

Stat 201: Statistics I

Chapter 8



Chapter 8

Hypothesis Testing

Section 8.1

Basics of Hypothesis Testing

Previously, statistics from random samples were used to learn something about populations by estimating population parameters. “Truths” about populations were inferred from the data of the sample.

A similar question can be posed: Is a sample drawn from a population that is the same in an important way to a known population, or is the sample drawn from a population that is significantly different?

Hypothesis tests

An **hypothesis test** is a formal statistical procedure to test claims about population parameters based on samples drawn from populations. Such claims, or **hypotheses**, are often written as simple mathematical statements.

It is important to be clear as to which population the claims or the tests are about,

Hypotheses, example

Example

- Male Metro State students are shorter on average than the national mean of 69.2 inches.
 - Population: Male Metro State students
 - $\mu < 69.2$
- The proportion of teen drivers who text or email while driving is 40%.
 - Population: Teen drivers
 - $p = 0.4$
- A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A new experimental treatment will extend the survival time.
 - Population: Patients with disease on experimental treatment
 - $\mu > 36$

Hypotheses

The first step to conduct an hypothesis test is to identify two hypotheses.

The **null hypothesis** is the claim that nothing interesting has occurred, that a sub-population is **not** different than the general population or that population parameters did **not** change after treatment.

Conversely, the **alternative hypothesis** is the claim that something interesting has occurred, that a sub-population is different or that parameters did change after treatment.

Hypotheses, cont.

Null hypothesis:

- Denoted by H_0
- Always a statement that a parameter is **equal to** some value
- That value, denoted p_0 or μ_0 , is called the proportion or mean under the null hypothesis

Alternative hypothesis:

- Denoted by H_1 or H_a
- Can be a statement that a parameter is **less than**, **greater than** or **not equal to** some value
- Is usually a statement representing the research question

One-sided vs. two sided tests

If an alternative hypothesis has the form of a parameter being less than or greater than some value, the hypothesis test is called a **one-sided test**.

If an alternative hypothesis has the form of a parameter being not equal to some value, the hypothesis test is called a **two-sided test**.

Hypotheses, example

Example

Identify the null and alternative hypotheses, and whether it is a one-sided or two-sided test.

In the United States, adult men have a mean height of 69.2 inches. The Metro State administration want to do a study to see if male Metro State students are shorter than the general US population.

- $H_0 : \mu = 69.2$ $H_a : \mu < 69.2$ One-sided

A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A clinical trial is conducted to see if a new experimental treatment will change the survival time.

- $H_0 : \mu = 36$ $H_a : \mu \neq 36$ Two-sided

Structure of hypothesis tests

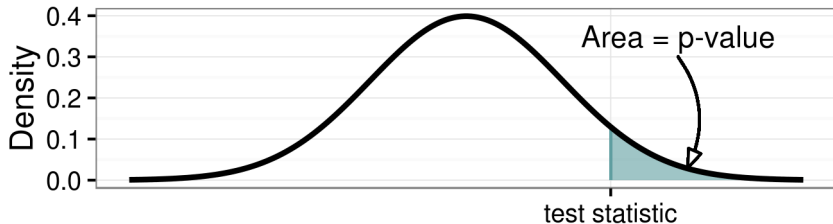
Starting with null and alternative hypotheses derived from the research question and a random sample, all hypothesis tests have the same basic structure.

- A test statistic is calculated which indicates the location of the sample within the sampling distribution, assuming the null hypothesis is true.
- The probability of getting a test statistic equal to or more extreme than the statistics belonging to the sample is calculated.
- If the calculated probability is below a pre-specified threshold, the null hypothesis is **rejected** and it is said that there is evidence to support the alternative hypothesis.
- If the calculated probability is not below the pre-specified threshold, the null hypothesis is **not rejected** and it is said that there is not evidence to support the alternative hypothesis.

P-values

In a hypothesis test, the **p-value** is the probability of getting a sample with the test statistic or one more extreme, assuming the null hypothesis is true.

- Not to be confused with the population proportion p or the probability function $P(A)$, though a p-value does represent a probability.



Calculating p-values

Calculating a p-value is that same as calculating probabilities in sampling distributions already learned.

- Identify sampling distribution:
 - z distribution for proportions
 - t distribution for means
- Calculate **test statistic**: z -score or t -score
- Find probability of test statistic or more extreme values in sampling distribution
- That probability is the p-value

Luckily, all these steps can be accomplished easily with technology.

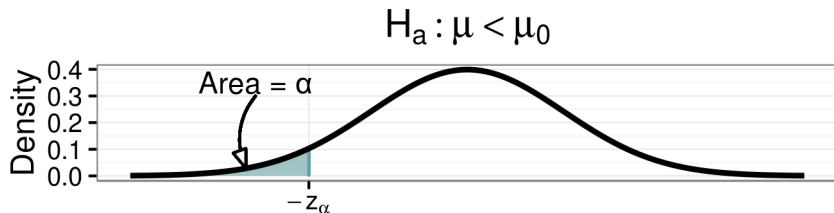
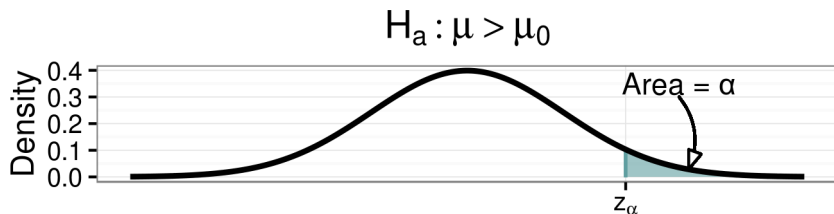
Significance level

Once the p-value is calculated, it is compared against a pre-specified threshold. This threshold is called the **significance level** of the test.

- Denoted by α
- This is the same α used for critical values and confidence intervals
- Thus, significance level can be thought of as an area in a sampling distribution
- Sometimes referred to as the rejection region

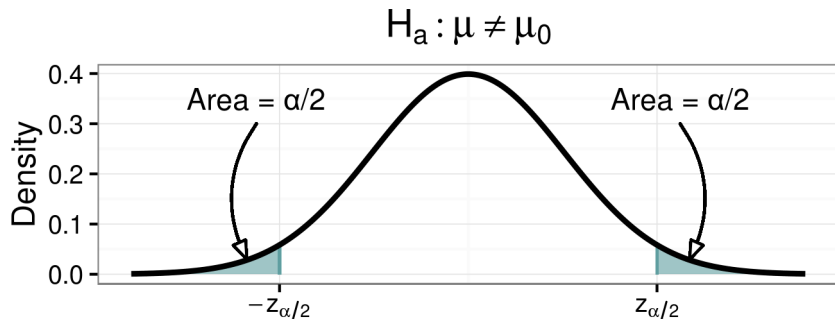
Significance level for one-sided test

In a one-sided test, the entire rejection region is located at one end of the distribution or the other.



Significance level for two-sided test

In a two-sided test, the rejection region is split between the lower and upper extremes of the distribution.



Stating conclusions

When reporting the results of an hypothesis tests, the following elements should be included:

- Report the test statistic and p-value
- Report a decision on the null hypothesis based on the p-value and significance level (α).
 - State the decision as “Reject H_0 ” or “Do not reject H_0 ”.
 - The null hypothesis is never “accepted”.
- State the conclusion in terms of the research question
 - “There is evidence for...”
 - “There is not evidence for...”

Stating conclusions, example

Example

A study is conducted to test the claim that male Metro State students are shorter than the general population height of 69.2 inches. The test at a $\alpha = 0.05$ level of significance produces a test statistic $t = -1.859$ and a p-value of 0.0358. State the conclusion of the test.

- $H_0 : \mu = 69.2, \quad H_a : \mu < 69.2$
- $p = 0.0358 < \alpha = 0.05$. Reject the null hypothesis. There is evidence to conclude that male Metro students are shorter than the general population.

Stating conclusions, example

Example

A patient diagnosed with a particular rare disease has an expected survival time of 36 months. A clinical trial is conducted to see if a new experimental treatment will change the survival time. The hypothesis test at $\alpha = 0.01$ level of significance produces a p-value of 0.098. State the conclusion of the test.

- $H_0 : \mu = 36$ $H_a : \mu \neq 36$
- $p = 0.098 > \alpha = 0.01$. Do not reject the null hypothesis. There is not evidence to conclude that the experimental treatment changes survival time.

Steps for hypothesis test

- 1 Identify null and alternative hypotheses from research question
- 2 Determine appropriate sampling distribution
- 3 Calculate test statistic
- 4 Calculate p-value
- 5 Compare p-value to significance level α and report decision
- 6 State conclusion in terms of original research question

Note: Steps 3 and 4 are often accomplished with technology

Making a decision with p-value

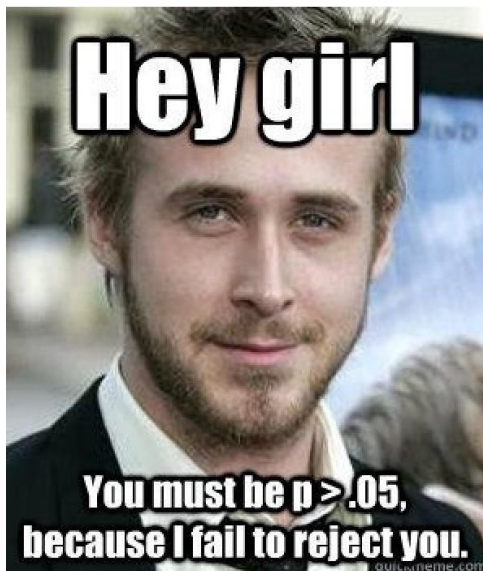
A **small** p-value ($p < \alpha$) means a low probability of getting the observed sample if the null hypothesis is true. Thus, the null hypothesis is rejected.

A **large** p-value ($p > \alpha$) then means that the sample is not unlikely under the null hypothesis. Thus, the null hypothesis is not rejected.

To remember...

If p is low, the null must go.

Or if this helps...



Critical value method

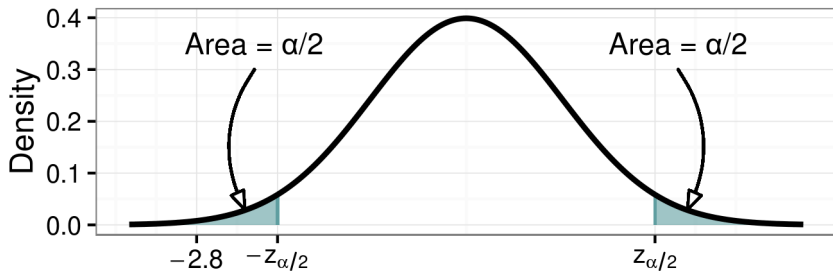
The hypothesis test procedure discussed thus far is known as the p-value method. An alternative method, known as the **critical value** method, does not use p-values, but rather compares test statistics directly to appropriate critical values. If the test statistic is more extreme than the critical value, the null hypothesis is rejected.

Critical value method, example

Example

A two-sided test with $\alpha = 0.05$ level of significance is conducted and results in a test statistic $z = -2.8$.

- Recall, critical value $z_{\alpha/2} = \pm 1.96$.
- Since -2.8 is more extreme than -1.96 ($-2.8 < -1.96$), reject the null hypothesis.



Types of errors in hypothesis tests

In an hypothesis test, either the null hypothesis is true or it is not, and either the null hypothesis is rejected or it is not. Thus, there are four possible outcomes.

	Reject H_0	Do not reject H_0
H_0 is true	Type I error (α)	Correct decision
H_0 is not true	Correct decision	Type II error (β)

- A type I error occurs when the null hypothesis is rejected when it is in fact true. The probability of making a type I error in a test is designated α .
- A type II error occurs when the null hypothesis is not rejected when it is in fact not true. The probability of making a type I error in a test is designated β .

Types of errors in hypothesis tests, cont.

- The acceptable probability of committing a type I error (α), or level of significance, is chosen prior to conducting a hypothesis test.
- The probability of committing a type II error (β) is determined by a number of factors, including the chosen α and the sample size.
- The probability of **not** committing a type II error ($1 - \beta$) is known as the **power** of a test.
- There is a trade-off between α and β . Smaller α result in larger β (and lower power) and vice versa.

Analogy to legal system

It can be useful to use a legal system analogy to help understand how hypothesis tests work.

Consider a criminal trial:

- A criminal defendant enters a trial with the **presumption of innocence**. A defendant is assumed innocent until proven guilty.
- A prosecutor presents **evidence** to show that the defendant should be considered guilty.
- The jury considers whether the evidence, when judged against a particular standard (**beyond a reasonable doubt**), is sufficient to abandon the presumption of innocence.
- The jury returns a **verdict**: either guilty or not guilty. A defendant is generally not declared innocent.
- A jury can make two kinds of mistakes: they can convict an innocent defendant or they can fail to convict a guilty defendant.

Analogy to legal system, cont.

- The presumption of innocence is comparable to the **null hypothesis**. An hypothesis tests is entered assuming that there is nothing interesting about the population(s) being studied.
- The evidence is comparable to the **sample** collected to answer the research the question.
- The “beyond a reasonable doubt” standard is comparable to the **significance level** or α of the test.
- The verdict is comparable to the **conclusion** of the test: either to reject the null or to fail to reject the null. The null is never accepted or proven.
- A conclusion can be in error one of two ways: a type I error (convict an innocent defendant) or a type II error (fail to convict a guilty defendant).

Section 8.2

Testing a Claim About a Proportion

Testing claims about population proportions

Claims about a population proportion p are tested using a sample proportion \hat{p} .

Recall, sampling distributions of proportion data follow a binomial distribution which, under certain conditions, approximates a normal distribution.

Test statistic for proportion tests

Tests for population proportions use a standard normal sampling distribution. The test statistic is a z -score.

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

where \hat{p} is the sample proportion, p_0 is the population proportion under the null hypothesis, $q_0 = 1 - p_0$ and n is the sample size.

P-values for proportion tests

Given a z test statistic, the p-value is the probability of getting z -scores more extreme on a standard normal distribution.

- For two-sided test:

$$H_a : p \neq p_0, \text{ p-value} = P(Z < -z) + P(Z > z) = 2 \times P(Z > z)$$

- For one-sided test:

$$H_a : p < p_0, \text{ p-value} = P(Z < z)$$

$$H_a : p > p_0, \text{ p-value} = P(Z > z)$$

Requirements for a proportion test

- Like all hypothesis tests, sample must be a simple random sample.
- Samples of proportion data follow a binomial distribution and must satisfy the binomial requirements:
 - Fixed number of trials
 - Trials are independent
 - Two possible outcomes (success/failure)
 - Probability of each outcome is constant
- For the binomial to approximate a normal distribution, np and nq must both be at least 5.

Steps for a proportion hypothesis test

- 1 Identify null and alternative hypotheses from research question
 $H_0 : p = p_0$
 $H_a : p \neq p_0, p < p_0, p > p_0$
- 2 Check the requirements for using normal distribution
- 3 Calculate z test statistic
- 4 Calculate p-value
- 5 Compare p-value to significance level α and report decision
- 6 State conclusion in terms of original research question

Hypothesis tests for a proportion in StatCrunch

- Stat → Proportion Stats → One Sample → With Summary
- Enter “# of successes” and “# of observations”
- Select “Hypothesis test for p”
- Enter the appropriate values for null and alternative hypotheses.
- Click “Compute!”
- The test statistic and p-value are found in “Z-Stat” and “P-value”

Hypothesis test for a proportion, example

Example

The proportion of teen drivers who text or email while driving is 40%. A school district adds an education program for all high school students in hopes of lowering that proportion. Two months after the program, the district surveys a random sample of teen drivers. The survey finds that 62 out of 212 teens surveyed had texted or emailed while driving during the previous month. Test whether the program was successful at a 0.05 level of significance.

- 1 Identify null and alternative hypotheses from research question

$$H_0 : p = 0.4$$

$$H_a : p < 0.4$$

Population: teen drivers who had attended the education program

Hypothesis test for a proportion, example

Example

- 2 Check the requirements for using normal distribution
 - Simple random sample
 - Requirements of a binomial satisfied (fixed sample size, two outcomes, independent, constant probability of success)
 - $p = 0.4$, $q = 0.6$, $n = 212$
 $np > 5$, $nq > 5$
- 3 Calculate z test statistic
 $z = -3.1964013$
- 4 Calculate p-value
 $p = 0.0007$

Hypothesis test for a proportion, example

Example

- ⑤ Compare p-value to significance level α and report decision
 $p = 0.0007 < \alpha = 0.05$. Reject null hypothesis.
- ⑥ State conclusion in terms of original research question
There is evidence that teen drivers who attended the program text and email while driving at a lower rate than 40%.

Section 8.3

Testing a Claim About a Mean

Testing claims about population means

Claims about a population mean μ are tested using a sample mean \bar{x} . Recall, if population standard deviations are unknown, sampling distributions of sample means follow a t distribution with the appropriate degrees of freedom.

Test statistic for mean tests

Tests for population means use the t sampling distribution. The test statistic is a t -score.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where \bar{x} is the sample mean, μ_0 is the population mean under the null hypothesis, s is the sample standard deviation and n is the sample size.

P-values for mean tests

Given a t test statistic, the p-value is the probability of getting t -scores more extreme on a t distribution.

- For two-sided test:

$$H_a : \mu \neq \mu_0, \text{ p-value} = P(T < -t) + P(T > t) = 2 \times P(T > t)$$

- For one-sided test:

$$H_a : \mu < \mu_0, \text{ p-value} = P(T < t)$$

$$H_a : \mu > \mu_0, \text{ p-value} = P(T > t)$$

Steps for a mean hypothesis test

- 1 Identify null and alternative hypotheses from research question

$$H_0 : \mu = \mu_0$$

$$H_a : \mu \neq \mu_0, \mu < \mu_0, \mu > \mu_0$$

- 2 Use a t distribution
- 3 Calculate t test statistic
- 4 Calculate p-value
- 5 Compare p-value to significance level α and report decision
- 6 State conclusion in terms of original research question

Hypothesis tests for a mean in StatCrunch

- Stat → T Stats → One Sample → With Summary
(or → With Data)
- Enter “Sample mean”, “Sample std. dev.” and “Sample size”
(or select column which contains data)
- Select “Hypothesis test for μ ”
- Enter the appropriate values for null and alternative hypotheses.
- Click “Compute!”
- The test statistic and p-value are found in “T-Stat” and
“P-value”

Hypothesis test for a mean, example

Example

A study is conducted of the heights of Metro State students. The heights of a sample of 35 male students are measured. The mean height from the sample is $\bar{x} = 68.1$ with a standard deviation of $s = 3.5$. Conduct a test at $\alpha = 0.05$ level of significance of the claim that male Metro State students are shorter than the general population of 69.2 inches.

- 1 Identify null and alternative hypotheses from research question

$$H_0 : \mu = 69.2$$

$$H_a : \mu < 69.2$$

Population: Male Metro State students

Hypothesis test for a mean, example

Example

- ② Use a t distribution
- ③ Calculate t test statistic
 $t = -1.8593394$
- ④ Calculate p-value
 $p = 0.0358$
- ⑤ Compare p-value to significance level α and report decision
 $p = 0.0358 < \alpha = 0.05$. Reject the null hypothesis.
- ⑥ State conclusion in terms of original research question
There is evidence that male Metro State students are shorter than the general population.

Hypothesis test for a mean, example

Example

A clinical trial of an experimental drug for a rare disease is conducted. The survival times in months of 10 patients given the new drug are below. Test the claim that the drug changes the survival time for the expected survival time of 36 months at a 0.01 level of significance.

35.5, 40.5, 44.7, 37.2, 34.7, 41.3, 37.8, 34.5, 39.8, 37.1

- 1 Identify null and alternative hypotheses from research question

$$H_0 : \mu = 36$$

$$H_a : \mu \neq 36$$

Population: Patients with disease taking experiment drug

Hypothesis test for a mean, example

Example

- ② Use a t distribution
- ③ Calculate t test statistic
 $t = 2.2462$
- ④ Calculate p-value
 $p = 0.05132$
- ⑤ Compare p-value to significance level α and report decision
 $p = 0.05132 > \alpha = 0.01$. Do not reject the null hypothesis.
- ⑥ State conclusion in terms of original research question
There is not evidence that the experimental drug changes survival time.