ANOVA

One-Way ANOVA

- One independent variable
- Only one 'p' value is obtained.

Two way ANOVA

- Two independent Variables.
- Three different 'p' values are obtained.
- Outcome of factorial Design.

Why do an ANOVA?

- when there are 3 or more means being compared, statistical significance can be ascertained by conducting one statistical test, ANOVA, or by repeated t-tests.

- Why not conduct repeated t-tests?

- Each statistical test is conducted with a specified chance of making a type I–error—the alpha level.

Why do an ANOVA?

- **Why not several t-tests?**
- Imagine we have a design with three groups that have to be compared:

  G1, G2, G3
- We will have to run several separate t-tests

  (one to compare G1 with G2, one to compare G1 with G3, and one to compare G2 with G3)
- For every test we use a general α-level of 0.05

- **α -level=0.05**
- 5% possibility to make Type I error, i.e. rejecting H_0, when H_0 is actually true.
- Our scope is too reduce the possibilities to have Type I error
- If we were to run 3 separate t-tests to compare G1, G2 and G3, each with a α-level of 0.05, the overall possibility not to make Type I error would be 0.857 [i.e. (0.95)3]
- Therefore subtracting that from the overall possibility not to make Type 1 error (1=100%)

  1-0.857=0.14
- We have 14% of possibilities to make Type 1 error.
- 14% >> than the usual 5%
- We can't be happy with that!

Assumptions of one way ANOVA

- The data are randomly sampled and independently chosen from the populations
- The variances of each sample are assumed equal.
- The residuals are normally distributed.

State the null hypothesis and alpha level

The null hypothesis is that all the groups have equal means.

$$\boxed{H_0: \mu_1 = \mu_2 = \mu_3}$$

The alternative hypothesis is that there is at least one significant difference between the means

Level of significance α is selected as 0.05

Computing a one-way ANOVA

Here is the basic one-way ANOVA table

Source	SS	df	MS	F	p	F_{crit}
Between						
Within						
Total						

Sum of squares (SS)

- Sum of squares are nothing but sum of squared deviations around the mean i.e. $(X- \bar{x})^2$

- To calculate the various sums of squares,

  we need:

  i) the ΣX^2 (sum of each score squared) in the whole sample

  ii) the ΣX (sum of all scores) for the whole sample

  iii) and the Group ΣX (sum of scores separately for each group): represented as ΣX_g in as follows

	FRONT ROW	MIDDLE ROW	BACK ROW	
	8	13	9	
	8	11	11	
	6	15	14	
	9	14	12	
	10	14	9	
	13	12	11	
	9	11	12	
	11	11	11	
	8	11	9	
	11	11	9	
	10	17	11	
	9	13	11	
$\Sigma X_g =$	112	153	129	= 394
$\Sigma X^2 =$	1082	1993	1413	= 4488
Mean =	9.33	12.75	10.75	

	FRONT ROW	MIDDLE ROW	BACK ROW	TOTAL
ΣX_g	112	153	129	394
ΣX^2	1082	1993	1413	4488

- $SS_{total} = \Sigma X^2 - \dfrac{(\Sigma X)^2}{N}$

$$= 4488 - 394^2 / 36$$

$$= 4488 - 4312.11$$

$$= 175.88$$

	FRONT ROW	MIDDLE ROW	BACK ROW	TOTAL
ΣX_g	112	153	129	394
ΣX^2	1082	1993	1413	4488

- $$SS_{between} = \frac{(\Sigma X_g)^2}{n} - \frac{(\Sigma X)^2}{N}$$

$$= \frac{112^2 + 153^2 + 129^2}{12} - (394^2 / 36)$$

$$= 4382.83 - 4312.11$$

$$= \mathbf{70.72}$$

$$SS_{total} = 175.88$$

$$SS_{between} = 70.72$$

$$SS_{within} = SS_{total} - SS_{between}$$

$$= 175.88 - 70.72$$

$$= 105.16$$

After filling in the sum of squares, we have …

Source	SS	df	MS	F	p	F_{crit}
Between	70.72					
Within	105.16					
Total	175.88					

Degrees of Freedom

- The degrees of freedom, noted in are calculated as N_i -1 for the total(Ni is the total number of observations);

- number of groups minus one for the between groups;

- And for the within error, subtract d.f. for groups from the total degrees of freedom.

So ,

Total degrees of freedom = total number of observations - 1

= 36-1

= **35**

Degrees of freedom between groups = Total number of groups – 1

= 3-1

= **2**

Degrees freedom within groups = Total d.f – Between group d.f

= 35-2

= **33**

- Filling in the degrees of freedom gives this …

Source	SS	df	MS	F	p	F_{crit}
Between	70.72	2				
Within	105.16	33				
Total	175.88	35				

Mean Squares-MS (Variance)

- **Why variance estimates are called Mean Squares?**

  A population variance is equal to the sum of the squared deviations about the mean divided by N.

$$\frac{\Sigma(X - \bar{x})^2}{N}$$

So the population variance is really the mean of the squared deviations about the mean, or the mean square(d) deviation about the mean. This is why the term mean square is used in place of variance.

It is obtained by dividing each sum of squares with corresponding degrees of freedom.

So,

$Ms_{between}$	MS_{within}
$= 70.7222 / 2$	$= 105.1667 / 33$
$= 35.36$	$= 3.18$

Source	SS	df	MS	F	p	F_{crit}
Between	70.72	2	35.36			
Within	105.16	33	3.18			
Total	175.88	35				

$F_{calculated}$ value

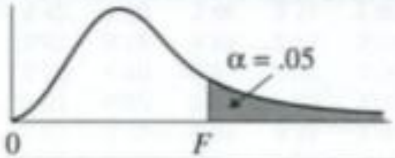
- It is the ratio of $MS_{between}$ and MS_{within}

$$F = MS(B) / MS(W)$$

- So, $F = \dfrac{35.3611}{3.1869} = 11.10$

Source	SS	df	MS	F	p	F_{crit}
Between	70.72	2	35.36	11.10		
Within	105.16	33	3.18			
Total	175.88	35				

Critical Values of the F Distribution for $\alpha = .05$



df NUMERATOR

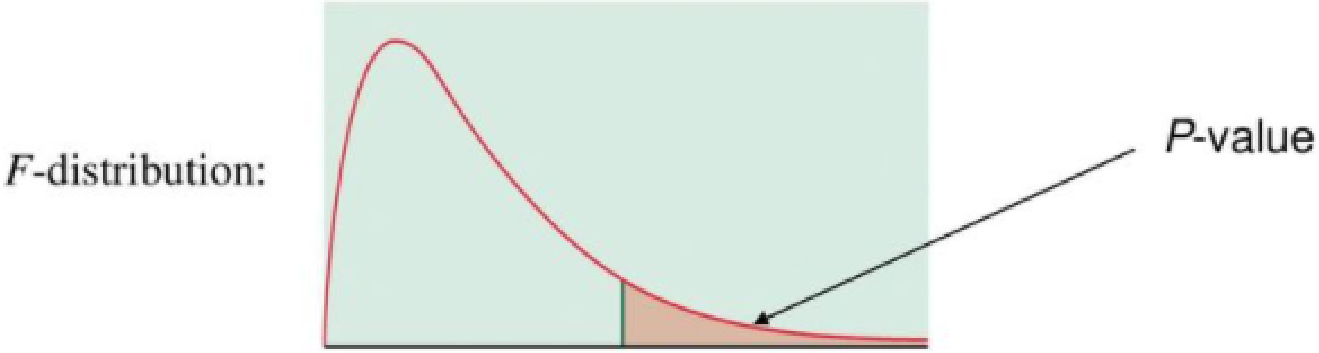
df Denominator	1	2	3	4	5	6	7	8	9	10	12	15	20	25	30	40	60	120	∞
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.14	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.07	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.04	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.02	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.00	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.97	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.95	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.94	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.92	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.88	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.78	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.69	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.60	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.51	1.46	1.39	1.32	1.22	1.00

Putting value of F crit to ANOVA table

Source	SS	df	MS	F	p	F_{crit}
Between	70.72	2	35.36	11.10		3.32
Within	105.16	33	3.18			
Total	175.88	35				

P- value

- It is the area under the F distribution curve that is to the right of your observed F-statistic.
- This area may be obtained by integral calculus



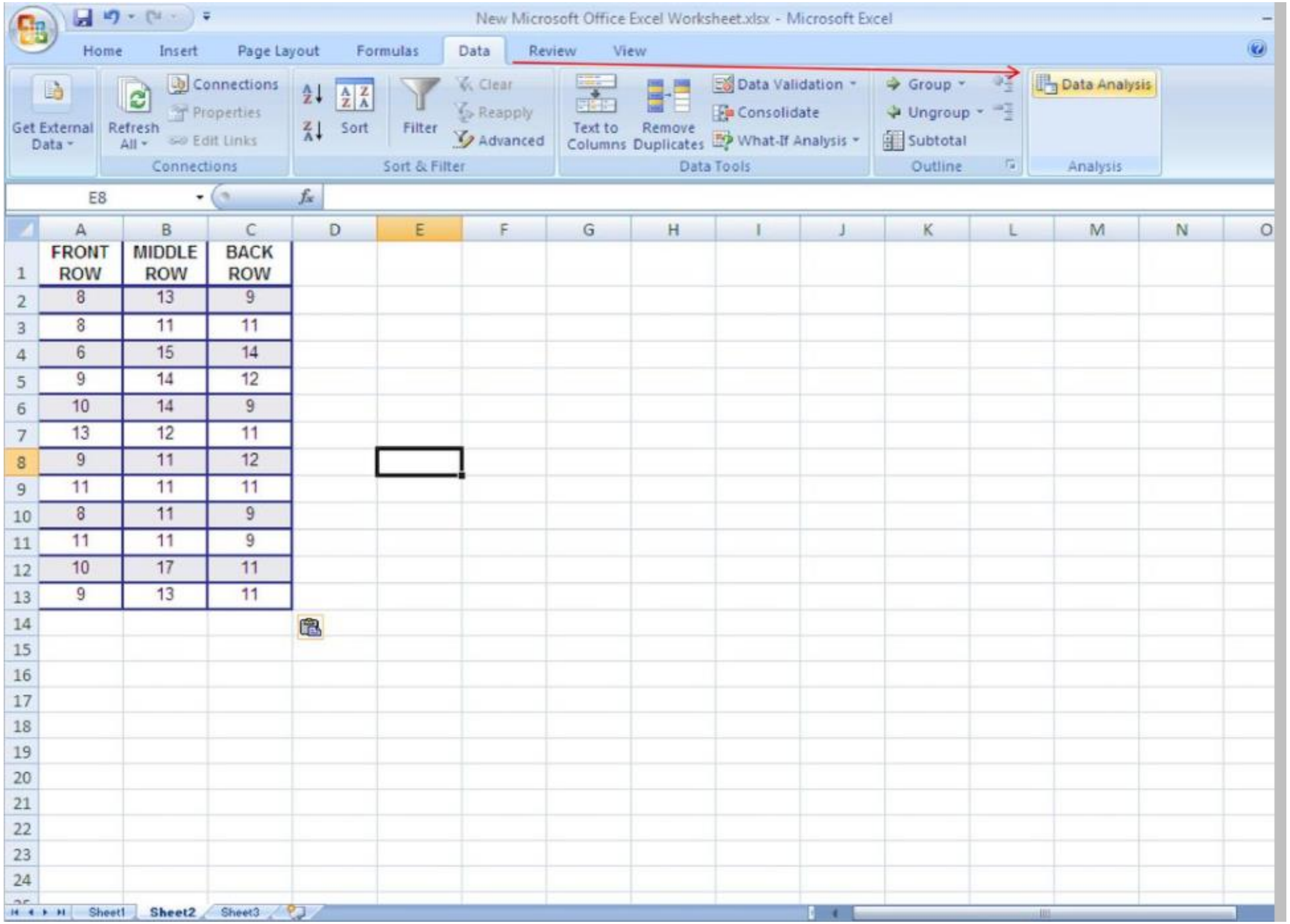
- If the *P*-value is less than or equal to a, reject H_0.
- It the *P*-value is greater than a, fail to reject H_0.

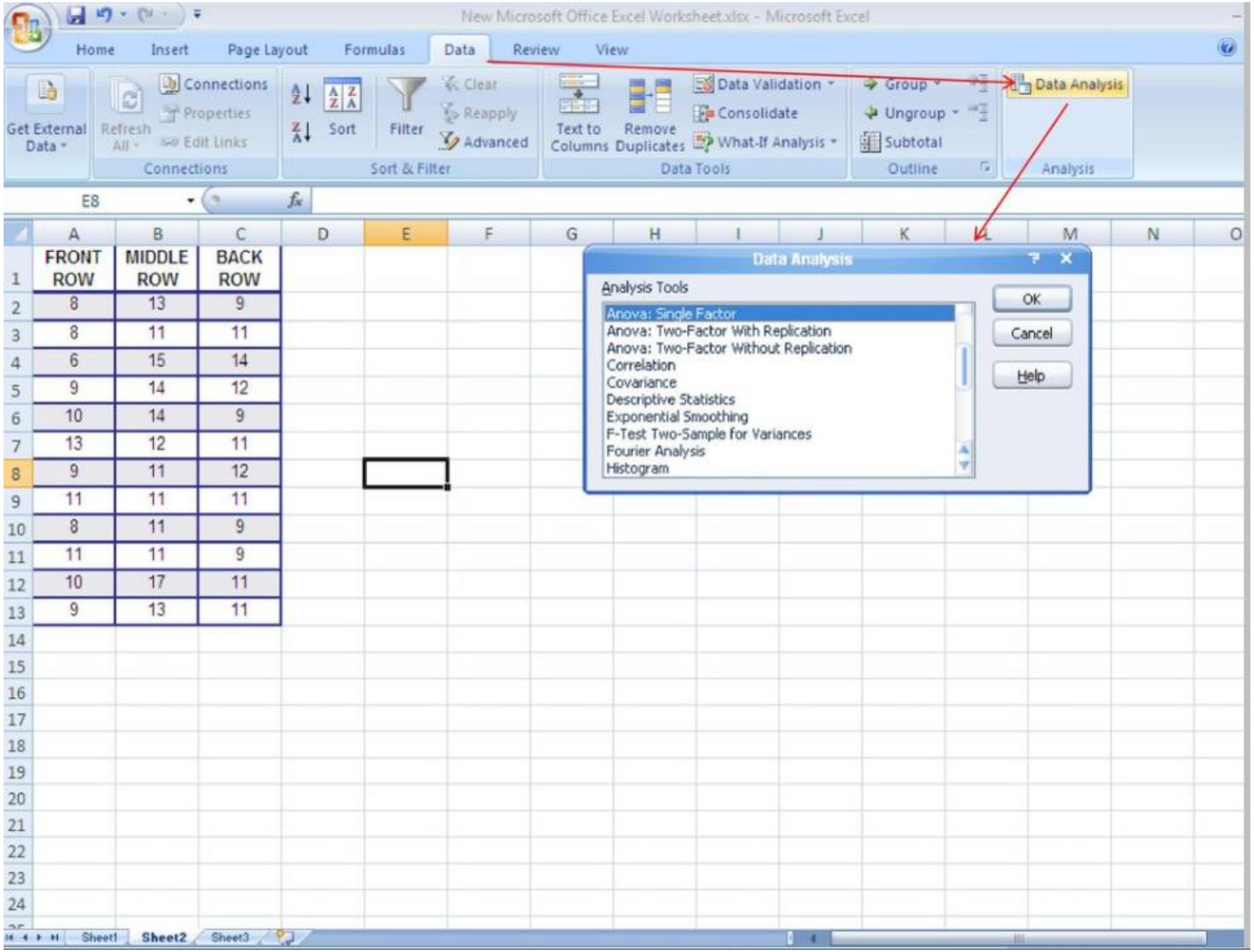
One-Way ANOVA: Interpretation

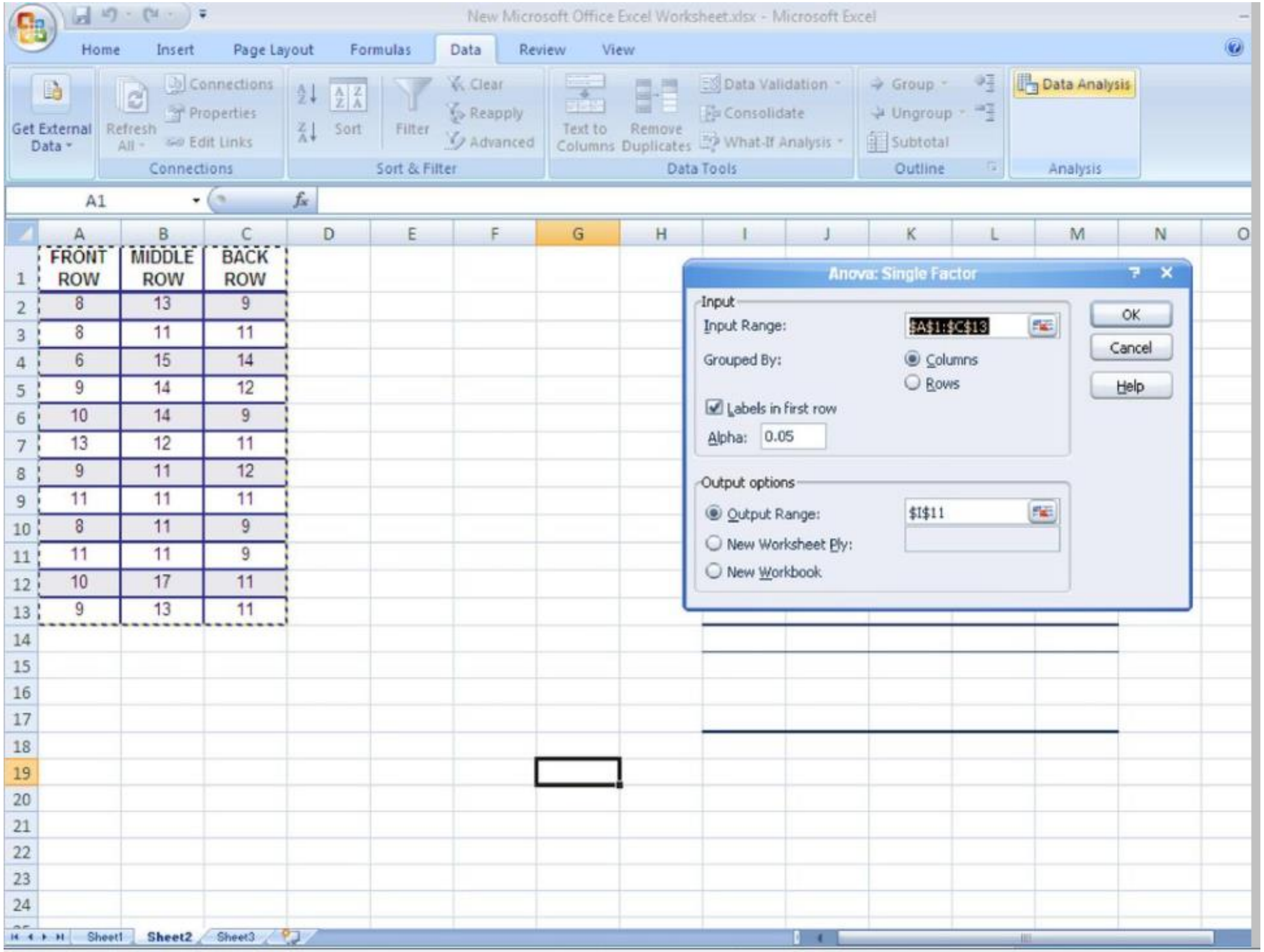
- Our obtained F of 11.10 is greater than the critical F of 3.32 i.e. $F_{calculated} > F_{critical}$ and p- value is less than 0.05

- We reject the null hypothesis and accept the alternative hypothesis that there is at least a difference between two of the group means.

- If $F_{cal} < F_{crit}$, We fail to reject the null hypothesis and conclude that there are no significant differences between the group means.


New Microsoft Office Excel Worksheet.xlsx - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Get External Data
Refresh All
Connections
Properties
Edit Links
Connections
Sort
Filter
Clear
Reapply
Advanced
Sort & Filter
Text to Columns
Remove Duplicates
Data Validation
Consolidate
What-If Analysis
Data Tools
Group
Ungroup
Subtotal
Outline
Data Analysis
Analysis
E8
FRONT ROW
MIDDLE ROW
BACK ROW
8 13 9
8 11 11
6 15 14
9 14 12
10 14 9
13 12 11
9 11 12
11 11 11
8 11 9
11 11 9
10 17 11
9 13 11
Sheet1
Sheet2
Sheet3


New Microsoft Office Excel Worksheet.xlsx - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Get External Data
Refresh All
Connections
Properties
Edit Links
Connections
Sort
Filter
Clear
Reapply
Advanced
Sort & Filter
Text to Columns
Remove Duplicates
Data Validation
Consolidate
What-If Analysis
Data Tools
Group
Ungroup
Subtotal
Outline
Data Analysis
Analysis
E8
FRONT ROW
MIDDLE ROW
BACK ROW
8 13 9
8 11 11
6 15 14
9 14 12
10 14 9
13 12 11
9 11 12
11 11 11
8 11 9
11 11 9
10 17 11
9 13 11
Data Analysis
Analysis Tools
Anova: Single Factor
Anova: Two-Factor With Replication
Anova: Two-Factor Without Replication
Correlation
Covariance
Descriptive Statistics
Exponential Smoothing
F-Test Two-Sample for Variances
Fourier Analysis
Histogram
OK
Cancel
Help
Sheet1
Sheet2
Sheet3


New Microsoft Office Excel Worksheet.xlsx - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Get External Data
Refresh All
Connections
Properties
Edit Links
Connections
Sort
Filter
Clear
Reapply
Advanced
Sort & Filter
Text to Columns
Remove Duplicates
Data Validation
Consolidate
What-If Analysis
Data Tools
Group
Ungroup
Subtotal
Outline
Data Analysis
Analysis
A1
FRONT ROW
MIDDLE ROW
BACK ROW
8 13 9
8 11 11
6 15 14
9 14 12
10 14 9
13 12 11
9 11 12
11 11 11
8 11 9
11 11 9
10 17 11
9 13 11
Anova: Single Factor
Input
Input Range:
$A$1:$C$13
Grouped By:
Columns
Rows
Labels in first row
Alpha:
0.05
Output options
Output Range:
$I$11
New Worksheet Ply:
New Workbook
OK
Cancel
Help
Sheet1
Sheet2
Sheet3

FRONT ROW	MIDDLE ROW	BACK ROW
8	13	9
8	11	11
6	15	14
9	14	12
10	14	9
13	12	11
9	11	12
11	11	11
8	11	9
11	11	9
10	17	11
9	13	11

Anova: Single Factor

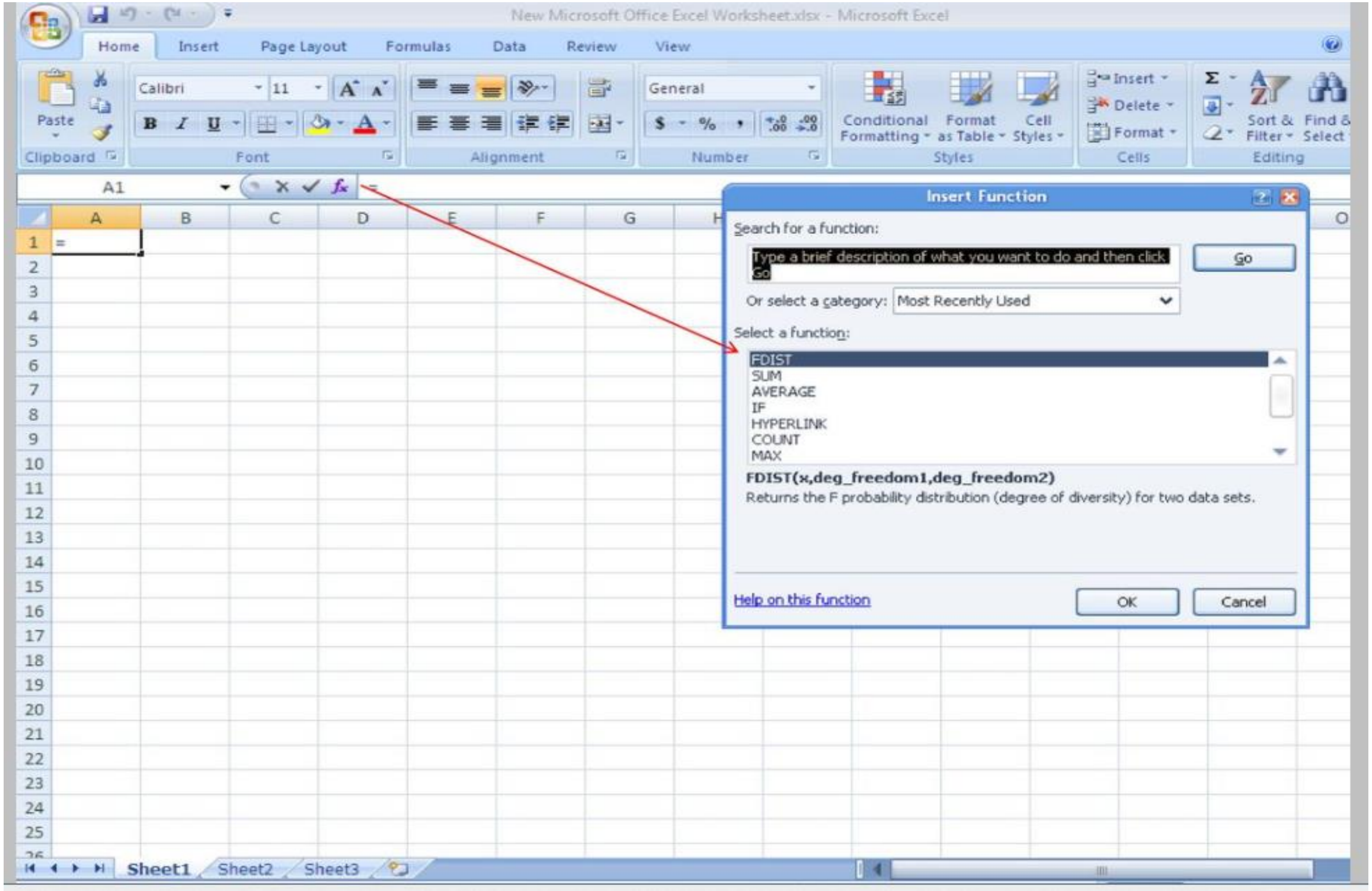
SUMMARY

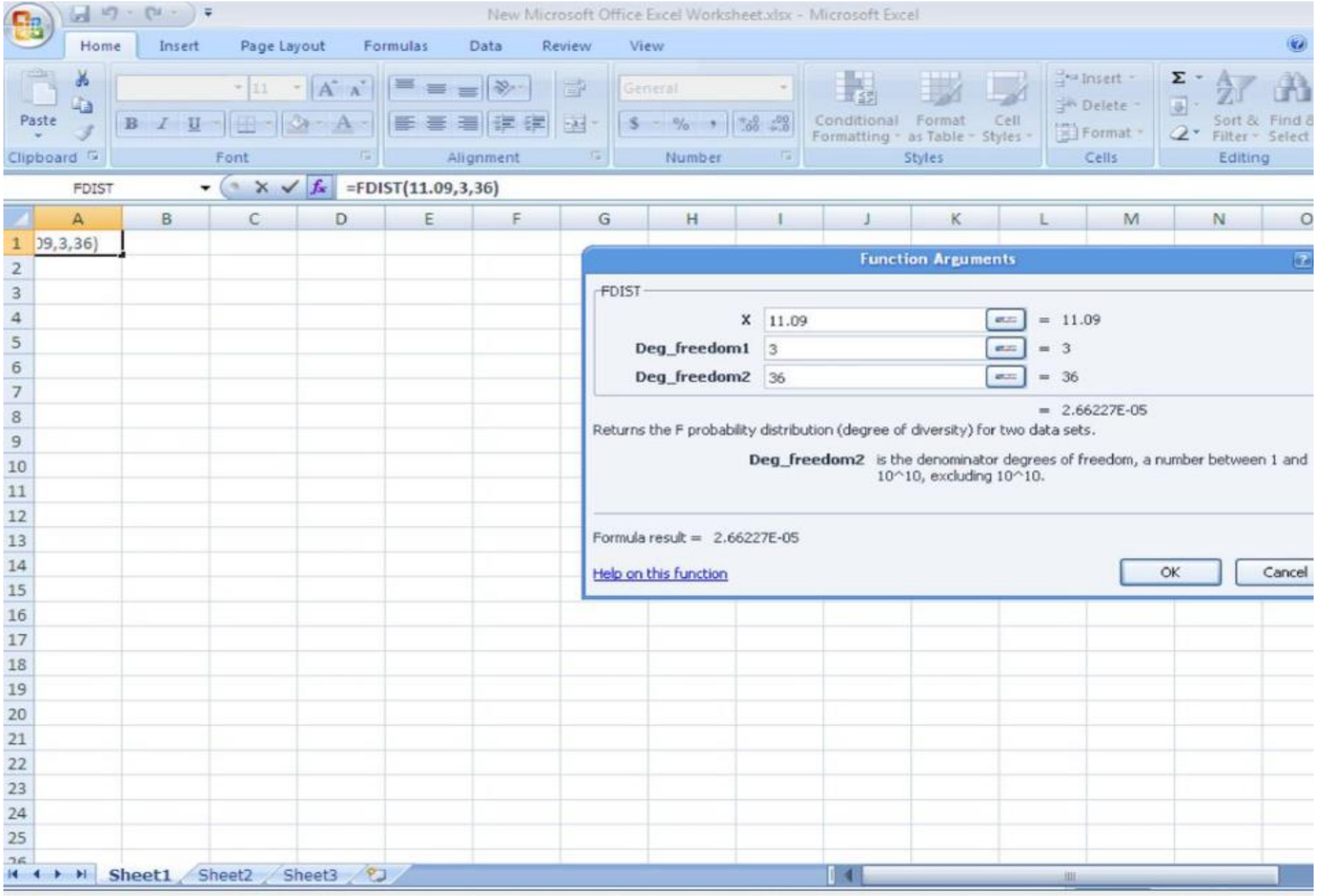
Groups	Count	Sum	Average	Variance
FRONT RC	12	112	9.333333	3.333333
MIDDLE R(	12	153	12.75	3.840909
BACK ROV	12	129	10.75	2.386364

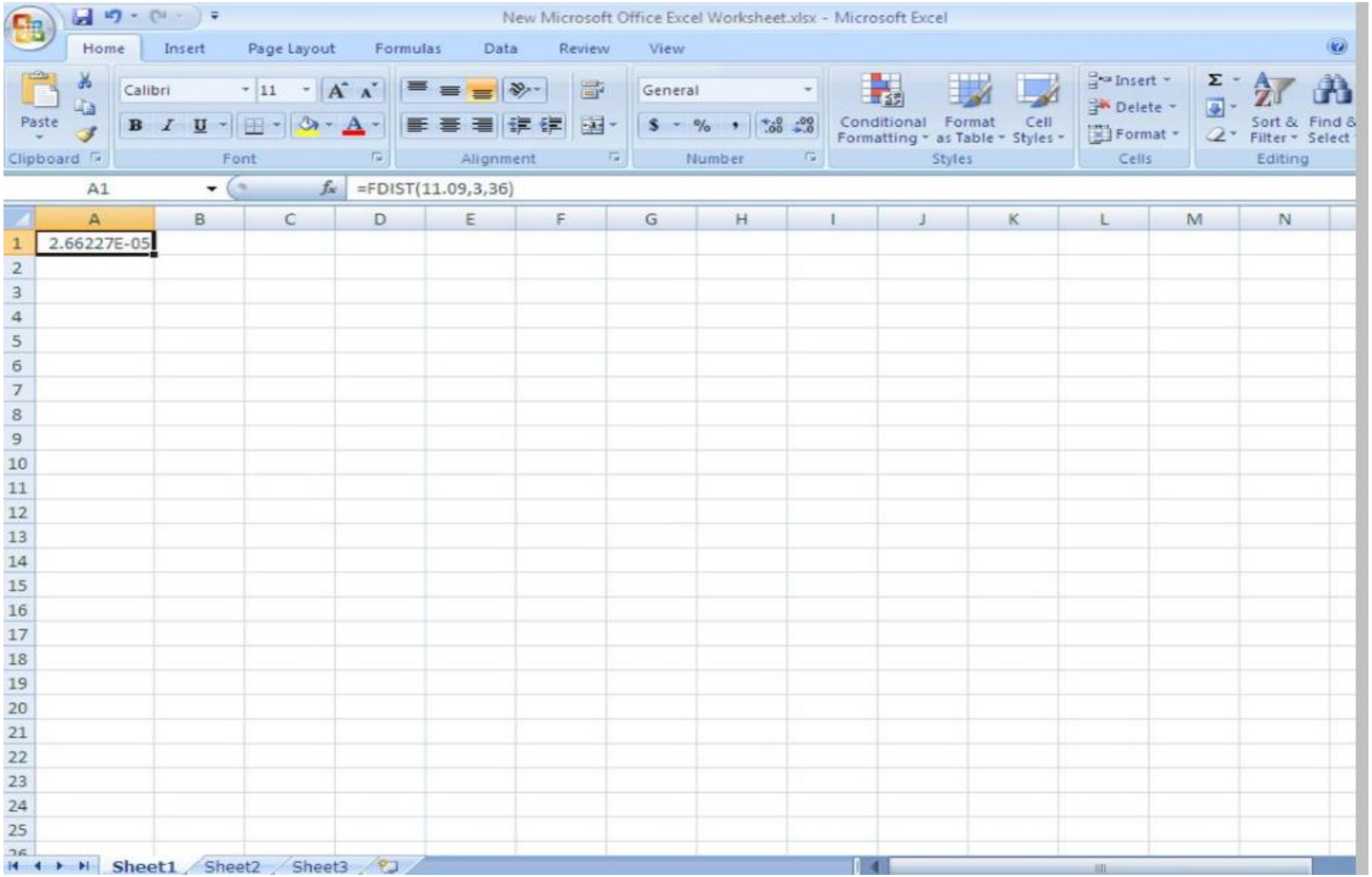
ANOVA

Source of Varic	SS	df	MS	F	P-value	F crit
Between (	70.72222	2	35.36111	11.09588	0.000206	3.284918
Within Gr	105.1667	33	3.186869			
Total	175.8889	35				

HOW TO CALCULATE
P- VALUE IN EXCEL?


New Microsoft Office Excel Worksheet.xlsx - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Paste
Clipboard
Calibri
11
Font
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
Sort & Filter
Find & Select
Editing
A1
Insert Function
Search for a function:
Type a brief description of what you want to do and then click Go
Go
Or select a category: Most Recently Used
Select a function:
FDIST
SUM
AVERAGE
IF
HYPERLINK
COUNT
MAX
FDIST(x,deg_freedom1,deg_freedom2)
Returns the F probability distribution (degree of diversity) for two data sets.
Help on this function
OK
Cancel
Sheet1
Sheet2
Sheet3


New Microsoft Office Excel Worksheet.xlsx - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Paste
Clipboard
Font
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
Sort & Filter
Find & Select
Editing
FDIST
=FDIST(11.09,3,36)
09,3,36)
Function Arguments
FDIST
X 11.09 = 11.09
Deg_freedom1 3 = 3
Deg_freedom2 36 = 36
= 2.66227E-05
Returns the F probability distribution (degree of diversity) for two data sets.
Deg_freedom2 is the denominator degrees of freedom, a number between 1 and 10^10, excluding 10^10.
Formula result = 2.66227E-05
Help on this function
OK
Cancel
Sheet1
Sheet2
Sheet3


New Microsoft Office Excel Worksheet.xlsx - Microsoft Excel
Home
Insert
Page Layout
Formulas
Data
Review
View
Paste
Clipboard
Calibri
11
Font
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
Sort & Filter
Find & Select
Editing
A1
=FDIST(11.09,3,36)
2.66227E-05
Sheet1
Sheet2
Sheet3