

**CSG1132 Communicating in an IT Environment**

Statistical literacy is about understanding and evaluating statistics. This workshop will introduce inferential methods and how to use the Equation Editor to present statistical formulae and findings.

**Activity 1 – Metacognitive reflection**

Responding appropriately to a learning challenge involves critical self reflection (metacognition) on whether learning outcomes are appropriate to progress to the next stage of problem solving. Check off the following learning outcomes from last week’s workshop for working with our dataset**.**

**If you cannot check off any item on the list, go back to last week’s workshop and complete the required activities:**

|  |  |
| --- | --- |
| **Learning outcome/objective** | **Completed?**  **(Check the box)** |
| 1. E-book readings on centrality and variance completed? |  |
| 1. RQs and thesis statements formulated? |  |
| 2. Assignment 2 dataset downloaded? |  |
| 3. Thesis statement test variables identified? |  |
| 4. Summary statistics (Centrality & Variability) for all variables completed? |  |
| 4. Frequency histograms and normality conclusions (data screening)? |  |
| 5. Assignment 2 research paper template downloaded and personalized? |  |
| 6. RQs, statements and stats added to sections 4.0 and 5.0 of the paper template |  |

**Activity 2 –** **Inference testing: Correlation**

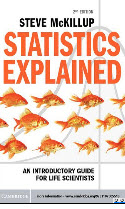
Whether two variables are changing together and how strong such a relationship might be (if it exists) is a commonly occurring problem in research. For example, how does the incidence of Type 2 diabetes change with obesity?

A common kind of inference test, is a test of *correlation* between two scale variables, to see if they are related:

RQ. Is age related to FB friends?

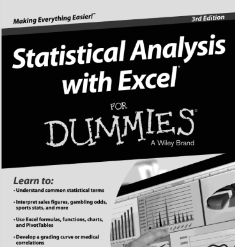
Thesis Statement: Age is positively related to FB friends

To answer our research question, we would perform a test of correlation for these two variables with data from the data set.

Review the lecture notes and video on correlation and correlation testing. Correlations are explained in Chapter 16 of Steve McKillup’s (2011) *Statistics Explained: An Introductory Guide for Life Scientists*. 2nd Ed. Cambridge : Cambridge University Press, 2011.

<http://library.ecu.edu.au/record=b2232524~S1>

Take time out now, to review McKillup’s explanation of correlation. Chapter 16 also explains how you should report a result from correlation testing.



Chapter 15 of Schmuller’s Statistical analysis in Excel for Dummies will tell you how to perform correlation testing in Excel.

<http://library.ecu.edu.au/record=b2528117~S1>

Correlation coefficients are paired to data types and assumptions of normality. For Pearsons variables must be interval or ratio and distribution is normal. In Excel, **Pearsons** is standard. If data is not normally distributed (non-parametric), **Spearman’s** is used.

Since Excel only supports Pearson as standard, it is usual to use a more sophisticated tool such as IBM’s SPSS for correlation testing where data is not normally distributed. Where data screening has shown one or more variables to be non-normally distributed, and you do not have access to SPSS, you should report the Excel result subject to the reservation that the goodness of fit test for normality was not satisfied and that further investigation of the result is required with a **non-parametric test.**

Going the extra distance here involves discovering how you can apply a non-parametric test of correlation in Excel, such as Spearman’s Rho. There are YouTube and other instructional resources on this.

**ACTIVITY 2**

Review the readings and videos on correlation. Open a new worksheet in Excel, copy the variable data for Age and FB friends. Next perform a correlation test, and record the result below.



Provide an interpretation of the result in terms of the thesis statement. What kind of relationship exists between age and FB friends?

Interpretation:

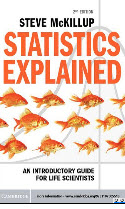
Any limitations/reservations?

**Activity 3 –** **Inference testing: Independent samples testing**

Do females spend more time on FB than males? This kind of RQ lends itself to the following thesis statement:

Thesis statement = Facebook hours is related to gender.

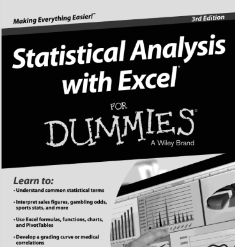
With this kind of question, we are interested in whether at the level of individual behaviour significant differences exist between males and females. An independent samples test would (t-test or Mann Whitney) is used for this purpose. The following readings will help you to understand independent samples testing:



Steve McKillup’s (2011) *Statistics Explained: An Introductory Guide for Life Scientists*. 2nd Ed. Cambridge : Cambridge University Press, 2011.

-Chapter 9 - Comparing the means of one and two samples of normally distributed data

<http://library.ecu.edu.au/record=b2232524~S1>



Schmuller’s Statistical analysis in Excel for Dummies explains how to do sample testing in Excel:

Chapter 11 - Two sample hypothesis testing

<http://library.ecu.edu.au/record=b2528117~S1>

Before undertaking Activity 3, you should also review the learning schedule videos in independent samples testing.

**ACTIVITY 3 TASK**

Open a new datasheet in Excel, copy the variable data for Gender and Facebook hours. Next perform an independent samples test and record the result below.



Now provide an interpretation of the result in terms of the thesis statement.

Interpretation:

Any limitations?

**Activity 4 – The Equation Editor**

Reporting the results of statistical analysis, requires the use of a special symbol set and sometimes you may need to use the Equation Editor in Excel. Review 1) the Youtube video on how to use the Equation Editor and 2) Module 7 on Descriptive Statistics. Create equations in Excel or Word for the following measure of central tendency and variation:

The sample mean ():

Explanation:

The sample standard deviation ():

Explanation:

Dr M. Brogan

Unit Coordinator