STATISTICS

WHAT IS STATISTICS?

- Statistics is the application of **what we know** to **what we want to know**.

  Is the Nifty a good model of the entire Indian economy?

  Does the population of Karnataka reflect the entire Indian population?

POPULATION VS. SAMPLE

- These terms come up again and again
- Population is every member of a group we want to study
- Sample is a small set of (hopefully) random members of the population

PARAMETER VS. STATISTIC

- A parameter is a characteristic of a population. Often we want to understand parameters.
- A statistic is a characteristic of a sample. Often we apply statistical inferences to the sample in an attempt to describe the population.

VARIABLE

- A variable is a characteristic that describes a member of the sample.
- Variables can be discrete, or continuous

age
gender

salary
birthplace

SAMPLING

- One of the great benefits of statistical models is that a reasonably sized (>30) random sample will almost always reflect the population.
- The challenge becomes, how do we select members randomly, and avoid bias?

SAMPLING BIAS

- There are several forms of bias:

  Selection Bias

  Perhaps the most common, this type of bias favors those members of a population who are more inclined and able to answer polls.

SAMPLING BIAS

Selection Bias

Undercoverage Bias: making too few observations or omitting entire segments ofa population

SAMPLING BIAS

Selection Bias

- Self-selection Bias: people who volunteer may differ significantly from those in the population who don't

SAMPLING BIAS

Selection Bias

Healthy-user Bias: the sample may come from a healthier segment of the overall population – people who walk/jog, work outside, follow healthier behaviors, etc.

UNDER COVERAGE BIAS

- A hospital survey of employees conducted during daytime hours
- Neglects to poll people who work the night shift.

SELF-SELECTION BIAS

- An online survey about a sports team
- Only people who feel strongly about the team will answer the survey.

HEALTHY-USER BIAS

- Polling customers at a fruit stand to study a connection between diet and health.
- Those polled likely do *other* things that have greater impact on their health.

SAMPLING BIAS

Survivorship Bias

If a population improves over time,
it may be due to lesser members leaving the population due to death, expulsion, relocation, etc.

A CLASSIC PUZZLE

- At the start of World War I, British soldiers wore cloth caps.
- The war office became alarmed at the high number of head injuries, so they issued metal helmets to all soldiers.

A CLASSIC PUZZLE

- They were surprised to find that the number of head injuries *increased* with the use of metal helmets.
- If the intensity of fighting was the same before and after the change, why should the number of head injuries increase?

A CLASSIC PUZZLE

- Answer: You have to consider *all* of the data
- Before the switch, many things that gave head injuries to soldiers wearing metal helmets would have caused fatalities for those wearing cloth caps!

ANOTHER SURVIVORSHIPEXAMPLE

- In World War II, statistician Abraham Wald worked for America's Statistical Research Group (SRG)

Adapted from https://en.wikipedia.org/wiki/Abraham_Wald

ANOTHER SURVIVORSHIPEXAMPLE

- One problem the SRG worked on was to examine the distribution of damage to aircraft by enemyfire and to advise the best placement of additional armor.

ANOTHER SURVIVORSHIP EXAMPLE

- Common logic was to provide greater protection to parts that received more damage.

ANOTHER SURVIVORSHIP EXAMPLE

- Wald saw it differently – he felt that damage must be more uniformly distributed and that aircraft that could return had been hit in less vulnerable parts.

ANOTHER SURVIVORSHIP EXAMPLE

- Wald proposed that the Navy reinforce the areas where returning aircraft were undamaged, since those were areas that, if hit, would cause the plane to be lost!

TYPES OF SAMPLING

- Random
- Stratified Random
- Cluster

RANDOM SAMPLING

- As its name suggests, random sampling means every member of a population has an equal chance of being selected.
- However, since samples are usually much smaller than populations, there’s a chance that entire demographics might be missed.

STRATIFIED RANDOM SAMPLING

- Stratified random sampling ensures that groups within a population are adequately represented.
- First, divide the population into segments based on some characteristic.
- Members cannot belong to two groups at once.

STRATIFIED RANDOM SAMPLING

- Next, take random samples from each group
- The size of each sample is based on the size of the group relative to the population.

STRATIFIED RANDOM SAMPLING EXAMPLE

- A company wants to conduct a survey of customer satisfaction
- They can only survey 10% of their customers
- They want to ensure that every age group is fairly represented

STRATIFIED RANDOM SAMPLING EXAMPLE

- The customer breakdown by age group is as follows:

20-29	30-39	40-49	50+	TOTAL
1400	4450	3200	950	10,000

stratum

strata

STRATIFIED RANDOM SAMPLING EXAMPLE

- To obtain a 10% sample, take 10% from each group:

20-29	30-39	40-49	50+	TOTAL
1400	4450	3200	950	10,000
140	**445**	**320**	**95**	**1,000**

CLUSTERING

- A third – and often less precise – method of sampling is clustering
- The idea is to break the population down into groups and sample a random selection of groups, or *clusters*.
- Usually this is done to reduce costs.

CLUSTERING EXAMPLES

- A marketing firm sends pollsters to a handful of neighborhoods
  (instead of canvassing an entire city)
- A researcher samples fishing boats that are in port on a particular day (also known as convenience sampling)

CENTRAL LIMIT THEOREM

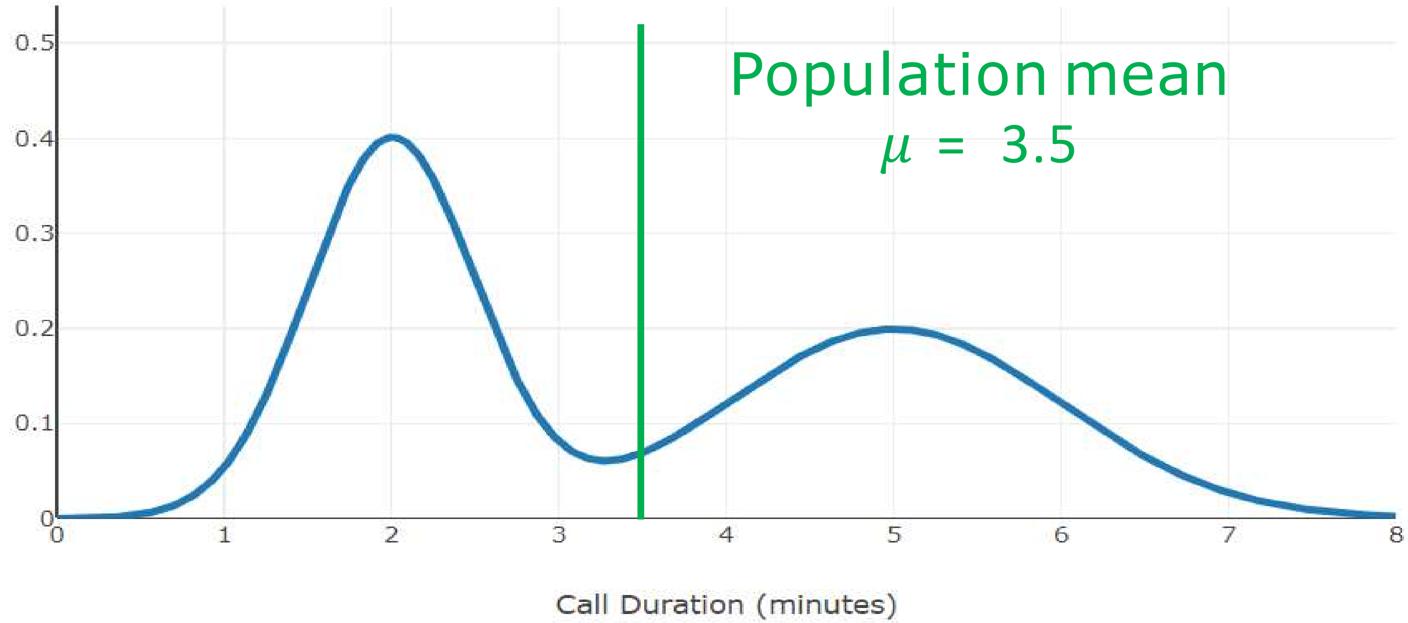
CENTRAL LIMIT THEOREM

- What makes sampling such a good statistical tool is the Central Limit Theorem
- Recall that a sample mean often varies from the population mean.
- The CLT considers a large number of random sample tests.

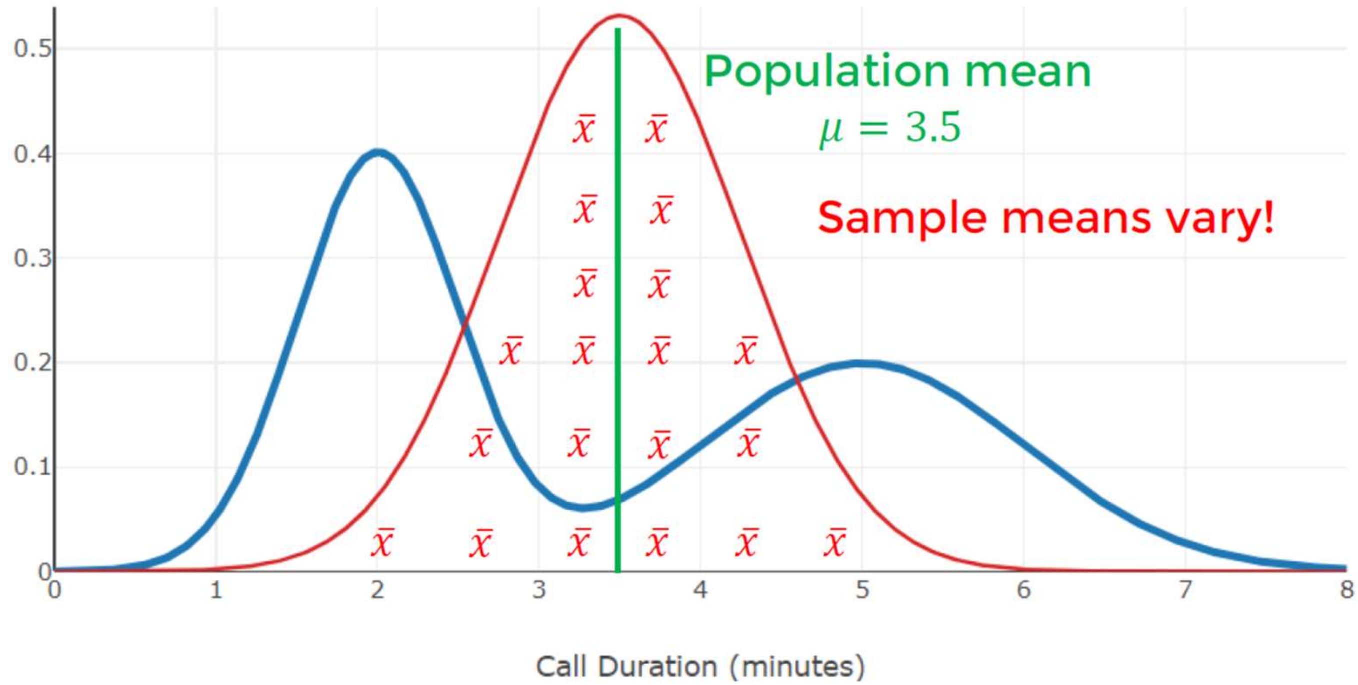
CENTRAL LIMIT THEOREM

- The CLT states that the mean values from a group of samples will be *normally distributed* about the population mean, even if the population itself is not normally distributed.
- That is, 95% of all sample means should fall within 2σ of the population mean

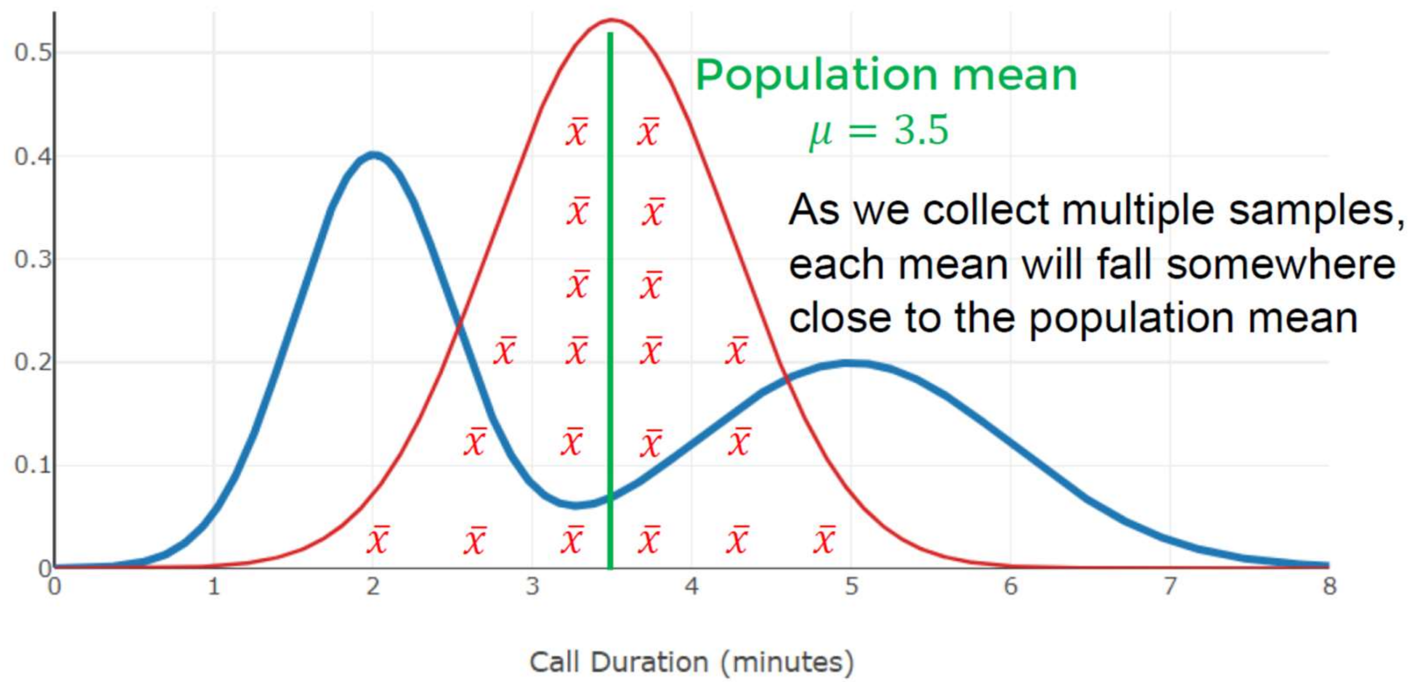
CENTRAL LIMITTHEOREM



CENTRAL LIMIT THEOREM

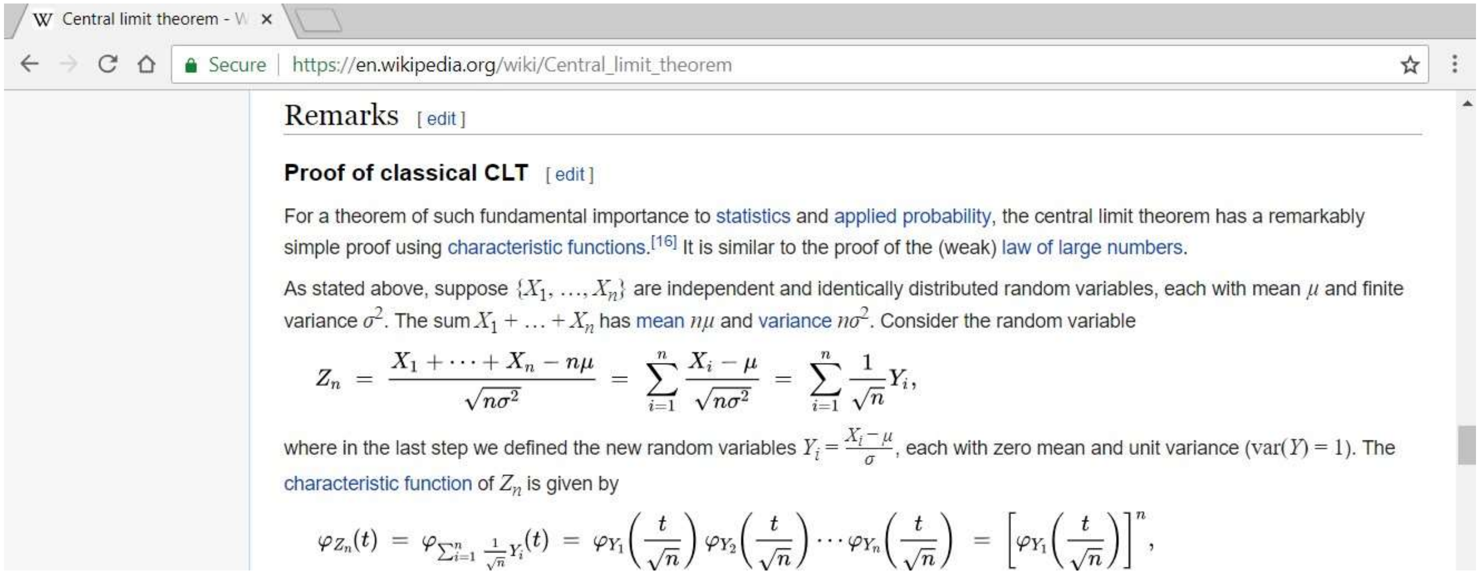


CENTRAL LIMIT THEOREM



PROOF OF CLT

https://en.wikipedia.org/wiki/Central_limit_theorem



Central limit theorem - W

Secure | https://en.wikipedia.org/wiki/Central_limit_theorem

Remarks [edit]

Proof of classical CLT [edit]

For a theorem of such fundamental importance to statistics and applied probability, the central limit theorem has a remarkably simple proof using characteristic functions.[16] It is similar to the proof of the (weak) law of large numbers.

As stated above, suppose $\{X_1, \ldots, X_n\}$ are independent and identically distributed random variables, each with mean μ and finite variance σ^2. The sum $X_1 + \ldots + X_n$ has mean $n\mu$ and variance $n\sigma^2$. Consider the random variable

$$Z_n = \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} Y_i,$$

where in the last step we defined the new random variables $Y_i = \frac{X_i - \mu}{\sigma}$, each with zero mean and unit variance ($\mathrm{var}(Y) = 1$). The characteristic function of Z_n is given by

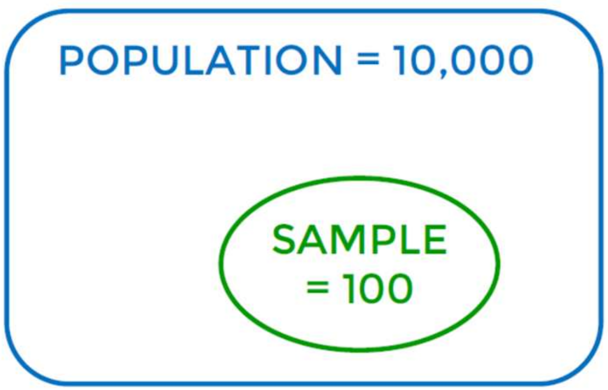
$$\varphi_{Z_n}(t) = \varphi_{\sum_{i=1}^{n} \frac{1}{\sqrt{n}} Y_i}(t) = \varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right) \varphi_{Y_2}\left(\frac{t}{\sqrt{n}}\right) \cdots \varphi_{Y_n}\left(\frac{t}{\sqrt{n}}\right) = \left[\varphi_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n,$$

STANDARD
ERROR

STANDARD ERROR

- Let's quickly review terminology
- Let's say we have a population of voters
- It is unrealistic to poll the entire population, so we poll a sample
- We calculate a statistic from that sample that lets us estimate a parameter of the population

STANDARD ERROR



N = # population members
P = population parameter
σ = pop. standard deviation

n = # sample members
$\hat{p}$ = sample statistic
$SE_{\hat{p}}$ = standard error of the sample

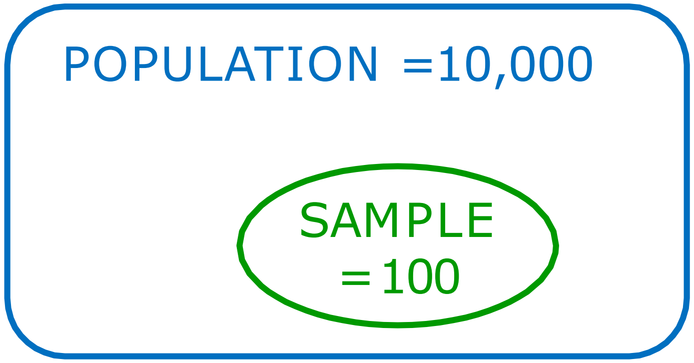
STANDARD ERROR

- If for the population of Australia, the mean height is 5′9″, and for our 100-person sample the mean height is 5′10″, then

$$P = 5'9"$$

$$\hat{p} = 5'10"$$

$$SE_{\hat{p}} = Standard\ Error\ of\ the\ Mean$$



STANDARD ERROR OF THE MEAN

- Where the population standard deviation describes how wide individual values stray from the population mean, the Standard Error of the Mean describes how far a sample mean may stray from the population mean.

STANDARD ERROR OF THE MEAN

- If the population standard deviation σ is known, then the sample standard error of the mean can be calculated as:

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

STANDARD ERROR EXERCISE

- An IQ Test is designed to have a mean score of 100 with a standard deviation of 15 points.
- If a sample of 10 scores has a mean of 104, can we assume they come from the general population?

STANDARD ERROR EXERCISE

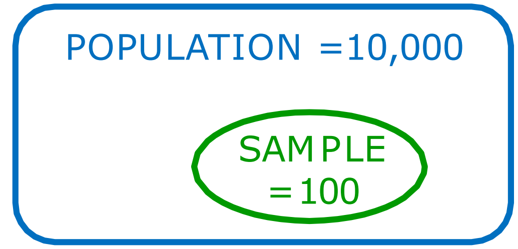
- Sample of 10 IQ Test scores:

$$n = 10 \quad \bar{x} = 104 \quad \sigma = 15$$

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{10}} = 4.743$$

- 68% of 10-item sample means are expected to fall between 95.257 and 104.743

CONFIDENCE INTERVALS



"We can say with a 95% confidence level that the population parameter lies within a confidence interval of plus-or-minus two standard errors of the sample statistic"

N = # population members

P = population parameter

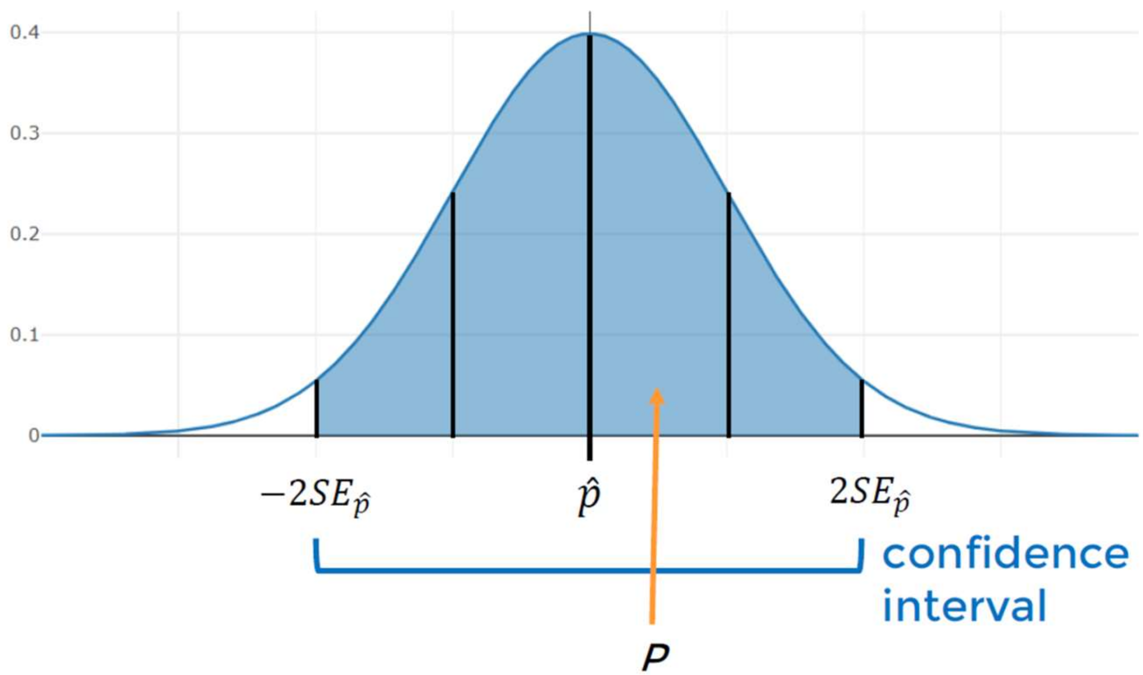
σ = pop. standard deviation

n = # sample members

$\hat{p}$ = sample statistic

$SE_{\hat{p}}$ = standard error of the sample

CONFIDENCE INTERVALS



In the above example, the sample statistic $\hat{\boldsymbol{p}}$ is a **point estimator** of the population parameter $\boldsymbol{P}$.

HYPOTHESIS TESTING

HYPOTHESIS TESTING

- Hypothesis Testing is the application of statistical methods to real-world questions.
- We start with an assumption, called the null hypothesis
- We run an experiment to test this null hypothesis

HYPOTHESIS TESTING

- Based on the results of the experiment, we either reject or fail to reject the null hypothesis
- If the null hypothesis is rejected, then we say the data supports another, mutually exclusive alternate hypothesis
- We never "PROVE" a hypothesis!

FRAMING THE HYPOTHESIS

- How do we frame the question that forms our null hypothesis?
- At the start of the experiment, the null hypothesis is assumed to be true.
- If the data fails to support the null hypothesis, only then can we look to an alternative hypothesis

FRAMING THE HYPOTHESIS

If testing something assumed to be true,
the null hypothesis can reflect the assumption:

Claim: *"Our product has an average shipping weight of 3.5kg."*

Null hypothesis: average weight = 3.5kg

Alternate hypothesis: average weight ≠ 3.5kg

FRAMING THE HYPOTHESIS

If testing a claim we *want* to be true, but can't assume, we test its opposite:

Claim: *"This prep course improves test scores."*

Null hypothesis: old scores < new scores

Alternate hypothesis: old scores >= new scores

FRAMING THE HYPOTHESIS

The null hypothesis should contain an equality ($=, \leq, \geq$):

average shipping weight $=$ 3.5kg

$$\boxed{H_0\colon \mu = 3.5}$$

The alternate hypothesis should not have an equality ($\neq, <, >$):

average shipping weight $\neq$ 3.5kg

$$\boxed{H_1\colon \mu \neq 3.5}$$

FRAMING THE HYPOTHESIS

The null hypothesis should contain an equality (=, ≤, ≥):

old scores ≥ new scores

$$H_0: \mu_0 \geq \mu_1$$

The alternate hypothesis should not have an equality (≠, <, >):

old scores < new scores

$$H_1: \mu_0 < \mu_1$$

HYPOTHESIS TESTING

- So what lets us reject or fail to reject the null hypothesis?

HYPOTHESIS TESTING

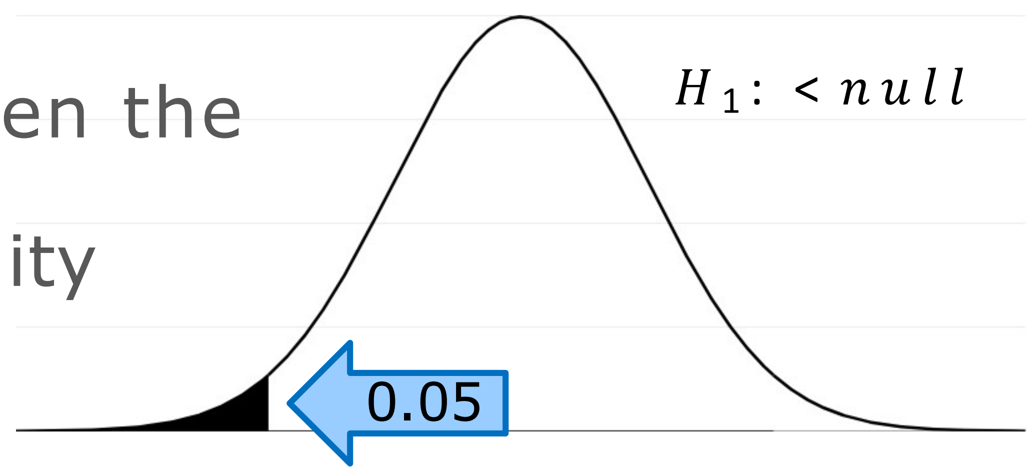
- We run an experiment and record the result.
- Assuming our null hypothesis is valid, if the probability of observing these results is very small (inside of 0.05) then we reject the null hypothesis.
- Here 0.05 is our level of significance

$$\alpha = 0.05$$

HYPOTHESIS TESTING -TAILS

- The level of significance α is the area inside the tail(s) of our null hypothesis.
- If $\alpha = 0.05$ and the alternative hypothesis is
  less than the null, then the

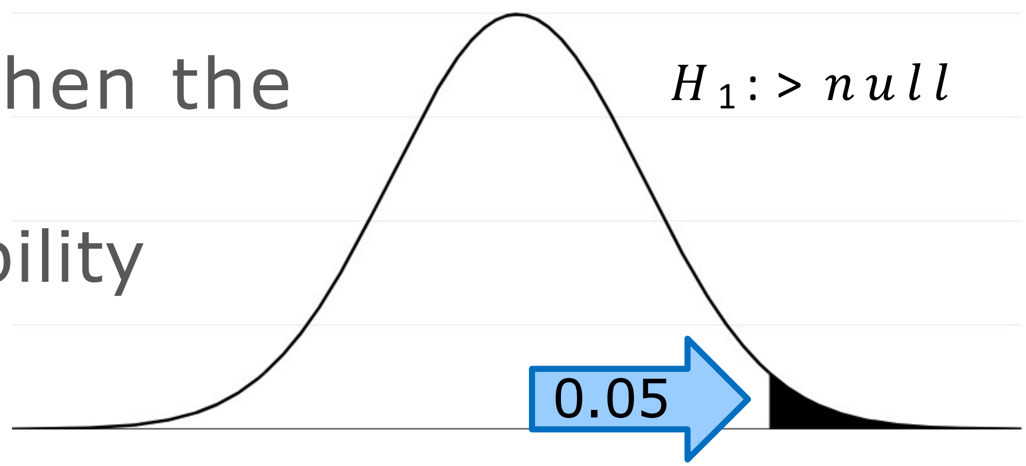
Left – tail of the probability curve has the area 0.05



HYPOTHESIS TESTING -TAILS

- The level of significance α is the area inside the tail(s) of our null hypothesis.
- If $\alpha = 0.05$ and the alternative hypothesis is more than the null, then the

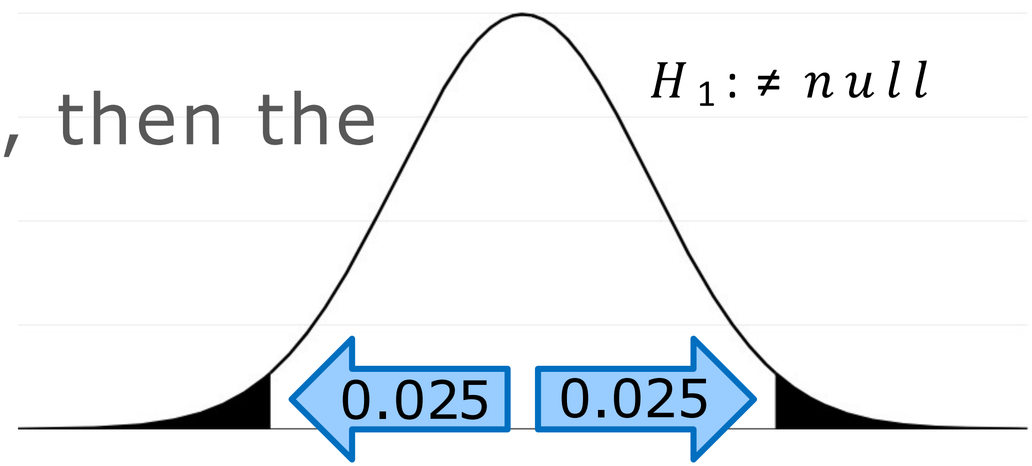
Right – tail of the probability curve has the area 0.05



HYPOTHESIS TESTING -TAILS

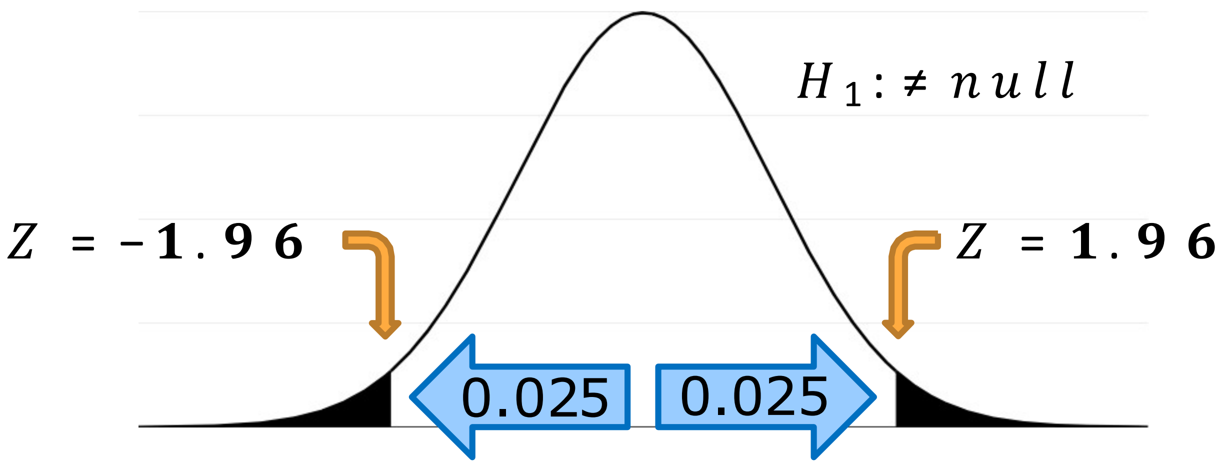
- The level of significance α is the area inside the tail(s) of our null hypothesis.
- If $\alpha = 0.05$ and the alternative hypothesis is not equal to the null, then the

Two tails of our probability curve *share* an area of 0.05



HYPOTHESIS TESTING -TAILS

- These areas establish our critical values or Z-scores:



TESTS OF MEAN VS. PROPORTION

- In the next two lectures, we'll work through full examples of Hypothesis Testing.
- There are two main types of tests:
- Test of Means
- Test of Proportions

TESTS OF MEAN VS. PROPORTION

- Each of these two types of tests has their own test statistic to calculate.
- Let's review the situation for each test before we work through some examples in the upcoming lectures.

TESTS OF MEAN VS. PROPORTION

- Mean
  when we look to find an average, or specific value in a population we are dealing with means
- Proportion
  whenever we say something like "35%" or "most" we are dealing with proportions

TEST STATISTICS

- When working with means:

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

assumes we know the population standard deviation

- When working with proportions:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot (1 - p)}{n}}}$$

HYPOTHESIS TESTING —P-VALUETEST

In a traditional test:

- take the level of significance α
- use it to determine the critical value
- compare the test statistic to the critical value

In a P-value test:

- take the teststatistic
- use it to determine the P-value
- compare the P-value to the level of significance α

HYPOTHESIS TESTING – P-VALUE TEST

"If the P-value is low,

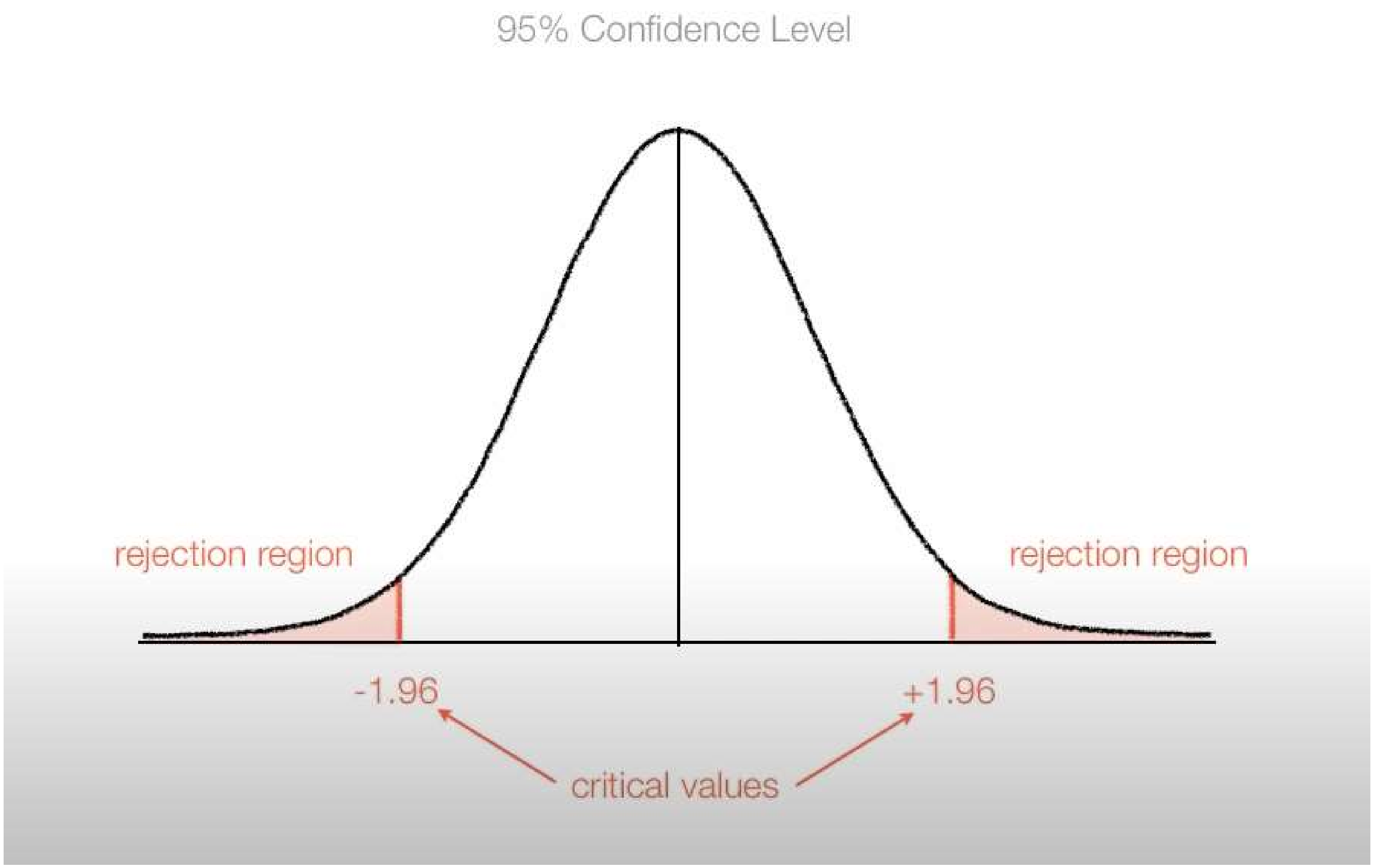
the null must go!"

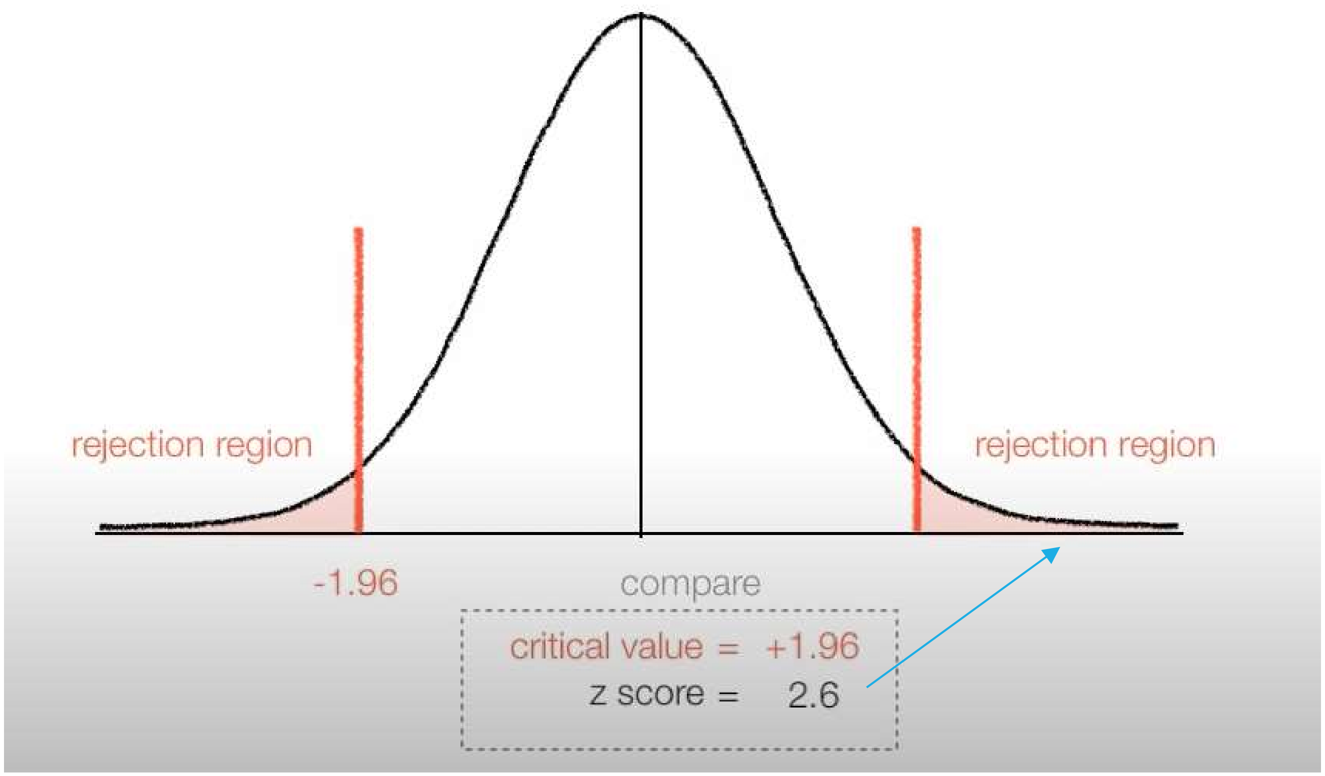
reject H_0

"If the P-value is high,

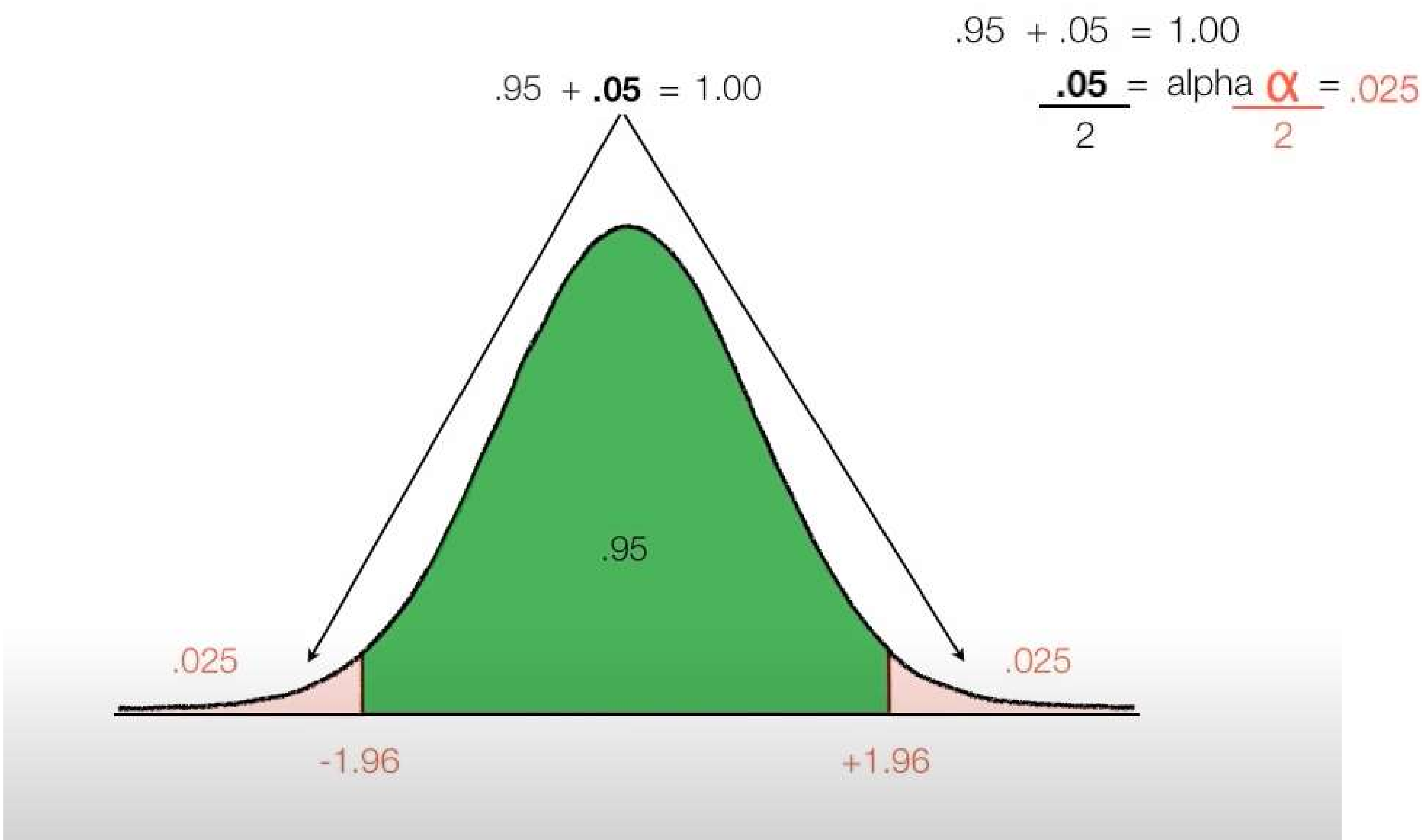
the null must fly!"

fail to reject H_0


95% Confidence Level
rejection region
rejection region
-1.96
+1.96
critical values



Reject the null hypothesis


.95 + .05 = 1.00
.05 / 2 = alpha α / 2 = .025
.95 + .05 = 1.00
.95
.025
.025
-1.96
+1.96

TESTING EXAMPLE

TESTING EXERCISE #1 - MEAN

- For this next example we'll work in the left-hand side of the probability distribution, with negative z-scores
- We'll show how to run the hypothesis test using the traditional method, and then with the P-value method

TESTING EXERCISE #1 - MEAN

$$\mu = 3.125$$
$$\sigma = 0.700$$

- A company is looking to improve their website performance.
- Currently pages have a mean load time of 3.125 seconds, with a standard deviation of 0.700 seconds.
- They hire a consulting firm to improve load times.

TESTING EXERCISE #1 - MEAN

- Management wants a 99% confidence level
- A sample run of 40 of the new pages has a mean load time of 2.875 seconds.

- Are these results statistically faster than before?

$$\mu = 3.125$$
$$\sigma = 0.700$$
$$\alpha = 0.01$$
$$n = 40$$
$$\bar{x} = 2.875$$

TESTING SOLUTION #1-MEAN

1. State the null hypothesis:

   $H_0: \mu \geq 3.125$

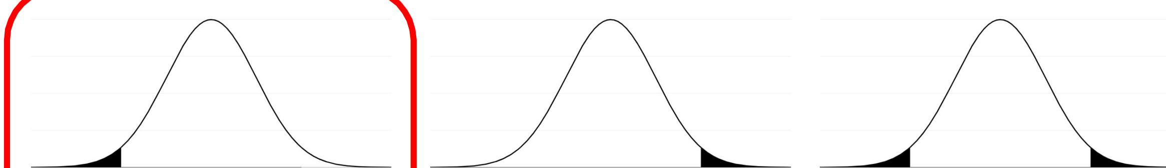
2. State the alternative hypothesis:

   $H_1: \mu < 3.125$

3. Set a level of significance:

   $\alpha = 0.01$

4. Determine the test type:



left tail   right tail   two tail

TESTING SOLUTION #1 - MEAN

TRADITIONAL METHOD:

5. Test Statistic:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{2.875 - 3.125}{0.7/\sqrt{40}} = -2.259$$

6. Critical Value:

z-table lookup on 0.01 $z = -2.325$

$$\mu = 3.125$$
$$\sigma = 0.700$$
$$\alpha = 0.01$$
$$n = 40$$
$$\bar{x} = 2.875$$

$$Z = -2.259$$
$$z = -2.325$$

TESTING SOLUTION #1 - MEAN

TRADITIONAL METHOD:

7. Fail to Reject the Null Hypothesis

Since $-2.259 > -2.325$, the test statistic falls outside the rejection region

We can't say that the new web pages are statistically faster.

$$\mu = 3.125$$
$$\sigma = 0.700$$
$$\alpha = 0.01$$
$$n = 40$$
$$\bar{x} = 2.875$$

$$Z = -2.259$$
$$z = -2.325$$

TESTING SOLUTION #1 - MEAN

P-VALUE METHOD:

5. Test Statistic:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{2.875 - 3.125}{0.7/\sqrt{40}} = -2.259$$

6. P-Value:

z-table lookup on -2.26 $P = 0.0119$

$\mu = 3.125$
$\sigma = 0.700$
$\alpha = 0.01$
$n = 40$
$\bar{x} = 2.875$

$Z = -2.259$
$P = 0.0119$

TESTING SOLUTION #1 - MEAN

P-VALUE METHOD:

7. Fail to Reject the Null Hypothesis

Since $0.0119 > 0.01$, the P-value is greater than the level of significance α

We can't say that the new web pages are statistically faster.

$$\mu = 3.125$$
$$\sigma = 0.700$$
$$\alpha = 0.01$$
$$n = 40$$
$$\bar{x} = 2.875$$

$$Z = -2.259$$
$$P = 0.0119$$

TESTING EXAMPLE

TESTING EXERCISE #2 - PROPORTION

- A video game company surveys 400 of their customers and finds that 58% of the sample are teenagers.
- Is it fair to say that most of the company's customers are teenagers?

TESTING SOLUTION #2 - PROPORTION

1. Set the null hypothesis: $H_0: P \leq 0.50$
2. Set the alternative hypothesis: $H_1: P > 0.50$
3. Calculate the test statistic:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{0.58 - 0.50}{\sqrt{\frac{0.50(1 - 0.50)}{400}}} = \frac{0.08}{0.025} = \mathbf{3.2}$$

TESTING SOLUTION #2 - PROPORTION

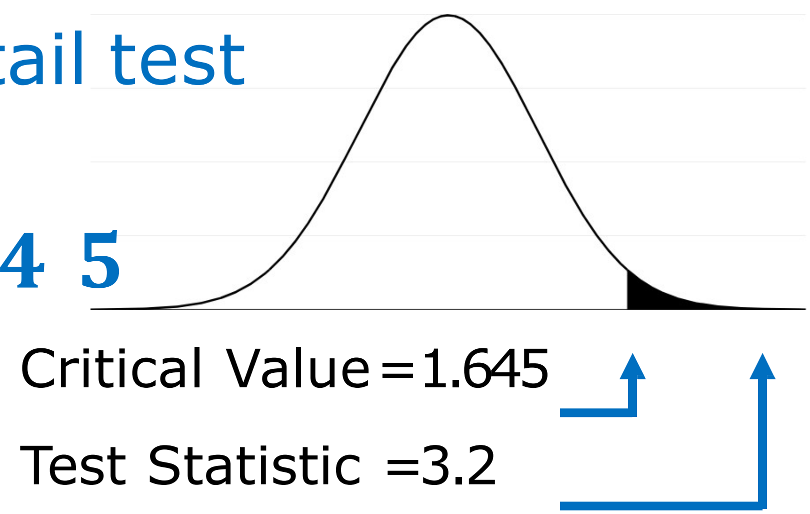
4. Set a significance level: $\alpha = 0.05$

5. Decide what type of tail is involved:

   $H_1: P > 0.50$ means a right-tail test

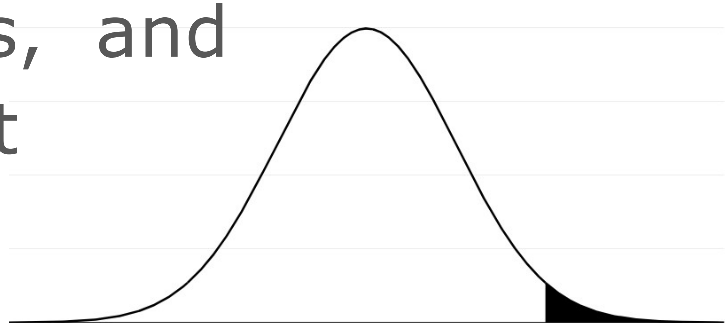
6. Look up the critical value:

$$Z = 1.645$$



TESTING SOLUTION #2 - PROPORTION

7. Based on the sample,
   we reject the null hypothesis, and support the claim that most customers are teenagers.

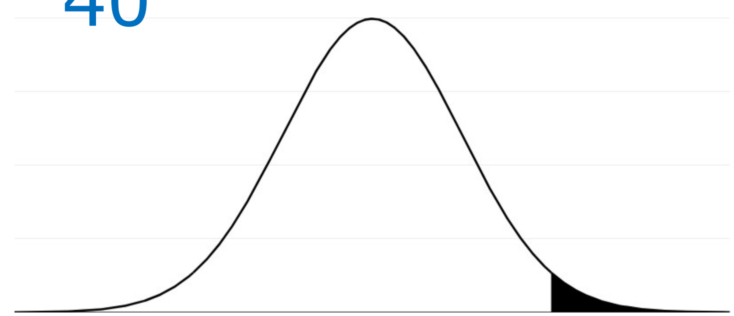


Critical Value = 1.645

Test Statistic = 3.2

TESTING SOLUTION #2 - PROPORTION

NOTE: The size of the sample matters! If we had started with a sample size of 40 instead of 400, our test statistic would have been only 1.01, and we would fail to reject the null hypothesis



Critical Value = 1.645

Test Statistic = 3.2

TYPE 1 AND TYPE 2 ERRORS

TYPE 1 AND TYPE 11 ERRORS

- Often in medical fields (and other scientific fields) hypothesis testing is used to test against results where the "truth" is already known.
- For example, testing a new diagnostic test for cancer for patients you have already succesfully diagnosed by other means.

TYPE 1 AND TYPE 11 ERRORS

- In this situation, you already know if the Null Hypothesis is True or False.
- In these situations where you already know the "truth", then you would know its possible to commit an error with your results .

TYPE 1 AND TYPE 11 ERRORS

- This type of analysis is common enough that these errors already have specific names:
- Type I Error
- Type II Error

TYPE 1 AND TYPE 11 ERRORS

- If we reject a null hypothesis that should have been supported, we've committed a Type I Error

$H_0 : There\ is\ no\ fire$

Pull the fire alarm, only to find out there really was nofire.

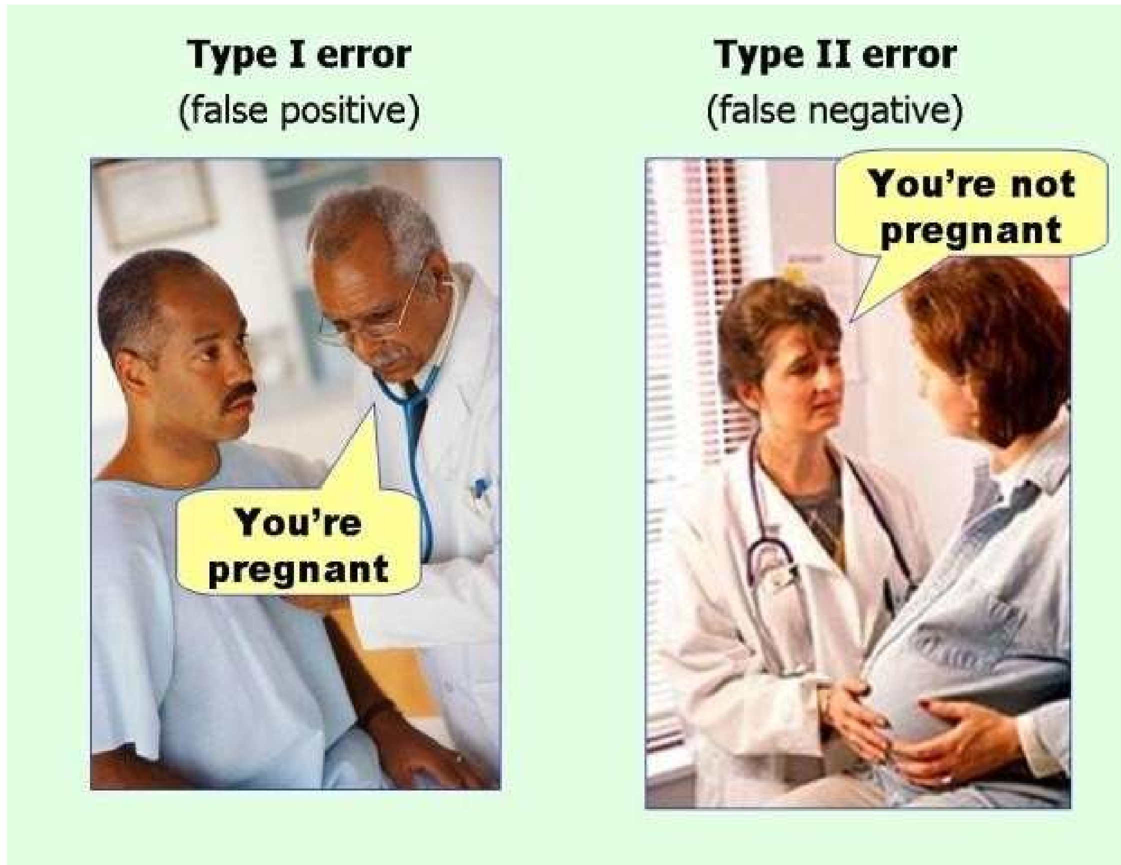
TYPE 1 AND TYPE 11 ERRORS

- If we fail to reject a null hypothesis that should have been rejected we've committed a Type II Error

$\boldsymbol{H_0}: There\ is\ no\ fire$

Don't pull the fire alarm, only to find there really is a fire.

H_0: Not pregnant
H_1: Are pregnant



REAL LIFE EXAMPLES

BIOLOGY

Hypothesis tests are often used in biology to determine whether some new treatment, fertilizer, pesticide, chemical, etc. causes increased growth, stamina, immunity, etc. in plants or animals.

For example, suppose a biologist believes that a certain fertilizer will cause plants to grow more during a one-month period than they normally do, which is currently 20 inches. To test this, she applies the fertilizer to each of the plants in her laboratory for one month.

BIOLOGY

She then performs a hypothesis test using the following hypotheses:

- **H_0:** μ = 20 inches (the fertilizer will have no effect on the mean plant growth)
- **H_A:** μ > 20 inches (the fertilizer will cause mean plant growth to increase)

If the p-value of the test is less than some significance level (e.g. α = .05), then she can reject the null hypothesis and conclude that the fertilizer leads to increased plant growth.

CLINICAL TRAILS

Hypothesis tests are often used in clinical trials to determine whether some new treatment, drug, procedure, etc. causes improved outcomes in patients.

For example, suppose a doctor believes that a new drug is able to reduce blood pressure in obese patients. To test this, he may measure the blood pressure of 40 patients before and after using the new drug for one month.

CLINICAL TRIALS

He then performs a hypothesis test using the following hypotheses:

H0: $\mu_{after} = \mu_{before}$ (the mean blood pressure is the same before and after using the drug)

HA: $\mu_{after} < \mu_{before}$ (the mean blood pressure is less after using the drug)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then he can reject the null hypothesis and conclude that the new drug leads to reduced blood pressure.

AD SPEND

Hypothesis tests are often used in business to determine whether or not some new advertising campaign, marketing technique, etc. causes increased sales.

For example, suppose a company believes that spending more money on digital advertising leads to increased sales. To test this, the company may increase money spent on digital advertising during a two-month period and collect data to see if overall sales have increased.

AD SPEND

hey may perform a hypothesis test using the following hypotheses:

- **H_0:** $\mu_{after} = \mu_{before}$ (the mean sales is the same before and after spending more on advertising)
- **H_A:** $\mu_{after} > \mu_{before}$ (the mean sales increased after spending more on advertising)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then the company can reject the null hypothesis and conclude that increased digital advertising leads to increased sales.

MARKETING

Hypothesis tests are also used often in manufacturing plants to determine if some new process, technique, method, etc. causes a change in the number of defective products produced.

For example, suppose a certain manufacturing plant wants to test whether or not some new method changes the number of defective widgets produced per month, which is currently 250. To test this, they may measure the mean number of defective widgets produced before and after using the new method for one month.

MARKETING

They can then perform a hypothesis test using the following hypotheses:

- **H_0:** $\mu_{after} = \mu_{before}$ (the mean number of defective widgets is the same before and after using the new method)
- **H_A:** $\mu_{after} \neq \mu_{before}$ (the mean number of defective widgets produced is different before and after using the new method)

If the p-value of the test is less than some significance level (e.g. $\alpha = .05$), then the plant can reject the null hypothesis and conclude that the new method leads to a change in the number of defective widgets produced per month.

STUDENT'S T-DISTRIBUTION

STUDENT'S T-DISTRIBUTION

- Developed by William Sealy Gossett while he was working at Guinness Brewery
- Published under the pseudonym "Student" as Guinness wouldn't let him use his name.
- Goal was to select the best barley from small samples, when the population standard deviation wasunknown!

PURPOSE OF A T-TEST

- Using the t-table, the Student's t-test determines if there is a significant difference between two sets of data
- Due to variance and outliers, it's not enough just to compare mean values
- A t-test also considers sample variances

TYPES OF STUDENT'S T-TEST

- One-sample t-test
  Tests the null hypothesis that the population mean is equal to a specified value μ based on a sample mean $\overline{x}$

TYPES OF STUDENT'S T-TEST

- Independent two-sample t-test Tests the null hypothesis that two sample means $\overline{x}_1$ and $\overline{x}_2$ are equal

TYPES OF STUDENT'S T-TEST

- Dependent, paired-sample t-test

  Used when the samples are dependent:
  - one sample has been tested twice (repeated measurements)
  - two samples have been matched or "paired"

ONE-SAMPLE STUDENT'ST-TEST

- **Calculate the t-statistic**

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

$\bar{x}$ = sample mean
μ = population mean
s = sample standard error
n = sample size

ONE-SAMPLE STUDENT'ST-TEST

- **Compare to a t-score**

$$t \lesseqgtr t_{n-1,\alpha}$$

t = t-statistic
$t_{n-1,\alpha}$ = t-critical
$n - 1$ = degrees of freedom
α = significance level

INDEPENDENT TWO-SAMPLET-TEST

The calculation of the t-statistic differs slightly for the following scenarios:

- equal sample sizes, equal variance
- unequal sample sizes, equal variance
- equal or unequal sample sizes, unequal variance

INDEPENDENT TWO-SAMPLET-TEST

- **Calculate the t-statistic**

$$t = \frac{signal}{noise} = \frac{difference\ in\ means}{sample\ variability} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

$\overline{x_1},\ \overline{x_2}$ = sample means

${s_1}^2, {s_2}^2$ = sample variances

$n_1,\ n_2$ = sample sizes

INDEPENDENT TWO-SAMPLET-TEST

- **Compare to a t-score**

$$t \lesseqgtr t_{df,\alpha}$$

t = t-statistic
$t_{df,\alpha}$ = t-critical
df = degrees of freedom
α = significance level

Since we have two, potentially unequal-sized samples with different variances, determining the degrees of freedom is a little more complicated.

DEGREES OF FREEDOM

- The Satterthwaite Formula:

$$df = \frac{\left(\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{{s_1}^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{{s_2}^2}{n_2}\right)^2}$$

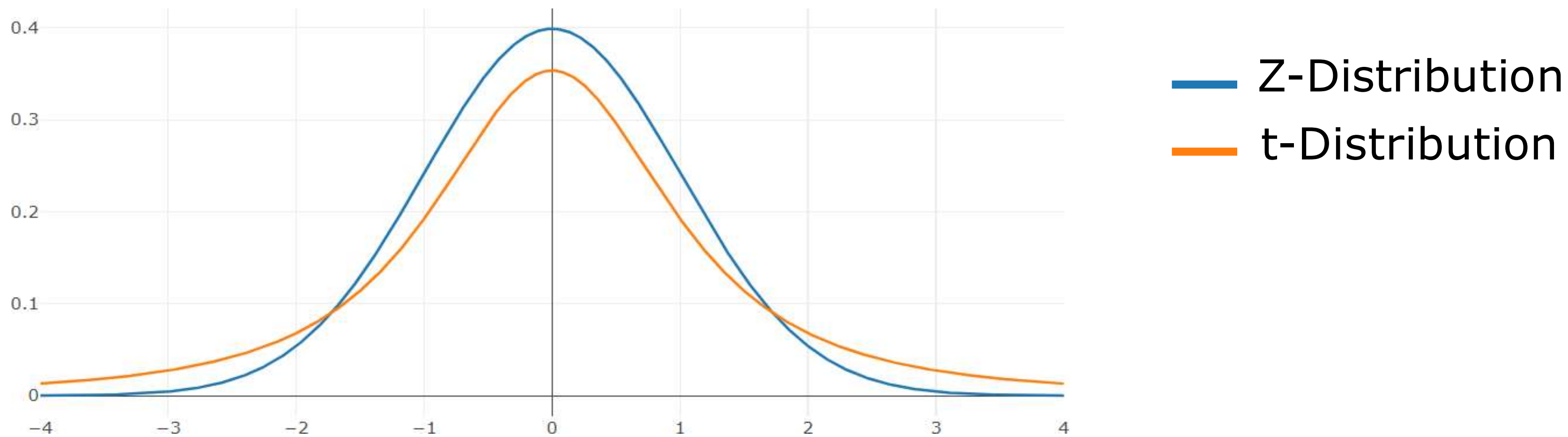
DEGREES OFFREEDOM

- The General Formula:

$$df = n_1 + n_2 - 2$$

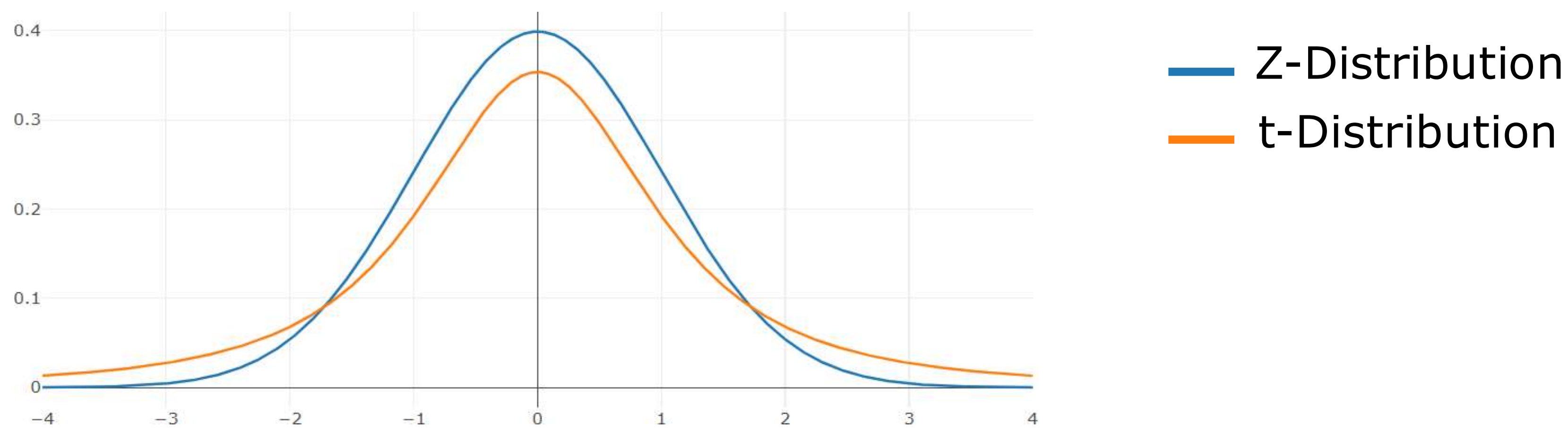
STUDENT'S T-DISTRIBUTION

- t-Distributions have fatter tails than normal Z-Distributions



STUDENT'S T-DISTRIBUTION

- They approach a normal distribution as the degrees of freedom increase.



EXAMPLE

An auto manufacturer has two plants that producethe same car.

EXAMPLE

They are forced to close one of the plants.

STUDENT'S T-TESTEXAMPLE

The company wants to know if there's a significant difference in production between the two plants.

STUDENT'S T-TESTEXAMPLE

Daily production over the same 10 days is as follows:

Plant A	Plant B
1184	1136
1203	1178
1219	1212
1238	1193
1243	1226
1204	1154
1269	1230
1256	1222
1156	1161
1248	1148

STUDENT'S T-TEST EXAMPLE

First compare sample means

$$\bar{x}_A - \bar{x}_B = 1222 - 1186 = \mathbf{36}$$

From this sample, it looks like Plant A produces 36 more cars per day than Plant B

Plant A	Plant B
1184	1136
1203	1178
1219	1212
1238	1193
1243	1226
1204	1154
1269	1230
1256	1222
1156	1161
1248	1148

	$\bar{x}_A$	$\bar{x}_B$
Mean	1222	1186

STUDENT'S T-TESTEXAMPLE

Is 36 more cars enough to say that the plants are different?

$$H_0: X_A \le X_B$$
$$H_1: X_A > X_B$$
$$one\text{-}tailed\ test$$

Plant A	Plant B
1184	1136
1203	1178
1219	1212
1238	1193
1243	1226
1204	1154
1269	1230
1256	1222
1156	1161
1248	1148

	$\bar{x}_A$	$\bar{x}_B$
Mean	1222	1186

$$(10 + 10 - 2) = 18\ degrees\ of\ freedom$$

STUDENT'S T-TESTEXAMPLE

Compute the variance

A	(x-1222)	$(x-1222)^2$
1184	-38	1444
1203	-19	361
1219	-3	9
1238	16	256
1243	21	441
1204	-18	324
1269	47	2209
1256	34	1156
1156	-66	4356
1248	26	676
		11232

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$\Sigma(x-1222)^2$	11232
$\frac{\Sigma(x-1222)^2}{9}$	**1248**

	Plant A	Plant B
	1184	1136
	1203	1178
	1219	1212
	1238	1193
	1243	1226
	1204	1154
	1269	1230
	1256	1222
	1156	1161
	1248	1148
	$\bar{x}_A$	$\bar{x}_B$
Mean	1222	1186
Variance	1248	1246

STUDENT'S T-TEST EXAMPLE

Compute the t-value

$$= \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

$$= \frac{36}{\sqrt{\frac{1248}{10} + \frac{1246}{10}}} = \frac{36}{15.792}$$

$$= \mathbf{2.28}$$

	Plant A	Plant B
	1184	1136
	1203	1178
	1219	1212
	1238	1193
	1243	1226
	1204	1154
	1269	1230
	1256	1222
	1156	1161
	1248	1148
	$\overline{x_A}$	$\overline{x_B}$
Mean	1222	1186
Variance	1248	1246

STUDENT'S T-TEST EXAMPLE

Look up our critical value from a t-table

a one-tailed test

95% confidence

18 degrees of freedom

critical value = 1.734

cum. prob	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
one-tail	0.10	0.05	0.025	0.01	0.005
two-tails	0.20	0.10	0.05	0.02	0.01
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861

STUDENT'S T-TEST EXAMPLE

Compare our t-value (2.28) to the critical value (1.734):

2.28 > 1.734

since our computed t-value is *greater* than the critical value, we reject the null hypothesis.

Plant A	Plant B
1184	1136
1203	1178
1219	1212
1238	1193
1243	1226
1204	1154
1269	1230
1256	1222
1156	1161
1248	1148

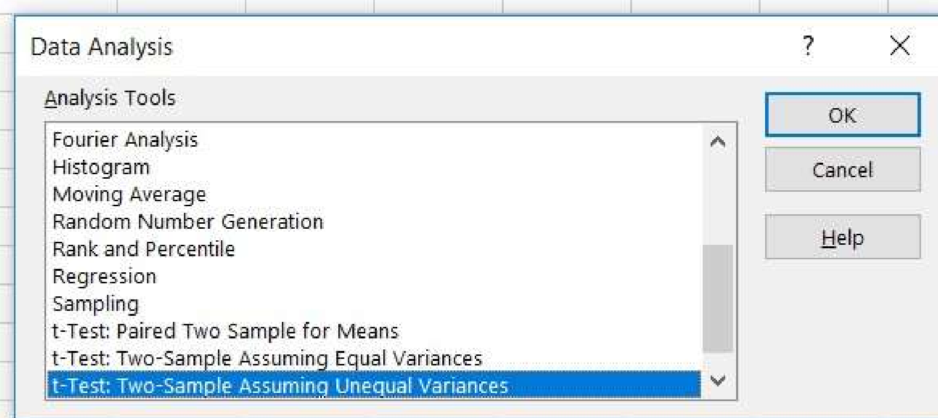
STUDENT'S T-TEST EXAMPLE

We believe with 95% confidence that Plant A produces more cars per day than Plant B.

We decide to close Plant B.

STUDENT'S T-TEST WITH EXCEL

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		*Variable 1*	*Variable 2*
4	Mean	1186	1222
5	Variance	1246	1248
6	Observations	10	10
7	Hypothesized Mean Difference	0	
8	df	18	
9	t Stat	-2.279577051	
10	P(T<=t) one-tail	0.017522528	
11	t Critical one-tail	1.734063607	
12	P(T<=t) two-tail	0.035045056	
13	t Critical two-tail	2.10092204	
14			



STUDENT'S T-TEST WITH PYTHON

```
>>> from scipy.stats import ttest_ind
>>> a = [1184, 1203, 1219, … 1248]
>>> b = [1136, 1178, 1212, … 1148]
>>> ttest_ind(a,b).statistic
2.2795770510504845
>>> ttest_ind(a,b).pvalue/2
0.017522528133638322
```