-tailed test
- Directional difference

- Must have a reason to choose this over two-tailed test

Steps in Hypothesis Testing (4)

Step 4: Selecting the test

- Two-tailed test
- Without specifying a direction of the difference

- One-tailed test
- Directional difference

- Must have a reason to choose this over two-tailed test

- A statistical test uses the data obtained from a sample to make a decision about whether or not the null hypothesis should be rejected.

Steps in Hypothesis Testing (5)

Step 5: Calculating the obtained value

- Use analysis procedure
- Chi-square, Z-tests, t-tests, ANOVA

Steps in Hypothesis Testing (6)

Step 6: Determining significance and critical regions

- Critical value
- Level of certainty (probability of making Type I error), degrees of freedom, one- or two-tailed test
- Level of significance

Steps in Hypothesis Testing (7)

Step 7: Making a decision

- Value obtained from the statistical test is compared to the critical value

Steps in Hypothesis Testing (7)

Step 7: Making a decision

- Value obtained from the statistical test is compared to the critical value
- If obtained value is greater than critical value, reject null hypothesis

Steps in Hypothesis Testing (7)

Step 7: Making a decision

- Value obtained from the statistical test is compared to the critical value
- If obtained value is greater than critical value, reject null hypothesis
- If obtained value is less than critical value, null hypothesis cannot be rejected at that level of significance

Steps In Hypotheses Testing (8)

- The critical or rejection region is the range of values of the test value that indicates that there is a significant difference and that the null hypothesis should be rejected.
- The noncritical or non-rejection region is the range of values of the test value that indicates that the difference was probably due to chance and that the null hypothesis should not be rejected.

Steps In Hypotheses Testing (8)

- The critical or rejection region is the range of values of the test value that indicates that there is a significant difference and that the null hypothesis should be rejected.
- The noncritical or non-rejection region is the range of values of the test value that indicates that the difference was probably due to chance and that the null hypothesis should not be rejected.
- Remember: the area under the curve represents the probability (or likelihood) that the sample test value is a chance finding.

Test Value

Definition (definition)

- The numerical value obtained from a statistical test is called the test value.

Outcomes and errors

In the hypothesis-testing situation, there are two possible actions.

- Reject null hypothesis when there is difference between variables
- Do not reject null hypothesis when no statistically significant difference between variables

Outcomes and errors

In the hypothesis-testing situation, there are two possible actions.

- Reject null hypothesis when there is difference between variables
- Do not reject null hypothesis when no statistically significant difference between variables
- If we can do two things, we can make two errors

Errors

Definition (Type I and Type II Errors)

- *Type I error*
  - Rejection of a null hypothesis that is true
  - Probability associated with Type I error is the significance value of the test is α
- *Type II error*
  - Failing to reject a false null hypothesis
  - Probability of committing a Type II error is β

Which is better?

Type I or Type II Error?

Which is better?

Type I or Type II Error?

- Depends on researcher's goal

Which is better?

Type I or Type II Error?

- Depends on researcher's goal
- Type I errors are brought out in future research
  - standard value of α is 5% or 1%

Power of tests

Definition (Power of a statistical tests)

- Probability that it will reject H_0 if false

Power of tests

Definition (Power of a statistical tests)

- Probability that it will reject H_0 if false
- linked with probability of Type II error

Power of tests

Definition (Power of a statistical tests)

- Probability that it will reject H_0 if false
- linked with probability of Type II error
- Sample size defines the power of the test

Power of tests

Definition (Power of a statistical tests)

- Probability that it will reject H_0 if false
- linked with probability of Type II error
- Sample size defines the power of the test
- Balance risk of committing Type II error and possibility of having a sample that is too large

Alpha level, critical probability, or significance level influence power

Determine falseness of null hypothesis

- Significance Levels in Hypotheses Testing
- Typical significance levels are:

Alpha level, critical probability, or significance level influence power

Determine falseness of null hypothesis

- Significance Levels in Hypotheses Testing
- Typical significance levels are:
- $0.10 = 10\%$ chance of rejecting a true null hypothesis, or 90% confidence that our sample statistic is true in the population, meaning it's not due to mere chance.

Alpha level, critical probability, or significance level influence power

Determine falseness of null hypothesis

- Significance Levels in Hypotheses Testing
- Typical significance levels are:
- 0.10 = 10% chance of rejecting a true null hypothesis, or 90% confidence that our sample statistic is true in the population, meaning it's not due to mere chance.
- 0.05 = 5% chance of rejecting a true null hypothesis, and 95% confidence that our sample statistic is true.

Alpha level, critical probability, or significance level influence power

Determine falseness of null hypothesis

- Significance Levels in Hypotheses Testing
- Typical significance levels are:
- 0.10 = 10% chance of rejecting a true null hypothesis, or 90% confidence that our sample statistic is true in the population, meaning it's not due to mere chance.
- 0.05 = 5% chance of rejecting a true null hypothesis, and 95% confidence that our sample statistic is true.
- 0.01 = 1% chance of rejecting a true null hypothesis, or 99% confidence that our sample statistic is true.