ISSN No. 0976-5697

Volume 8, No. 6, July 2017 (Special Issue III)

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

ONE-WAY ANOVA CALCULATIONS USING MS-EXCEL

S. Mahendra
Student in Department of Computers,
SSDC & PG College,
Kodur, India

S. Upendra
Research scholar in Department of Statistics,
SVU, Tirupati, India

Abstract - This paper proposes the concept of ANOVA, one-way ANOVA and assumptions of ANOVA. This accommodates linear model in one-way ANOVA with estimation of parameters in linear model, different sum of squares, degrees of freedom's, mean sum of squares, ANOVA, the inference about the analysis and also explain the MS-EXCEL procedure for calculations of one-way ANOVA.

Keywords – ANOVA, Linear model, Estimation of Parameters, Sum of Squares, degrees of freedom and ANOVA table.

ANOVA: The **Analysis Of Variance** (ANOVA) was introduced by Prof. R. A. Fisher in the year 1920's; to deal with the problem in the analysis of agronomical data variation is inherent in nature. The main aim of ANOVA is used to test the homogeneity of several means.

"According to Prof. R. A. Fisher, ANOVA is the separation of one group of causes from the variation of another group of causes".

In other words, splitting up of total variation into variation due to several factors, variation due to error and compare the factors variation against the error variation by using F- test statistic is called ANOVA.

ANOVA Assumptions

1. The experimental errors of your data are normally distributed
2. Equal variances between treatments, i.e., Homogeneity of variances, Homoscedasticity
3. Independence of samples Each sample is randomly selected and independent

One-way ANOVA

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

The data of one-way classification as follows:

(see table-1)

Where

GT=Grand Total= $\sum_i \sum_j X_{ij}$; i =1, 2, -----, k, j=1, 2, ----, n.

Linaer model:

The mathematical linear model for one-way ANOVA is given by

$X_{ij} = \mu + t_i + \varepsilon_{ij}$; i =1, 2, -----, k, j=1, 2, ----, n.

Where X_{ij} = the jth observation in ith sub group

μ = group mean

t_i = the ith sub group effect

ε_{ij} = the random error

Null-hypothesis:

The one-way ANOVA compares the means between the groups you are interested in and determines whether any of those means are statistically significantly different from each other. Specifically, it tests the null hypothesis:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

Where μ = group mean and k = number of sub-groups

Estimation of parameters:

We estimate the parameters by using principles of least squares as follows:

Consider the residual as

$$\varepsilon_{ij} = y_{ij} - \mu - t_i$$

The residual sum of squares

$$\varepsilon_{ij}^{\ 2} = \left(y_{ij} - \mu - t_i\right)^2$$

We can estimate the parameters by minimizing the above residual sum of square, then we get

$$\frac{\partial \varepsilon_{ij}^{\ 2}}{\partial \mu} = 0 \text{, then } \hat{\mu} = \frac{\sum_i \sum_j x_{ij}}{nk} = \frac{GT}{nk}$$

$$\hat{t}_i = \frac{\sum_i (x_{ij})}{k} - \frac{\sum_i \sum_j x_{ij}}{nk}$$

And $$\frac{\partial \varepsilon_{ij}^{\ 2}}{\partial t_i} = 0 \quad \hat{t}_i = \frac{\sum_i (x_{ij})}{k} - \frac{GT}{nk}$$

$$\hat{t}_i = \frac{\sum_i (x_{ij})}{k} - \hat{\mu}$$

$$\because \hat{\mu} = \frac{GT}{nk}$$

Sum of Squares:

To test the above null-hypothesis by using the following various Sum of Squares(SS) as follows:

CONFERENCE PAPER
National Conference dated 27-28 July 2017 on
Recent Advances in Graph Theory and its Applications (NCRAGTA2017)
Organized by Dept of Applied Mathematics
Sri Padmawati Mahila Vishvavidyalayam (Women's University) Tirupati, A.P., India

$$TSS = \sum_i \sum_j (X_{ij} - \mu)^2$$

$$TSS = \sum_i \sum_j (X_{ij} - \overline{X}_i + \overline{X}_i - \mu)^2$$

$$TSS = \sum_i \sum_j \left[(X_{ij} - \overline{X}_i)^2 + (\overline{X}_i - \mu)^2 + 2(X_{ij} - \overline{X}_i)(\overline{X}_i - \mu) \right]$$

$$TSS = \sum_i \sum_j (X_{ij} - \overline{X}_i)^2 + \sum_i \sum_j (\overline{X}_i - \mu)^2 + 2\sum_i \sum_j (X_{ij} - \overline{X}_i)(\overline{X}_i - \mu)$$

$$TSS = \sum_i \sum_j (X_{ij} - \overline{X}_i)^2 + \sum_i \sum_j (\overline{X}_i - \mu)^2$$

$$TSS = SS_b + SS_w$$

Where, TSS= Total Sum of Squares

SS_b= Sum of squares between Sub groups

SS_w= Sum of Squares with in sub groups (error Sum of Squares)

Then, $SS_w = TSS - SS_b$

We calculate TSS and SS_b by using following formulae,

$$TSS = \sum_i \sum_j X_{ij}^2 - c.f.$$

$$SSb = \sum_i T_i^2 - c.f.$$

Here, $c.f = \dfrac{GT^2}{nk} = \dfrac{\sum_i \sum_j X_{ij}^2}{nk}$

Degrees Of Freedom:

The degrees of freedom for different sum of squares are given by

Degrees of freedom for TSS is (nk-1)

Degrees of freedom for SS_b is (k-1)

Degrees of freedom for SS_w is k (n-1)

Mean Sum of Squares:

The Mean Sum of Squares [MSS] is calculated as the Sum of Squares dividing with their degrees of freedoms. They are given by

MS_b=MSS for sub groups= $\dfrac{SSb}{(k-1)}$

MS_w =MSS for error= $\dfrac{SSw}{k(n-1)}$

ANOVA Table:

To test the null-hypothesis

H_0 : All the sub groups' effects are equal.

$t_1 = t_2 = ---- = t_i = -- = t_k$

(see table-2)

Inference:

We compare F-calculated value with F-Critical (table) value at specified level of significance for [(k-1), k (n-1)] degrees of freedom and conclusions are drawn accordingly.

MS-EXCEL procedure for calculations of one-way ANOVA

The ANOVA analysis tools provide different types of variance analysis. The tool that you should use depends on the number of factors and the number of samples that you have from the populations that you want to test.

ANOVA: Single Factor

This tool performs a simple analysis of variance on data for two or more samples. The analysis provides a test of the hypothesis that each sample is drawn from the same underlying probability distribution against the alternative hypothesis that underlying probability distributions are not the same for all samples. If there are only two samples, you can use the worksheet function t-Test. With more than two samples, there is no convenient generalization of t-Test, and the Single Factor ANOVA model can be called upon instead.

ANOVA: Single Factor dialog box

Input Range:

Enter the cell reference for the range of data that you want to analyze. The reference must consist of two or more adjacent ranges of data arranged in columns or rows.

Grouped by To indicate whether the data in the input range is arranged in rows or in columns, click **Rows** or **Columns**.

Labels in First Row/Labels in First Column:

If the first row of your input range contains labels, select the **Labels in First Row** check box. If the labels are in the first column of your input range, select the **Labels in First Column** check box. This check box is clear if your input range has no labels. Microsoft Office Excel generates the appropriate data labels for the output table.

Alpha:

Enter the level at which you want to evaluate the critical values for the F statistic. The alpha level is a significance level that is related to the probability of having a type I error (rejecting a true hypothesis).

Output Range:

Enter the reference for the upper-left cell of the output table. Excel automatically determines the size of the output area and displays a message if the output table will replace existing data or extend beyond the bounds of the worksheet.

New Worksheet Ply:

Click to insert a new worksheet in the current workbook, and paste the results starting at cell A1 of the new worksheet. To name the new worksheet, type a name in the box.

New Workbook:

Click to create a new workbook in which results are added to a new worksheet.

ANOVA Tables in Excel—using Different Methods of Teaching Example

1. You need the methods of teaching Data format

2..You may need to install an add-in to Excel to do an AN OVA table:

2.1.If you click the "Data" tab on the menu bar, do you hav e a tab that says "Analysis", with a "Data Analysis" icon on it?

2.2. If yes, you're good to go.

3. Go to "Data" tab on menu bar.

CONFERENCE PAPER
National Conference dated 27-28 July 2017 on
Recent Advances in Graph Theory and its Applications (NCRAGTA2017)
Organized by Dept of Applied Mathematics
Sri Padmawati Mahila Vishvavidyalayam (Women's University) Tirupati, A.P., India

4. Go to “Analysis” tab
5. Click the “Data Analysis” icon
6. Select “ANOVA: Single Factor (One-way ANOVA)”
7. Click the table icon beside “Source data”
8. Select all rows and columns of methods of teaching data, plus the row and column containing the label:

8.1. In the example, this is A3to D11

8.2. Enter

9. For “Rows per sample” enter number of replicates

10.

For alpha, enter your pre-chosen significance level. Tell me in your report what you want that to be.

11. Click the table icon for “Output range”

11.1.

Point to one cell where you want the output table to begin.

11.2. Enter

12. OK

Tables and figures:

Sub groups	**Observations**						**total**	**mean**
	1	**2**	**------**	**j**	**-----**	**n**		
1	X_{11}	X_{12}	------	X_{1j}	------	X_{1n}	T_1	$\bar{X}_1$
2	X_{21}	X_{22}	------	X_{2j}	------	X_{2n}	T_2	$\bar{X}_2$
⋮	⋮	⋮	------	⋮	------	⋮	⋮	⋮
i	X_{i1}	X_{i2}	------	X_{ij}	------	X_{in}	T_i	$\bar{X}_i$
⋮	⋮	⋮	------	⋮	------	⋮	⋮	⋮
k	X_{k1}	X_{k2}	------	X_{kj}	------	X_{kn}	T_K	$\bar{X}_k$
							GT	

Table-1

Source	Sum of Squares	Degrees of freedom	Mean Sum of Squares	F- Ratio	F- cri	*Sig.*
Between	SS_b	k-1	MS_b	MS_b/MS_w	$F_{[(k-1),k(n-1)]}$	*p* value
Within	SS_w	k(n-1)	MS_w			
Total	$SS_b + SS_w$	nk-1				

Table-2

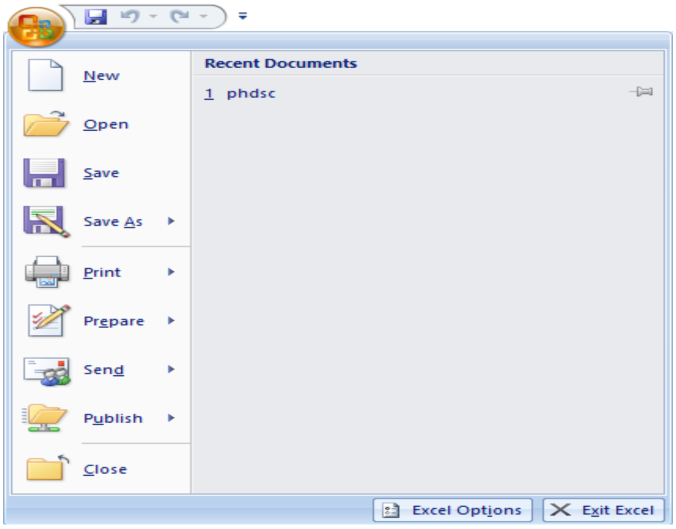

Figure-1: Install an add-in to Excel

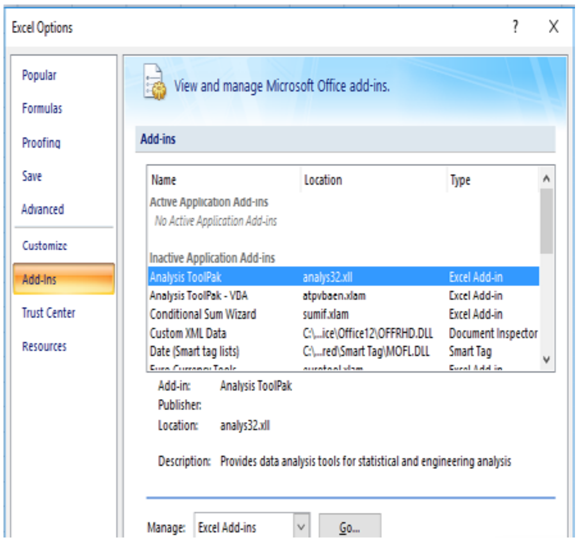

Figure-2.1: Install Analysis Tool pack

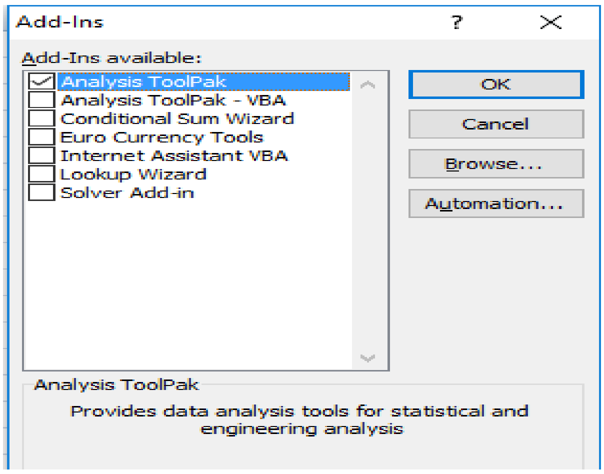

Figure-2.2: Install Analysis Tool pack

CONFERENCE PAPER
National Conference dated 27-28 July 2017 on
Recent Advances in Graph Theory and its Applications (NCRAGTA2017)
Organized by Dept of Applied Mathematics
Sri Padmawati Mahila Vishvavidyalayam (Women's University) Tirupati, A.P., India

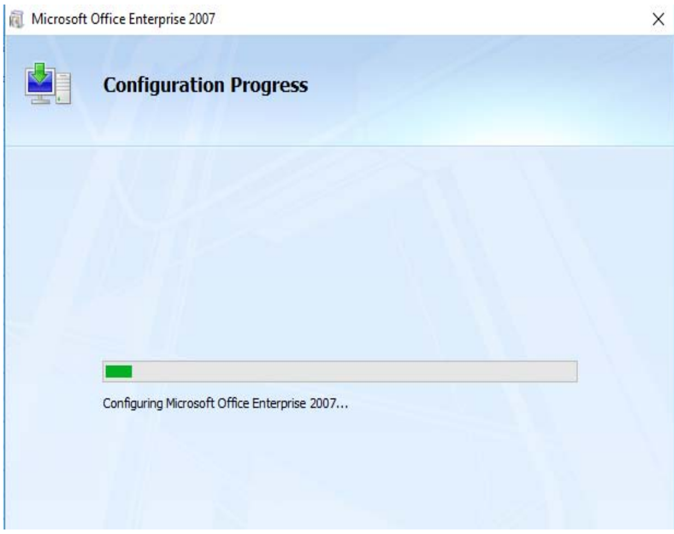

Figure-2.3: Install Analysis Tool pack

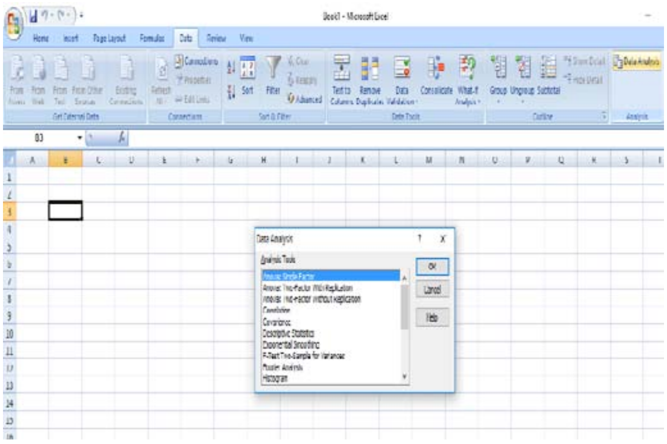

Figure-3: Selection Data Analysis of Analysis Tool pack in Data menu

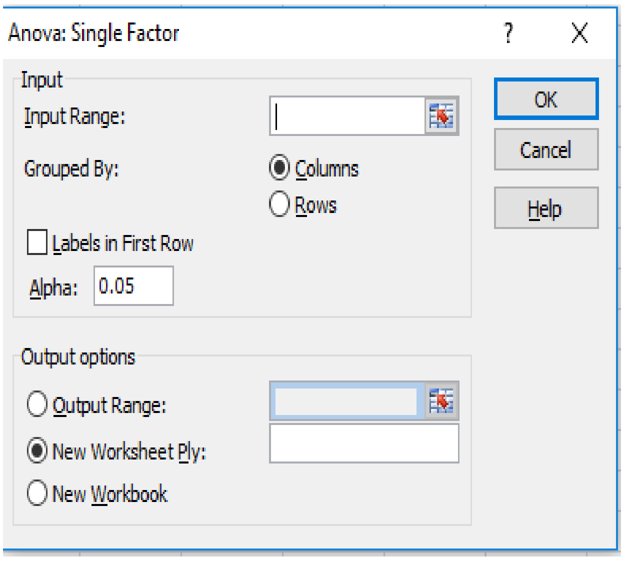

Figure-4: Fields for ANOVA: Single Factor

	A	B	C	D	E	F	G	H	I	J	K	L
1	One-way ANOVA											
2												
3	Method 1	Method 2	Method 3	Method 4		Anova: Single Factor						
4	51	82	79	85								
5	87	91	84	80		SUMMARY						
6	50	92	74	65		Groups	Count	Sum	Average	Variance		
7	48	80	98	71		Method 1	8	483	60.375	214.2679		
8	79	52	63	67		Method 2	8	623	77.875	157.5536		
9	61	79	83	51		Method 3	8	624	78	164.5714		
10	53	73	85	63		Method 4	8	575	71.875	181.5536		
11	54	74	58	93								
12												
13						ANOVA						
14						Source of Variation	SS	df	MS	F	P-value	F crit
15						Between Groups	1645.34	3	548.448	3.055648	0.04466	2.94669
16						Within Groups	5025.63	28	179.487			
17												
18						Total	6670.97	31				

Figure 5– Data and output from Anova: Single Factor data analysis tool

	A	B	C	D	E	F	G	H	I	J	K
1	Error terms										
2											
3		x_1	x_2	x_3	x_4			e_1	e_2	e_3	e_4
4		51	82	79	85			-9.375	4.125	1	13.125
5		87	91	84	80			26.625	13.125	6	8.125
6		50	92	74	65			-10.375	14.125	-4	-6.875
7		48	80	98	71			-12.375	2.125	20	-0.875
8		79	52	63	67			18.625	-25.875	-15	-4.875
9		61	79	83	51			0.625	1.125	5	-20.875
10		53	73	85	63			-7.375	-4.875	7	-8.875
11		54	74	58	93	$\bar{x}$		-6.375	-3.875	-20	21.125
12	$\bar{x}_j$	60.375	77.875	78	71.875	72.03125		0	0	0	0
13	a_j	-11.6563	5.84375	5.96875	-0.15625	0					

Figure 6– Error terms for Example

Example:A school district uses four different methods of teaching their students how to read and wants to find out if there is any significant difference between the reading scores achieved using the four methods. It creates a sample of 8 students for each of the four methods. The reading scores achieved by the participants in each group are as follows: (see fig-5)

This time the p-value = .04466 < .05 = α, and so we reject the null hypothesis, and conclude that there are significant differences between the methods.

i.e. all four methods don’t have the same mean.

Note that although the variances are not the same, as we will see shortly, they are close enough to use ANOVA.

Observation: We next review described using above Example.(see fig-6)

From Figure 6, we see that

- x□ = total mean = AVERAGE(B4:E11) = 72.03 (cell F12)
- mean of the group means = AVERAGE(B12:E12) = 72.03 = total mean
- $\sum_j a_j$= 0 (cell F13)
- $\sum_i e_{ij}$ = 0 for all j (cells H12 through K12)

CONFERENCE PAPER
National Conference dated 27-28 July 2017 on
Recent Advances in Graph Theory and its Applications (NCRAGTA2017)
Organized by Dept of Applied Mathematics
Sri Padmawati Mahila Vishvavidyalayam (Women’s University) Tirupati, A.P., India

We also observe that *Var* (*e*) = VAR (H4:K11) = 162.12, and so by Property 3,

$$Var(e) = \frac{n-k}{n-1} MS_W$$

And

$$MS_W = \frac{n-1}{n-k} Var(e) = \frac{31}{28}(162.12) = 179.49$$

This agrees with the value given in Figure 5.

REFERENCES:

1. Excel Statistics: A Quick Guide-Neil J. Salkind
2. Black, K. Business Statistics: For Contemporary Decision Making (6th ed.). Hoboken, NJ: John Wiley & Sons, Inc., 2010.
3. Weiers, R.M. Introduction to Business Statistics (7th ed.). Mason, OH: South-Western Cengage Learning, 2011.

4. Interpretation and Uses of Medical Statistics by Leslie E Daly and Geoffrey J. Bourke, Blackwell science, 2000

5. Practical Statistics for Medical Research by Douglas G. Altman, Chapman and Hall, 1991

6. R.A. Olshen (1973). “The conditional level of the F-test,” Journal of the American Statistical Association, 68, pp.692–698.

CONFERENCE PAPER
National Conference dated 27-28 July 2017 on
Recent Advances in Graph Theory and its Applications (NCRAGTA2017)
Organized by Dept of Applied Mathematics
Sri Padmawati Mahila Vishvavidyalayam (Women’s University) Tirupati, A.P., India