HYPOTHES
IS
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

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 ≠

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 (=, ≤ ,≥):

average shipping weight = 3.5kg

$$H_0: \quad \mu = 3.5$$

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

average shipping weight ≠ 3.5kg

$$H_1: \quad \mu \neq 3.5$$

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

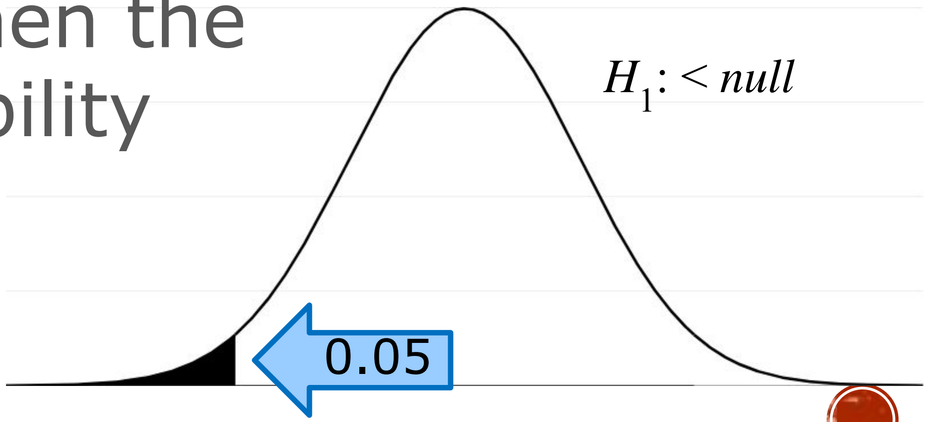
  the probability of observing these results

  is very small (inside of 0.05) then we reject the null hypothesis.
- Here 0.05 is our level of significance

$$\alpha = 0.05$$

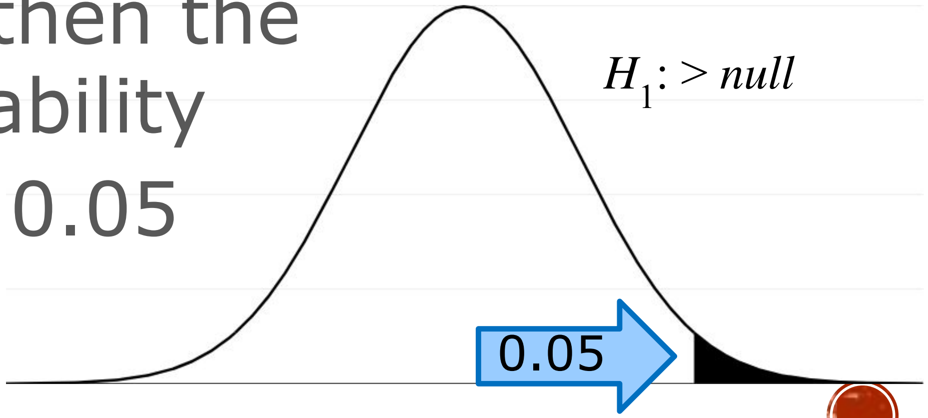
HYPOTHESIS TESTING - TAILS

- 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 our probability curve has an area of 0.05



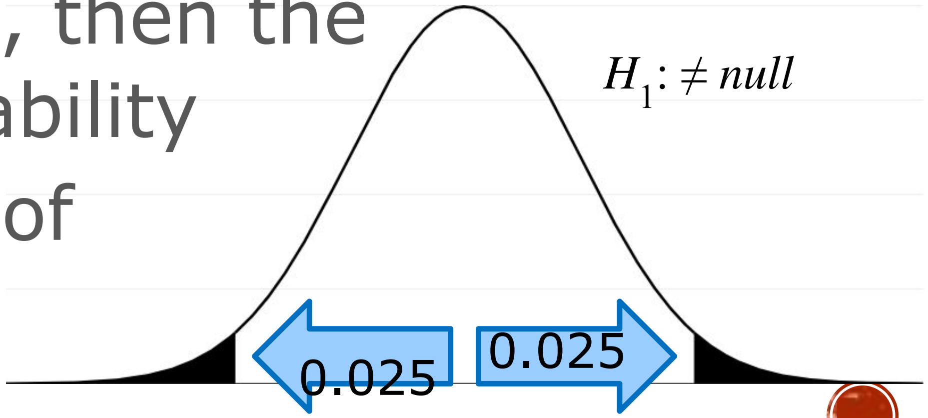
HYPOTHESIS TESTING - TAILS

- 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 our probability curve has an area of 0.05



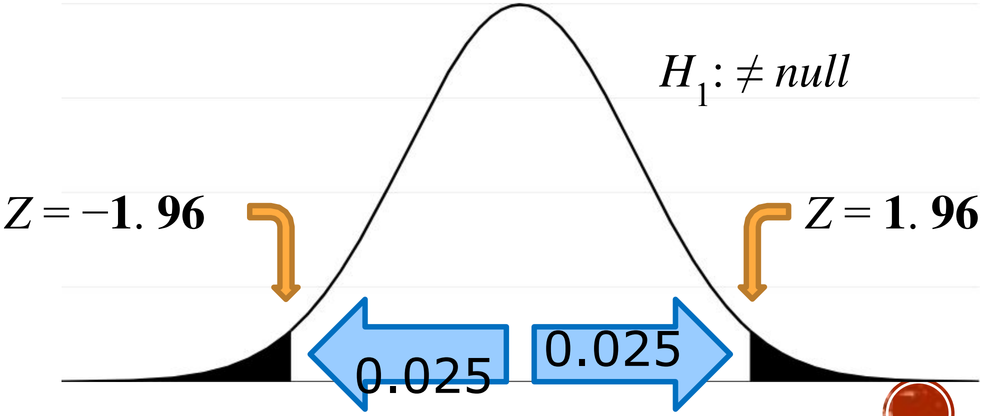
HYPOTHESIS TESTING - TAILS

- 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

- 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:

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

HYPOTHESIS TESTING – P-VALUE TEST

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 test statistic
- 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

TESTING EXAMPLE EXERCISE #1

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?

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

TESTING SOLUTION #1 - MEAN

1. State the null hypothesis:

$$H_0 : \mu \geq 3.125$$

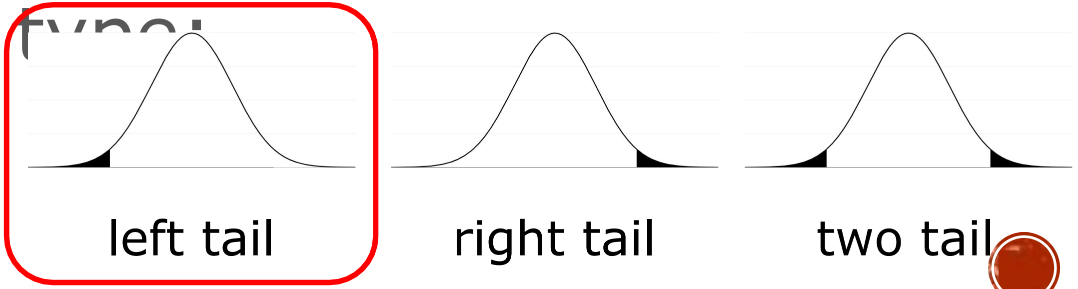
2. State the alternative hypothesis:

$$H_1 : \quad \mu < 3.125$$

3. Set a level of significance:

$$\alpha = 0.01$$

4. Determine the test type:



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

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

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

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

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

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

TESTING EXAMPLE EXERCISE #2

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:

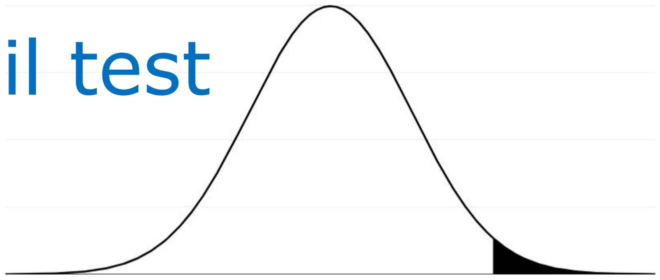
$$Z = \frac{p\Sigma - p}{\sqrt{\frac{p \cdot q}{n}}} \quad \frac{0.58 - 0.50 \quad 0.08}{\sqrt{\frac{0.50(1-0.50)}{400}}} \quad \frac{}{0.025} = \mathbf{3.2}$$

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:

   $$Z = \mathbf{1.645}$$



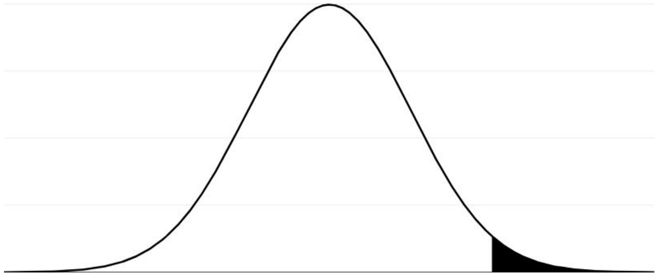
Critical Value =

1.645 Test Statistic

= 3.2

TESTING SOLUTION #2 - PROPORTION

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

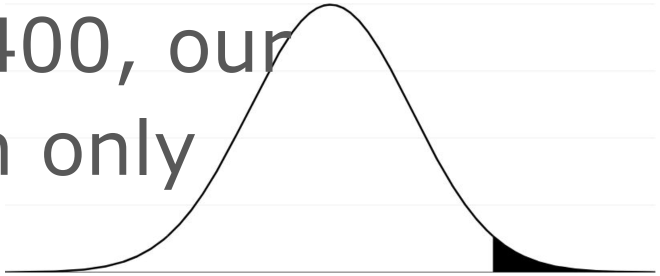


Critical Value =
1.645 Test Statistic
= 3.2

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 I AND TYPE II 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 I AND TYPE II 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 I AND TYPE II ERRORS

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

TYPE I AND TYPE II 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 no fire.

TYPE I AND TYPE II ERRORS

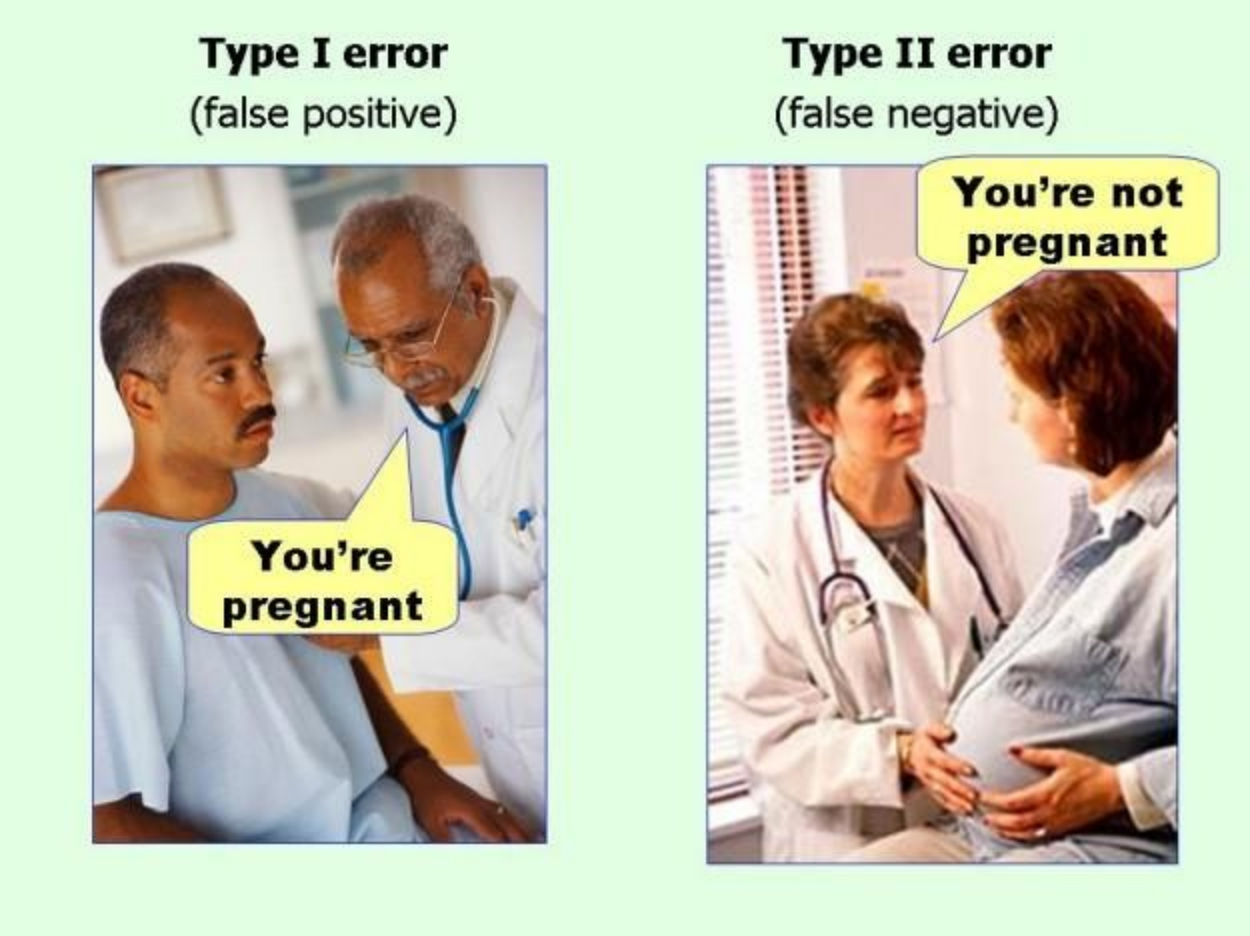
- If we fail to reject a null hypothesis that should have been rejected we've committed a Type II Error

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*



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 was

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 $\bar{x}$

TYPES OF STUDENT'S T-TEST

- Independent two-sample t-test

  Tests the null hypothesis that two sample means $\bar{x}_1$ and $\bar{x}_2$ 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'S T-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'S T-TEST

- Compare to a t-score

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

t = t-statistic

$t_{n-1,\alpha}$ = t-critical

$n-1$ = degrees of freedom

α = significance level

INDEPENDENT TWO-SAMPLE T-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-SAMPLE T-TEST

- Calculate the t-statistic

$$t = \frac{signal}{noise} = \frac{difference\ in\ means}{sample\ variability} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\bar{x}_1, \bar{x}_2$ = sample means

s_1^2, s_2^2 = sample variances

n_1, n_2 = sample sizes

INDEPENDENT TWO-SAMPLE T-TEST

- Compare to a t-score

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

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:

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

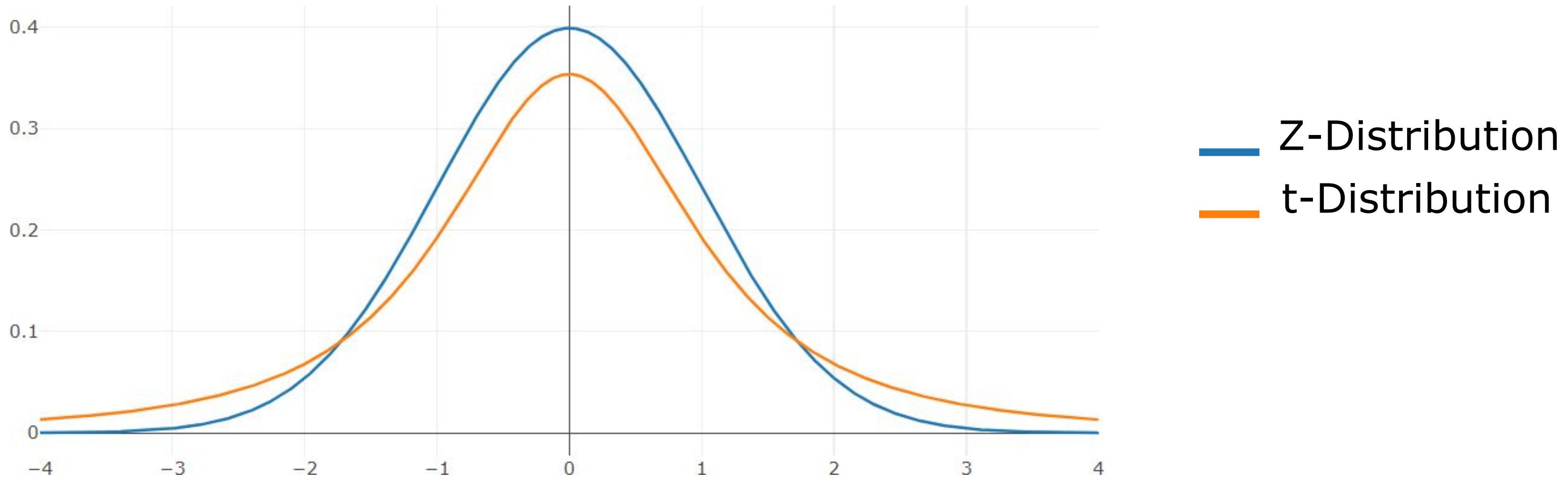
DEGREES OF FREEDOM

- The General Formula:

$$df = n_1 + n_2 - 2$$

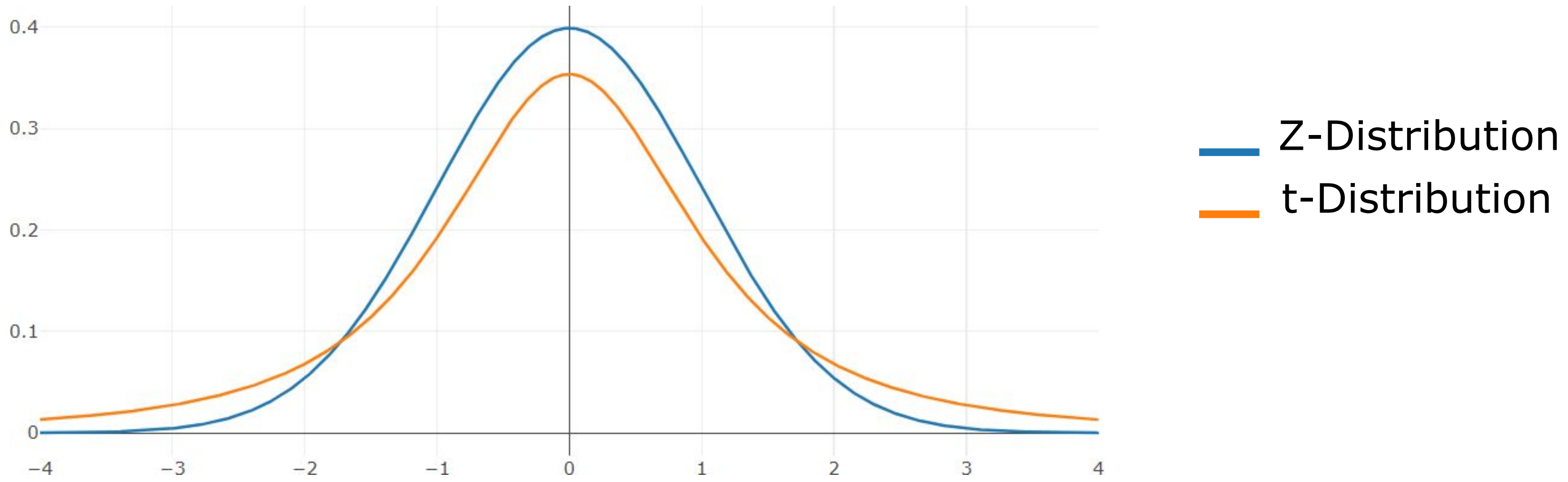
STUDENT'S T-DISTRIBUTION

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



STUDENT'S T-DISTRIBUTION

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



STUDENT’S T-DISTRIBUTION EXAMPLE EXERCISE

PIERIAN DATA

STUDENT'S T-TEST EXAMPLE

An auto manufacturer has two plants that produce the same car.

STUDENT'S T-TEST EXAMPLE

They are forced to close one of the plants.

STUDENT'S T-TEST EXAMPLE

- The company wants to know if there's a
- significant difference in production between

- the two plants.

STUDENT'S T-TEST EXAMPLE

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-TEST EXAMPLE

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

$$H_0 : X_A \leq X_B$$

$$H_1 : X_A > X_B$$

one-tailed test

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

	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-TEST EXAMPLE

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

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

$\Sigma(x-1222)^2$	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

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

critical value = 1.734

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.

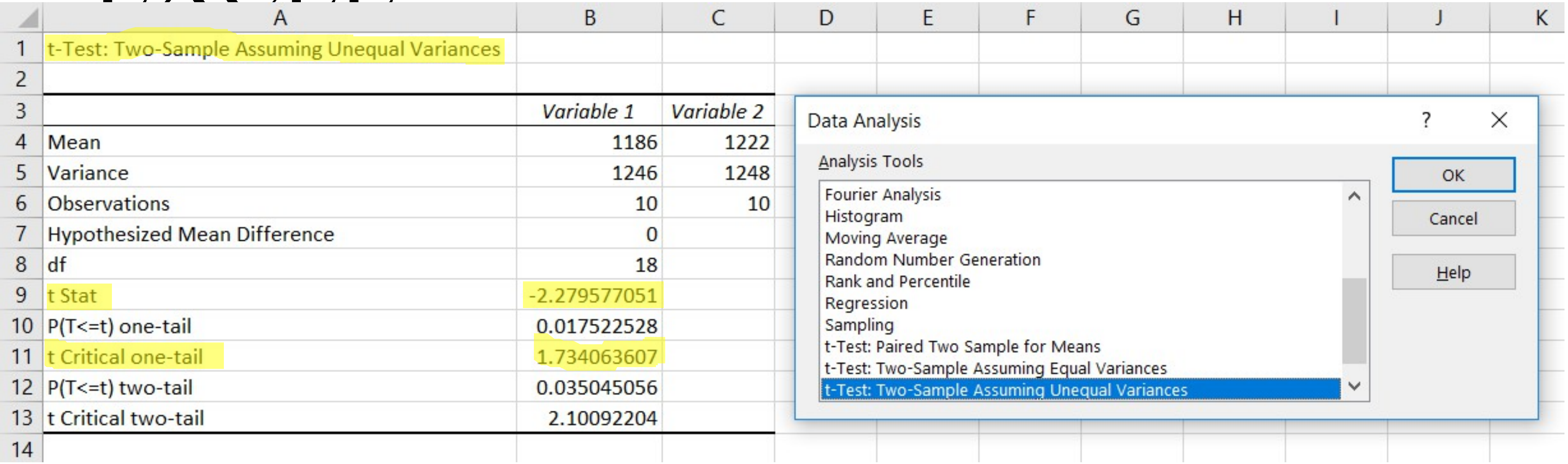
Plant A	Plant B
1184	1136
1203	1178
1219	1212
1238	1193
1243	1226
1204	1154
1269	1230
1256	1222
1156	1161
1248	1148

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

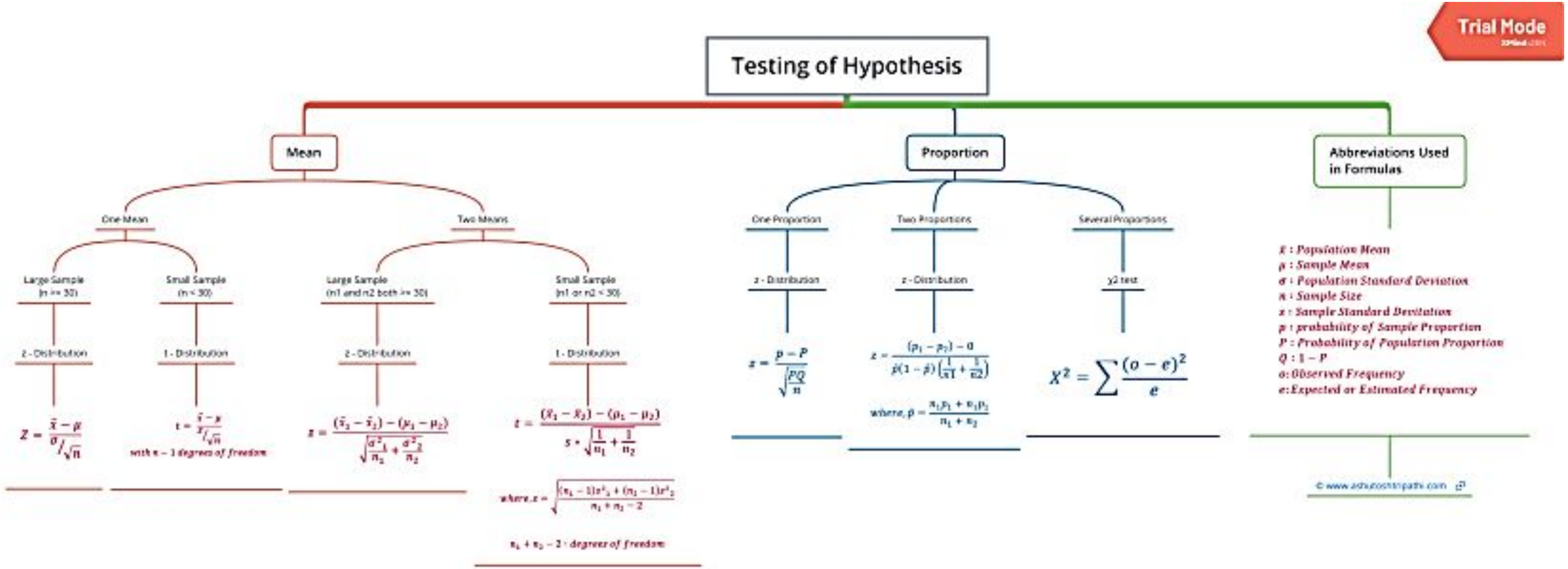


	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		*Variable 1*	*Variable 2*
4	Mean	1186	1222
5	Variance	1246	1248
6	Observations	10	10
7	Hypothesized Mean Difference	0	
8	df	18	
9	t Stat	-2.279577051	
10	P(T<=t) one-tail	0.017522528	
11	t Critical one-tail	1.734063607	
12	P(T<=t) two-tail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

NEXT UP: ANOVA


Trial Mode
Testing of Hypothesis
Mean
Proportion
Abbreviations Used in Formulas
One Mean
Two Means
Large Sample (n >= 30)
Small Sample (n < 30)
Large Sample (n1 and n2 both >= 30)
Small Sample (n1 or n2 < 30)
z - Distribution
t - Distribution
z - Distribution
t - Distribution
with n – 1 degrees of freedom
$n_1 + n_2 - 2$ degrees of freedom
One Proportion
Two Proportions
Several Proportions
z - Distribution
z - Distribution
χ2 test
$X^2 = \sum \frac{(o-e)^2}{e}$
x̄ : Population Mean
μ : Sample Mean
σ : Population Standard Deviation
n : Sample Size
s : Sample Standard Deviation
p : probability of Sample Proportion
P : Probability of Population Proportion
Q : 1 – P
o: Observed Frequency
e: Expected or Estimated Frequency
© www.ashutoshtripathi.com