



Introduction

In the previous tutorial, we used hypothesis testing to conclude possible differences between two-population means/proportions. Frequently, however, we will need to evaluate the differences between several different populations or groups.

We will complete most of the work using Excel, but you will see that some thought is first required about organising the data for analysis.

Specifically, the aims of this tutorial are to:

- Apply the one-way analysis of variance (ANOVA) to test the difference between the means of several groups
- Interpret the results of the analysis.

Scenario

Conrobar is a manufacturing company which employs over 3,000 people. Management is concerned about the wide variation in productivity between employees and whether employee job satisfaction and their sense of job security play any part in this problem. The company also has staff planning issues that need addressing, including planning to replace workers who retire, undertaking a recruitment drive for new employees and providing adequate promotion opportunities within the company.

The company has surveyed 48 of its staff, collecting data on these and related issues.

Open the data file and install the Data Analysis Tool Pak

- Download the file **Conrobar.xls** from Cloud Deakin. Please **save it** to the hard drive.
- Open the file in Excel.
- Install the Data Analysis Tool Pak. [See below for instructions].

Instruction:

From the top of *Excel (Microsoft Office Ribbon)*, click on **File** tab (Figure 1a), select **Options** (Figure 1b), choose **Add-ins** (Figure 1c), and then press **Go...** button to *manage excel add-ins* (Figure 1c). Finally, select **Analysis Tool Pack** and press **OK** (Figure 1d).

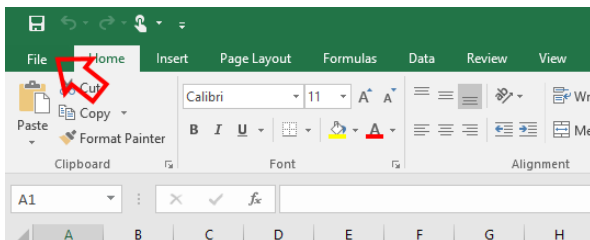


Figure 1a.

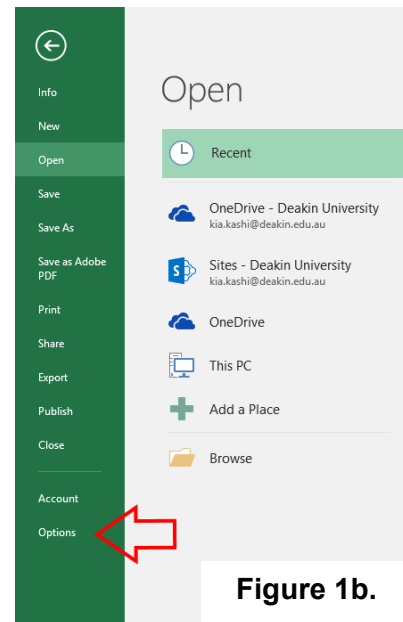


Figure 1b.

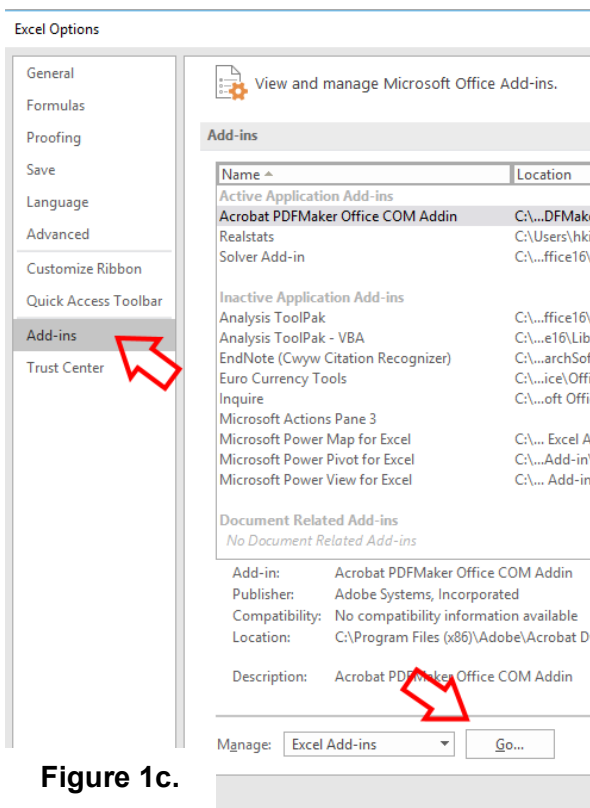


Figure 1c.

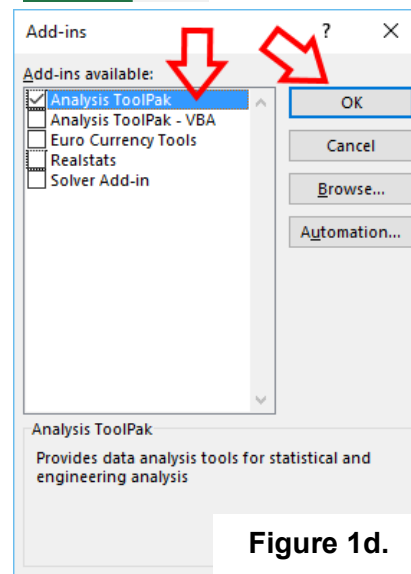


Figure 1d.

Q1. ANOVA

(a) Discuss the suitability of multiple t-tests for multiple groups/populations.

<Positive points>

- When we have multiple groups it is easier to do one test to establish whether there are differences rather than multiple t-tests (save energy and time.) When the number of groups comparing is more than three the number of possible comparisons and therefore number of possible t-tests increases dramatically: $n(n-1)/2$.
- When we reject the null hypotheses again and again in multiple t-tests, the type I error will multiply, therefore the type I error rate in multiple t-test can be much higher than a single ANOVA, where it remains at the 0.05 level.

<Negative points>

- We do not know however whether all the means are different from one another or only some of them are different. To figure this additional specificity we need to conduct a multiple comparison procedure. That is another additional test.

(b) What assumptions do we make about data when using ANOVA?

- Randomness and independence: selection of a random sample.
- Normality: Sample comes from normally distributed population.
- Homogeneity of variance: the population variance of the groups are similar. Equal size groups are recommended.

Q2. “Productivity” across “Departments”

The analysis of the 40 employees surveyed found that there were differences in the average Productivity across the three departments. A Comparative Summary Measure of the three departments is as follows:

	<i>Admin</i>	<i>Prod'n</i>	<i>Dist'n</i>
Mean	99.71	100.43	96.21
Standard Error	1.31	0.94	0.85
Median	98.6	101	96.3
Mode	#N/A	95.4	#N/A
Standard Deviation	4.35	4.32	3.42
Sample Variance	18.96	18.62	11.68
Kurtosis	-1.23	-1.10	0.25
Skewness	0.10	-0.10	0.54
Range	13.2	14.5	12
Minimum	93	92.5	91.2
Maximum	106.2	107	103.2
Sum	1096.8	2109.1	1539.4
Count	11	21	16

(a) Identify the “response” variable and describe the data type [Is it numerical or categorical data?]

Response variable = “Productivity”, Numerical

(b) Identify the “factor” variable and describe the different factors/groups.

Department has three categories “factor/Groups” -> Admin, Dist’n and Prod’n.

(c) Write the null and alternative hypothesis for this scenario using both notation and words.

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

where 1 = Admin, 2 = Prod’n, 3 = Dist’n

H_1 : Not all μ_j are equal

where $j = 1, 2, 3$

H_0 : Mean productivity rate of employees between the three departments is no different

H_1 : At least one of the departments differ with respect to the mean productivity rate.

(d) Perform an ANOVA test using Excel (Use $\alpha = 5\%$) – See Appendix-1 for steps.

(e) Based on the computer output created in (f), briefly advise Conrobar management about the findings.

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Admin	11	1096.8	99.70909	18.96291		
Prod'n	16	1539.4	96.2125	11.68383		
Dist'n	21	2109.1	100.4333	18.62333		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	172.0615	2	86.03077	5.25038	0.008924	3.204317
Within Groups	737.3533	45	16.38563			
Total	909.4148	47				

The decision rule: if the p-value < 0.5. Reject the H_0 .

Because the p-value is 0.008924 which is less than the $\alpha = 0.05$. We reject the H_0

We have sufficient evidence to conclude that the mean productivity rates of the employees in the three departments are not all equal.

(f) Use multiple comparisons to determine which groups are different. See Appendix-2 for steps.

To determine which of the means are significantly different from one another, we used Tuley-Karmer Multiple comparison.

Tukey Kramer Multiple Comparisons						
Group	Sample Mean	Sample Size	Comparison	Absolute Difference	Std. Error of Difference	Critical Range Results
Admin	99.71	11	Group 1 to Group 2	0.72	1.06547553	3.6652 Means are not different
Prod'n	100.43	21	Group 1 to Group 3	3.5	1.12124373	3.8571 Means are not different
Dist'n	96.21	16	Group 2 to Group 3	4.22	0.94996084	3.2679 Means are different
Other Data						
Level of significance	0.05					
Numerator d.f.	3					
Denominator d.f.	45					
MSW	16.39					
Q Statistic	3.44					

(g) Based on the computer output created in (h), now refine your response to Conrobar management.

We conclude that there is a significant difference between the mean productivity rate of Prod'n (Group 2) and Dist'n (Group 3). The other pairwise differences are small enough that they may be due to chance (There is no significant difference between Admin (Group1) and Prod'n (Group 2) and Admin (Group1) and Dist'n (Group 3).

Q3. "Unpaid Overtime" across "Departments"

The analysis of the 40 employees surveyed found that there were differences in the average unpaid overtime across the three departments. In particular, the Prod'n reported a higher average unpaid overtime than both Admin and Dist'n departments. A Comparative Summary Measure of the three department is as follows:

	<i>Admin</i>	<i>Prod'n</i>	<i>Dist'n</i>
Mean	6.636	8.463	6.295
Standard Error	0.710	0.718	0.568
Median	6.3	9.75	6.5
Mode	#N/A	10	3.5
Standard Deviation	2.355	2.873	2.605
Sample Variance	5.547	8.255	6.784
Kurtosis	-0.428	-1.543	-0.760
Skewness	0.375	-0.329	-0.023
Range	7.7	8	9.5
Minimum	3	4	2
Maximum	10.7	12	11.5
Sum	73	135.4	132.2
Count	11	16	21

(a) Identify the "response" variable and describe the data type [Is it numerical or categorical data?]

Unpaid Overtime is a numerical data type

(b) Identify the "factor" variable and describe the different factors/groups.

Department has three categories "factor/Groups" -> Admin, Dist'n and Prod'n.

(c) Write the null and alternative hypothesis for this scenario using both notation and words.

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

where 1 = Admin, 2 = Prod'n, 3 = Dist'n

H₁ : Not all μ_j are equal

where j = 1,2,3

H₀ : Mean unpaid overtime hours of employees between the three departments is no different

H₁ : At least one of the departments differ with respect to the mean unpaid overtime hours.

(d) Perform an ANOVA test using Excel (Use $\alpha = 5\%$) and multiple comparisons if necessary.

Appendix 1:

Steps for ANOVA:

1. Organise the data in a column format

- Copy variables "Productivity" and "Department" into a Working sheet tab.
- Set up filters and sort by "Department".

	A	B
1	Prdtvty	Departm
2	105.0	Admin
3	98.6	Admin
4	96.4	Admin
5	104.3	Admin
6	98.0	Admin
7	102.1	Admin
8	106.2	Admin

- To set up the data in a column format create column headings for each department.
- Then copy the "Productivity" values belonging to that particular department.

D	E	F
Admin	Prod'n	Dist'n
105.0	103.2	95.0
98.6	102.1	97.2
96.4	92.5	98.2
104.3	99.0	91.4
98.0	104.0	102.5
102.1	95.4	94.3
106.2	101.0	92.4
93.0	96.6	97.6
101.3	95.4	98.0
96.9	105.9	96.5
95.0	98.3	91.2
	98.1	95.3
	98.0	97.0
	102.5	93.5
	103.4	103.2
	105.1	96.1
	107.0	
	97.2	
	102.9	
	94.5	
	107.0	

Anova: Single Factor

Input

Input Range:

Grouped By: ☒ Columns ☐ Rows

☒ Labels in First Row

Alpha:

Output options

☐ Output Range:

☒ New Worksheet Ply:

☐ New Workbook

OK Cancel Help

2. Use Data Analysis Toolpak for ANOVA

- Select Data -> Data Analysis.
- In the Data Analysis, Select "ANOVA Single Factor".
- Select all the data including the Headings as the Input range.
- Click Column, Check Labels in First Row and enter 0.5 as Alpha.
- Click New Worksheet Ply -> Enter name "ANOVA".
- Click Ok.

Appendix 2:

Steps for Multiple Comparison: The Tukey-Kramer Procedure. Please note that there are different worksheets to choose depending on the size of the group/factors.

1. Return to the ANOVA tab you created earlier.
2. Note the Names of the Groups, sample size (in the Count column), and the sample mean (in the Average column) of each group (you need these details).
3. Record the MSW value (found in the cell that is the intersection of the MS column and Within Groups row), and the denominator degrees of freedom (found in the cell that is the intersection of the df Column and Within Group row.)
4. Open **Tukey-Kramer.xls** and then select the **Tukey-Kramer3** worksheet.
5. In the **Tukey-Kramer3**, in the cell range A5:C7 – enter the Name, Sample mean, and Sample size of each group from the ANOVA tab.
6. Enter 0.05 as the level of significance in cell B10
7. Enter 3 as the Numerator d.f. (equal to the number of groups in cell B11.)
8. Enter 45 as the Denominator d.f. in Cell B12.
9. Enter 16.385 as the MSW in cell B13.
10. Look up the Studentised range Q statistic using Table E.10. A copy of the table is available in CloudDeakin. You need Denominator d.f. (in this case it is 45) and Numerator d.f. (in this case 3). An approximation of Q statistics is good enough (select the nearest Denominator d.f.)
11. Enter 3.44 as the Q statistic in Cell B14.

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Admin	11	1096.80	99.71	18.96		
Prod'n	21	2109.10	100.43	18.62		
Dist'n	16	1539.40	96.21	11.68		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	172.0615	2	86.03077	5.25038	0.008924	3.204317
Within Groups	737.3533	45	16.39			
Total	909.4148	47				

Tukey Kramer Multiple Comparisons						
Group	Sample Mean	Sample Size	Comparison	Absolute Difference	Std. Error of Difference	Critical Range Results
Admin	99.71	11	Group 1 to Group 2	0.72	1.06547553	3.6652 Means are not different
Prod'n	100.43	21	Group 1 to Group 3	3.5	1.12124373	3.8571 Means are not different
Dist'n	96.21	16	Group 2 to Group 3	4.22	0.94996084	3.2679 Means are different
Other Data						
Level of significance	0.05					
Numerator d.f.	3					
Denominator d.f.	45					
MSW	16.39					
Q Statistic	3.44					