3.1 Confidence intervals

Estimation and confidence intervals

Estimation is the process of obtaining information about a parameter by using a statistic. An ***estimator*** is a statistical method used to calculate an estimate based on observable data. A good estimator gives estimates that are both accurate and precise. Accuracy is measured in terms of bias. Numerically, bias is the distance between the mean of the sampling distribution and the population mean. Precision is measured in terms of standard error.

PARTICIPATION ACTIVITY | 3.1.1: Accuracy and precision.

Animation captions:

1. Accuracy is a measure of closeness to the true value of a parameter. Precision is a measure of closeness to the average value of the measurements.
2. Measurements that are neither accurate nor precise are neither close to the true value nor tightly clustered.
3. Measurements that are accurate but not precise are loosely spread out around the true value.
4. Measurements that are precise but not accurate are tightly clustered around an incorrect value.
5. Measurements that are both accurate and precise are tightly clustered around the true value.

Two types of estimates exist: point estimates and interval estimates. A ***point estimate*** is a single value estimate for a parameter. An ***interval estimate*** is a range of values that is likely to contain the parameter being estimated. Combined with a probability statement, an interval estimate is called a ***confidence interval***. The percentage in which the confidence interval contains the parameter is called the ***confidence level***, which is denoted by c.

A confidence interval is ***accurate*** if the confidence interval contains the true population parameter. A confidence interval's ***precision*** refers to the width of the confidence interval.

A confidence interval is constructed by looking at the sample statistic and margin of error. A ***margin of error***, denoted by m, is the range of values above and below the point estimate. Numerically,

$$m = (\text{critical value})(\text{standard error})$$

where the critical value, which depends on c and the underlying distribution of the statistic, is the number of standard errors to be added to the point estimate. Thus,

$$\text{estimate} \pm m = \text{estimate} \pm (\text{critical value})(\text{standard error})$$

The resulting interval is referred to as the $c(100)\%$ confidence interval. That is, $c(100)\%$ of the time, the true value of the population parameter will be in the $c(100)\%$ confidence interval when the same estimator is used, as shown in the animation below.

PARTICIPATION ACTIVITY | 3.1.2: A 95% confidence interval for a population parameter.

Animation captions:

1. Suppose the true value of the population parameter is 4.
2. A sample is drawn from the population. The sample statistic is 3.5.
3. A 95% confidence interval of $[1, 6]$ implies that if samples are drawn repeatedly, the probability that the true population parameter is in the interval is 95%.
4. The sample statistic is one possible outcome of repeated drawing of samples.
5. The size of the confidence interval is such that on average, only 1 of every 20 potential 95% confidence intervals will not include the true value of the population parameter.

Example 3.1.1: How millennials acquire news.

A survey by the Media Insight Project asked a random sample of $1,045$ American adults between the ages of 18 and 34 questions about technology use when acquiring news and frequency[1]. The results of the survey have a margin of error of ± 3.8 percentage points.

Percent of millennials who ...	
Say keeping up with the news is at least somewhat important	85%
Get news daily	69%
Regularly follow five or more 'hard' news topics	45%
Usually see diverse opinions through social media	86%
Pay for at least one news-specific service, app, or digital subscription	40%

a. What is the point estimate for the percent of millennials who get their news daily?
b. What is the 95% confidence interval for the percent of millennials who get their news daily?
c. What is the interpretation of this interval?

Solution

a. 69%

b. $69\% \pm 3.8\%$
c. A 95% confidence exists that the true population percentage of millennials who get news daily is between 65.2% and 72.8%. The 95% represents the probability that the true population percentage is contained in a confidence interval if the estimator is used repeatedly.

PARTICIPATION ACTIVITY 3.1.3: Estimation and confidence intervals.

1) A confidence interval estimate of a population parameter is constructed around the sample statistic.
- True
- False

2) A confidence interval always gives a correct estimate of the population parameter.
- True
- False

3) The true population parameter is $p = 0.5$. An estimator that produces a confidence interval of $[0.42, 0.49]$ is accurate.
- True
- False

4) A poll has a margin of error of $\pm 2.5\%$. The width of the corresponding confidence interval is 5%.
- True
- False

PARTICIPATION ACTIVITY 3.1.4: Americans' view of socialism and capitalism.

According to the results of a Gallup poll dated May 2-4, 2016 on American views involving Capitalism and Socialism, young people view the federal government and socialism more positively than older Americans do[2].

The table below gives the percentage of people who view the following positively according to age group: small business, entrepreneurs, free enterprise, capitalism, big business, federal government, and socialism.

Just off the top of your head, would you say you have a positive or negative image of each of the following?				
	18-29 years	30-49 years	50-64 years	65+ years
Small business	98	94	96	93
Entrepreneurs	90	87	87	83
Free enterprise	78	84	89	91
Capitalism	57	54	69	63
Big business	57	49	52	53
Federal government	58	43	38	40
Socialism	55	37	27	37

The poll has a $\pm 4\%$ margin of error at a 95% confidence level.

Select the definition that matches each term

1) 58%

- 95% confidence interval for the percent of Americans ages 18-29 who view socialism positively
- Point estimate for the percent of Americans ages 18-29 who view socialism positively
- Point estimate for the percent of Americans ages 18-29 who view the federal government positively
- 95% confidence interval for the percent of Americans ages 18-29 who view the federal government positively

2) 55%

- 95% confidence interval for the percent of Americans ages 18-29 who view socialism positively
- Point estimate for the percent of Americans ages 18-29 who view socialism positively
- Point estimate for the percent of Americans ages 18-29 who view the federal government positively
- 95% confidence interval for the percent of Americans ages 18-29 who view the federal government positively

3) [54%, 62%]

- 95% confidence interval for the percent of Americans ages 18-29 who view socialism positively
- Point estimate for the percent of Americans ages 18-29 who view socialism positively
- Point estimate for the percent of Americans ages 18-29 who view the federal government positively
- 95% confidence interval for the percent of Americans ages 18-29 who view the federal government positively

4) [51%, 59%]

- 95% confidence interval for the percent of Americans ages 18-29 who view socialism positively
- Point estimate for the percent of Americans ages 18-29 who view socialism positively
- Point estimate for the percent of Americans ages 18-29 who view the federal government positively

©zyBooks 01/31/23 17:54 1267703
Traver Yates
MAT-243-J3996-OL-TRAD-UG.23EW3

- 95% confidence interval for the percent of Americans ages 18-29 who view the federal government positively

Reset

References

(*1) The Media Insight Project. "How Millennials Get News: Inside the Habits of America's First Digital Generation." *NORC at the University of Chicago*, American Press Institute, 2017, mediainsight.org/Pages/how-millennials-get-news-inside-the-habits-of-americas-first-digital-generation.aspx

(*2) Newport, Frank. "Americans' Views of Socialism, Capitalism Are Little Changed." *Gallup*, American Press Institute, 6 May 2016, news.gallup.com/poll/191354/americans-views-socialism-capitalism-little-changed.aspx

3.2 Confidence intervals for population means

z confidence intervals

Choosing an appropriate estimator for the population mean μ depends on what is known about the population. Although the population mean is usually unknown, the confidence interval that quantifies the range within which the true population mean lies can be calculated by sampling the population. Ex: A 95% confidence interval means that the interval calculated has a 95% probability of containing the population mean. If such a population is sampled 20 times, 19 of the 20 calculated intervals are expected to contain the population mean.

PARTICIPATION ACTIVITY 3.2.1: The z-confidence interval.

Animation content:

undefined

Animation captions:

1. A confidence interval for the population mean of a normal distribution is calculated from the sample mean $\bar{x}$, standard deviation σ, and number of samples n.

2. The margin of error m for the population mean is equal to $z^* \frac{\sigma}{\sqrt{n}}$.
3. The confidence level c is the area under the normal distribution curve in the interval $[\bar{x} - m, \bar{x} + m]$.
4. The significance level α is the area under the curve outside the confidence interval.

The critical values z^* for common confidence levels are shown in the following table.

Table 3.2.1: Critical values for common confidence levels.

Confidence level	Critical value
$c = 0.90$	$z^* = 1.645$
$c = 0.95$	$z^* = 1.960$
$c = 0.99$	$z^* = 2.576$

Example 3.2.1: Confidence interval for the mean final grade.

Suppose the mean final grade is estimated for all introductory statistics classes taught at a particular college. The population standard deviation is 4. The final grades for a randomly selected statistics class with 9 people are: 76, 80, 82, 83, 83, 85, 85, 87, and 88. Find the 95% confidence interval for the mean final grade.

Solution

The sample mean is

$$\bar{x} = \frac{76 + 80 + 82 + 83 + 83 + 85 + 85 + 87 + 88}{9} = 83.222$$

At the 95% confidence level, the margin of error is

$$z^* = 1.960$$
$$z^* \frac{\sigma}{\sqrt{n}} = 1.960 \frac{4}{\sqrt{9}}$$
$$\approx 2.613$$

Thus, the 95% confidence interval is $[83.222 - 2.613, 83.222 + 2.613] = [80.609, 85.835]$.

Analysis

The population mean final grade is generally unknown. The obtained interval means that a 95% confidence exists that $[80.609, 85.835]$ contains the true population mean final grade.

PARTICIPATION ACTIVITY 3.2.2: Confidence interval.

The weights of 5 squash (in pounds) are $10, 17, 17.5, 18.5$, and 19.5. The sample weights have a mean of $\bar{x} = 16.5$. The accepted population standard deviation for this type of squash is $\sigma = 1.25$.

1) What is the margin of error at the 90% confidence level? Type as #.###

Check Show answer

2) What is the margin of error at the 99% confidence level? Type as: #.###

Check Show answer

3) What is the 90% confidence interval? Type as: [##.###, ##.###]

Check Show answer

4) What is the 99% confidence interval? Type as: [##.###, ##.###]

Check Show answer

Python-Function 3.2.1: norm.interval().

The `norm.interval()` function is used to find a confidence interval for a normally distributed variable. The function takes the desired confidence level, the sample mean, and the standard error as parameters.


```
import scipy.stats as st
print(st.norm.interval(0.95, 0, 1))
```

```
(-1.959963984540054, 1.959963984540054)
```

A confidence interval can also be calculated from raw data. The following example imports data from ExamScores.csv into a DataFrame, calculates the sample mean, calculates the standard error based on a given population standard deviation, and calculates the 99% confidence interval.

```
scores = pd.read_csv('http://data-
analytics.zybooks.com/ExamScores.csv')
sigma = 2.5
mean = scores['Exam1'].mean()
stderr = sigma/math.sqrt(len(scores['Exam1']))
print(st.norm.interval(0.99, mean, stderr))
```

```
(81.78930681614078,
83.610693183859226)
```

Run example

Margin of error and sample size for means when σ is known

The width of the confidence interval is twice the margin of error. Recall that the margin of error depends on the confidence level and the standard error. Thus, given a confidence level, the width of the confidence interval changes by changing the standard error. Increasing the sample size decreases the standard error. Similarly, decreasing the sample size increases the standard error. The size of the sample needed to guarantee a confidence interval with a specified margin of error is given by the formula

$$n = \left(\frac{z^* \sigma}{m}\right)^2$$

Example 3.2.2: Microbeads in a water reservoir.

Scientists study the amount of microbeads (in μg/liter) in a water reservoir. The scientists need to know how many water samples should be taken to be 95% sure that their estimate $\bar{x}$

differs from the actual value μ by at most 0.8 μg/liter given a population standard deviation of $\sigma = 2.5$ μg/liter. How many water samples should the scientists take?

Solution

The margin of error is $m = 0.8$ and the population standard deviation is $\sigma = 2.5$. Since the confidence level is 95%, the critical value is $z^* = 1.960$. Using the formula above,

$$n = \left(\frac{z^*\sigma}{m}\right)^2 = \left(\frac{1.960 \cdot 2.5}{0.8}\right)^2 \approx 37.516$$

Thus, the scientists should take at least 38 water samples.

PARTICIPATION ACTIVITY 3.2.3: Tax assessment.

A tax assessor assesses the mean property tax bill for all property owners in a certain city. A recent survey obtained a sample mean of 1400 dollars. The population standard deviation is known to be 1000 dollars.

1) How many tax records should be obtained at a 90% confidence level to have a margin error of 100 dollars? Round up to the nearest whole number.

Check Show answer

2) If the population standard deviation goes up to 1500, would the margin of error be equal, greater than, or less than 100? Type as: equal, greater than, or less than

Check Show answer

t confidence intervals

The confidence interval for a population mean is defined as

$$\left[\bar{x} - z^* \frac{\sigma}{\sqrt{n}}, \bar{x} + z^* \frac{\sigma}{\sqrt{n}}\right]$$

In practice, however, the population standard deviation is rarely known. Thus, confidence intervals using the t-distribution are more useful because the sample standard deviation can always be computed.

The t-confidence interval is given by

$$\left[\bar{x} - t^* \frac{s}{\sqrt{n}}, \bar{x} + t^* \frac{s}{\sqrt{n}}\right]$$

where t^* is the critical value that depends on the degrees of freedom and significance level. The values for common significance levels given selected degrees of freedom are shown in the table below.

Table 3.2.2: Critical values t^* for selected degrees of freedom df and significance level α.

	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$df = 5$	2.015	2.571	4.032
$df = 10$	1.812	2.228	3.169
$df = 15$	1.753	2.131	2.947
$df = 24$	1.711	2.064	2.797
$df = 32$	1.309	1.694	2.449

Example 3.2.3: Circumference of basketballs.

The mean circumference of basketballs produced in a manufacturing facility is supposed to be 29 inches. A random sample of 25 basketballs has a mean of 29.1 inches with a sample standard deviation of 0.217 inches. Find the confidence interval at the $\alpha = 0.01$ significance level.

Solution

At a significance level of $\alpha = 0.01$, the confidence level is 99%. The critical value that corresponds to a 99% confidence level and degrees of freedom $df = 25 - 1 = 24$ is $t^* = 2.797$ from the table above. Thus, the margin of error is

$$m = t^* \frac{s}{\sqrt{n}} = 2.797 \frac{0.217}{\sqrt{25}} \approx 0.121$$

Thus, the 99% confidence interval is

$$[\bar{x} - m, \bar{x} + m] = [29.1 - 0.121, 29.1 + 0.121] = [28.979, 29.221]$$

Analysis

The obtained confidence interval suggests that a 99% confidence exists that the true population mean circumference of basketballs produced in the facility is between 28.979 and 29.221 inches. Since the 99% confidence interval contains the population mean of 29 inches, insufficient evidence exist to believe that the mean circumference of basketballs is not 29 inches.

PARTICIPATION ACTIVITY 3.2.4: Weights of pumpkins.

The weight (in pounds) of 6 pumpkins are $5, 7, 7.5, 8, 8.5,$ and 8.75. The weights of the sample have a mean of $\bar{x} = 7.458$ and a standard deviation of $s = 1.364$.

1) What is the degrees of freedom?

Check Show answer

2) What is the critical value when the significance level is $\alpha = 0.1$? Type as: #.###

Check Show answer

3) What is the margin of error? Type as: #.###

Check Show answer

4) What is the 90% t-confidence interval for pumpkin weights? Type as: [#.### , #.###]

Check Show answer

Python-Function 3.2.2: t.interval().

The `t.interval()` function is used to find a confidence interval for a variable with a t-distribution. The function takes the desired confidence level, the degrees of freedom, the mean, and the standard error as parameters.

In the following example, the degrees of freedom and standard error are calculated from the number of samples and the standard deviation given in a data summary.

```
import scipy.stats as st

n = 100

# Degrees of freedom is number of samples minus 1
df = n - 1

mean = 219

# The standard error is standard deviation/sqrt(number of samples)
stderr = 35.0/(n ** 0.5)

print(st.t.interval(0.95, df, mean, stderr))
```

```
(212.05524066971961,
225.94475933028039)
```


A confidence interval can also be calculated from raw data. The following example imports data from ExamScores.csv into a DataFrame and finds the confidence interval.

```
import pandas as pd
import scipy.stats as st
scores = pd.read_csv('http://data-
analytics.zybooks.com/ExamScores.csv')

# Let n be the number of students who took Exam 1.
n = scores[['Exam1']].count()

# Degrees of freedom is number of samples minus 1
df = n - 1

# The mean of Exam1 scores are obtained
mean = scores[['Exam1']].mean()

# The standard error is standard deviation/sqrt(number
of samples)
stdev = scores[['Exam1']].std()
stderr = stdev/(n ** 0.5)

print(st.t.interval(0.95, df, mean, stderr))
```

```
(array([ 80.05931209]),
array([ 85.34068791]))
```


Run example

Margin of error and sample size for means when σ is unknown

The z-distribution is commonly used to estimate the necessary sample size for a survey. Since the population standard deviation is generally unknown, the t-distribution should be used instead. For a situation with an unknown population standard deviation, the sample standard deviation should be used. Thus,

$$n = \left(\frac{t^* s}{m}\right)^2$$

where t^* is the t critical value at a specified confidence level, s is the sample standard deviation based on a preliminary study, and m is the margin of error.

However, t^* depends on the degrees of freedom $n - 1$, which is also in terms of n. To find t^*, z^* is used to find a preliminary sample size n^*. That is,

$$n^* = \left(\frac{z^* s}{m}\right)^2$$

Then n can be calculated using the formula above with $df = n^* - 1$.

Example 3.2.4: Quality control.

A quality control department in a company wants to estimate the average length of a part to within a margin of error ± 0.35 cm at 95% confidence interval. A preliminary pilot study found that the sample standard deviation of the length of parts is 2.50 cm. How many parts should the department use?

Solution

The formula above uses t^*, and thus a preliminary sample size n^* is needed. To find n^*, z^* is used in the formula instead of t^*. Since $z^* = 1.960$ at the 95% confidence level,

$$n^* = \left(\frac{z^* s}{m}\right)^2 = \left(\frac{1.960 \cdot 2.50}{0.35}\right)^2 = 196$$

The critical t value that corresponds to $df = n^* - 1 = 196 - 1 = 195$ is $t^* = 1.972$, which yields

$$n = \left(\frac{t^* s}{m}\right)^2 = \left(\frac{1.972 \cdot 2.50}{0.35}\right)^2 \approx 199$$

Thus, the quality control department should use a sample size of 199 parts.

PARTICIPATION ACTIVITY 3.2.5: Machine output.

A company that manufactures tools wants to estimate the average output of a machine to within a margin of error of ± 2.1 units with 90% confidence. A preliminary pilot study found that the sample standard deviation of the machine's output is 7.3 units.

1) To the nearest whole number, what is the preliminary estimate n^*?

Check Show answer

2) What is the sample size n? The critical value when $df = 32$ is $t^* = 1.694$.

Check Show answer

CHALLENGE ACTIVITY 3.2.1: Confidence intervals for population means.

456500.2535406.qx3zqy7

3.3 Confidence intervals for population proportions

Confidence interval for population proportions

Constructing a confidence interval for a population proportion is similar to constructing a confidence interval for population means. Suppose the sample proportion $\hat{p}$ and the number of samples n are known from sampling the population. The margin of error m is given by $m = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where z^* is the critical value corresponding to the desired confidence level c. The confidence interval for the population proportion p is given by $[\hat{p} - m, \hat{p} + m]$.

Example 3.3.1: Confidence interval for the proportion.

In a survey of 1200 randomly selected registered voters in a particular city, 348 people are in favor of banning public smoking. State and interpret the 95% confidence interval for the population proportion of voters in favor of banning public smoking.

Solution

The sample proportion is

$$\hat{p} = \frac{348}{1200} = 0.29$$

At the 95% confidence level, the margin of error is

$$m = z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.960\sqrt{\frac{0.29(1-0.29)}{1200}} \approx 0.026$$

Thus, the 95% confidence interval is

$$[\hat{p} - m, \hat{p} + m] = [0.29 - 0.026, 0.29 + 0.026] = [0.264, 0.316]$$

The 95% confidence level suggests that if the same estimator was used to construct the interval, then the true population parameter would be within the interval 95% of the time. The obtained confidence interval suggests a 95% confidence exists that the true proportion of voters who favor banning smoking is between 26.4% and 31.6%.

PARTICIPATION ACTIVITY 3.3.1: Confidence interval for the proportion.

In a poll of 1000 randomly selected prospective voters in a local election, 281 voters were in favor of a school bond measure.

1) What is the sample proportion? Type as: #.###

Check Show answer

2) What is the margin of error for the 90% confidence level? Type as: #.###

Check Show answer

3) What is the margin of error for the 95% confidence level? Type as: #.###

Check Show answer

4) What is the 95% confidence interval? Type as: [#.###, #.###]

Check Show answer

Python-Function 3.3.1: norm.interval().

The `norm.interval()` function can also be used to find the confidence interval for population proportion. In the following example, the standard error is calculated from the number of samples and the proportion given in a data summary and the result is used as input in the `norm.interval()` function.


```
# In a survey of 1200 randomly selected registered voters,
# 348 were in favor of banning public smoking. Find the 95%
# confidence interval for the proportion of voters in favor
# of banning public smoking.

import scipy.stats as st

# Let n be the number of voters surveyed.
n = 1200

# Let p be the proportion of voters that voted in favor
p = 348.0/1200.0

# The standard error is sqrt(p * (1-p)/n)
stderr = (p * (1 - p)/n) ** 0.5

print(st.norm.interval(0.95, p, stderr))
```

```
(0.26432646675431226, 0.31567353324568877)
```

A confidence interval can also be calculated from raw data. The following example imports data from ExamScores.csv into a DataFrame, filters based on a minimum score, and finds the confidence interval.

```python
# In the Exam Scores data set, find a 99% confidence interval
# for the proportion of students who scores more than 90 in Exam 1.

import pandas as pd
import scipy.stats as st
scores = pd.read_csv('http://data-analytics.zybooks.com/ExamScores.csv')

# Let n be the number of students who took Exam 1.
n = scores[['Exam1']].count()

# Let x be the total of all Exam 1 scores greater than 90
x = (scores[['Exam1']] > 90).values.sum()

# Let p be x/n, the proportion of all students that scored over 90 on Exam 1
# Multiplying by 1.0 is needed for correct floating point arithmetic
p = x/n*1.0

# The standard error is sqrt(p * (1-p)/n)
stderr = (p * (1 - p)/n) ** 0.5

print(st.norm.interval(0.99, p, stderr))
```



```
(array([ 0.02645375]),
array([ 0.29354625]))
```

Run example

Margin of error and sample size for proportions

Finding the necessary sample size given a confidence level and margin of error is similar to finding the sample size for a confidence interval for population means. The size of the sample needed to guarantee a confidence interval with a specified margin of error is given by the formula

$$n = \left(\frac{z^*}{m}\right)^2 p(1-p)$$

where z^* is the critical value at a specified confidence level, p is the proportion, and m is the margin of error.

However, p also depends on n. Two possible options exist:

1. A preliminary value for p can be used based on a previous study.
2. The worst case scenario with $p = 0.50$ can be used, which provides the largest sample size needed to satisfy the required margin of error at a specified confidence level.

Example 3.3.2: Six months after Brexit.

Brexit is a popular term for the successful referendum on June 2016 for the withdrawal of the United Kingdom from the European Union. Six months later, a poll conducted by CNN and ComRes resulted in 47% of the respondents voting to leave the European Union had the referendum been held that day[1]. If the margin of error is ± 2.17 percentage points, find the sample size needed for a 90% confidence interval if

a. $p = 0.47$ is used
b. no previous estimate is given

Solution

a. The margin of error is $m = 0.0217$. Since a previous estimate is known, the sample proportion is $p = 0.47$. Since the confidence level is 90%, the critical value is $z^* = 1.645$. Using the formula above,

$$n = \left(\frac{z^*}{m}\right)^2 p(1-p) = \left(\frac{1.645}{0.0217}\right)^2 (0.47)(1-0.47) \approx 1432$$

Thus, the pollsters sampled at least 1432 voting aged adults.

b. When no previous estimate is known, $p = 0.5$ is used. Using the same formula gives,

$$n = \left(\frac{z^*}{m}\right)^2 p(1-p) = \left(\frac{1.645}{0.0217}\right)^2 (0.50)(1-0.50) \approx 1437$$

Thus, the pollsters would have sampled 1437 voting aged adults if no previous estimate is given in the news story.

PARTICIPATION ACTIVITY 3.3.2: Margin of error and sample size for proportions.

A poll reported a 36% approval rating for a politician with a margin of error of 1 percentage point.

1) How many voters should be sampled for a 90% confidence interval? Round up to the nearest whole number.

Check Show answer

2) How many voters should be sampled for a 95% confidence interval? Round up to the nearest whole number.

Check Show answer

3) How many voters should be sampled for a 95% confidence interval if no previous estimate is given? Round up to the nearest whole number.

Check Show answer

CHALLENGE ACTIVITY | 3.3.1: Confidence intervals for population proportions.

Critical values for quick reference during this activity.

Confidence level	Critical value
0.90	$z^* = 1.645$
0.95	$z^* = 1.960$
0.99	$z^* = 2.576$

456500.2535406.qx3zqy7

References

(*1) Greene, Richard Allen. "Brexit poll: Six months on, Brits stand by EU referendum decision." *CNN*, 19 December 2016. www.cnn.com/2016/12/19/europe/cnn-brexit-poll/index.html

3.4 Hypothesis testing

Hypothesis

A ***hypothesis*** is a proposed explanation of a phenomenon, usually as a starting point for further analysis. Ex: If a lamp doesn't illuminate when turned on, one hypothesis is: "Lamp is not plugged in". Subsequent analysis may result in rejecting the hypothesis. Ex: Observing the lamp is indeed plugged into an electric outlet causes one to reject the "Lamp is not plugged in" hypothesis.

PARTICIPATION ACTIVITY 3.4.1: Hypotheses.

Indicate which statements are likely to be hypotheses.

1) A researcher proposes collecting data for 1,000 people's daily sitting times and weights to determine if the following is true: The more one sits, the more weight one gains.

- Hypothesis
- Not a hypothesis

2) Upon obtaining data for 1,000 people's daily sitting times and weights and doing a calculation, a researcher writes: The 1,000 people had a mean sitting time of 5 hours per day.

- Hypothesis
- Not a hypothesis

3) A researcher proposes collecting data for 1,000 people's daily sitting times and genders, and computing the means for men and women separately, to determine if the following is true: In America, men sit longer than women each day.

- Hypothesis
- Not a hypothesis

4) People should stand more to gain health benefits.

- Hypothesis
- Not a hypothesis

Forming a hypothesis

Hypothesis for one population

A common hypothesis formed by researchers is whether the correlation of two variables in a sample suggests that those variables are correlated in the population.

PARTICIPATION ACTIVITY 3.4.2: A common hypothesis relates to correlation of variables in a sample.

Animation content:

undefined

Animation captions:

1. If a population is known, a line can be drawn showing the correlation of height and weight.
2. If a small sample is taken, a line can also be drawn. A hypothesis may ask if that correlation is strong enough to infer the correlation exists in the population.

PARTICIPATION ACTIVITY 3.4.3: Hypotheses, samples, and populations.

Refer to the animation above.

1) The population data shows that as height increases, weight _____.

- Increases
- Decreases

2) The sample consists of _____ members.

- 10
- 3

3) The sample's line indicates a ______ correlation.

- Positive
- Negative

4) A researcher hypothesizes: Greater height does NOT suggest greater weight. The steeper the line for the sample, the more likely is the analyst to _____ this hypothesis.

- reject
- not reject

5) The analyst usually could just compute the correlation for an entire population.

- True
- False

Hypothesis for two populations

Researchers also commonly investigate hypotheses relating to two populations. Ex: Given samples from two populations, a hypothesis might be that the two populations have the same mean (the difference in the two samples' means is statistically insignificant). The challenge of course is that the analyst only has data from the samples; the population's data is unknown. Ex: Salaries for 100,000 people who live in a city can't reasonably become known to a researcher, but salaries for samples of 50 males and 50 females in that city can be known.

PARTICIPATION ACTIVITY | 3.4.4: A common hypothesis asks whether two populations have the same mean vs. different means.

Animation content:

undefined

Animation captions:

1. In practice, hair color doesn't impact salary, so samples of blondes and brunettes are from populations with the same mean salary. The small difference in sample means is just due to chance.
2. In practice, smokers do have lower salaries, so smokers and non-smokers are from populations with different mean salaries. The larger difference in sample means is statistically significant.

PARTICIPATION ACTIVITY 3.4.5: Hypotheses, samples, and populations.

Refer to the above animation.

1) A researcher creates this hypothesis: Blondes and brunettes earn the same salary. Usually, the analyst can simply compute the mean for all blondes in the population, and for all brunettes, and then compare those two means.

- True
- False

2) A researcher creates this hypothesis: Blondes and brunettes earn the same salary. The analyst obtains data for a sample of 5 blondes and 5 brunettes, and determines a small difference exists. The hypothesis can be rejected.

- True
- False

3) Blondes and brunettes are shown to come from populations with the same mean salary in the animation's dot plots. Thus, for a different feature like hours spent at work, blondes and brunettes also come from populations with the same mean.

- True
- False

4) A researcher hypothesizes: Non-smokers earn the same as smokers. The large difference in mean salary for size-5 samples of non-smokers and smokers may cause the analyst to ______ that hypothesis.

- Reject
- Not reject

Null and alternative hypotheses

In statistics, a ***hypothesis*** is a statement that makes a claim about the parameters of one or more populations. ***Hypothesis testing*** is the formal process by which a hypothesis is retained or rejected. Hypothesis testing compares two competing hypotheses about a population, the null hypothesis and the alternative hypothesis.

A ***null hypothesis***, denoted H_0, is a statement assumed to be true unless sufficient data indicates otherwise. Typically, a null hypothesis is a statement of equality between the true value of the population parameter and the hypothesized value or a statement of no difference between the parameters of two populations. Ex: The statement "The average salary of the residents of San Francisco is not different than the average salary of the residents of Austin" or "The average salary of the residents of San Francisco is the same as the average salary of the residents of Austin" is a null hypothesis.

In contrast, an ***alternative hypothesis***, denoted H_a, is a statement that contradicts H_0. Typically, an alternative hypothesis asserts that the true value of the population parameter is not the same as the hypothesized value or that the parameters for two populations are different. Ex: The alternative hypothesis corresponding to the null hypothesis above is "The average salary of the residents of San Francisco is different from the average salary of the residents of Austin, Texas."

An alternative hypothesis may be left-tailed, right-tailed, or two-tailed depending on the nature of the difference from the null hypothesis.

- A left-tailed alternative hypothesis asserts that the value of a parameter is less than the value asserted in the null hypothesis.
- A right-tailed alternative hypothesis asserts that the value of a parameter is greater than the value asserted in the null hypothesis.
- A two-tailed alternative hypothesis asserts that the value of a parameter is not equal to, that is, either less than or greater than the value asserted in the null hypothesis.

Example 3.4.1: Null hypothesis and alternative hypothesis.

Identify the type of each alternative hypothesis and state the corresponding null hypothesis.

a. The average time spent sitting at a particular company's office is not 5.7 hours.
b. Patients given a drug recover from a disease in less time than patients not given the drug.

Solution

a. The alternative hypothesis asserts that a parameter is not equal to a value and is thus a two-tailed alternative hypothesis. The corresponding null hypothesis is "The average time spent sitting at a particular company's office is 5.7 hours".

b. The alternative hypothesis asserts that a parameter is less than a value and is thus a left-tailed hypothesis. The corresponding null hypothesis is "Patients given a drug recover from the disease in the same amount of time as or more time than patients not given the drug".

Example 3.4.2: Do non-smokers earn more than smokers?

Among Northern California's unemployed, 60 non-smokers and 29 smokers were surveyed after getting jobs, and those two samples' mean salaries were $32,000 and $24,000, respectively. Researchers posed two hypotheses:

- (Null hypothesis) H_0 : Non-smokers do NOT earn more than smokers.
- (Alternative hypothesis) H_a: Non-smokers earn more than smokers.

Through analysis of data, the researchers rejected the null hypothesis (do not earn more); the large salary difference among the two samples is unlikely explainable by chance. The researchers favored the alternative hypothesis, that non-smokers do earn more.[1] Note that the study makes no claim that smoking *causes* lower salary; for example, smokers might have another trait that causes both lower salary and smoking.

Source: Wikimedia/US Air Force, Anthony Sanchelli [2]

PARTICIPATION ACTIVITY 3.4.6: Null and alternative hypotheses.

1) Initial data from a sample of 100 people suggests that stress increases blood pressure (a positive correlation

is observed). What is the null hypothesis?

- Stress does not increase blood pressure.
- Stress reduces blood pressure.
- Stress increases blood pressure.

2) Initial data from a sample of 100 people suggests that sunlight improves moods. A null hypothesis is: Sunlight does not improve mood. What is the alternative hypothesis?

- Sunlight has no impact on mood.
- Sunlight has no impact on mood is false.
- Sunlight improves mood.

3) Initial data suggests mirrors at store entrances decrease shoplifting. Are the following reasonable null and alternative hypotheses?
Null: A small mirror at a store entrance decreases shoplifting slightly.
Alternative: A larger mirror decreases shoplifting further.

- Yes
- No

PARTICIPATION ACTIVITY | 3.4.7: Types of alternative hypothesis.

Classify each alternative hypothesis as left-tailed, right-tailed, or two-tailed.

Select the definition that matches each term

1) Right-tailed hypothesis

- The weight of a standardized part in an aircraft deviates from 15.4 kg.

- The percentage of student drivers in California that pass a drive test on the first try is less than 45%.
- A particular baseball player will hit more than 28 home runs in the 2018 season.

2) Two-tailed hypothesis

- The weight of a standardized part in an aircraft deviates from 15.4 kg.
- The percentage of student drivers in California that pass a drive test on the first try is less than 45%.
- A particular baseball player will hit more than 28 home runs in the 2018 season.

3) Left-tailed hypothesis

- The weight of a standardized part in an aircraft deviates from 15.4 kg.
- The percentage of student drivers in California that pass a drive test on the first try is less than 45%.
- A particular baseball player will hit more than 28 home runs in the 2018 season.

Reset

Comparative experiments

Sometimes a person creates hypotheses and analyzes existing data, such as examining U.S. Census Bureau data to compare employment statistics between 1970 and 2010. However, other times a person poses a hypothesis and then conducts a study to generate data to be analyzed. In one kind of study, the sample is divided into a ***treatment group*** who was given a treatment being studied (like doing daily exercise), and a ***control group*** who was not (like not doing daily exercise), with individuals randomly assigned to a group.

PARTICIPATION ACTIVITY | 3.4.8: Control and treatment groups.

A study seeks to determine whether American college freshmen students earn higher grades if those students exercise regularly. 200 students from University XYZ are selected, and randomly assigned to group A or B. Group A is required to spend 15 minutes on a treadmill every morning, while group B is required to watch a 15 minute video every morning, for the entire semester. Grades at the end of the semester are then analyzed.

Note that numerous studies do conclude that exercise improves academic performance, from young children to college students[3 4].

Select the definition that matches each term

1) Population

- 100 students using a treadmill daily.
- All American college freshmen students.
- The 200 University XYZ students.
- 100 students watching a daily video.

2) Sample

- 100 students using a treadmill daily.
- All American college freshmen students.
- The 200 University XYZ students.
- 100 students watching a daily video.

3) Control group

- 100 students using a treadmill daily.
- All American college freshmen students.
- The 200 University XYZ students.
- 100 students watching a daily video.

4) Treatment group

- 100 students using a treadmill daily.
- All American college freshmen students.

- The 200 University XYZ students.
- 100 students watching a daily video.

Reset

Test statistic

Sufficient data from samples of populations are needed to test a hypothesis. Ex: If the average salary of a sample of 50 people from San Francisco is $\$105,200$ and the average salary of a sample of 40 people from Austin is $\$98,700$, no decision can be immediately made about the null hypothesis. Factors such as the sampling method, the sample size, and the magnitude of the difference in the average salary affect whether the hypothesis should be rejected. In addition, the observed difference in average salary may be due to chance.

The central goal of hypothesis testing is to determine whether H_0 can be rejected in favor of H_a or whether H_0 fails to be rejected based on evidence in the form of sample data. A ***test statistic*** is a value calculated from sample data during hypothesis testing that measures the degree of agreement between the sample data and the null hypothesis. Mathematically, the test statistic is the difference between the estimate and the value asserted in the null hypothesis divided by the standard error.

Common hypothesis tests and associated test statistics are shown in the following table. The specific methods and formulas used to find these test statistics are introduced elsewhere.

Table 3.4.1: Hypothesis tests and associated test statistics.

Application	Hypothesis test	Test statistic
One or two population μ, p, known σ	Z-test	Z-statistic
One or two population μ, unknown σ	t-test	t-statistic
Comparing a parameter among three or more groups	ANOVA	F-statistic
Comparing categorical variables	Chi-square test	χ^2-statistic

p-value and statistical significance

Intuitively, inferences cannot be made from a trivially small difference between populations or a difference between populations in which the sample sizes are very small, as differences can be observed by chance in such cases. The concept of statistical significance formalizes the intuition of making inferences about populations from observed sample statistics. A ***statistically significant*** result differs enough from a null hypothesis that a conclusion can be inferred about the population.

In hypothesis testing, the probability of obtaining a result that is as extreme or more extreme than the data if the null hypothesis were true is known as the ***p-value***. The p-value of a result is determined from the test statistic. If the p-value is less than a specified significance level, denoted by α, then two possibilities exist.

- The null hypothesis is true and the observed data is relatively unusual with a sample statistic that is extreme simply due to chance.
- The null hypothesis is false and the alternative hypothesis provides a more reasonable explanation for the population parameter.

In most fields, $\alpha = 0.05$ is used most often as the significance level for hypothesis testing. Thus, the probability that a result with an extreme deviation from the null hypothesis is due to chance must be 5% or less for the result to be considered statistically significant.

A ***practically significant*** result implies that the magnitude of the difference between the hypothesized value of a parameter and the sample statistic is large enough to be meaningful in real life. Statistical significance does not necessarily imply practical significance of a result. Ex: A statistically significant difference in the number of employees at a company arriving to work at 8:00 am and the number of employees arriving at 7:59 am is likely of little practical significance.

PARTICIPATION ACTIVITY 3.4.9: The p-value.

1) The p-value is the same as the significance level α.

   - True
   - False

2) The same p-value can imply statistical significance in some hypothesis tests but not others.

   - True
   - False

3) Lower p-values imply greater practical significance of the result of a hypothesis test.

   - True

False

Type I and type II errors

In hypothesis testing, a researcher can decide to reject the null hypothesis or fail to reject the null hypothesis. In reality, the null hypothesis is either true or false. Thus, the researcher can make an error by rejecting a true null hypothesis or failing to reject a false null hypothesis. A ***type I error*** is the incorrect rejection of a true null hypothesis, and a ***type II error*** is the failure to reject a false null hypothesis. In other words, a type I error is a false positive and a type II error is a false negative.

The significance level of a hypothesis test, denoted α, is the probability of making a type I error. The probability of making a type II error is denoted β, and the probability $1-\beta$ of avoiding type II errors is the power of the hypothesis test.

The following animation illustrates the relationship between the truth of H_0, the decision made, and type I and type II errors.

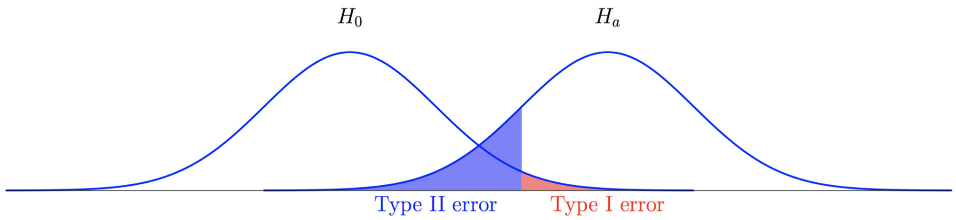
PARTICIPATION ACTIVITY | 3.4.10: Type I and type II errors.

Animation captions:

1. In hypothesis testing, the null hypothesis H_0 is either true or false.
2. Two decisions are possible. Either H_0 is rejected, or H_0 fails to be rejected.
3. Rejecting H_0 is the correct decision if H_0 is false.
4. However, rejecting H_0 if H_0 is true is known as a type I error. The probability of a type I error is the significance level of the test, denoted α.
5. Failing to reject H_0 is the correct decision if H_0 is true.
6. However, failing to reject H_0 if H_0 is false is known as a type II error. The probability of a type II error is denoted β.
7. The power of a statistical test is $1-\beta$. Ideally, a statistical test should have a low significance level (α) and high power $(1-\beta)$.

The following figure shows the regions of two normally distributed random variables corresponding to a null hypothesis and an alternative hypothesis along with the regions corresponding to type I and type II errors.

Figure 3.4.1: Regions corresponding to type I and type II errors for two normally distributed random variables corresponding to a null hypothesis and an alternative hypothesis.



PARTICIPATION ACTIVITY 3.4.11: Types of errors.

Match each scenario with the correct description.

Select the definition that matches each term

1) Type II error

- Null hypothesis: The suspect did not commit a crime.

  Actual outcome: The suspect committed the crime.

  Decision: The suspect was found not guilty.

- Null hypothesis: The suspect did not commit a crime.

  Actual outcome: The suspect committed the crime.

  Decision: The suspect was found guilty.

- Null hypothesis: The suspect did not commit a crime.

  Actual outcome: The suspect did not commit the crime.

  Decision: The suspect was found guilty.

2) No error

- Null hypothesis: The suspect did not commit a crime.

  Actual outcome: The suspect committed the crime.

  Decision: The suspect was found not guilty.

- Null hypothesis: The suspect did not commit a crime.

  Actual outcome: The suspect committed the crime.

  Decision: The suspect was found guilty.

- Null hypothesis: The suspect did not commit a crime.

  Actual outcome: The suspect did not commit the crime.

  Decision: The suspect was found guilty.

3) Type I error

- Null hypothesis: The suspect did not commit a crime.

  Actual outcome: The suspect committed the crime.

  Decision: The suspect was found not guilty.

- Null hypothesis: The suspect did not commit a crime.

  Actual outcome: The suspect committed the crime.

  Decision: The suspect was found guilty.

- Null hypothesis: The suspect did not commit a crime.

  Actual outcome: The suspect did not commit the crime.

  Decision: The suspect was found guilty.

Reset

Summary of hypothesis testing

The following are the basic steps of performing a hypothesis test.

1. *Specify the null hypothesis H_0 and the alternative hypothesis H_a.*
2. *Specify the significance level α.*
3. *Collect the data.*
4. *Calculate the test statistic and the corresponding p-value.*
5. *Compare the p-value with the significance level.*
6. *Determine whether to reject or fail to reject H_0.*

References

(*1) Judith J. Prochaska, Anne K. Michalek, Catherine Brown-Johnson, Eric J. Daza, Michael Baiocchi, Nicole Anzai, Amy Rogers, Mia Grigg, Amy Chieng. *Likelihood of Unemployed Smokers vs Nonsmokers Attaining Reemployment in a One-Year Observational Study*. JAMA Internal Medicine, 2016.

(*2) Sanchelli, Anthony. "Lighting a cigarette." Wikipedia: The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 26 Feb 2013, https://commons.wikimedia.org/wiki/File%3ALighting_a_cigarette.jpg.

(*3) Trost, Stewart G. "Active education: Physical education, physical activity and academic performance." (2007). https://folio.iupui.edu/bitstream/handle/10244/587/Active_Ed.pdf?sequence=2

(*4) Neubert, Amy Patterson. "College students working out at campus gyms get better grades." Purdue Today (2013). http://www.purdue.edu/newsroom/releases/2013/Q2/college-students-working-out-at-campus-gyms-get-better-grades.htm.

3.5 Hypothesis test for a population mean

z-test for population means

A z-test is a hypothesis test in which the z-statistic follows a normal distribution. The z-test for a population mean can be used to determine whether the population mean is the same as the hypothesized mean μ_0, assuming that the population standard deviation σ is known. When performing a hypothesis test involving the mean of a single population with a known population standard deviation, the distribution of the z-test statistic is assumed to be $N\left(\mu_0, \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$. In practice, the population standard deviation is rarely known, so a more useful test involves the t-distribution because the standard deviation of a sample can always be computed.

Conditions for performing the z-test for a population mean

- *Randomness. Data should be collected randomly by means of simple random sampling, stratified random sampling, or cluster sampling. For the examples presented in this material, this condition can be assumed.*
- *Independence. Each observation should not affect other observations. In most surveys, sampling is performed without replacement, which means that a person or an observation is not used twice. For sampling without replacement, the sample size should be less than 10% of the population size to guarantee independence.*
- *Normality. The sampling distribution should be approximately normal, which is guaranteed either when the sample size is at least 30 by the CLT or when the data's parent distribution is also normal.*

Procedure 3.5.1: Hypothesis testing for population means.

Given a randomly selected sample taken from a population with a known population standard deviation σ

- Set the null and alternative hypotheses

$$H_0 : \mu = \mu_0 H_a : \mu > \mu_0 (\text{right-tailed}) H_a : \mu < \mu_0 (\text{left-tailed}) H_a : \mu \neq \mu_0 (\text{two-tailed})$$

where μ is the population mean and μ_0 is the hypothesized population mean.

- Use statistical software to find the test-statistic

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

- Use statistical software to find the p-value that corresponds to z.
- Make a decision given a previously selected significance level α, typically 0.05.
  - If the p-value is less than the significance level, sufficient evidence exists to reject the null hypothesis H_0 in favor of the alternative hypothesis H_a.
  - If the p-value is greater than or equal to the significance level, insufficient evidence exists to reject the null hypothesis H_0.

Example 3.5.1: Mean battery life.

A popular electronics website wants to determine whether a smartphone has an 7.8 hour battery life as claimed by the manufacturer in response to user complaints of poor battery life. The website sampled 10 smartphones with a mean battery life of 7.6 hours. The population standard deviation of the battery life is $\sigma = 0.57$ hours. Does sufficient evidence exist that the battery life of the smartphone is actually lower than the manufacturer's claim at a significance level of $\alpha = 0.05$?

Solution

The null hypothesis is that the smartphone's mean battery life is $\mu_0 = 7.8$ hours. Because customers believe that the mean battery life is lower, the hypothesis test is left-tailed. That is, the alternative hypothesis is that the smartphone's mean battery life is less than 7.8 hours. Mathematically,

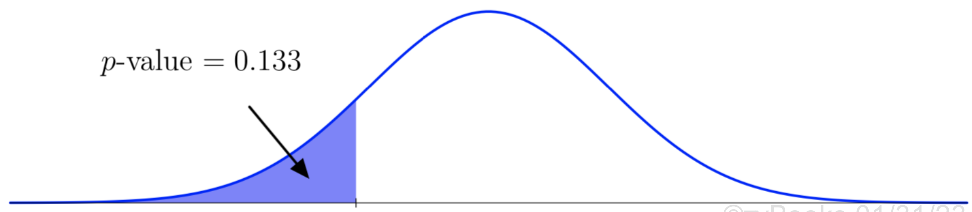
$$H_0 : \mu = 7.8 H_a : \mu < 7.8$$

Since $\bar{x} = 7.6$, $n = 10$, and $\sigma = 0.57$, the test statistic is

$$z = \frac{7.6 - 7.8}{\frac{0.57}{\sqrt{10}}} \approx -1.110$$

The p-value is

$$p\text{-value} = P(z \leq -1.110) \approx 0.133$$



Since the p-value is greater than the significance level $\alpha = 0.05$, insufficient evidence exists to support the hypothesis that the mean battery life of the smartphone is less than the manufacturer's claim.

Analysis

Although the mean battery life of the 10 sampled smartphones is less than the manufacturer's claim, the lower mean could have occurred due to chance. However, the probability that the sample mean battery life is at most 7.6 hours is 13.3%, which is much higher than the probability of incorrectly rejecting the manufacturer's claim that the smartphone has a mean battery life of 7.8 hours. Thus, the lower sample mean can be most likely be attributed to chance.

Example 3.5.2: Carry-on baggage volume.

An airline conducts a study on the volume of carry-on luggage. Many flight attendants believe that the average volume of carry-on luggage passengers bring onto airplanes exceeds the allowed 1.6 ft^3. The airline samples 73 pieces of carry-on luggage with an average volume of 1.7 ft^3. Does sufficient evidence exist that passengers routinely bring luggage that exceeds the maximum allowed volume given a population standard deviation of $\sigma = 0.29$ ft^3 at a significance level of $\alpha = 0.10$?

Solution

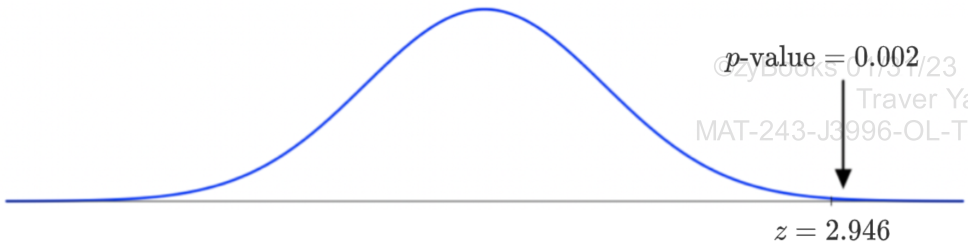
The null hypothesis is that the mean carry-on luggage volume is at most 1.6 ft^3. Since flight attendants believe that the luggage passengers carry routinely exceeds the limit, the hypothesis test is right-tailed. That is, the alternative hypothesis is that the mean volume of carry-on luggage is greater than 1.6 ft^3. Mathematically,

$$H_0 : \mu = 1.6 H_a : \mu > 1.6$$

Since $\bar{x} = 1.7$, $n = 73$, and $\sigma = 0.29$, the test statistic is

$$z = \frac{1.7 - 1.6}{\frac{0.29}{\sqrt{73}}} \approx 2.946$$

The p-value is $P(z \geq 2.946) \approx 0.002$.



Since the p-value is less than the significance level $\alpha = 0.1$, sufficient evidence exists to support the hypothesis that passengers routinely exceed the allowed volume limit for carry-on luggage.

Analysis

The maximum carry-on allowance of 1.6 ft^3 suggests that the null hypothesis should be $\mu \leq 1.6$. However, the null hypothesis is defined as a claim of equality. In addition, rejecting a null hypothesis of $\mu = 1.6$ implies than a population mean of less than 1.6 will also be rejected. Thus, the convention is to always express the null hypothesis as a statement of equality and so $H_0 : \mu = 1.6$.

Although the mean volume of the 73 sampled carry-on luggage is greater than the manufacturer's claim, the greater mean volume could have occurred due to chance. However, the probability that the sample mean luggage volume is at least 1.6 ft^3 is 0.2%, which is much lower than the probability of incorrectly rejecting the claim that the mean is less than 1.6 ft^3. Thus, the greater sample mean cannot be attributed to chance.

PARTICIPATION ACTIVITY 3.5.1: Intelligence quotient.

Intelligence quotient (IQ) test scores are believed to have a mean of 100 and a population standard deviation of 15. In a random sample of 36 students in a high school, the mean IQ test score is 105. Researchers claim that the mean IQ test scores at this high school is statistically higher than 100.

1) What is the hypothesized mean?

Check Show answer

2) Is the hypothesis test two-tailed, right-tailed, or left-tailed? Type as: two-tailed, left-tailed, or right-tailed

Check Show answer

3) What is the z-score?

Check Show answer

4) What is the p-value? Type as: #.###

Check Show answer

5) Should the null hypothesis that the IQ test scores of students at the high school is equal to 100 be rejected at the $\alpha = 0.05$ significance level? Type as: yes or no

Check Show answer

Python-Function 3.5.1: ztest(x1, value).

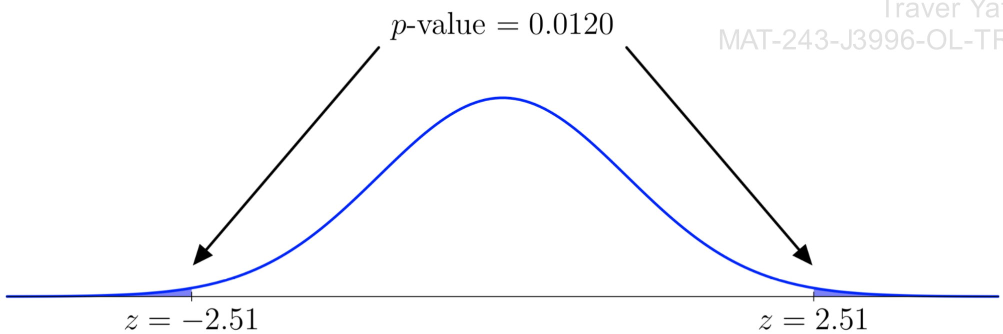
The `ztest(x1, value)` function is used to perform a one-sample z-test for means. The function requires the statsmodels.stats.weightstats library to be imported, and takes two inputs. The first input x1 is an array and the second input value is the hypothesized value of the population mean.

The function returns the z-score and the two-tailed p-value.

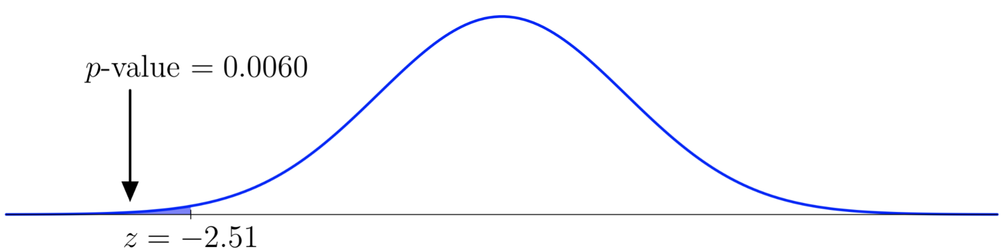
```
from statsmodels.stats.weightstats import ztest
import pandas as pd
scores = pd.read_csv('http://data-analytics.zybooks.com/ExamScores.csv')
print(ztest(x1 = scores['Exam1'],  value = 86))
```

```
(-2.5113146627890988, 0.012028242796839027)
```

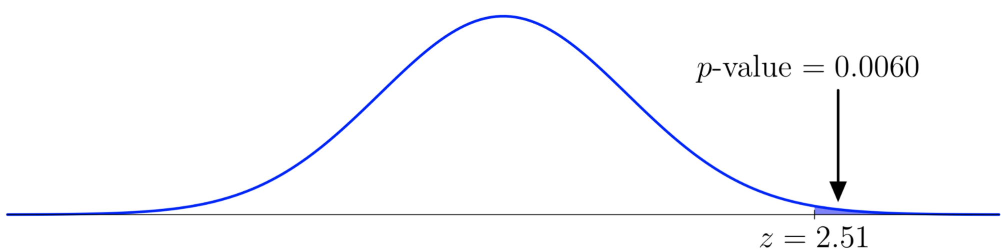
The graph for the two-tailed z-test corresponding to the output above is shown below.



The one-sided p-value is the two-sided p-value divided by 2. The graph for the corresponding left-tailed z-test is shown below.



The graph for the corresponding right-tailed z-test is shown below.



Run example

Student's t-test

Because the population standard deviation is rarely known, the t-test is commonly used to compare the observed sample mean to a hypothesized mean. The following conditions must be met to use the t-distribution.

Conditions for the t-distribution

- *Randomness. Data should be collected randomly by means of simple random sampling, stratified random sampling, or cluster sampling. For the examples presented in this material, this condition can be assumed.*
- *Independence. Each observation should not affect other observations.*
- *Normality. The t-distribution can be used if the underlying population distribution is approximately normal. In most cases, the normality of the underlying distribution cannot be determined. Thus, the sample size determines whether the t-distribution can be used.*
  - *If the sample size is less than 15, the t-distribution can be used if the data is not skewed or no outliers are present.*
  - *If the sample size is between 15 and 30, the t-distribution can be used even if the data is mildly skewed. If outliers can be removed, then using the t-distribution is also appropriate.*
  - *The t-distribution can always be used when the sample size is sufficiently large and extreme outliers can be removed.*

Procedure 3.5.2: The t-test.

1. Set the null and alternative hypotheses

$$H_0: \mu = \mu_0 H_a: \mu > \mu_0 \text{ (right-tailed)} H_a: \mu < \mu_0 \text{ (left-tailed)} H_a: \mu \neq \mu_0 \text{ (two-tailed)}$$

   where μ is the population mean and μ_0 is the hypothesized population mean.

2. Use statistical software to find the t-statistic and the degrees of freedom df:

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

   and

$$df = n - 1$$

3. Use statistical software to find the p-value that corresponds to t.
4. Make a decision given a previously selected significance level α, typically 0.05.
   - If the p-value is less than the significance level, sufficient evidence exists to reject the null hypothesis H_0 in favor of the alternative hypothesis H_a.
   - If the p-value is greater than or equal to the significance level, insufficient evidence exists to reject the null hypothesis H_0.

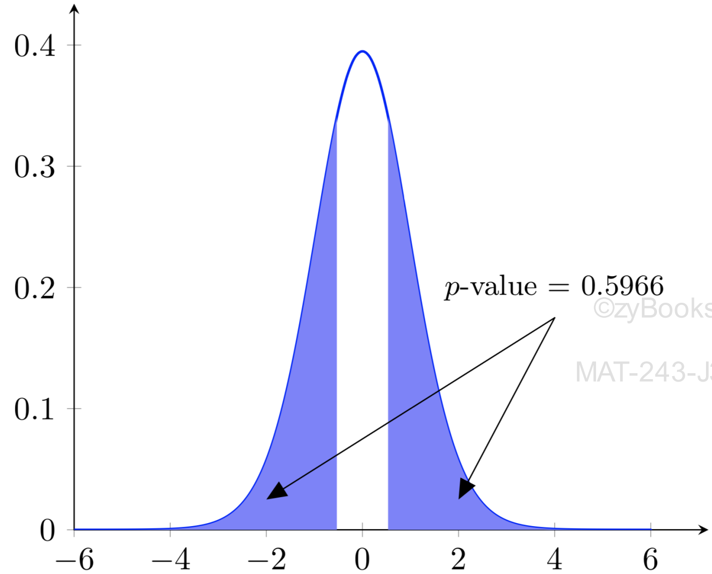
Python-Function 3.5.2: scipy.stats.ttest_1samp(a, popmean).

The `scipy.stats.ttest_1samp(a, popmean)` function takes in an array or a column of a DataFrame and the hypothesized population mean as inputs and gives the t-statistic and two-tailed p-value as outputs.

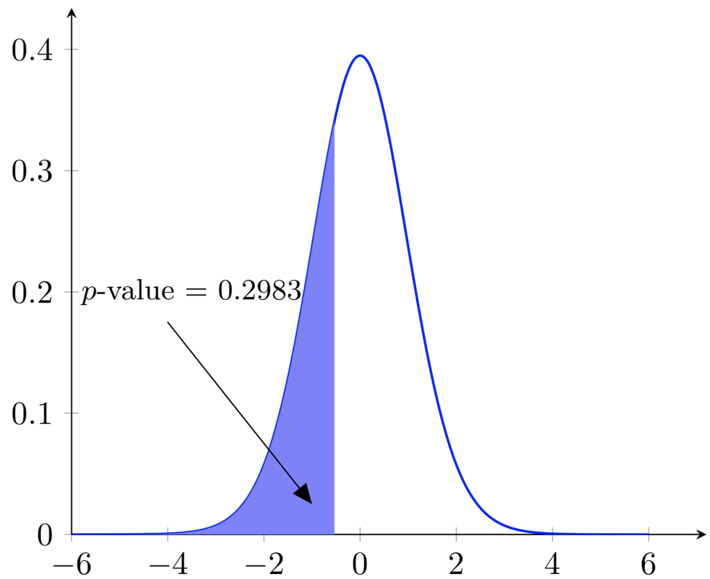
```
import pandas as pd
import scipy.stats as st
scores = pd.read_csv('http://data-analytics.zybooks.com/ExamScores.csv')
print(st.ttest_1samp(scores['Exam1'], 82))
```

```
Ttest_1sampResult(statistic=0.53270311028859929, pvalue=0.59664679344599247)
```

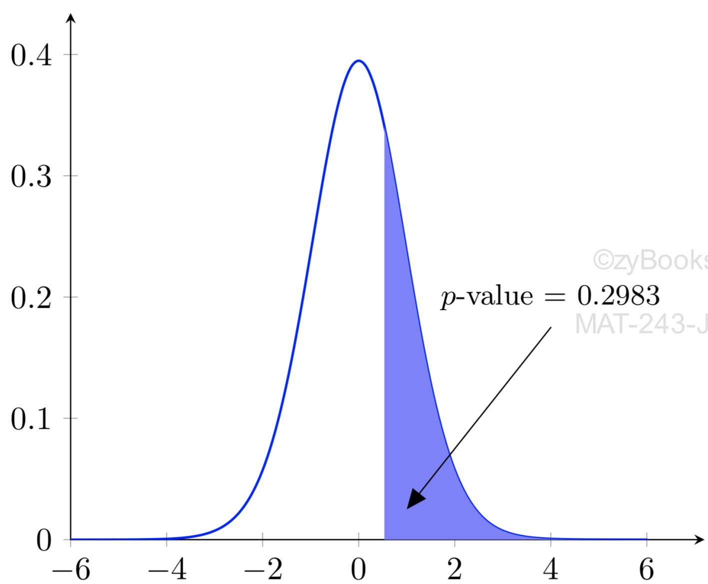

The graph for the two-tailed t-test corresponding to the output above is shown below.



The one-sided p-value is the two-sided p-value divided by 2. The graph for the corresponding left-tailed t-test is shown below.



The graph for the corresponding right-tailed t-test is shown below.



Run example

Example 3.5.3: Circumference of basketballs.

The mean circumference of basketballs produced in a manufacturing facility is supposed to be 29 inches. A random sample of 25 basketballs has a mean of 29.1 inches with a sample standard deviation of 0.217 inches. The quality control supervisor claims that the mean circumference of the basketballs produced in the facility is different from 29 inches. At the $\alpha = 0.01$ significance level, does sufficient evidence exist to support the supervisor's claim?

Solution

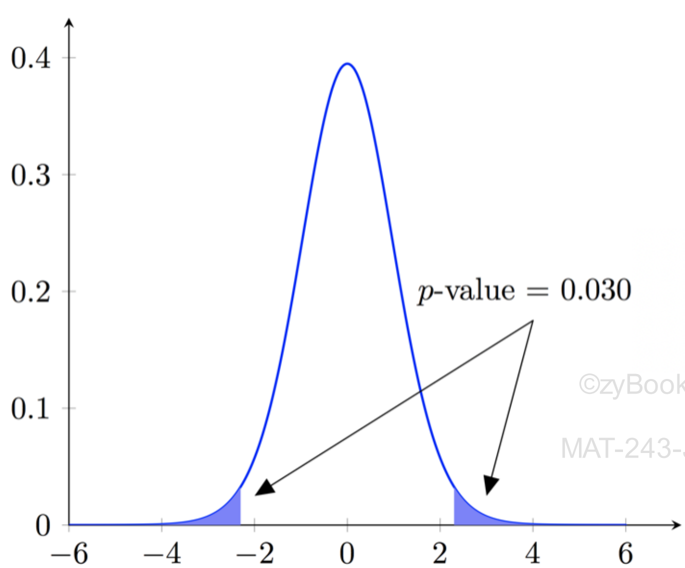
The null hypothesis is that the mean circumference of basketballs produced in the facility is 29 inches. Since the supervisor's claim is that the mean circumference is different from 29 inches, the hypothesis test is two-tailed. Mathematically,

$$H_0 : \mu = 29 H_a : \mu \neq 29$$

Since $\bar{x} = 29.1$, $n = 25$, and $s = 0.217$, the test statistic is

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{29.1 - 29}{\frac{0.217}{\sqrt{25}}} \approx 2.304$$

The two-tailed p-value is $P(t \leq -2.30 \text{ or } t \geq 2.30) = 0.030$.



Since the p-value is greater than the significance level $(0.030 > 0.01)$, insufficient evidence exists to support the claim that the mean circumference of basketballs produced in the facility is different than 29 inches.

Analysis

Because the population standard deviation σ is unknown, the t-test is used instead of the z-test.

PARTICIPATION ACTIVITY 3.5.2: Exam scores.

Using the ExamScores dataset, a teacher found that the mean score for Exam2 is 79.4, which is below the expected mean of 83. At the $\alpha = 0.05$ significance level, does sufficient evidence exist that the mean score of the class is lower than the expected mean? Use the output below.

```
import pandas as pd
import scipy.stats as st
scores = pd.read_csv('http://data-analytics.zybooks.com/ExamScores.csv')
print(st.stats.ttest_1samp(scores['Exam2'], 83))
```

```
Ttest_1sampResult(statistic=-1.7760577369106121, pvalue=0.081932948675890918)
```

Run example

1) What is the null hypothesis H_0?

- $\bar{x} = 79.4$
- $\mu = 83$
- $t = -1.776$

2) What is the alternative hypothesis H_a?

- $\mu \neq 83$
- $\mu < 83$
- $\mu > 83$

3) What is the p-value?

- 0.082
- 0.041

4) What is the conclusion for the t-test?

- Reject H_0
- Fail to reject H_0

CHALLENGE ACTIVITY 3.5.1: Hypothesis test for a population mean.

456500.2535406.qx3zqy7

3.6 Hypothesis test for a population proportion

The z-test can also be used to determine whether the population proportion is the same as the hypothesized proportion p_0. When performing a hypothesis test involving the proportion of a single population, the distribution of the z-test statistic is assumed to be $N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$.

The conditions that must be satisfied are similar to those of the z-test for a population mean. However, to satisfy the normality condition, $np_0 \geq 5$ and $n(1-p_0) \geq 5$ where n is the sample size and p_0 is the hypothesized proportion.

Procedure 3.6.1: Hypothesis testing for population proportion.

Given a randomly selected sample taken from a population

1. Set the null and alternative hypotheses

$$H_0: p = p_0 H_a: p > p_0 \text{ (right-tailed)} H_a: p < p_0 \text{ (left-tailed)} H_a: p \neq p_0 \text{ (two-tailed)}$$

   where p is the population proportion and p_0 is the hypothesized population proportion.

2. Use statistical software to find the test-statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

3. Use statistical software to find the p-value that corresponds to z.
4. Make a decision given a previously selected significance level α, typically 0.05.
   - If the p-value is less than the significance level, sufficient evidence exists to reject the null hypothesis H_0 in favor of the alternative hypothesis H_a.
   - If the p-value is greater than or equal to the significance level, insufficient evidence exists to reject the null hypothesis H_0.

Example 3.6.1: Human sex ratio.

The human sex ratio is the ratio of the number of males to the number of females within a certain age group. According to a 2002 study on sex ratios[1], the expected ratio of males to females is 106 to 100 or 0.515. Because of cultural norms and national health policies, some nations may have a much higher or much lower sex ratio. In a random sample of 189 people, 85 people are males. Does sufficient evidence exist that the sex ratio of males to females in the population is different than expected at the $\alpha = 0.05$ significance level?

Solution

Since the dataset contains binary categorical data, the hypothesis test involves population proportion rather than population mean. The null hypothesis is that the sex ratio of males to females is 0.515. Since the question asks if the ratio is different, the hypothesis test is two-tailed with an alternative hypothesis that the sex ratio of males to females is not 0.515. Mathematically,

$$H_0 : p = 0.515 H_a : p \neq 0.515$$

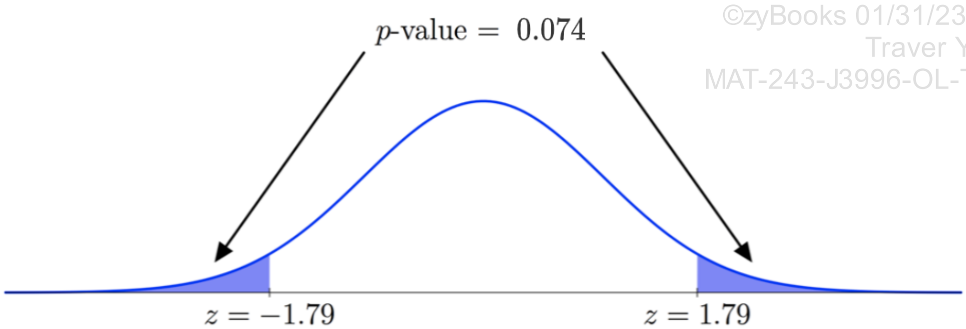
85 of 189 people in the sample are males, so the sample proportion is $\hat{p} = \frac{85}{189} \approx 0.450$. Since $p_0 = 0.515$ and $n = 189$, the test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.45 - 0.515}{\sqrt{\frac{0.515(1-0.515)}{189}}} \approx -1.79$$

The p-value is found using statistical software or a table.

$$P(z \leq -1.79 \text{ or } z \geq 1.79) \approx 0.074$$

.



Since the p-value is greater than the significance level $\alpha = 0.05$, insufficient evidence exists to support the claim that the sex ratio in the population from which the same is drawn is different than the expected sex ratio of 0.515.

Analysis

The p-value of a two-tailed test is twice that of a one-tailed test. Since the question is framed as a difference from the expected proportion, the area above $z = 1.79$ is included in the p-value. Note that had the question stated that the sex ratio of males to females is lower instead of different, then the results would have been that sufficient evidence exists to conclude that sex ratio of males to females is lower than the expected ratio of 0.515 because the p-value for a one-tailed test is 0.037.

Example 3.6.2: Customer satisfaction.

After a series of scandals, the percentage of satisfied customers of an airline dipped to an all time low of 47%. As a result, aggressive changes were implemented to improve customer experience and a public relations firm was hired to rehabilitate the airline's image. To determine if their efforts are successful, a survey was conducted to determine if customers were satisfied or unsatisfied with the airline. 132 of the 240 respondents answered that their experience was satisfactory. Does sufficient evidence exist that the company's efforts were succeeding at the $\alpha = 0.05$ significance level? That is, does statistically significant evidence exist that suggests that the proportion of satisfied customers is higher?

Solution

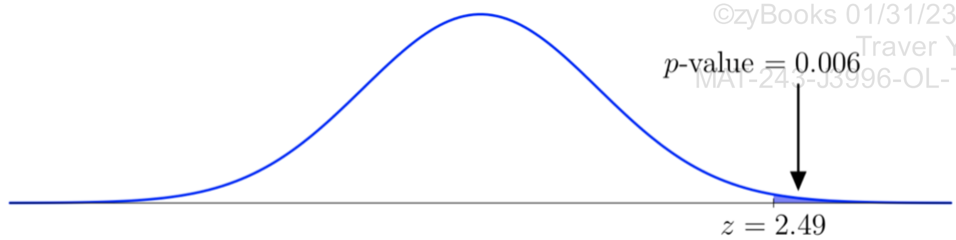
Since the survey asked if the customers were satisfied or unsatisfied, the dataset contains binary categorical data. Thus, the hypothesis test involves population proportion instead of population mean. The null hypothesis is that the customer satisfaction proportion holds steady. Since the board of directors want to know if the changes are successful in increasing customer satisfaction, the hypothesis test is right-tailed with an alternative hypothesis that the percentage of satisfied customers is higher. Mathematically,

$$H_0: p = 0.47 H_a: p > 0.47$$

132 of the 240 respondents viewed their experience as satisfactory, so the sample proportion is $\hat{p} = \frac{132}{240} = 0.55$. Since $p_0 = 0.47$ and $n = 240$, the test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.55 - 0.47}{\sqrt{\frac{0.47(1-0.47)}{240}}} \approx 2.483$$

The p-value is $P(z \geq 2.483) \approx 0.006$.



Since the p-value is less than the significance level $\alpha = 0.05$, sufficient evidence exists to support the claim that the customer satisfaction is higher than the all time low of 47%.

Python-Function 3.6.1: proportions_ztest(count, nobs, value, prop_var).

The `proportions_ztest(count, nobs, value, prop_var = value)` function is used to perform a one-sample z-test for proportions. The function requires the statsmodels.stats.proportion library to be imported, and takes four inputs. The first input count is the number of observations meeting some condition, the second input nobs is the total number of observations, the third input is the hypothesized value of the population proportion, and the fourth is the hypothesized value of the population proportion which is used to calculate the variance of the estimate.

The function returns the z-score and the two-tailed p-value.

```
from statsmodels.stats.proportion import proportions_ztest
counts = 31
nobs = 50
value = 0.50
print(proportions_ztest(counts, nobs, value, prop_var = value))
```

```
(1.697056274847714, 0.08968602177036457)
```


Run example

PARTICIPATION ACTIVITY 3.6.1: Exam scores proportion.

Using the ExamScores dataset, a teacher found that 31 of 50 students scored over 80 in Exam1, which is over the expected proportion of 0.5. At the $\alpha = 0.01$ significance level, does sufficient evidence exist that the proportion of scores over 80 is greater than 0.5? Use the output below.

```
print(proportions_ztest(31, 50, 0.5, prop_var =
0.5))
```

```
(1.697056274847714,
0.08968602177036457)
```


1) What is the null hypothesis H_0?

- $\hat{p} = 0.62$
- $p = 0.5$
- $p = 0.62$

2) What is the alternative hypothesis H_a?

- $p \neq 0.5$
- $p < 0.5$
- $p > 0.5$

3) What is the p-value?

- 0.08969
- 0.04485

4) What is the conclusion for the t-test?

- Reject H_0
- Fail to reject H_0

CHALLENGE ACTIVITY 3.6.1: Hypothesis test for a population proportion.

456500.2535406.qx3zqy7

References

(*1) Grech, Victor, et al. "Unexplained differences in sex ratios at birth in Europe and North America." *BMJ*. 324(7344). 27 April 2002. www.ncbi.nlm.nih.gov/pmc/articles/PMC102777/