University of Wollongong

School of Mathematics and Applied Statistics

STAT904 Statistical Consulting

Assignment 1, 2018

This Assignment must be lodged as a single PDF document on the Moodle site for this subject by 5pm on Thursday 15 March. It must have your name and student number on the front page and include your name in the PDF file name. It must include the statement "No part of this Assignment has been copied from anyone else, and I have not lent any part of it to anyone else." *Assignments without this statement will not be marked.*

This Assignment is worth 5% of the final mark for this subject. It may be typed or handwritten, but quality and clarity of presentation will be taken into account in deciding the mark. You should read the assignment before the lecture on Thursday 8 March and ask any questions or seek clarification at that lecture.

1. A researcher from the Nursing Department approaches you concerning her research. She has read a paper giving some results of an analysis of the relationship between the HSC (final year high school) results and the results in a first year university Nursing course. The correlations are given below.

Year	HSC*GPA	HSC*Bio	HSC*GenSci	HSC Eng*GPA	HSC Bio*Bio	*n*
2013	.41	.36	.30	.30	.37	109
2014	.37	.32	.26	.12	.45	85

She asks you to analyse these correlations. She seems interested in how the HSC scores predict university results and whether there was any change between the two years.

a) What more would you want to know about the study?
b) Carry out the required analysis and write a short report. Include any qualifications you think should be made concerning the analysis.
c) The researcher is planning to carry out her own study. Write a brief report on how she should plan and analyse her study.

2. A very common and widely applicable analysis method is the chi-squared test of independence. Write a report that covers the following:

a) Briefly describe the test, including a discussion of when it is should be used.
b) Many computer packages will give a warning message if the expected frequencies in some cells are less than 5. Briefly explain why this warning is given and what options are available for dealing with it.
c) Suppose we have the following table giving the number of people in a study that are overweight or not by their annual income:

	Not Overweight	Overweight
Less than $30,000	3	2
Greater than or equal to $30,000	4	1

Perform a test of independence using the chi-squared test and the exact test. Then repeat the tests with each cell frequency multiplied by 10. Comment on the results.

Be sure to appropriately reference the sources that you use to draft the report.