MIS771 Descriptive Analytics and Visualisation

Topic 2 Tutorial – Hypothesis Testing and comparing two samples

Introduction

In this tutorial you will cover hypothesis tests for drawing inferences about both single and two populations and confidence intervals for estimating a population mean/proportion and differences between two-population means/proportions. You will complete some of the work using Excel, but you will see that some thought is first required about setting up the null and alternative hypotheses first, and about other steps in the process, and you will benefit it you do some of this on paper.

Specifically, the aims of this tutorial are to:

- Conduct hypothesis tests and calculate confidence interval estimates for μ and π.
- Perform confidence intervals and hypothesis tests to situations involving two samples.
  - Inferential techniques for comparing means (μ_1, μ_2):
    - Two independent populations
    - Two dependent (related) populations
  - Inferential techniques for comparing proportions (π_1, π_2)
- Practice varying the level of confidence to see the effect on interval estimates.
- Practice varying the level of significance (α) to see the effects (if any) on your conclusions.
- Interpret the results from your analysis regarding the scenarios given.

Scenario

BLITZ management would like to further discuss their survey results with the research company. In particular, they would like more insights on, time spent on a visit, number of visits per month, activity in the online store and member behaviour among different loyalty tiers. They would like to discuss this year's results compared to last year's performance.

Step 1. Open the data file and install the Data Analysis Tool Pak

a) Download the file ***BLITS_Dataset_M1T2.xls*** from Cloud Deakin. **Save it** to your hard drive.

b) Open the file in Excel.

c) Install the data Analysis Tool Pak.

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).

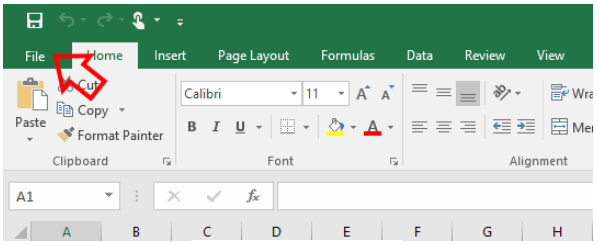

Figure 1a.

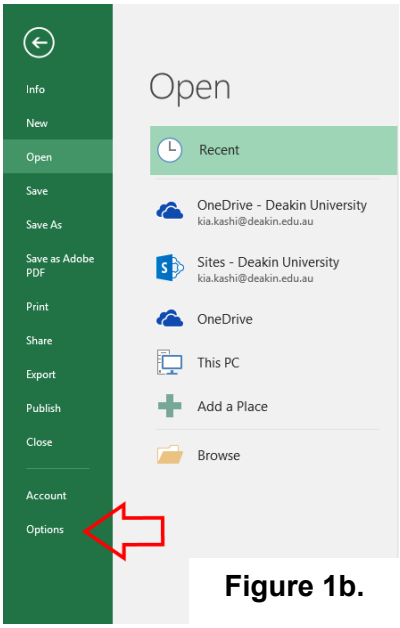

Figure 1b.

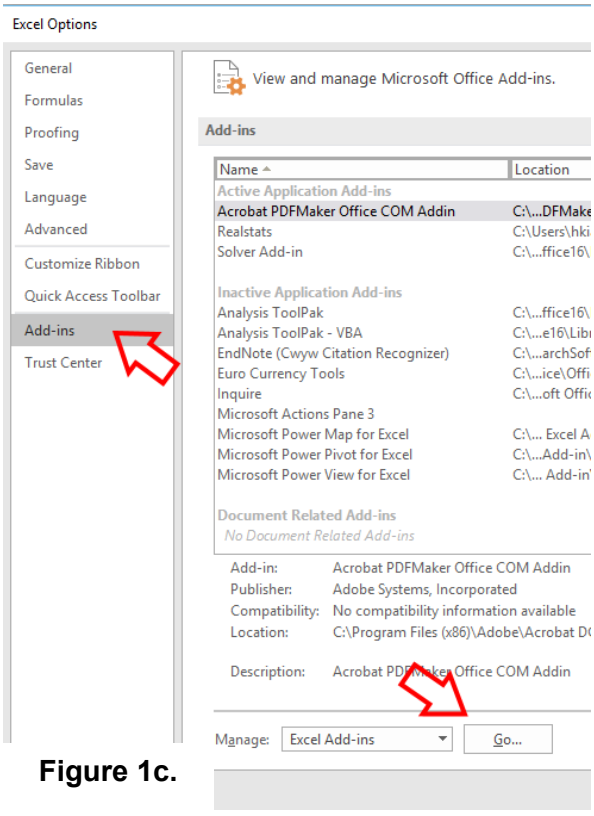

Figure 1c.

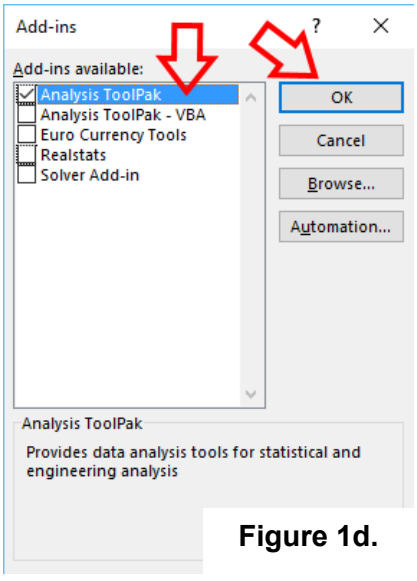

Figure 1d.

Note that instructions for using Excel to perform hypothesis tests and compare two samples (confidence intervals and hypothesis tests) are provided in the appendices.

Q1. Time Spent per visit

300 randomly selected customers were participated in the customer research survey and it is found that, on average, a BLITZ customer spends 40.44 minutes during a visit to any of the BLITZ stores with a standard deviation of 11.46 minutes. ($\bar{x}$ = 40.44, s = 11.46)

The corresponding figure for average time spent per customer visit last year was 38 minutes. Management would like to know if this year's average time spent is more than the last year's average.

(a) Explain why they should use hypothesis testing in this situation?
(b) What type of variable are we working with?
(c) Is this a one sample or two sample situation?
(d) Use Excel to complete the hypothesis test (use α=1%). Interpret your results.

Q2. BLTIZ Online Store

Out of the 300 customers participated in the survey it is found that only 66 members have visited BLITZ online store. BLITZ have set a KPI to have at least 25% of their customers who visit their stores to also visit their online store. They are very much interested to know whether the survey result is significantly lower than their expectation.

(a) Explain why we can use hypothesis testing in this situation.
(b) What type of variable are we working with?
(c) Is this a one sample or two sample situation?
(d) Perform a hypothesis test in Excel. (Use α=5%) What p-value did you obtain? Interpret this value.

Q3. BLITZ customers – Time spent shopping across cities

Following on from Q1, given we now know that the time spent shopping is more than last year, we wish to estimate *how much* more time customers are spending shopping.

(a) Calculate a 95% confidence interval for the average time spent shopping for all customers?
(b) Calculate a 90% confidence interval for the average time spent shopping for all customers?
(c) Calculate a 99% confidence interval for the average time spent shopping for all customers?
(d) Comment on the differences between the three confidence intervals.

BLITZ management would like to know whether the time spent differs across cities. Particularly they would like to know whether there is *a difference* in average time in Melbourne and Perth customers.

	Melbourne	**Perth**
$\bar{x}$	46.5	30.0
s	11.2	4.5
n	100	100

(a) Perform a hypothesis test using Excel to clarify BLITZ management on this. (Use α= 5%)
(b) Based on the computer output created above, briefly advise BLITZ management about the findings.

Q4. Loyalty program awareness across different age groups

BLITZ management also would like to know whether younger customers are more likely to be aware of the loyalty program compared to older customers. Out of the 71 customers aged under 30 years old, 31 were aware of the program. Of the 229 customers aged 30 or above, 83 were aware of the program.

(a) Explain why we can use hypothesis testing in this situation?
(b) Is this a one sample or two sample situation?
(c) Write down H0 and H1 that are appropriate for this problem. (Write them down in words and in symbols.)
(d) Use Excel to complete the hypothesis test (use α=5%).
(e) Would it be different if we use α=1% or α=10%?
(f) Interpret the output to BLITZ management?

Q5. Average spend across different loyalty tiers

The membership structure is set up in a way to reward members more according to their level of membership. To justify the extra costs associated with higher levels of membership, management want an estimate in the difference spent between levels (in particular between BLUE tier members and SILVER tier members).

This year's survey included 28 BLUE tier members and 24 SILVER tier members. Based on the survey results it is found out that BLUE tier average spend per visit is $97.32 with a standard deviation of $24.55 and SILVER tier average spend per visit is $109 with a standard deviation of $27.05.

(a) Explain why a confidence interval is appropriate in this case.
(b) What assumptions do you have to make about the data?
(c) Calculate a 95% confidence interval and interpret your answer.

Appendix 1:

Hypothesis Tests and confidence intervals Instructions

HYPOTHESIS TESTING – ONE SAMPLE
 Hypothesis Test tables for one sample are given in the 'HT' worksheet. You first need to decide whether it is a test about a mean (table 1) or proportion (table 2). Input appropriate values in the blue cells. This may require you first to do some preliminary calculations (such as summary measures). Important calculations and results are shown in the green cells. Note that the table is presented in the same order as the six step procedure discussed in lectures. When you have finished, Copy and Paste | Paste Values the whole table to the Working sheet.

HYPOTHESIS TESTING – TWO SAMPLE
 Hypothesis Test tables for two samples are given in the 'HT (2 Sample)' worksheet. You first need to decide whether it is a test comparing means (tables 1 – 3) or proportions (table 4). In the former case, decide between independent (equal variance), independent (unequal variance) or dependent. Input appropriate values in the blue cells. This may require you first to do some preliminary calculations (such as summary measures). Important calculations and results are shown in the green cells. When you have finished, Copy and Paste | Paste Values the whole table to the Working sheet.

CALCULATING CONFIDENCE INTERVALS – ONE SAMPLE
 Confidence Intervals for one sample are given in the 'CI' worksheet. You first need to decide whether it is a test about a mean (table 1) or proportion (table 2). Input appropriate values in the blue cells. This may require you first to do some preliminary calculations (such as summary measures). Important calculations and results are shown in the green cells. When you have finished, Copy and Paste | Paste Values the whole table to the Working sheet..

CALCULATING CONFIDENCE INTERVALS – TWO SAMPLE

Confidence Interval tables for two samples are given in the 'CI (2 Sample)' worksheet. You first need to decide whether it is a test comparing means (tables 1 – 3) or proportions (table 4). In the former case, decide between independent (eq variance), independent (unequal variance) or dependent.

Input appropriate values in the blue cells. This may require you first to do some preliminary calculations (such as summary measures). Important calculations and results are shown in the green cells.

When you have finished, Copy and Paste | Paste Values the whole table to the Working sheet.