ANOVA TESTING

David Baker, Luciana Piarulli

What is ANOVA?

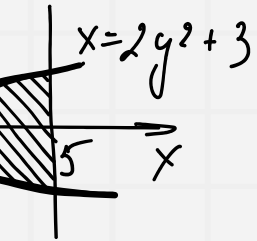
ANOVA (**AN**alysis **O**f **VA**riance) is used to compare three or more means to see if one is different. This is similar to a 2-sample t-test, but it can be used for three or more means. 2-sample t is for 2 means only. ANOVA is used for data in quasi-experiments, and field studies, and experiments (such as STEM projects). The result of an ANOVA test tells you if the means across each group are different or not

There are 2 types of ANOVA: One-way and Two-way. A one-way ANOVA is used when you want to compare the means of three or more groups or treatments based on a single independent variable (factor) to determine if there is a statistically significant difference between the groups. You would use a two-way ANOVA when you want to examine the effect of two independent variables on a dependent variable, particularly to assess both the individual (main) effects of each independent variable and the interaction between them

Hypothesis for ANOVA

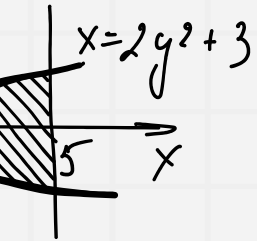
Ho: All μ are the same

Ha: At least one μ is different



Assumptions and Conditions

- Random sample / assignment / representative sample
- The independent (manipulated) variable must be categorical, and the dependent (measured) variable must be quantitative
- Each group's data follows a normal distribution
- At least two categorical independent groups in an independent variable (two independent variables for 2-way ANOVA
- Samples must be independent
- Homogeneity of the variance of the population



Statistic Used for ANOVA Testing

The statistic used for ANOVA tests is the F-statistic

- Named in honor of Sir Ronald Fisher
- Ratio of two variances (square of SD)
- numerator-"the extent to which the sample averages differ"
- denominator "overall variability in the samples"

$$F = \frac{s_1^2}{s_2^2}$$

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

Via https://sixsigmastudyguide.com/f-distributionf-statistic-f-test/

One-Way ANOVA Example

In this example, data was collected to try and see if the average number of lyrics in four different music albums differed overall

- A random sample of 6 songs was taken from each album (generated using the RNG on a TI-84)
- The total number of words in that song were found
- A One-Way ANOVA was performed on this data set to see if the true average number of words in each album differed

One-Way ANOVA Data

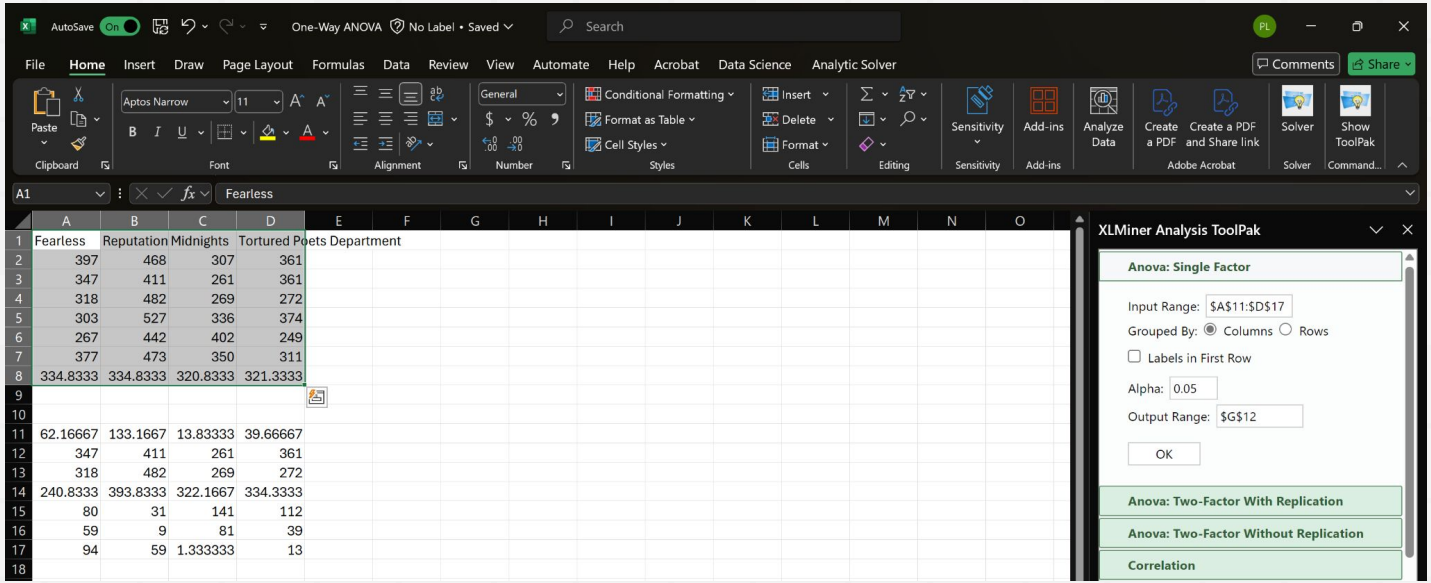
Fearless	Reputation	Midnights	Tortured Poets Department
397	468	307	361
347	411	261	361
318	482	269	272
303	527	336	374
267	442	402	249
377	473	350	311

Assumptions

- stated random sample
- independent variable (album) is categorical, and dependent variable (word count) is quantitative
- Number of independent groups is greater than 2
- Samples were independent
- Distribution of each group is approximately normal (histograms of each group were made)
- variances of groups are homogenous (explained after)

How to test the data in excel

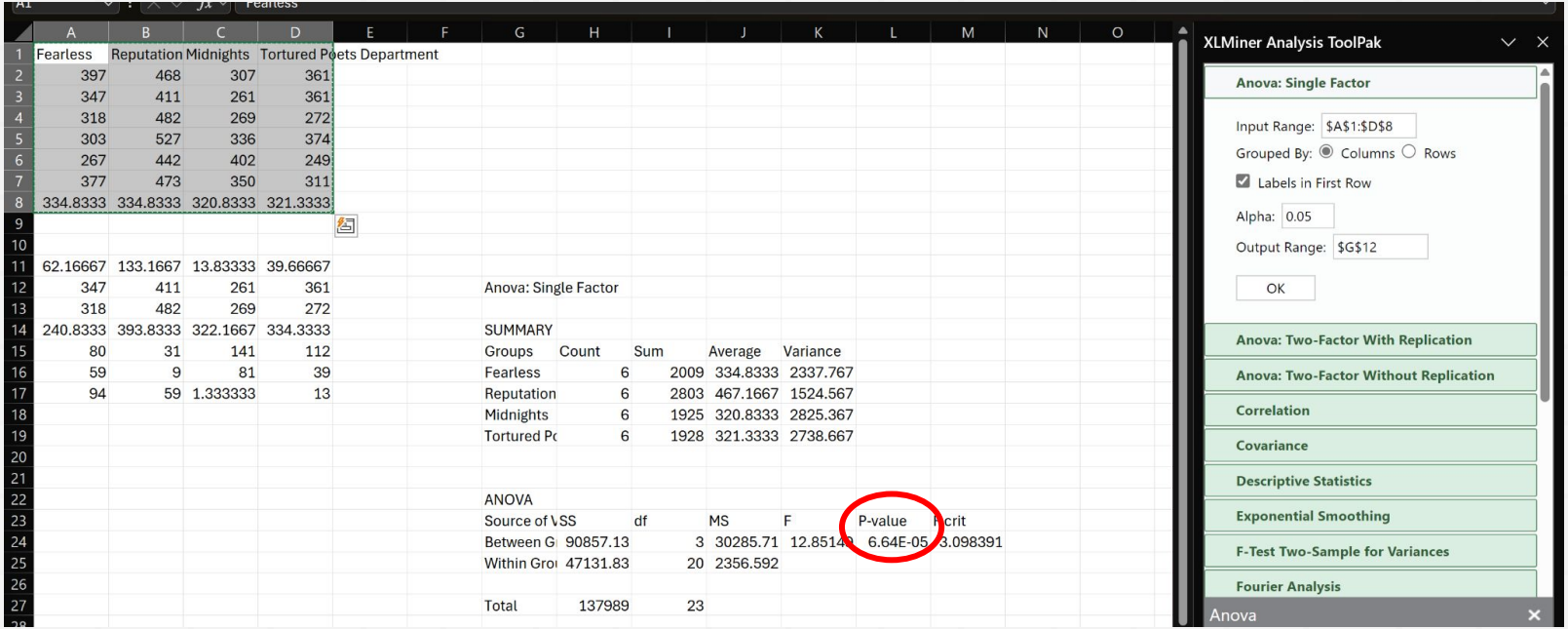
(Ho: All groups are evenly distributed; Ha: At least one mu is different)



	A	B	C	D
1	Fearless	Reputation	Midnights	Tortured Poets Department
2	397	468	307	361
3	347	411	261	361
4	318	482	269	272
5	303	527	336	374
6	267	442	402	249
7	377	473	350	311
8	334.8333	334.8333	320.8333	321.3333
9				
10				
11	62.16667	133.1667	13.83333	39.66667
12	347	411	261	361
13	318	482	269	272
14	240.8333	393.8333	322.1667	334.3333
15	80	31	141	112
16	59	9	81	39
17	94	59	1.333333	13
18				

How to test the data in excel

(Ho: All groups are evenly distributed; Ha: At least one mu is different)



	A	B	C	D
1	Fearless	Reputation	Midnights	Tortured Poets Department
2	397	468	307	361
3	347	411	261	361
4	318	482	269	272
5	303	527	336	374
6	267	442	402	249
7	377	473	350	311
8	334.8333	334.8333	320.8333	321.3333
9				
10				
11	62.16667	133.1667	13.83333	39.66667
12	347	411	261	361
13	318	482	269	272
14	240.8333	393.8333	322.1667	334.3333
15	80	31	141	112
16	59	9	81	39
17	94	59	1.333333	13
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Fearless	6	2009	334.8333	2337.767
Reputation	6	2803	467.1667	1524.567
Midnights	6	1925	320.8333	2825.367
Tortured Pc	6	1928	321.3333	2738.667

ANOVA

Source of V	SS	df	MS	F	P-value	Fcrit
Between G	90857.13	3	30285.71	12.85141	6.64E-05	3.098391
Within Gro	47131.83	20	2356.592			
Total	137989	23				

XLMiner Analysis ToolPak

Anova: Single Factor

Input Range: $A$1:$D$8

Grouped By: Columns Rows

Labels in First Row

Alpha: 0.05

Output Range: $G$12

OK

Anova: Two-Factor With Replication

Anova: Two-Factor Without Replication

Correlation

Covariance

Descriptive Statistics

Exponential Smoothing

F-Test Two-Sample for Variances

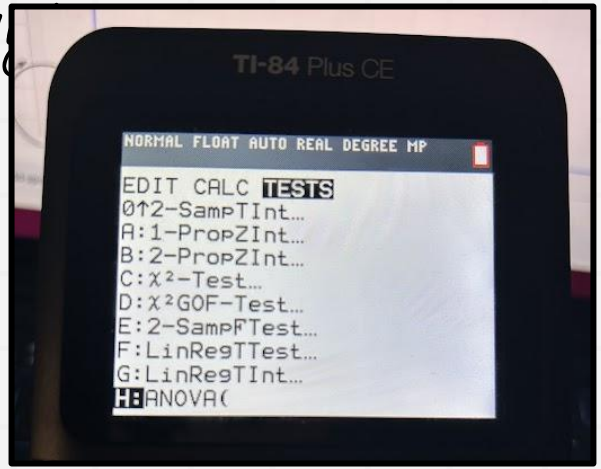
Fourier Analysis

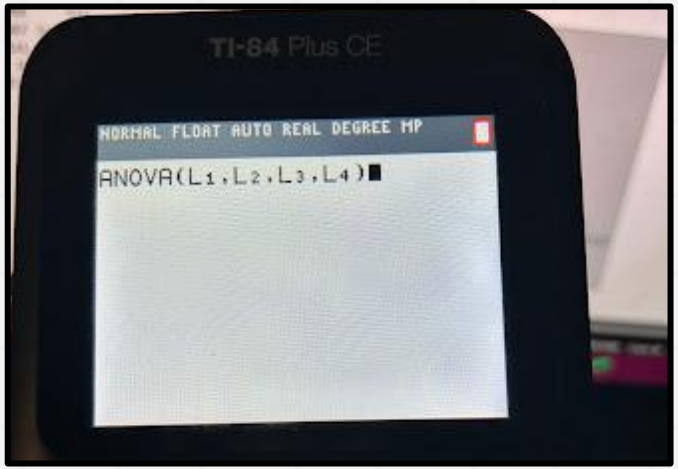
Anova

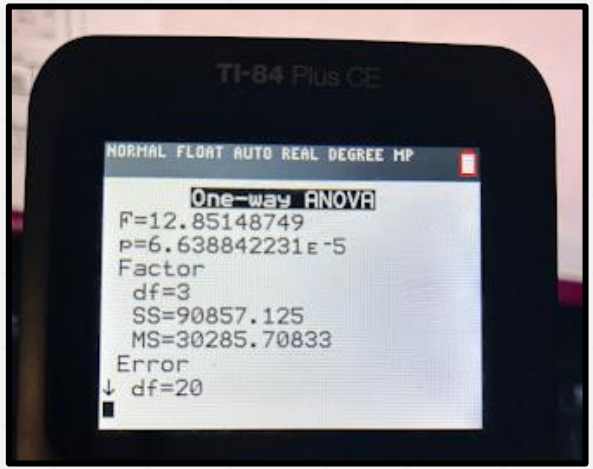
How to test the data in graphing calculator

- Enter each group of data into a separate list (in this case, each album group would go into its own list)
- Go to stat -> test -> H: ANOVA (last option)
- Enter in lists used separated by commas
- A menu will pop up with the test statistic and p-value

How to test the data in graphing calculator







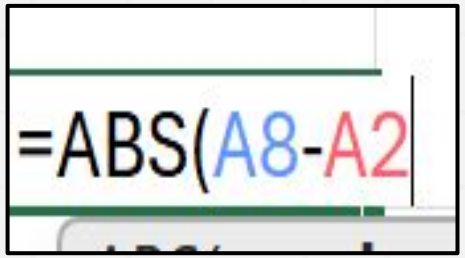
Conclusion

- Since the p-value is less than .05, reject the null hypothesis
- There is convincing evidence that at least one of the mus is different
- Post hoc tests can be used to determine which means are statistically significant from the rest (or using 2-sample t-tests between all the different means, but this can take a while)

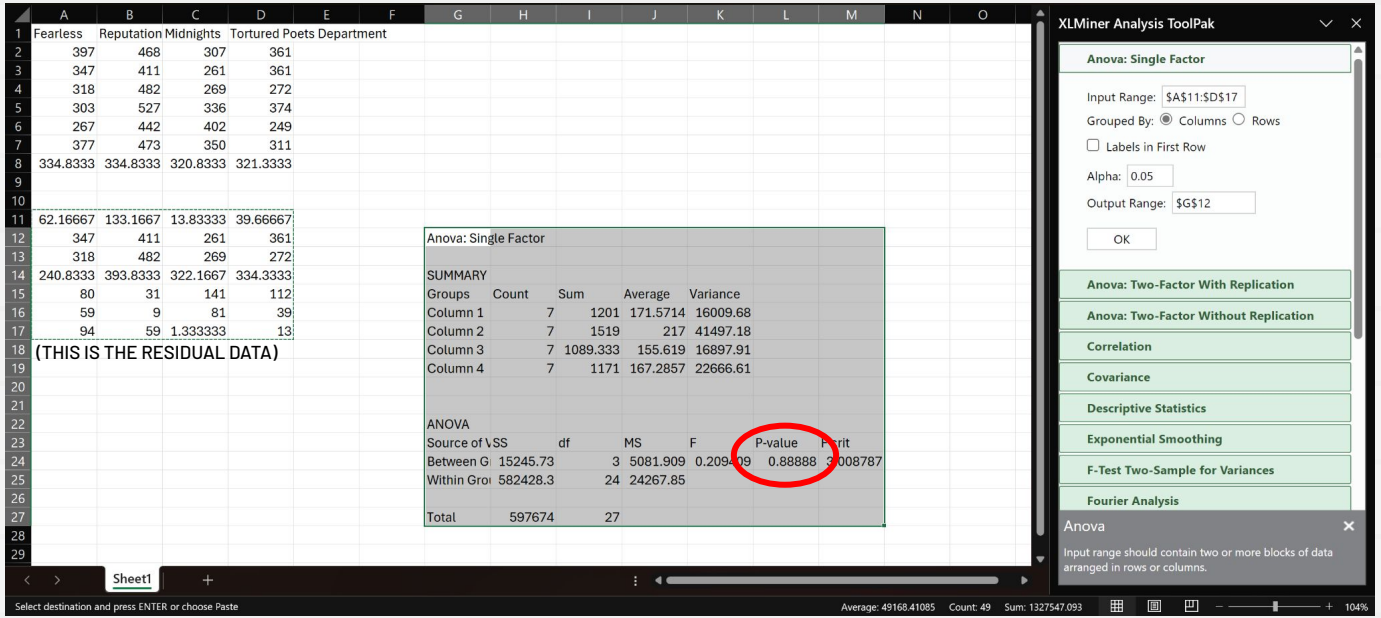
Back to this...

- Homogeneity assumption can be calculated using an ANOVA of the absolute residuals of each value (|mean-value|)
- If the p-value of the test is greater than .05, the samples of the variances are homogenous
- This is called Levene's test

Example formula:



Back to this...



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Fearless	Reputation	Midnights	Tortured Poets Department											
2	397	468	307	361											
3	347	411	261	361											
4	318	482	269	272											
5	303	527	336	374											
6	267	442	402	249											
7	377	473	350	311											
8	334.8333	334.8333	320.8333	321.3333											
9															
10															
11	62.16667	133.1667	13.83333	39.66667											
12	347	411	261	361											
13	318	482	269	272											
14	240.8333	393.8333	322.1667	334.3333											
15	80	31	141	112											
16	59	9	81	39											
17	94	59	1.333333	13											
18	(THIS IS THE RESIDUAL DATA)														
19															
20															
21															
22															
23															
24															
25															
26															
27															
28															
29															

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	7	1201	171.5714	16009.68
Column 2	7	1519	217	41497.18
Column 3	7	1089.333	155.619	16897.91
Column 4	7	1171	167.2857	22666.61

ANOVA

Source of V	SS	df	MS	F	P-value	F crit
Between G	15245.73	3	5081.909	0.209409	0.88888	3.008787
Within Gro	582428.3	24	24267.85			
Total	597674	27				

XLMiner Analysis ToolPak

Anova: Single Factor

Input Range: $A$11:$D$17

Grouped By: Columns Rows

Labels in First Row

Alpha: 0.05

Output Range: $G$12

OK

Anova: Two-Factor With Replication

Anova: Two-Factor Without Replication

Correlation

Covariance

Descriptive Statistics

Exponential Smoothing

F-Test Two-Sample for Variances

Fourier Analysis

Anova

Input range should contain two or more blocks of data arranged in rows or columns.

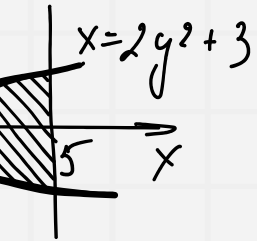
Sheet1

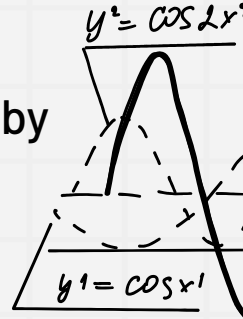
Select destination and press ENTER or choose Paste

Average: 49168.41085 Count: 49 Sum: 1327547.093 104%

Two-Way ANOVA

Does age affect how satisfied you are with life? A study was conducted with Male and Female adults, and groups were split by gender and age





Two-Way ANOVA

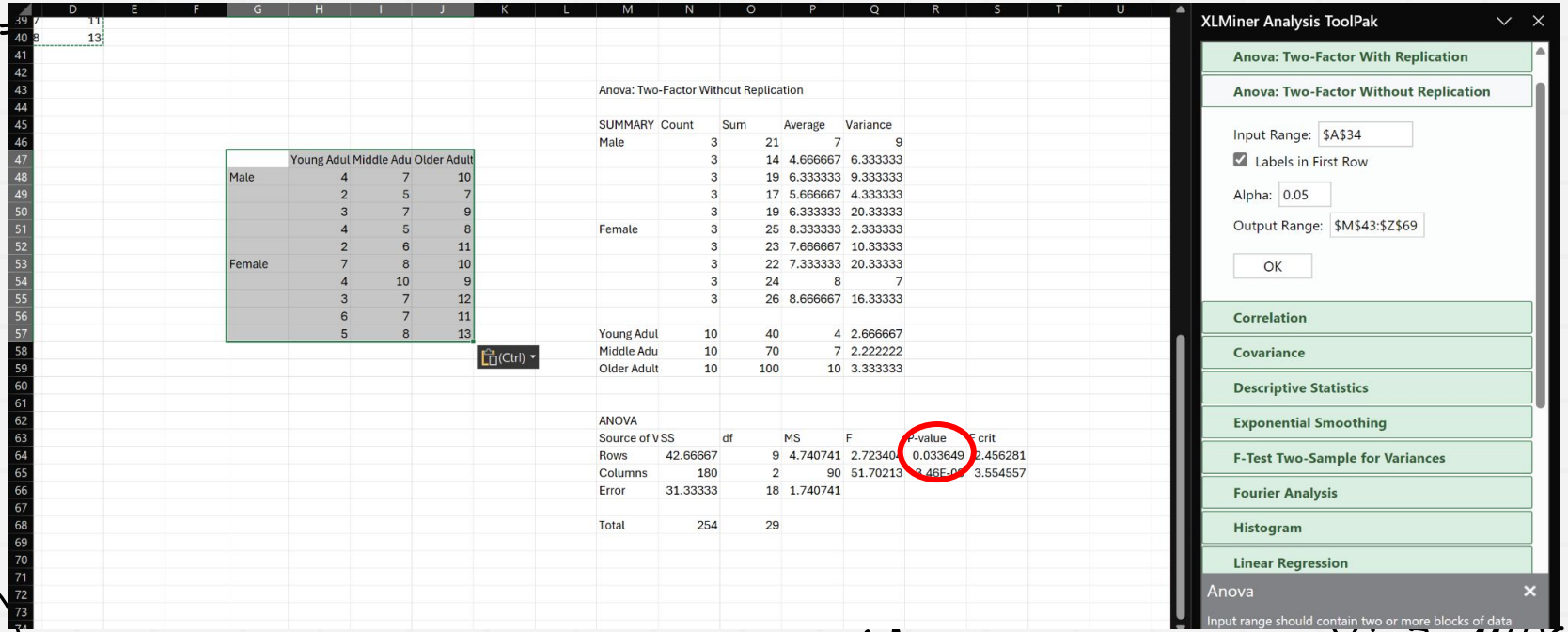
	Young Adul	Middle Adu	Older Adult
Male	4	7	10
	2	5	7
	3	7	9
	4	5	8
	2	6	11
Female	7	8	10
	4	10	9
	3	7	12
	6	7	11
	5	8	13

Assumptions

- stated random sample
- independent variable (album) is categorical, and dependent variable (word count) is quantitative
- Number of independent groups is greater than 2
- 2 different independent variables used
- Samples were independent
- Distribution of each group is approximately normal (histograms of each group were made in excel)
- variances of groups are homogenous (also conducted in excel)

How to do Two-Way ANOVA in Excel

(Ho: All groups are evenly distributed; Ha: At least one mu is different)



	Young Adul	Middle Adu	Older Adult
Male	4	7	10
	2	5	7
	3	7	9
	4	5	8
	2	6	11
Female	7	8	10
	4	10	9
	3	7	12
	6	7	11
	5	8	13

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
Male	3	21	7	9
	3	14	4.666667	6.333333
	3	19	6.333333	9.333333
	3	17	5.666667	4.333333
	3	19	6.333333	20.33333
Female	3	25	8.333333	2.333333
	3	23	7.666667	10.33333
	3	22	7.333333	20.33333
	3	24	8	7
	3	26	8.666667	16.33333
Young Adul	10	40	4	2.666667
Middle Adu	10	70	7	2.222222
Older Adult	10	100	10	3.333333

ANOVA

Source of V	SS	df	MS	F	P-value	F crit
Rows	42.66667	9	4.740741	2.723404	0.033649	2.456281
Columns	180	2	90	51.70213	2.46E-08	3.554557
Error	31.33333	18	1.740741			
Total	254	29				

XLMiner Analysis ToolPak

- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
  - Input Range: $A$34
  - Labels in First Row
  - Alpha: 0.05
  - Output Range: $M$43:$Z$69
  - OK
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram
- Linear Regression

Anova

Input range should contain two or more blocks of data

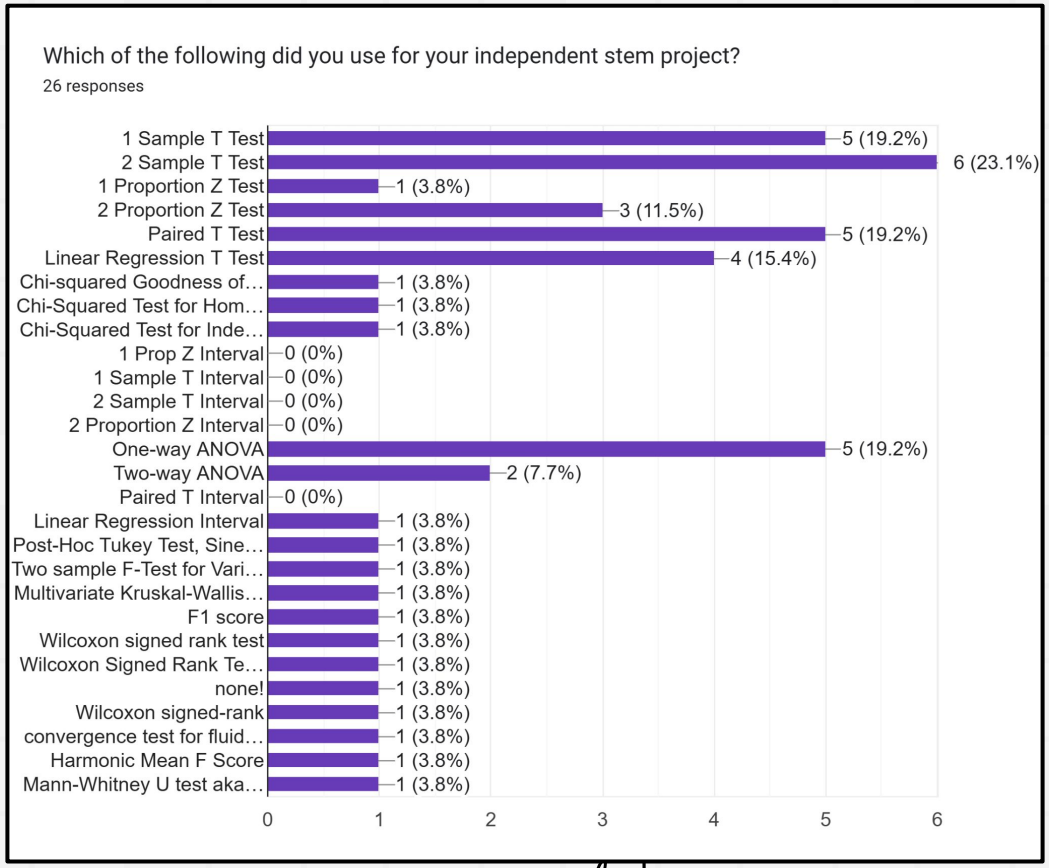
How to do Two-Way ANOVA in calculator

Two-way ANOVA can be conducted in a graphing calculator, however, a special program must be downloaded onto the calculator to conduct the test.

Conclusion

- Since the p-value is less than .05, reject the null hypothesis
- There is convincing evidence that at least one of the mus is different
- Post hoc tests can be used to determine which means are statistically significant from the rest (or using 2-sample t-tests between all the different means, but this can take a while)

Tests used last year by the seniors



Sources:

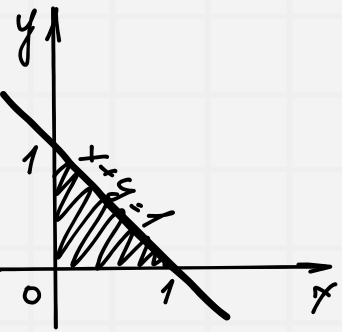
One-Way ANOVA - SPSS Tutorials - LibGuides at Kent State University

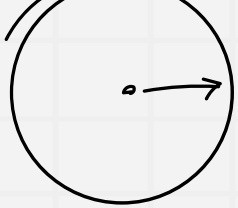
Analysis of variance (ANOVA) | Statistics and probability | Khan Academy

https://www.bing.com/videos/riverview/relatedvideo?q=what+is+anova+used+for&mid=3973B52CA7C308380D253973B52CA7C308380D25&FORM=VIRE&ajaxhist=0

https://www.bing.com/videos/riverview/relatedvideo?q=two+way+anova+excel&mid=5697B25327954A3FDC665697B25327954A3FDC66&FORM=VIRE

How to Perform Levene's Test in Excel



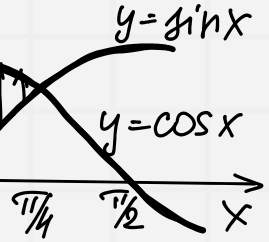


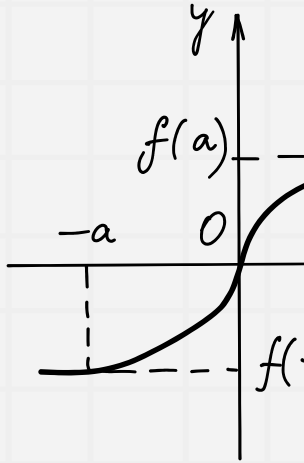
$S = 2\pi R$

Thanks!

Do you have any questions?

youremail@freepik.com
+34 654 321 432
yourwebsite.com





$2\sqrt{y^2 - x^2}$