RESEARCH SCHOOL OF FINANCE, ACTUARIAL STUDIES AND APPLIED STATISTICS

QUANTITATIVE RESEARCH METHODS (STAT1008)

SEMESTER 1, 2016

Assignment 1

This assignment is worth 15% of your overall course mark. Solutions should be placed in the STAT1008 assignment box labelled with your tutor's name by 3pm 23 Mar 2016. The boxes are located adjacent to the office of the **Research School of Finance, Actuarial Studies and Applied Statistics**. Please present your assignment in a single stack of pages held together by a staple in the upper left hand corner. (That is, do not use folders, plastic sheaths or other devices to join pages together.)

On a cover sheet, a copy of which is available on the Wattle site for the course, clearly state the details required, including your **name** and **student number** and your tutorial group details **(tutorial group number, day, time and tutor's name)**.

Data relating to each question (if relevant), can be found in a comma delimited file (".csv") on Wattle site. You will be assessed on your understanding of statistics and your ability to use RStudio, not on your typing and word-processing expertise.

In the interests of equity, **late assignments** will not be accepted. Please see the course outline for more information on this matter.

RStudio should be used to produce the relevant graphics and statistics, unless the question directs you otherwise. Answers should be presented as an edited Microsoft Word file, obtained as illustrated in Tutorial 4. You should include all relevant graphs and values generated in RStudio (e.g. if a question directs you to create and discuss a graph, you should include the graph itself as well as your discussion in your submitted assignment).

All numeric answers should be rounded to 4 decimal places as appropriate, clearly stating where rounding has been used.

You should show all relevant calculations and working for questions which you are instructed to do "by hand". These can be typed in Word, or handwritten. Marks may be deducted for failure to show working or calculations. Additionally, you should be careful to define all notation used. Marks may be deducted for unexplained notation. Further, you should be careful to specify and include units where appropriate.

Question 1

Air quality has been tracked in New York for five months, with daily readings of Ozone, Solar R, Wind and Temperature. This information has been entered into a comma delimited file, available as "airquality.csv" on Wattle site, with variables as follows:

X: Overall Day Number

Ozone: Ozone density

Solar.R: Solar Radiation

Wind: Wind speed measured in Miles per Hour

Temp: temperature measured in Fahrenheit

Month: the month in numerical form

Day: day of the month

(a) (5 marks) In this spreadsheet, to what does each row correspond? What is the type (nominal, ordinal, continuous or count) of each variable? You should give sufficient explanation and discussion to show your thoughts on the data.
(b) (3 marks) Create a histogram of temperature and comment on its shape.
(c) (4 marks) Create boxplots of ozone for each month, ensuring you can make appropriate comparisons. Comment on the boxplots – consider the shape, important features etc. of each boxplot, and compare the information provided for each month.
(d) (3 marks) It is desired to investigate the relationship between temperature and ozone. Create an appropriate scatterplot for this aim. Describe the relationship, or lack thereof, that you see in the scatterplot. Based on your scatterplot what sort of value (considering sign and magnitude) would you expect for the correlation between these two variables?
(e) (2 marks) Use RStudio to calculate the covariance between ozone and wind speed. You should include the RStudio session window in your assignment, and also report the value in a sentence, e.g. "covariance is…".
(f) (4 marks) BY HAND (that is, without RStudio, and using only a hand calculator), and using only the values you found in part (e), calculate the correlation between ozone and wind speed. Show all work, and be careful to define any notation you use. Comment on the value you find.
(g) (3 marks) Someone suggests that the relationship between temperature and ozone is different for July and August than for the other months. Calculate the correlation between these two variables separately for each month group and comment appropriately.

Question 2

A sport shop plans to offer 2 specially priced skate board models at a sidewalk sale. The basic model will return a profit of $120 and the sports model $150. Past experience indicates that sales of the basic model will have a mean of 5.4 skate boards with a standard deviation of 1.2, and sales of the sports model will have a mean of 3.2 skate boards with a standard deviation of 0.8. The cost of setting up for the sidewalk sale is $200.

(a) (3 marks) Define random variables and use them to express the shop's net profit
(b) (3 marks) What's the mean of the net profit (writing down formula and answer)?
(c) (3 marks) What's the standard deviation of the net profit (writing down formula and answer)?
(d) (4 marks) Do you need to make any assumptions in calculating the mean? How about the standard deviation? Please provide your reasoning for your answers.

END OF ASSIGNMENT