Starting Off

Imagine that Trump sends a tweet claiming that his approval ratings are 55% and that everyone thinks he is doing a good job.

He doesn’t provide evidence from any poll to support his tweet. You saw a previous poll that estimated his approval ratings to be 43%. How can you prove that the Trump’s true approval rating is actually below 55%?

Hypothesis Testing

Data Science Immersive

FLATIRON SCHOOL

Introduction to Inferences

There are two types of statistical inferences:

1. Estimation - Use information from the sample to estimate (or predict) the parameter of interest.

Using the result of a poll about the president's current approval rating to estimate (or predict) his or her true current approval rating nationwide.

2. Statistical Test - Use information from the sample to determine whether a certain statement about the parameter of interest is true.

The president’s job approval rating is above 50%.

When Normal Distribution Breaks Down

In the previous lesson, we learned that if the population is normal with mean μ and standard deviation, σ, then the distribution of the sample mean will be Normal with mean μ and standard error $\frac{\sigma}{\sqrt{n}}$.

What do we do when the population standard deviation is not know?

Recall that if X comes from a normal distribution with mean, μ, and variance, σ^2, or if $n \geq 30$, then the sampling distribution will be approximately normal with mean μ and standard error, $SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$

You must use the T-Distribution!!

When Normal Distribution Breaks Down

What if the conditions are not met?

What will you do if you cannot use the t-interval? What do we do when the above conditions are not satisfied?

1. If you do not know if the distribution comes from a normally distributed population and the sample size is small (i.e $n<30$), you can use the Normal Probability Plot to check if the data come from a normal distribution.

2. You may want to consider what is known as nonparametric statistical methods. A procedure such as the one-sample Wilcoxon procedure.

Student's or T-Distribution

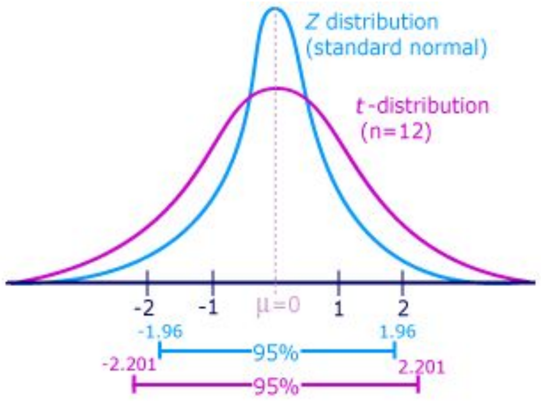
- William Sealy Gosset, a statistician at Guinness Brewing Company, was running experiments to determine the highest yielding strains of barley

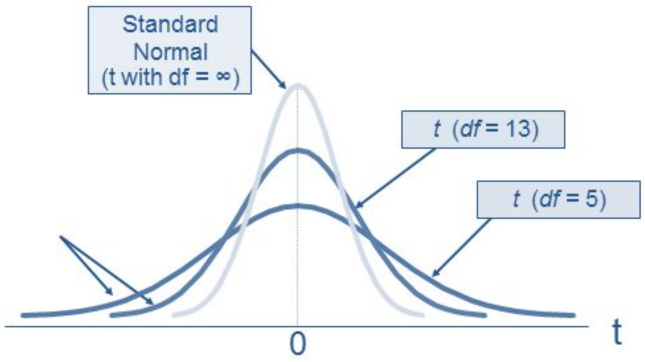
- Published a paper detailing the t-distribution under the pseudonym "Student" because Guinness had a policy that its employees could not publish research

DF = degrees of freedom = *n-1*

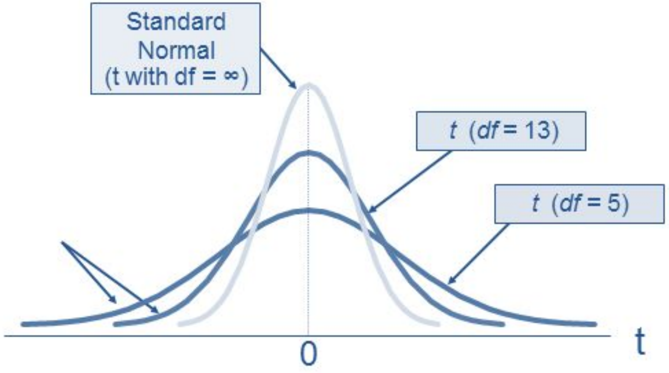
[Interactive T-Distribution]

What happens to the shape of our t-distribution as our sample size increases?





T-Distribution



PDF:

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Parameters: $\nu > 0$ where ν is degrees of freedom (n-1)

T-Score v. Z-Score

95% DISTRIBUTION COMPARISON

Z-distribution, ± 1.96

Student's t-distribution

n	df	Interval
10	9	± 2.262
30	29	± 2.045
75	74	± 1.993
100	99	± 1.984

Brief Deviation

When calculating for the sample standard deviation or variance, you must divide by the degrees of freedom instead of n.

DF = n - 1

This is due to something called Bessel's Correction.

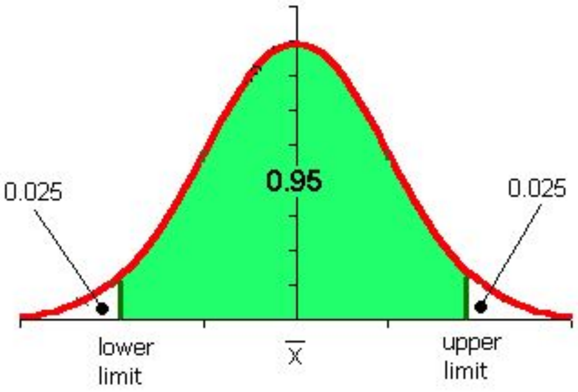
$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

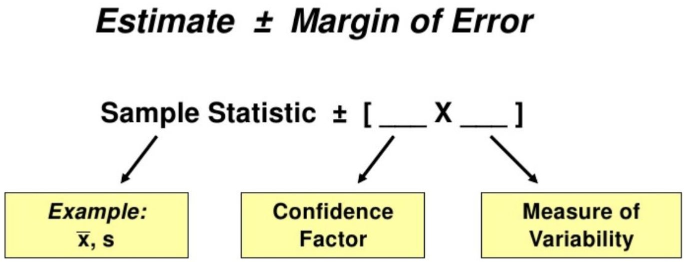
$$s^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$$

Confidence Intervals

Our level of confidence that if we obtained a sample of equal size through the same process, our sample would contain the population mean.

IT IS NOT: The % chance the population mean lies within our sample interval. (Many people will say this!)





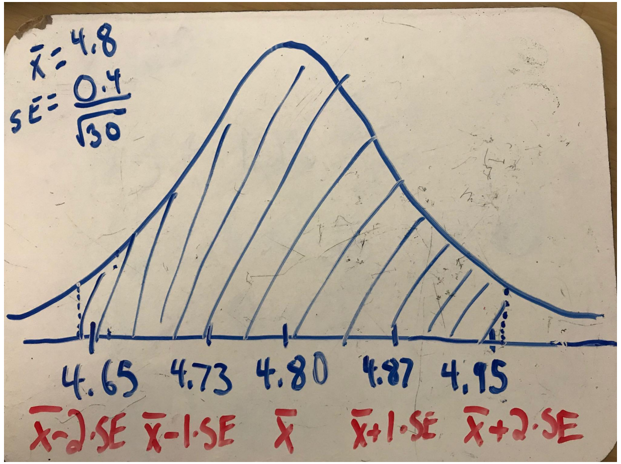
Confidence Interval Factory Example

You are inspecting a hardware factory and want to construct a 95% confidence interval of acceptable screw lengths. You draw a sample of 30 screws and calculate their mean length as 4.8 centimeters and the standard deviation as 0.4 centimeters. Calculate the bounds of your confidence interval?

\

Confidence Interval Factory Example

You are inspecting a hardware factory and want to construct a 90% confidence interval of acceptable screw lengths. You draw a sample of 30 screws and calculate their mean length as 4.8 centimeters and the standard deviation as 0.4 centimeters. Calculate the bounds of your confidence interval?



Confidence Interval Factory Example

You are inspecting a hardware factory and want to construct a 90% confidence interval of acceptable screw lengths. You draw a sample of 30 screws and calculate their mean length as 4.8 centimeters and the standard deviation as 0.4 centimeters. Calculate the bounds of your confidence interval?

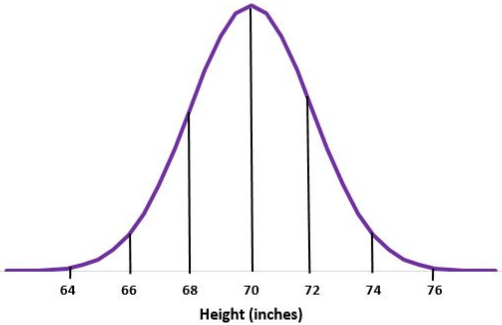
```
import scipy.stats as scs
n = 30
mean = 4.8
t_value = scs.t.ppf(0.95,n-1)
margin_error = t_value* 0.4/(n**0.5)
confidence_interval = (mean - margin_error, mean + margin_error)
```

```
In [2]: confidence_interval
Out[2]: (4.675913306601235, 4.924086693399876)
```

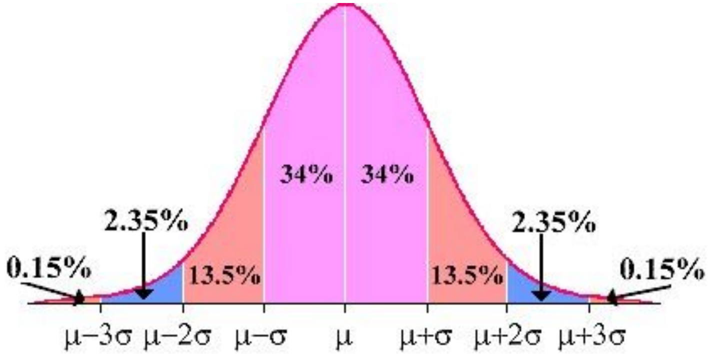
But first

If you had a sampling distributions with a mean of 70 inches and a standard deviation of 2 inches. What is the probability that you would draw the following samples:

A) X ≤ 68 inches
B) X ≥ 76 inches
C) 66 inches ≤ X ≤ 74 inches



So, if you pulled a sample from a population and found it's average to be 77 inches, how likely is it that this sample came from a population with a mean of 70?



Hypothesis Testing

A hypothesis, in statistics, is a statement about a population where this statement typically is represented by some specific numerical value.

In testing a hypothesis, we use a method where we gather data in an effort to gather evidence about the hypothesis.

Below these are summarized into seven steps to conduct a test of a hypothesis.

1. **Setting up two competing hypotheses and check conditions.**
2. **Set some level of significance called alpha.**
3. **Identify the sampling distribution.**
4. **Calculate a test statistic.**
5. **Calculate probability value (p-value), or find rejection region**.
6. **Make a test decision about the null hypothesis**.
7. **State an overall conclusion**.

The Null and Alternative Hypothesis

The two hypotheses are named the null hypothesis and the alternative hypothesis.

Null hypothesis

The null hypothesis is typically denoted as *H0*. The null hypothesis states the "status quo". This hypothesis is assumed to be true until there is evidence to suggest otherwise.

Alternative hypothesis

The alternative hypothesis is typically denoted as *Ha* or *H1*. This is the statement that one wants to conclude. It is also called the research hypothesis.

We usually set the hypothesis that one wants to conclude as the alternative hypothesis, also called the research hypothesis.

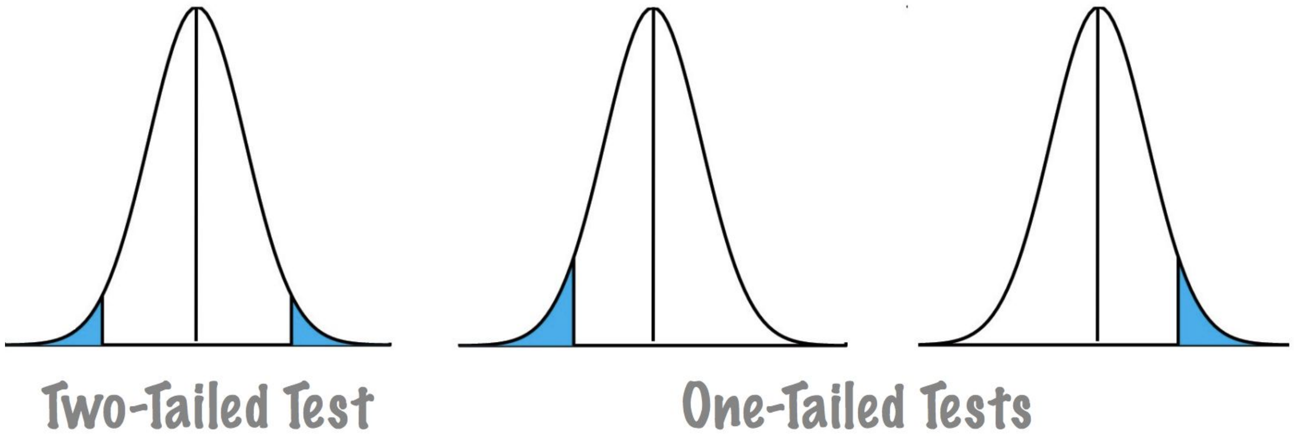
The Null and Alternative Hypothesis

There are three types of alternative hypotheses:

Two-sided test - The population parameter is not equal to a certain value.

Left-tailed test - The population parameter is less than a certain value.

Right-tailed test - The population parameter is greater than a certain value.



The Null and Alternative Hypothesis

Below are the possible alternative hypothesis from which we would select only one of them based on the research question. The symbols p_0 and μ_0 are just used in these general statements. In practice, these get replaced by the parameter value being tested.

Two-sided test - $H_0 : \mu = \mu_0$ $\quad H_a : \mu \neq \mu_0$

Left-tailed test - $H_0 : \mu \geq \mu_0$ $\quad H_a : \mu < \mu_0$

Right-tailed test - $H_0 : \mu \leq \mu_0$ $\quad H_a : \mu > \mu_0$

Practice: Null and Alternative Hypothesis

When determining the marketing plan for Flatiron School in NYC, and employee said that most of the students come from out of the city and move here for the program. The marketing director did not believe this, so he asks you to determin if it is true or not

To answer this question, we can set it up as a hypothesis testing problem and use data collected to answer it. This example is about a population proportion and thus we set up the hypotheses in terms of *p*. Here the value p_o is 0.5 since more than 0.5 constitute a majority.

The hypothesis set up would be a right-tailed test:

$$H_0 : p \leq 0.50 \qquad\qquad H_a : p > 0.5$$

Practice: Null and Alternative Hypothesis

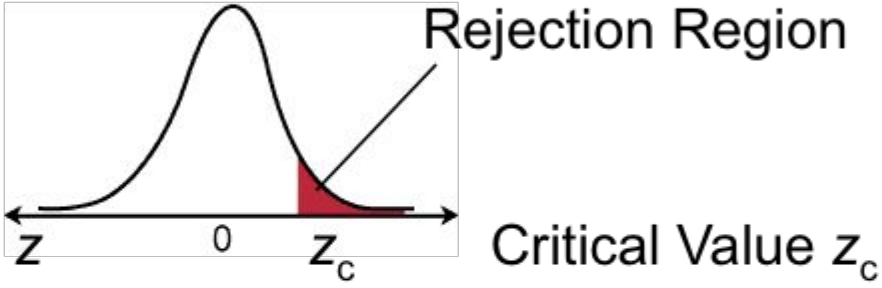
1. A tire manufacturer claims that their tires last an average of 42,000 miles. A consumer protection agency wants to test to see if that claim is true.

2. The length of a cut of lumber from a store is supposed to exactly be 8.5 feet. A builder wants to check whether the shipment of lumber she receives has a mean length different from 8.5 feet.

3. A political news company believes the national approval rating for the current president has fallen below 45%.

Rejection Regions

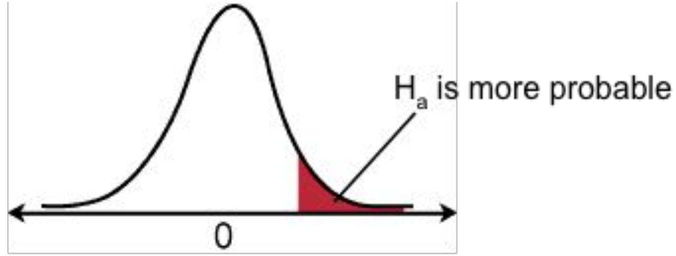
Sampling distribution for $\bar{x}$



The **rejection region** is the range of values for which the *null hypothesis is not probable*. It is always in the direction of the alternative hypothesis. Its area is equal to α.

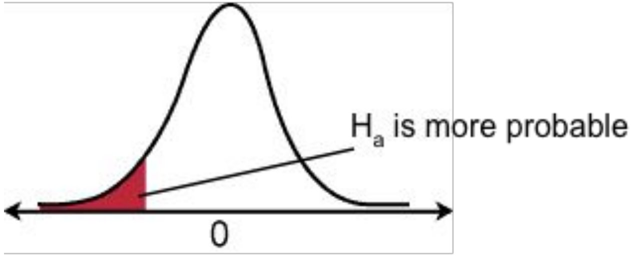
A **critical value** separates the rejection region from the non-rejection region.

The Null and Alternative Hypothesis



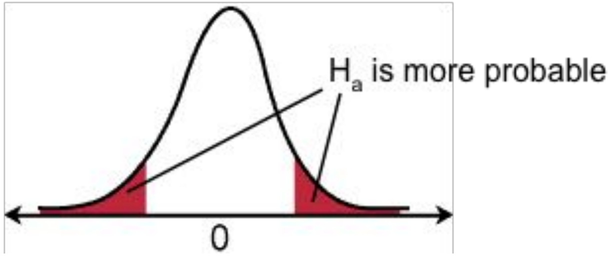
Right-tail test

H_a: $\mu >$ value



Left-tail test

H_a: $\mu <$ value



Two-tail test

H_a: $\mu \neq$ value

Type I and Type II Errors

How do we determine whether to reject the null hypothesis? It begins with the level of significance α, which is the probability of the Type I error.

What is Type I error and what is Type II error?

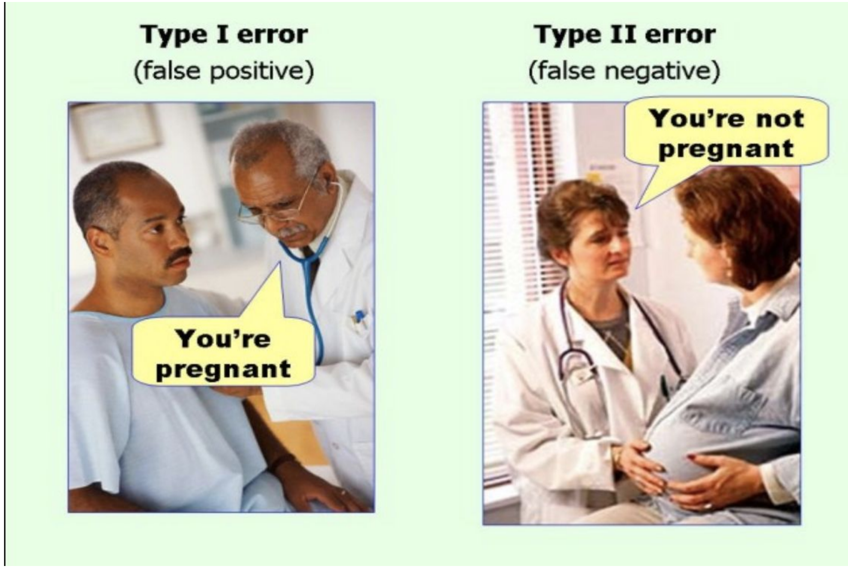
Decision	*Reality*	
	H_0 is true	H_0 is false
Reject H_0	Type I error	Correct
Fail to Reject H_0	Correct	Type II error

The probability of Type II error is denoted by: β

Errors

Let's set up the null and alternative hypotheses so that a Type I error is more serious.

- Type I error: false positive
- Type II error: false negative



Important Terms

Test statistic: The sample statistic one uses to either reject H_o (and conclude H_a) or not to reject H_o.

Critical values: The values of the test statistic that separate the rejection and non-rejection regions.

Rejection region: the set of values for the test statistic that leads to rejection of H_o.

Non-rejection region: the set of values not in the rejection region that leads to non-rejection of H_o.

***P*-value:** The *p*-value (or probability value) is the probability that the test statistic equals the observed value or a more extreme value under the assumption that the null hypothesis is true.

Hypothesis Testing

Below these are summarized into seven such steps to conducting a test of a hypothesis.

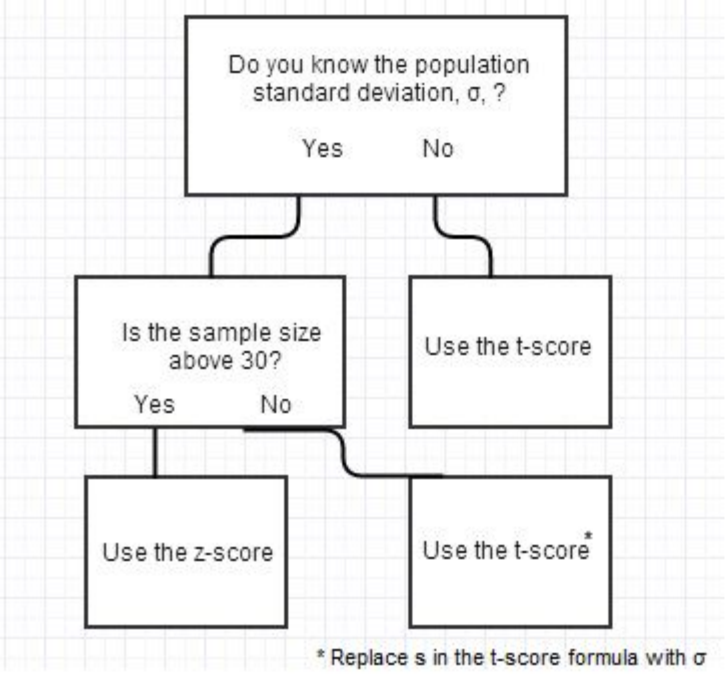
1. **Setting up two competing hypotheses and check conditions**
2. **Set some level of significance called alpha.**
3. **Identify the sampling distribution.**
4. **Calculate a test statistic.**
5. **Calculate probability value (p-value), or find rejection region**.
6. **Make a test decision about the null hypothesis**.
7. **State an overall conclusion**.

Hypothesis testing

A **z-score** and a **t-score** are both used in **hypothesis testing**.

The general rule of thumb for *when* to use a t-score is when your sample:

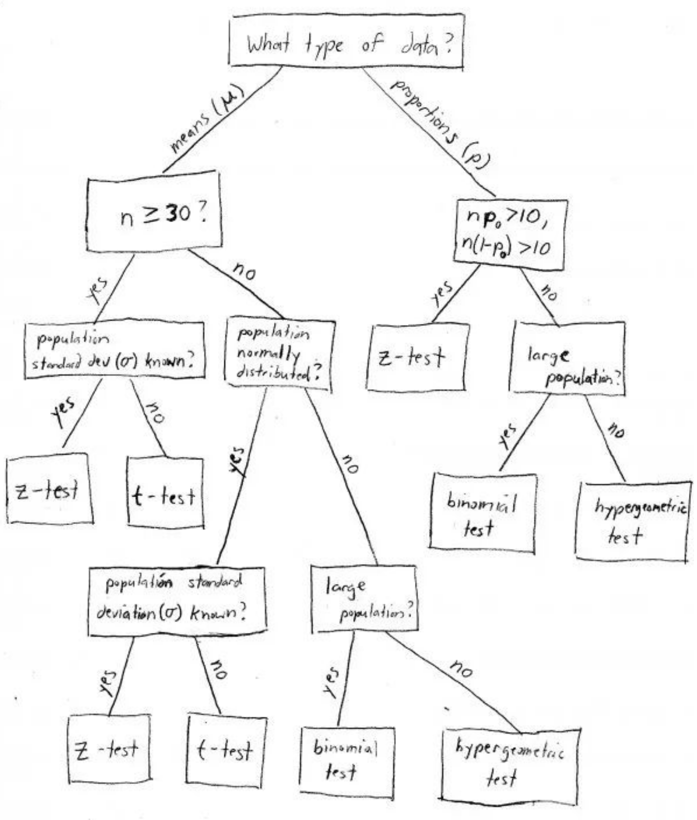
- Has a sample size below 30,
- Has an unknown population standard deviation.



*

Hypothesis testing

Test For	Null Hypothesis (H_0)	Test Statistic	Distribution	Use When
Population mean (μ)	$\mu = \mu_0$	$\dfrac{(\bar{x}-\mu_o)}{\sigma/\sqrt{n}}$	Z	Normal distribution or $n > 30$; σ known
Population mean (μ)	$\mu = \mu_0$	$\dfrac{(\bar{x}-\mu_o)}{s/\sqrt{n}}$	t_{n-1}	$n < 30$, and/or σ unknown
Population proportion (p)	$p = p_0$	$\dfrac{\hat{p}-p_o}{\sqrt{\dfrac{p_o(1-p_o)}{n}}}$	Z	$n\hat{p},\ n(1-\hat{p}) \geq 10$
Difference of two means ($\mu_1 - \mu_2$)	$\mu_1 - \mu_2 = 0$	$\dfrac{(\bar{x}_1-\bar{x}_2)-0}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$	Z	Both normal distributions, or $n_1, n_2 \geq 30$; σ_1, σ_2 known
Difference of two means ($\mu_1 - \mu_2$)	$\mu_1 - \mu_2 = 0$	$\dfrac{(\bar{x}_1-\bar{x}_2)-0}{\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}}$	t distribution with df = the smaller of n_1-1 and n_2-1	$n_1, n_2 < 30$; and/or σ_1, σ_2 unknown
Mean difference μ_d (paired data)	$\mu_d = 0$	$\dfrac{(\bar{d}-\mu_d)}{s_d/\sqrt{n}}$	t_{n-1}	$n < 30$ pairs of data and/or σ_d unknown
Difference of two proportions ($p_1 - p_2$)	$p_1 - p_2 = 0$	$\dfrac{(\hat{p}_1-\hat{p}_2)-0}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$	Z	$n\hat{p},\ n(1-\hat{p}) \geq 10$ for each group



One-Sided *T*-Test for a Mean

A food manufacturer claims there is less than 230 mg of sodium in one serving of a cereal. You work for a national health service and are asked to test this claim. You find that a random sample of 52 servings has a mean sodium content of 232 mg and a sample standard deviation of 10 mg. At $\alpha = 0.05$, what can you say about the manufacturer's claim?

1. Write the null and alternative hypothesis.

2. State the level of significance. $\alpha = 0.05$

3. Determine the sampling distribution.

Since the sample size is at least 30, the sampling distribution is normal, but the population standard deviation is unknown.

4. Find the test statistic and standardize it. $n = 52$ $s = 10$ $\bar{x} = 232$

$$t_{stat} = \frac{\bar{x} - \mu}{\frac{s}{n}}$$

$$\frac{s}{n} = \frac{10}{\sqrt{52}} = 1.387$$

$$t_{stat} = \frac{232 - 230}{1.387} = 1.44$$

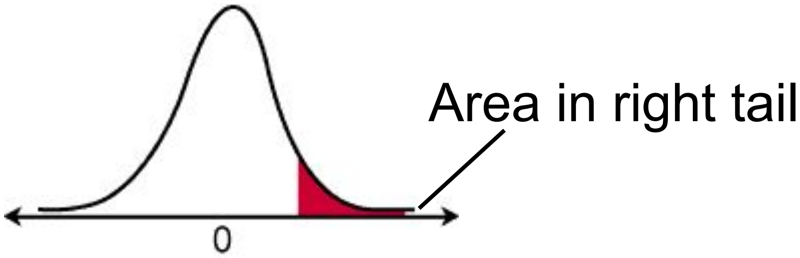
5a. Calculate the P-value for the test statistic.

Since this is a right-tail test, the P-value is the area found to the right of $t = 1.44$ in the normal distribution.

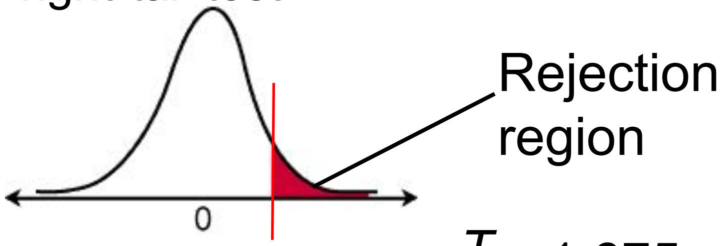
From the table P = 1 – 0.9220

P = 0.078

*



5b. Find the critical values. Since H_a contains the > symbol, this is a right-tail test.



$T_c = 1.675$

$\alpha = 0.05$ so find 0.0500 in the T table which corresponds to a T_c of 1.675

6. Make your decision.

$t = 1.44 < T_c = 1.675$

The test statistic does not fall in the rejection region, so fail to reject H_0

7. Interpret your decision.

There is not enough evidence to reject the null hypothesis that the average sodium content is less than or equal to than 230 mg.

*

Testing μ –Small Sample

A university says the mean number of classroom hours per week for full-time faculty is 11.0. A random sample of the number of classroom hours for full-time faculty for one week is listed below. You work for a student organization and are asked to test this claim. At α = 0.01, do you have enough evidence to reject the university's claim?Assume that the population is normally distributed

11.8 8.6 12.6 7.9 6.4 10.4 13.6 9.1

1. Write the null and alternative hypothesis

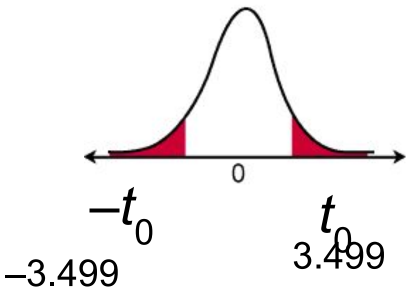
2. State the level of significance $\alpha = 0.01$

3. Determine the sampling distribution

Since the sample size is 8, the sampling distribution is a t-distribution with 8 – 1 = 7 d.f.

*

Since H_a contains the ≠ symbol, this is a two-tail test.



4. Find the critical values.

5a. Find the rejection region.

5 b. Find the test statistic and standardize it

$n = 8$ $\bar{x}$ 10.050 $s = 2.485$

$$t = \frac{10.050 - 11.0}{\frac{2.485}{\sqrt{8}}} = \frac{-0.95}{0.878} = -1.08$$

6. Make your decision.

$t =$ –1.08 does not fall in the rejection region, so fail to reject H_0 at $\alpha = 0.01$

7. Interpret your decision.

There is not enough evidence to reject the university's claim that faculty spend a mean of 11 classroom hours.

Practice

The mean length of the lumber is supposed to be 8.5 feet. A builder wants to check whether the shipment of lumber she receives has a mean length different from 8.5 feet. If the builder observes that the sample mean of 61 pieces of lumber is 8.3 feet with a sample standard deviation of 1.2 feet. What will she conclude using a 99% confidence level?

1. Setting up two competing hypotheses and check conditions
2. Set some level of significance called alpha.
3. Identify the sampling distribution.
4. Calculate a test statistic.
5. Calculate probability value (p-value), or find rejection region.
6. Make a test decision about the null hypothesis.
7. State an overall conclusion.

*

Hypothesis test for One-Sample Proportion

We can find probabilities associated with values of p^ by using the following formula:

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

Consider the example from earlier. A random sample of 1500 U.S. adults is taken. They are asked whether they approve or disapprove of the current president's performance so far (i.e. an approval rating). Of the 1500 surveyed, 660 respond with "approve". Use this poll to evaluate Trump's tweet that he has at least a 55% approval rating.

The 95% confidence interval found in Lesson 5 for the population proportion who approve the president's performance so far is (0.415, 0.465).

*