MIS771 Descriptive Analytics and Visualisation	DEAKIN UNIVERSITY
Topic 4 Tutorial **Two-way analysis of variance and Chi Square Test**	

Introduction

This tutorial consists of two parts.

Part A: In the previous tutorial, we used one way ANOVA to test the differences between the means of several groups. In this tutorial, we extended the analysis of variance to the two-factor factorial design, in which two factors are simultaneously evaluated. In this week's tutorial, we learn how to apply the two-way analysis of variance (ANOVA) and interpret the results of the analysis.

Part B: In previous tutorials, we used the hypothesis-testing procedure to analyse both numerical and categorical data in two or more samples/groups. This tutorial extends the hypothesis testing to analyse the difference between population proportions in two or more populations using chi square test.

Similar to the previous week, we will complete most of the work using Excel and we need to organise the data for analysis in a particular way.

Scenario

Conrobar is a manufacturing company with over 3,000 employees. Management is concerned about the wide variation in productivity between employees. The company has collected data on these and related issues.

PART - A

Open the data file and install the Data <u>Analysis Tool Pak</u>

a) Download the file **ConrobarT4.xls** from Cloud Deakin. Please **save it** to the hard drive.

b) Open the file in Excel.

c) Install the Data <u>Analysis Toolpak</u>. [See the previous tutorial for instructions].

Q1. TWO-WAY ANOVA

(a) What are the advantages of Two-way ANOVA?

(b) For each of the following statements, explain what is wrong and why.

    i.   The two-way ANOVA is used when there are two dependent (outcome) variables.

    ii.   In a 2 X 3 ANOVA, each level of factor A appears with only two levels of Factor B.

    iii.   You can perform a two-way ANOVA only when the sample sizes are the same in each cell.

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

Q2. Job satisfaction, Gender and Departments

The management team would like to broaden the scope of the analysis to compare the employee job satisfaction from different Departments and Gender of the employee.

(a) What is the dependent "response" variable in this scenario?

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

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

(d) Perform a Two-way ANOVA test using Excel (Use α= 5%) – See Appendix-1 for steps.

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

(f) To understand the interaction effect better, draw a cell means plot See Appendix-2 for steps.

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

Q3. Unpaid Overtime, Gender and Job Security

The effect of the employees' gender and their perception of Job security on unpaid overtime were being studied.

(a) Is there a significant interaction between Gender and Job Security?

(b) Is there an effect due to Job Security?

(c) Is there an effect due to Gender?

(d) What can you conclude about the effect of Gender and the perception of Job Security on employee's Unpaid Overtime?

HINT: follow the steps in Q2.

Appendix 1:

Steps for 2Way-ANOVA:

1. Organise the data in a r X c format

Gender	Admin	Dis	Prod
Female	11	12	10
	13	15	10
	8	12	10
	11	12	11
	11	15	14
	11	13	11
	11	18	11
Male	15	17	16
	17	13	18
	16	16	16
	10	17	16
	17	14	13
	16	14	13
	15	9	13

2. Use Data Analysis Toolpak for ANOVA
   a. Select Data -> Data Analysis.
   b. In the Data Analysis, Select "ANOVA Two-Factor with Replication".
   c. Select all the data including the Headings as the Input range.
   d. Enter 7 for "Rows per sample" and enter 0.5 as Alpha.
   e. Click New Worksheet Ply -> Enter name "2Way-ANOVA".
   f. Click Ok.
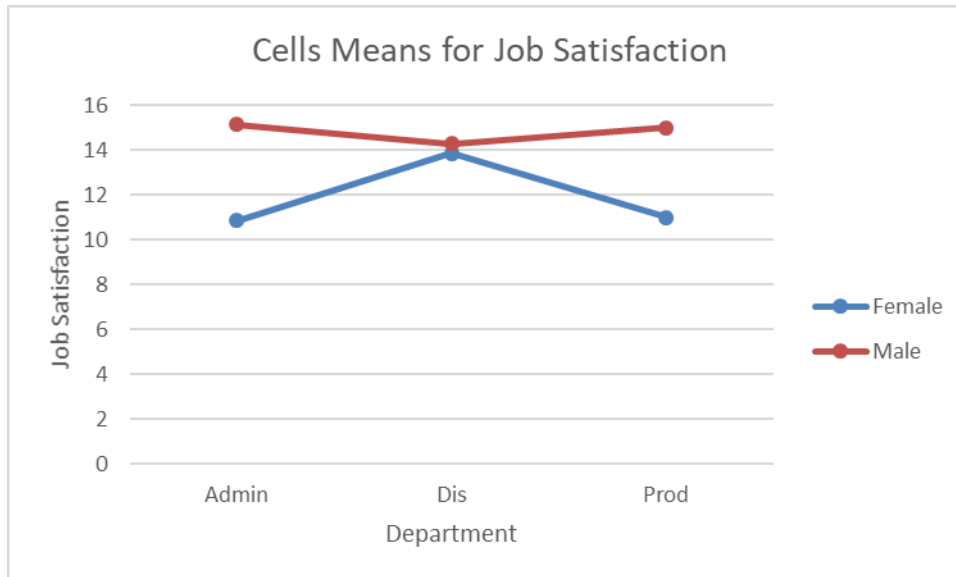
Appendix 2:

1. Insert "Line with Marker" chart (Blank)
2. Add Female average as a data series
3. Add Male averages as a data series
4. Add department labels as X Axis.

Cells Means for Job Satisfaction

PART – B

Open the data file and install the Data <u>Analysis Tool Pak</u>

    d)   Download the file ***BLITZT5.xls*** from Cloud Deakin. Please **save it** to the hard drive.

    e)   Open the file in Excel.

    f)   Install the Data <u>Analysis Toolpak</u>. [See the previous tutorial for instructions].

Q1. Chi-Square Test for the difference between two or more proportions (independent sample)

    (d) Discuss the advantages and disadvantages of using z-test and chi-square test for the difference between two proportions.

    (e) Discuss each of the following statements.
         i.   For the chi-square test, we must have equal sample size for each group/population.
        ii.   The expected frequency is at least five for each cell in the table.

Q2. Awareness of the Loyalty program across the three cities

The BLITZ management team would like to broaden the scope of the analysis to investigate whether there is a significant difference in the proportion of loyalty program awareness between the three cities.

    (h) Write the null and alternative hypothesis for this scenario using both notation and words.
    (i)   Conduct a hypothesis test (Chi-Square Test Use α= 5%) that will determine if the population proportion of loyalty program awareness is the same for all three cities – See Appendix-3 for steps.
    (j)   Based on the computer output created in (b), briefly advise BLITZ management about the findings.
    (k) Use the multiple comparison procedure to determine which population proportions differ significantly (if). See Appendix-4 for steps.
    (l)   Based on the computer output created in (d), now refine your response to BLITZ management.

Q3. Online Shopping across the three cities

Is there a significant difference in the proportion of online shoppers between the three cities?

    (a) Write the null and alternative hypothesis for this scenario using both notation and words.
    (b) Conduct a hypothesis test (Chi-Square Test Use α= 5%) that will determine if the population proportion of BLITZ online shoppers is the same for all three cities – See Appendix-3 for steps.
    (c) Based on the computer output created in (b), briefly advise BLITZ management about the findings.
    (d) Use the multiple comparison procedure to determine which population proportions differ significantly (if). See Appendix-4 for steps.
    (e) Based on the computer output created in (d), now refine your response to BLITZ management.

1. Summarise the data to show observed frequencies in a tabular format.

Excel's PivotTable can be used to summarise the data.

Instruction I (Pivot Table):
While in the 'Workings' worksheet, select the **Insert** tab and click on the **Pivot Table**. From the pop-up window, enter the **Table/Range** and **Location** as shown in Figure 1a. This step creates a blank pivot table.
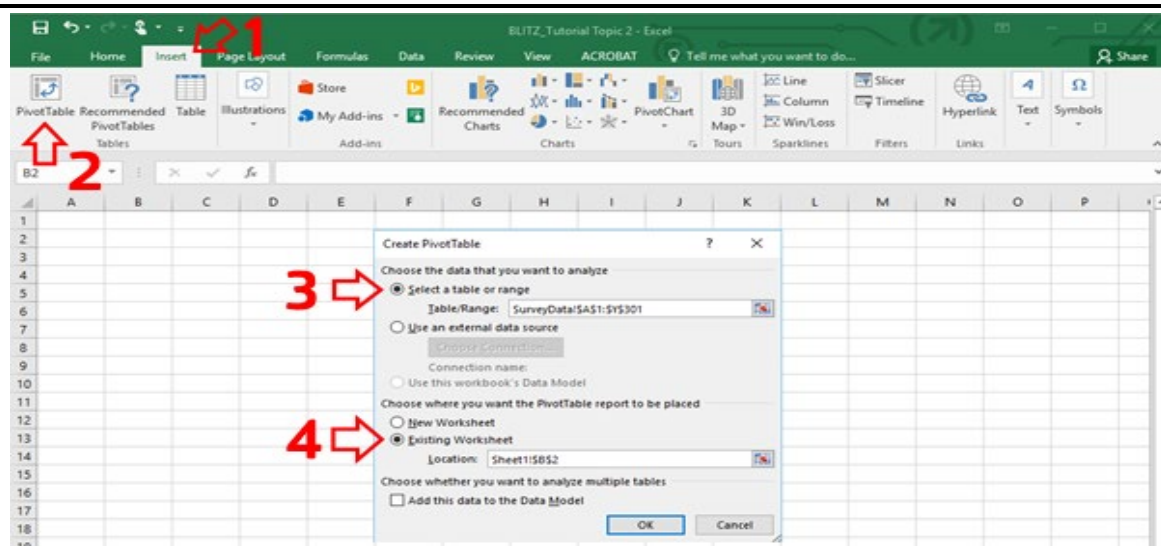

Figure 1a.

Instruction II (Pivot Table Cont'd):
*From the **Pivot Table Field List:***
   a) Click-and-drag the **Loyalty Program Awareness** variable to the **'row labels'** of the table.
   b) Click-and-drag the **City** variable to the **'Column labels'** of the table.
   c) Click-and-drag the **Same** Categorical variable to the **'∑ values'** of the table.

Then copy the table and paste it as "values" *(see Figure 1b)*

Count of Loyalty Program Awarenes	Column Labels			
Row Labels	Melbourne	Perth	Sydney	Grand Total
NO	57	73	56	186
YES	43	27	44	114
Grand Total	**100**	**100**	**100**	**300**

Count of Loyalty Program Awareness	Column Labels			
Row Labels	Melbourne	Perth	Sydney	Grand Total
NO	57	73	56	186
YES	43	27	44	114
Grand Total	100	100	100	300

Figure 1b

2. Compute expected frequencies.
   a. Delete expected values so we can compute them.

Observed Frequencies

Count of Loyalty Program Awarenes Column Labels	Melbourne	Perth	Sydney	Grand Total
Row Labels				
NO	57	73	56	186
YES	43	27	44	114
Grand Total	**100**	**100**	**100**	**300**

Expected Frequencies

Count of Loyalty Program Awareness Column Labels	Melbourne	Perth	Sydney	Grand Total
Row Labels				
NO	57	73	56	186
YES	43	27	44	114
Grand Total	100	100	100	300

→

Observed Frequencies

Count of Loyalty Program Awarenes Column Labels	Melbourne	Perth	Sydney	Grand Total
Row Labels				
NO	57	73	56	186
YES	43	27	44	114
Grand Total	**100**	**100**	**100**	**300**

Expected Frequencies

Count of Loyalty Program Awareness Column Labels	Melbourne	Perth	Sydney	Grand Total
Row Labels				
NO				186
YES				114
Grand Total	100	100	100	300

   b. Delete expected values so we can compute them as below.
   c. Insert Excel function CHISQ.TEST. The format of the function is as follows:

=CHISQ.TEST(**Observed frequency Cells**, **Expected Frequency Cells**)

B20			fx	=CHISQ.TEST(B5:D6,B15:D16)			
	A		**B**	**C**	**D**	**E**	
1							
2	**Observed Frequencies**						
3	**Count of Loyalty Program Awareness**	**Column Labels**					
4	**Row Labels**		**Melbourne**	**Perth**	**Sydney**	**Grand Total**	
5	NO		57	73	56	186	
6	YES		43	27	44	114	
7	**Grand Total**		**100**	**100**	**100**	**300**	
8							
9							
10							
11				Step 1			
12	**Expected Frequencies**						
13	Count of Loyalty Program Awareness		Column Labels				
14	Row Labels		Melbourne	Perth	Sydney	Grand Total	
15	NO		=E15*B17/E17	=E15*C17/E17	=E15*D17/E17	186	
16	YES		=E16*B17/E17	=E16*C17/E17	=E16*D17/E17	114	
17	Grand Total		100	100	100	300	
18							
19							
20	P value		=CHISQ.TEST(B5:D6,B15:D16)	Step 2			
21							

<Alternatively use the template☺>

Appendix 4: Post-hoc test – Pairwise comparison: Marascuilo procedure

1. Use the corresponding template and enter the observed frequencies into the table.