

What is the process for conducting a two-tailed hypothesis test of a mean in statistics?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the process for conducting a two-tailed hypothesis test of a mean in statistics?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159337>

The process for conducting a two-tailed hypothesis test of a mean in statistics involves the following steps:

1. Determine the null and alternative hypotheses: The null hypothesis (H_0) states that there is no significant difference between the mean of the population and a specified value, while the alternative hypothesis (H_a) states that there is a significant difference.
2. Set the significance level: This is the probability of rejecting the null hypothesis when it is actually true. This is often denoted as alpha (α) and is typically set at 0.05 or 0.01.
3. Collect the sample data: A random sample is taken from the population under study and the sample mean, sample standard deviation, and sample size are calculated.
4. Calculate the test statistic: The test statistic is a measure of how far the sample mean is from the hypothesized population mean. In a two-tailed test, the test statistic will be compared to a critical value from the t-distribution.
5. Determine the critical value: The critical value is the value on the t-distribution that corresponds to the chosen significance level and degrees of freedom ($n-1$).
6. Compare the test statistic to the critical value: If the absolute value of the test statistic is greater than the critical value, the null hypothesis can be rejected in favor of the alternative hypothesis.
7. Interpret the results: If the null hypothesis is rejected, it can be concluded that there is a significant difference between the sample mean and the hypothesized population mean. If the null hypothesis is not rejected, there is insufficient evidence to conclude a significant difference.
8. Calculate the p-value: The p-value is the probability of observing a test statistic at least as extreme as the one calculated, assuming the null hypothesis is true. If the p-value is less than the chosen significance level, the null hypothesis can be rejected.
9. Draw a conclusion: Based on the results of the hypothesis test, a conclusion can be made regarding the significance of the difference between the sample mean and the hypothesized population mean.

Statistics - Hypothesis Testing a Mean (Two Tailed)

A population mean is an average of value a population.

Hypothesis tests are used to check a claim about the size of that population mean.

Hypothesis Testing a Mean

The following steps are used for a hypothesis test:

Check the conditions
Define the claims
Decide the significance level
Calculate the test statistic
Conclusion

For example:

Population: Nobel Prize winners
Category: Age when they received the prize.

And we want to check the claim:

"The average age of Nobel Prize winners when they received the prize is **not** 60"

By taking a sample of 30 randomly selected Nobel Prize winners we could find that:

The mean age in the sample (\bar{x}) is 62.1
The standard deviation of age in the sample (s) is 13.46

From this sample data we check the claim with the steps below.

1. Checking the Conditions

The conditions for calculating a confidence interval for a proportion are:

The sample is randomly selected
And either:

The population data is normally distributed
Sample size is large enough

A moderately large sample size, like 30, is typically large enough.

In the example, the sample size was 30 and it was randomly selected, so the conditions are fulfilled.

Note: Checking if the data is normally distributed can be done with specialized statistical tests.

2. Defining the Claims

We need to define a **null hypothesis** (H_0) and an **alternative hypothesis** (H_1) based on the claim we are checking.

The claim was:

"The average age of Nobel Prize winners when they received the prize is **not** 60"

In this case, the **parameter** is the mean age of Nobel Prize winners when they received the prize (μ).

The null and alternative hypothesis are then:

Null hypothesis: The average age was 60.

Alternative hypothesis: The average age is **not** 60.

Which can be expressed with symbols as:

(H_0): ($\mu = 60$)

(H_1): ($\mu \neq 60$)

This is a '**two-tailed**' test, because the alternative hypothesis claims that the proportion is **different** from the null hypothesis.

If the data supports the alternative hypothesis, we **reject** the null hypothesis and **accept** the alternative hypothesis.

3. Deciding the Significance Level

The significance level (α) is the **uncertainty** we accept when rejecting the null hypothesis in a hypothesis test.

The significance level is a percentage probability of accidentally making the wrong conclusion.

Typical significance levels are:

($\alpha = 0.1$) (10%) ($\alpha = 0.05$) (5%) ($\alpha = 0.01$) (1%)

A lower significance level means that the evidence in the data needs to be stronger to reject the null hypothesis.

There is no "correct" significance level - it only states the uncertainty of the conclusion.

Note: A 5% significance level means that when we reject a null hypothesis:

We expect to reject a **true** null hypothesis 5 out of 100 times.

4. Calculating the Test Statistic

The test statistic is used to decide the outcome of the hypothesis test.

The test statistic is a standardized value calculated from the sample.

The formula for the test statistic (TS) of a population mean is:

$$\frac{\bar{x} - \mu}{s} \cdot \sqrt{n}$$

$(\bar{x} - \mu)$ is the **difference** between the **sample** mean ((\bar{x})) and the claimed **population** mean ((μ)).

(s) is the sample standard deviation.

(n) is the sample size.

In our example:

The claimed ((H_0)) population mean ((μ)) was (60)

The sample mean ((\bar{x})) was (62.1)

The sample standard deviation ((s)) was (13.46)

The sample size ((n)) was (30)

So the test statistic (TS) is then:

$$\frac{62.1 - 60}{13.46} \cdot \sqrt{30} = \frac{2.1}{13.46} \cdot \sqrt{30} \approx 0.156 \cdot 5.477 = \underline{0.855}$$

You can also calculate the test statistic using programming language functions:

Example

With Python use the scipy and math libraries to calculate the test statistic.

```
import scipy.stats as stats
```

```
import math
```

```
# Specify the sample mean (x_bar), the sample standard deviation (s), the mean claimed in the null-hypothesis (mu_null), and the sample size (n)
```

```
x_bar = 62.1
```

```
s = 13.46
```

```
mu_null = 60

n = 30

# Calculate and print the test statistic

print((x_bar - mu_null)/(s/math.sqrt(n)))
```

Example

With R use built-in math and statistics functions to calculate the test statistic.

Specify the sample mean (x_bar), the sample standard deviation (s), the mean claimed in the null-hypothesis (mu_null), and the sample size (n)

```
x_bar <- 62.1

s <- 13.46

mu_null <- 60

n <- 30

# Output the test statistic

(x_bar - mu_null)/(s/sqrt(n))
```

5. Concluding

There are two main approaches for making the conclusion of a hypothesis test:

The **critical value** approach compares the test statistic with the critical value of the significance level. The **P-value** approach compares the P-value of the test statistic and with the significance level.

Note: The two approaches are only different in how they present the conclusion.

The Critical Value Approach

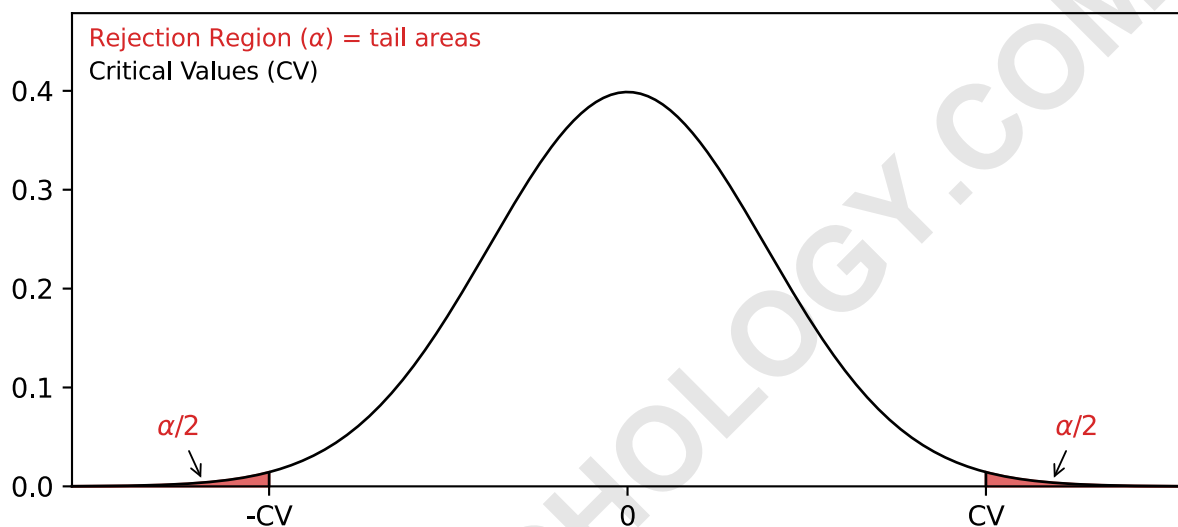
For the critical value approach we need to find the **critical value** (CV) of the significance level ((alpha)).

For a population mean test, the critical value (CV) is a **T-value** from a student's t-distribution.

This critical T-value (CV) defines the **rejection region** for the test.

The rejection region is an area of probability in the tails of the standard normal distribution.

Because the claim is that the population proportion is **different** from 60, the rejection region is split into both the left and right tail:



The size of the rejection region is decided by the significance level (α).

The student's t-distribution is adjusted for the uncertainty from smaller samples.

This adjustment is called degrees of freedom (df), which is the sample size ($(n) - 1$)

In this case the degrees of freedom (df) is: $(30 - 1 = \underline{29})$

Choosing a significance level (α) of 0.05, or 5%, we can find the critical T-value from a T-table, or with a programming language function:

Note: Because this is a two-tailed test the tail area (α) needs to be split in half (divided by 2).

Example

With Python use the Scipy Stats library `t.ppf()` function find the T-Value for an $(\alpha)/2 = 0.025$ at 29 degrees of freedom (df).

```
import scipy.stats as stats
```

```
print(stats.t.ppf(0.025, 29))
```

Example

With R use the built-in `qt()` function to find the t-value for an $(\alpha)/2 = 0.025$ at 29 degrees of freedom (df).

```
qt(0.025, 29)
```

Using either method we can find that the critical T-Value is (approx -2.045)

For a **two-tailed** test we need to check if the test statistic (TS) is **smaller** than the negative critical value (-CV), **or bigger** than the positive critical value (CV).

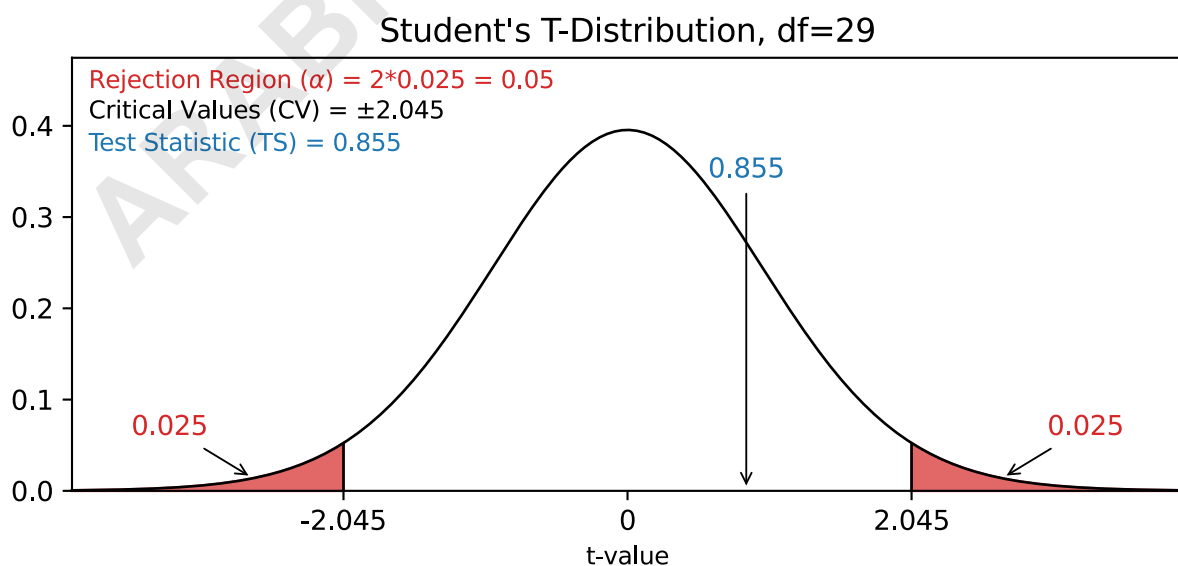
If the test statistic is smaller than the **negative** critical value, the test statistic is in the **rejection region**.

If the test statistic is bigger than the **positive** critical value, the test statistic is in the **rejection region**.

When the test statistic is in the rejection region, we **reject** the null hypothesis (H_0).

Here, the test statistic (TS) was (approx 0.855) and the critical value was (approx -2.045)

Here is an illustration of this test in a graph:



Since the test statistic is **between** the critical values we **keep** the null hypothesis.

This means that the sample data does not support the alternative hypothesis.

And we can summarize the conclusion stating:

The sample data does **not** support the claim that "The average age of Nobel Prize winners when they received the prize is not 60" at a **5% significance level**.

The P-Value Approach

For the P-value approach we need to find the **P-value** of the test statistic (TS).

If the P-value is **smaller** than the significance level ((alpha)), we **reject** the null hypothesis ((H_0)).

The test statistic was found to be (approx 0.855)

For a population proportion test, the test statistic is a T-Value from a student's t-distribution.

Because this is a **two-tailed** test, we need to find the P-value of a T-value **bigger** than 0.855 and **multiply it by 2**.

The student's t-distribution is adjusted according to degrees of freedom (df), which is the sample size ((30) - 1 = 29)

We can find the P-value using a T-table, or with a programming language function:

Example

With Python use the Scipy Stats library `t.cdf()` function find the P-value of a T-value bigger than 0.855 for a two tailed test at 29 degrees of freedom (df):

```
import scipy.stats as stats
```

```
print(2*(1-stats.t.cdf(0.855, 29)))
```

Example

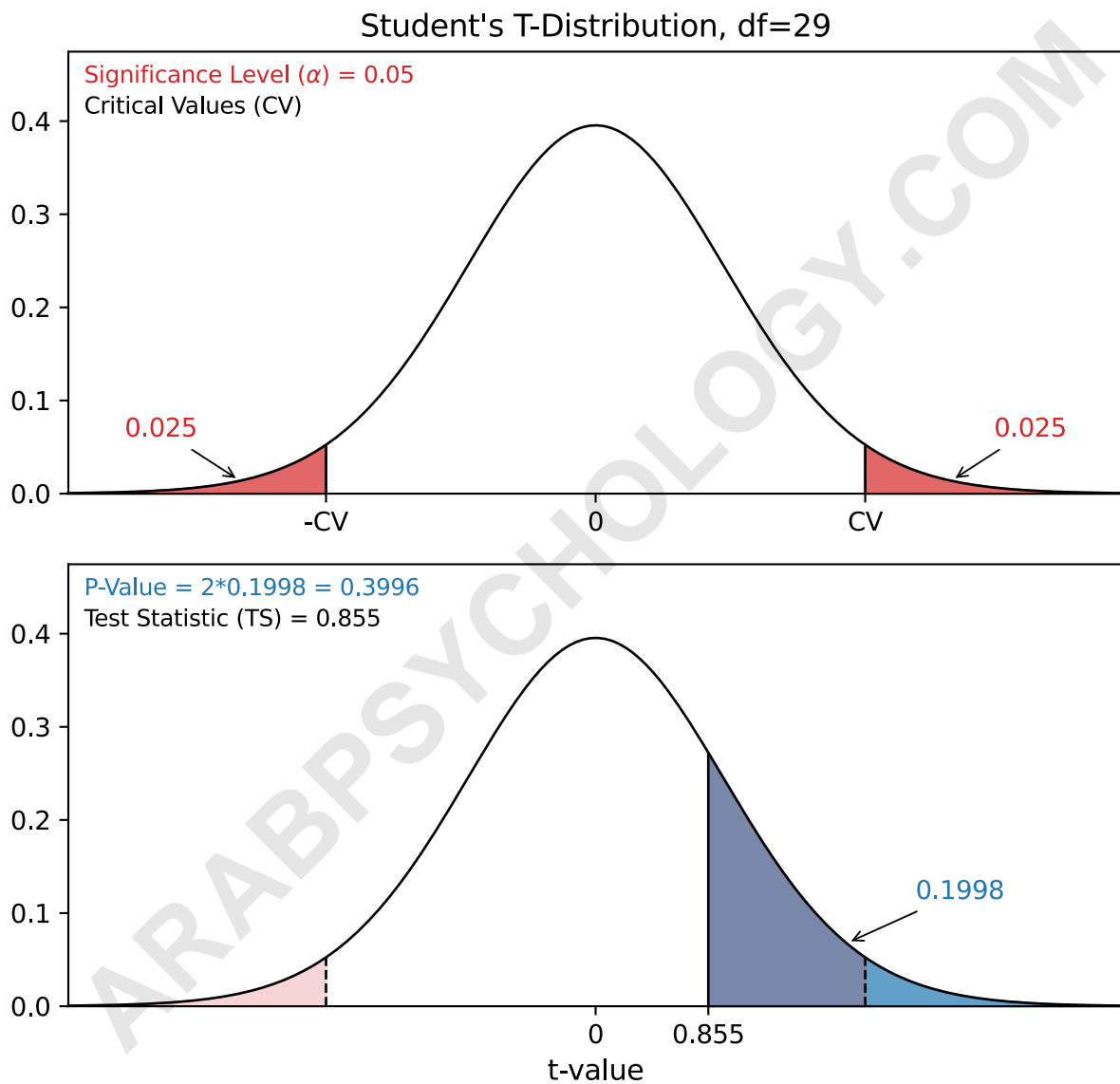
With R use the built-in `pt()` function find the P-value of a T-Value bigger than 0.855 for a two tailed test at 29 degrees of freedom (df):

```
2*(1-pt(0.855, 29))
```

Using either method we can find that the P-value is (approx 0.3996)

This tells us that the significance level (α) would need to be smaller 0.3996, or 39.96%, to **reject** the null hypothesis.

Here is an illustration of this test in a graph:



This P-value is **bigger** than any of the common significance levels (10%, 5%, 1%).

So the null hypothesis is **kept** at all of these significance levels.

And we can summarize the conclusion stating:

The sample data does **not** support the claim that "The average age of Nobel Prize winners when they received the prize is not 60" at a **10%, 5%, or 1% significance level**.

Calculating a P-Value for a Hypothesis Test with Programming

Many programming languages can calculate the P-value to decide outcome of a hypothesis test.

Using software and programming to calculate statistics is more common for bigger sets of data, as calculating manually becomes difficult.

The P-value calculated here will tell us the **lowest possible significance level** where the null-hypothesis can be rejected.

Example

With Python use the `scipy` and `math` libraries to calculate the P-value for a two tailed hypothesis test for a mean.

Here, the sample size is 30, the sample mean is 62.1, the sample standard deviation is 13.46, and the test is for a mean different from 60.

```
import scipy.stats as stats
```

```
import math
```

```
# Specify the sample mean (x_bar), the sample standard deviation (s), the mean claimed in the null-hypothesis (mu_null), and the sample size (n)
```

```
x_bar = 62.1
```

```
s = 13.46
```

```
mu_null = 60
```

```
n = 30
```

```
# Calculate the test statistic
```

```
test_stat = (x_bar - mu_null)/(s/math.sqrt(n))
```

```
# Output the p-value of the test statistic (two tailed test)
```

```
print(2*(1-stats.t.cdf(test_stat, n-1)))
```

Example

With R use built-in math and statistics functions find the P-value for a two tailed hypothesis test for a mean.

Here, the sample size is 30, the sample mean is 62.1, the sample standard deviation is 13.46, and the test is for a mean different from 60.

```
# Specify the sample mean (x_bar), the sample standard deviation (s), the mean claimed in the null-hypothesis (mu_null), and the sample size (n)
```

```
x_bar <- 62.1
```

```
s <- 13.46
```

```
mu_null <- 60
```

```
n <- 30
```

```
# Calculate the test statistic
```

```
test_stat = (x_bar - mu_null)/(s/sqrt(n))
```

```
# P-value the p-value of the test statistic (two tailed test)
```

```
2*(1-pt(test_stat, n-1))
```

Left-Tailed and Two-Tailed Tests

This was an example of a **left** tailed test, where the alternative hypothesis claimed that parameter is **smaller** than the null hypothesis claim.

You can check out an equivalent step-by-step guide for other types here:

Right-Tailed Test
Two-Tailed Test

★+1 W3schools PathfinderTrack your progress - it's free!

Log in

Sign Up