Chapter 8—Hypothesis Testing

A hypothesis is a claim made about a population parameter (such as a population mean or population proportion).

Examples:

1. An advertisement claims that a certain brand of truck gets an average of 35 miles per gallon.
2. A company states that the average salary of all women in the company is $60,000.
3. A politician says that she has the support of more than 50% of people in a state.

A hypothesis test is a method for using sample data to decide between two competing claims (hypotheses) about an unknown population mean or population proportion.

Steps to Hypothesis Testing:

1. Create two contradictory statements (hypotheses).
2. Collect sample data.
3. Analyze the sample data to see which statement the data supports.
4. Make a decision.

null hypothesis (H_0)- the statement that the value of a population parameter is equal to some claimed value. It is assumed to be true unless it can be shown to be incorrect beyond a reasonable doubt.

alternative hypothesis (H_1)– the statement that the parameter has a value that is different than the value in the null hypothesis.

Types of Tests

Two-tailed:	$H_0: \mu = k$ $H_1: \mu \neq k$	or	$H_0: p = k$ $H_1: p \neq k$
Left-tailed:	$H_0: \mu = k$ $H_1: \mu < k$	or	$H_0: p = k$ $H_1: p < k$
Right-tailed:	$H_0: \mu = k$ $H_1: \mu > k$	or	$H_0: p = k$ $H_1: p > k$

Example: For a certain year a study reports that the nationwide percentage of college students using credit cards was 83%. If you want to set up a statistical test to challenge the claim that the proportion p = 0.83, what would the null and alternative hypotheses be?

H_0: p = 0.83 (claim)
H_1: p $\neq$ 0.83

If you want to set up a statistical test to test the claim that the proportion is less than 0.83, what would the null and alternative hypotheses be?

H_0: p = 0.83
H_1: p < 0.83 (claim)

If you want to set up a statistical test to test the claim that the proportion is more than 0.83, what would the null and alternative hypotheses be?

H_0: p = 0.83
H_1: p > 0.83 (claim)

Section 8-3 Testing a Claim about a Proportion

Recall: if x is a binomial random variable for which np $\geq$ 5 and nq $\geq$ 5, the distribution of $\hat{p}$ can be approximated by a normal distribution with a mean μ = p and and a standard deviation of $\sigma = \sqrt{\frac{pq}{n}}$

Example: When 109,857 arrests for federal offenses were randomly selected, it was found that 33,176 of them were drug offenses. Test the claim that 30% of federal crimes were for drug offenses. Use $\alpha = 0.01$.

1. State the null and alternative hypotheses.

H_0: p = 0.30 (claim)
H_1: p $\neq$ 0.30

Note that this is a <u>two-tailed test</u>. This will be important in step 3 where we will shade the probability.

2. Compute the test statistic under the assumption that H_o is true. The value of the test statistic will tell us whether or not the sample results were unusual.

The formula for the test statistic is:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$p = 0.30$ (from the null hypothesis)

$q = 1 - p$ = 1 - 0.30 = 0.70

$$\hat{p} = \frac{33{,}176}{109{,}857} = 0.302$$

$n = 109{,}857$

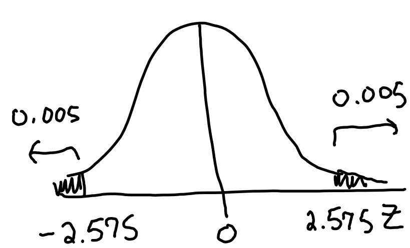
$$z = \frac{0.302 - 0.30}{\sqrt{\frac{(0.30)(0.70)}{109{,}857}}} = 1.45$$

z= = 1.45

3. Compute the critical value.

The critical value is based on the significance level of the test, or alpha. In this problem, $\propto = 0.01$.

Use the standard normal distribution and split the probability between the two tails.



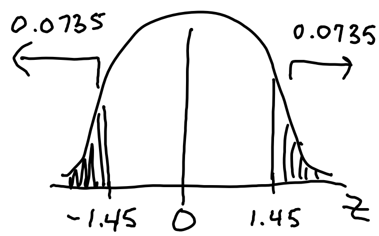
Critical values: $z_{\propto/2}$ = -2.575 and 2.575

Reject the null hypothesis if the test statistic is less than -2.575 or greater than 2.575 (in each tail of the distribution.)

4. State the conclusion in terms of the problem.

Since the Since 1.45 is not less than -2.575 or greater than 2.575, fail to reject the null hypothesis. There is not enough evidence to reject the claim that the drug offense rate is equal to 30%.

5. Compute the p-value. The p-value in this case is the probability of getting a value of the test statistic at least as extreme as what was observed in the sample data, assuming that the null hypothesis is true. Compute the p-value by plotting the test statistic and its opposite and shading the both tails because this is a two-tailed test.



area in one tail = normalcdf(-9999, -1.45, 0, 1) = 0.0735

p-value = 2*0.0735 = 0.1470

If the p-value is less than $\propto$, reject H_0. Using the p-value, we confirm that we should fail to reject H_0.

Example: In a survey of 703 randomly selected workers, 16% got their jobs through newspaper ads. Test the claim that less than 20% of workers get their jobs through newspaper ads. Use a 10% level of significance.

1. State the null and alternative hypothesis.

H_0: p = 0.20
H_1: p < 0.20

Note that this is a left-tailed test.

2. Compute the test statistic.

The formula for the test statistic is:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$p = 0.20$ (from the null hypothesis)

$q = 1 - p$ = 1 - 0.20 = 0.80

$\hat{p} = 0.16$

$n = 703$

$$z = \frac{0.16 - 0.20}{\sqrt{\frac{(0.20)(0.80)}{703}}} = -2.65$$

3. Compute the critical value.

The critical value is based on the significance level of the test, or alpha. In this problem, $\alpha = 0.10$.

Use the standard normal distribution and shade the left tail.

Critical value: z = -1.28

Reject the null hypothesis if the test statistic is less than -1.28 (in the left tail of the distribution.)

4. State the conclusion in terms of the problem.

Since -2.65 is less than -1.28, reject the null hypothesis. There is enough evidence to support the claim that less than 20% of workers get their jobs through newspaper ads.

5. Compute the p-value by plotting the test statistic and shading the left tail because this is a left-tailed test.

p-value = normalcdf(-9999, -2.65, 0, 1) = 0.0400

If the p-value is less than $\propto$, reject H_0. Using the p-value, we confirm that we should reject H_0.

Example: In one study of smokers who tried to quit smoking with nicotine patch therapy, 39 were smoking one year after the treatment and 32 were not smoking one year after the treatment. Use a 5% significance level to test the claim that among smokers who try to quite with nicotine patch therapy, the majority are smoking a year after treatment.

1. State the null and alternative hypothesis.

H_0: p = 0.50
H_1: p > 0.50

Note that this is a right-tailed test.

2. Compute the test statistic.

The formula for the test statistic is:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

$p = 0.50$ (from the null hypothesis)

$q = 1 - p$ = 1 - 0.50 = 0.50

$\hat{p} = 0.55$

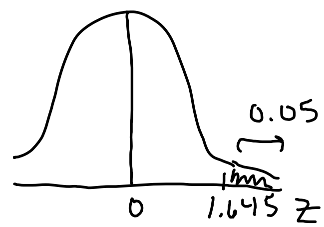
$n = 71$

$$z = \frac{0.55 - 0.50}{\sqrt{\frac{(0.50)(0.50)}{71}}} = 0.84$$

3. Compute the critical value.

The critical value is based on the significance level of the test, or alpha. In this problem, $\propto = 0.05$.

Use the standard normal distribution and shade the right tail.



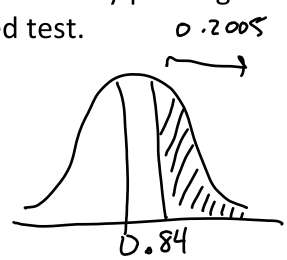
Critical value: z = 1.645

Reject the null hypothesis if the test statistic is greater than 1.645 (in the right tail of the distribution.)

4. State the conclusion in terms of the problem.

Since 0.84 is not greater than 1.645, fail to reject the null hypothesis. There is not enough evidence to support the claim that among smokers who try to quit with nicotine patch therapy, the majority are smoking a year after the treatment.

5. Compute the p-value by plotting the test statistic and shading the right tail because this is a right-tailed test.



p-value = normalcdf(0.84, 9999, 0, 1) = 0.2005

If the p-value is less than ∝, reject H_0. Using the p-value, we confirm that we should fail to reject H_0.

Errors in hypothesis testing

There is some chance that the use of hypothesis testing with sample data may lead us to the wrong conclusion about a population characteristic.

Suppose we look at these two competing hypotheses about an individual who was recently arrested:

H_0: Suspect is innocent
H_1: Suspect is guilty

The American system of justice says that an individual is "innocent until proven guilty." We will assume H_0 (innocence) is true, unless we collect overwhelming evidence against it, in which case we'll reject the null hypothesis in favor of the alternative that the suspect is guilty.

However, there's a chance that one of two types of errors may be made.

Type 1: An innocent person may be found guilty.
Type 2: A guilty person may fail to be convicted.

The first type of error, a Type 1, involves rejecting H_0 even though the null hypothesis is true. The second type of error, a Type 2, results from failing to reject H_0 when it is false.

		True State of Nature	
		H_0 is true	H_0 is false
Decision	Reject H_0	Type I error	Correct decision
	Fail to Reject H_0	Correct decision	Type II error

$\propto$ = The probability of a Type I error-- It is called the level of significance of the test.

Thus, a test with $\propto$= 0.01 is said to have a level of significance of 0.01 or to be a level 0.01 test. Common levels of significance are 0.01, 0.05, and 0.10. This can be controlled by the researcher.

β = The probability of a Type II error—this is not controlled by the researcher.

If we can select $\propto$, the probability of making a Type I error, why would we ever select $\propto$ = 0.05 instead of $\propto$= 0.01? Why not always select a very small value for $\propto$? To achieve a very small probability of making a Type I error, we would need the corresponding test procedure to require the evidence against H_0 to be very, very strong before the null hypothesis can be rejected.

It turns out that for a fixed sample size, n, a decrease in $\propto$ results in an increase in β. (e.g. if we make it really, really hard to incorrectly convict an innocent person (Type 1 error), that means that more guilty people are going to go free (Type 2 error)).

The usual recommendation is to choose the largest value of $\propto$ that you can tolerate, based on how serious a Type I error is.

power – (1-β) the probability of supporting a true alternative hypothesis. How good of a job does your test do to prove your claim (when it is, in fact, true). Ways to increase power include to increasing sample size or decreasing $\propto$.

Using the calculator for hypothesis testing

Testing a Claim about a Proportion:

STAT→ TESTS→1-Prop ZTest → enter p_0, x, n, and the type of test it is $\neq, < or > \rightarrow$ Calculate

Note: p_0 is the proportion that is assumed under H_0. Also, you may need to compute x if the problem gives you the sample proportion, $\widehat{p}$.

Section 8-4: Testing a Claim about a Mean, σ Known

Recall, if x has a normal distribution with mean μ and standard deviation σ, then $\bar{x}$ has a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. After you establish the null and alternate hypotheses, compute the test statistic based on the assumption that the null hypothesis is true:

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

Example: It has been reported that the average commute time to Washington D.C. is 46 minutes one way. Let x be a random variable representing commuting time. Assume x has a normal distribution with σ = 6. Suppose a sample of 5 commuters is selected, and their average one-way commuting times are shown below.

57 43 44 54 68

Do the data indicate that the average commuting time is more than 46 minutes? Let $\propto = 0.05$

1. State the null and alternative hypothesis.

H_0: μ = 46
H_1: μ > 46 (claim)

Note that this is a right-tailed test.

2. Compute the test statistic.

The formula for the test statistic is:

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

μ = 46 (from the null hypothesis)

$\bar{x}$ = 53.2 (Use 1-var stats on your calculator)

σ = 6 (given in problem)

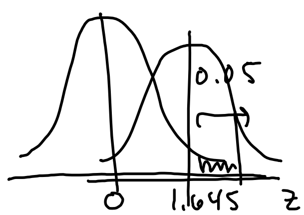
n = 5

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{53.2 - 46}{6 / \sqrt{5}} = 2.68$$

3. Compute the critical value.

The critical value is based on the significance level of the test, or alpha. In this problem, $\propto = 0.05$.

Use the standard normal distribution and shade the right tail.



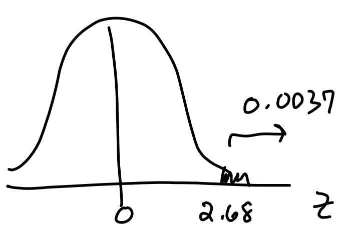
Critical value: z = 1.645

Reject the null hypothesis if the test statistic is greater than 1.645 (in the right tail of the distribution.)

4. State the conclusion in terms of the problem.

Since 2.68 is greater than 1.645, reject the null hypothesis. There is enough evidence to support the claim that the mean commuting time is greater than 46 minutes.

5. Compute the p-value by plotting the test statistic and shading the right tail because this is a right-tailed test.



p-value = normalcdf(2.68, 9999, 0, 1) = 0.0037

If the p-value is less than $\propto$, reject H_0. Using the p-value, we confirm that we should reject H_0.

Example: A special cable supposedly has a breaking strength of 800 pounds. A researcher selects a sample of 31 cables and finds that the average breaking strength is 793 pounds. At $\propto = 0.01$, test the claim that the average breaking strength is actually less than 800 pounds. Assume that the standard deviation of the population is known to be 12 pounds.

1. State the null and alternative hypothesis.

H_0: $\mu = 800$
H_1: $\mu < 800$ (claim)

Note that this is a left-tailed test.

2. Compute the test statistic.

The formula for the test statistic is:

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$\mu = 800$ (from the null hypothesis)

$\bar{x} = 793$ (given in the problem)

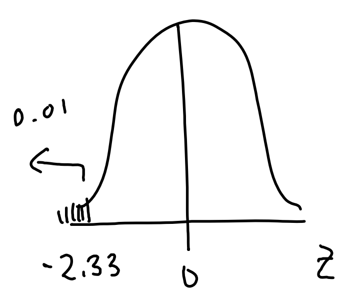
$\sigma = 12$ (given in problem)

n = 31

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{793-800}{12/\sqrt{31}} = -3.25$$

3. Compute the critical value.

The critical value is based on the significance level of the test, or alpha. In this problem, $\propto = 0.01$.



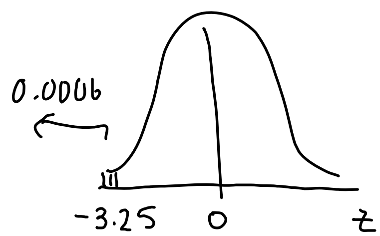
Critical value: $z = -2.33$

Reject the null hypothesis if the test statistic is less than -2.33 (in the left tail of the distribution.)

4. State the conclusion in terms of the problem.

Since -3.25 is less than -2.33, reject the null hypothesis. There is enough evidence to support the claim that the average breaking strength is less than 800 lbs.

5. Compute the p-value by plotting the test statistic and shading the left tail because this is a left-tailed test.



p-value = normalcdf(-3.25, 9999, 0, 1) = 0.0006

If the p-value is less than α, reject H_0. Using the p-value, we confirm that we should reject H_0.

Example: Developmental Psychology (March 2003) published a study on teacher perceptions of the behavior of elementary school children. Teachers rated the aggressive behavior of a sample of 500 New York City public school children by responding to the statement, "The child threatens or bullies others in order to get his/her own way." Responses were measured on a scale ranging from 1 (never) to 5 (always). The average score for the sample was 2.95. Suppose the standard deviation of the population is known to be 1.05.

Test the claim that the average score for all children in NYC elementary schools is equal to 3. Use $\alpha = 0.10$.

1. State the null and alternative hypothesis.

H_0: $\mu = 3$ (claim)
H_1: $\mu \neq 3$

Note that this is a two-tailed test.

2. Compute the test statistic.

The formula for the test statistic is:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$\mu = 3$ (from the null hypothesis)

$\bar{x} = 2.95$ (given in the problem)

$\sigma = 1.05$ (given in problem)

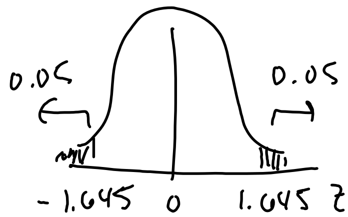
n = 500

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{2.95 - 3}{1.05 / \sqrt{500}} = -1.06$$

3. Compute the critical value.

The critical value is based on the significance level of the test, or alpha. In this problem, $\propto = 0.10$.

Use the standard normal distribution and split the probability between the two tails.



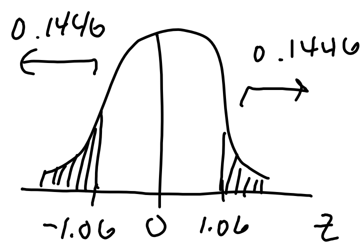
Critical value: z =+/- 1.645

Reject the null hypothesis if the test statistic is less than -1.645 or greater than 1.645.

4. State the conclusion in terms of the problem.

Since -1.06 is not less than -1.645, fail to reject the null hypothesis. There is not enough evidence to reject the claim that the average score for all NYC elementary school children is equal to 3.

5. Compute the p-value by plotting the test statistic and shading both tails.



p-value = 2* normalcdf(-9999, -1.06, 0, 1) = 2*0.1446 = 0.2892

If the p-value is less than ∝, reject H_0. Using the p-value, we confirm that we should fail to reject H_0.

Testing a Claim about a Mean: σ Known

With summary statistics,

STAT→TESTS→Z-Test→STATS→ enter μ_0, σ, $\bar{x}$, and n, and the type of test it is→ Calculate

With data in L1,

STAT→TESTS→Z-Test→STATS→ enter μ_0, σ, L1, and the type of test it is→Calculate

Section 8.5: Testing a Claim about a Mean: σ Unknown

1. State the null and alternate hypotheses.

2. Take a sample and compute the test statistic,

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

with degrees of freedom = n -1.

3. Use the t distribution to find the critical value(s).

4. Make a decision and state your conclusion in terms of the problem.

Example: In a British study, 12 healthy college students, deprived of one night's sleep, received an array of tests intended to measure creativity. Suppose the scores of the 12 sleep-deprived students had a mean of 63 and a standard deviation of 17. (Lower scores are associated with a decreased ability to think creatively.) Test the hypothesis that the true mean score of the sleep-deprived subjects is less than 80, the mean score of subjects who received sleep prior to taking the test. Use a significance level of 0.01.

1. State the null and alternative hypothesis.

H_0: μ = 80

H_1: $\mu < 80$ (claim)

Note that this is a left-tailed test.

2. Compute the test statistic.

The formula for the test statistic is:

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

$\mu = 80$ (from the null hypothesis)

$\bar{x} = 63$ (given in the problem)

$s = 17$ (given in problem)

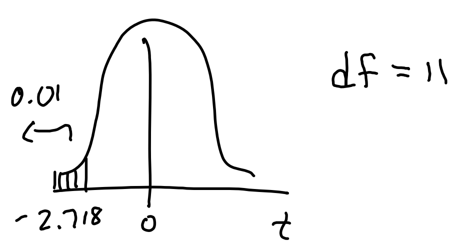
n = 12

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{63-80}{17/\sqrt{12}} = -3.46$$

3. Compute the critical value.

The critical value is based on the significance level of the test (α) and the degrees of freedom. In this problem, $\alpha = 0.01$, degrees of freedom = n – 1 = 12 – 1 = 11.

Use the t distribution and shade 0.01 in the left tail.



Critical value: t = - 2.718

Reject the null hypothesis if the test statistic is less than -2.718.

4. State the conclusion in terms of the problem.

Since -3.46 is less than -2.718, reject the null hypothesis. There is enough evidence to support the claim that the average score of all sleep deprived subjects is less than 80.

Example: The average family size in the United States is reported as 3.18. A random sample of 24 families in a particular school district resulted in the following family sizes:

5	4	5	4	4	4	3	6	4	3	3	5
6	3	3	2	7	4	5	2	2	2	3	5

For this sample, $\bar{x} = 3.9$ and s = 1.4. At $\alpha = 0.05$, test the claim that the average family size in this school district differs from the national average.

1. State the null and alternative hypothesis.

H_0: $\mu = 3.18$
H_1: $\mu \neq 3.18$ (claim)

Note that this is a two-tailed test.

2. Compute the test statistic.

The formula for the test statistic is:

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

$\mu = 3.18$ (from the null hypothesis)

$\bar{x} = 3.9$ (given in the problem)

$s = 1.4$ (given in problem)

n = 24

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{3.9-3.18}{1.4/\sqrt{24}} = 2.52$$

3. Compute the critical value.

The critical value is based on the significance level of the test (α) and the degrees of freedom. In this problem, $\alpha = 0.05$, degrees of freedom = n – 1 = 24 – 1 = 23.

Use the t distribution and shade 0.025 in each tail (divide α by two because this is a two-tailed test).

Critical value: t = +/-2.069

Reject the null hypothesis if the test statistic is less than -2.069 or greater than 2.069.

4. State the conclusion in terms of the problem.

Since 2.52 is greater than 2.069, reject the null hypothesis. There is enough evidence to support the claim that the average family size in this school district differs from the national average.

With summary statistics,

STAT→TESTS→T-Test→STATS→ enter μ_0, s, $\bar{x}$, and n, and the type of test it is→ Calculate

With data in L1,

STAT→TESTS→T-Test→STATS→ enter μ_0, s, L1, and the type of test it is→Calculate