Assignment 1 STA4164
Due: Feb, 23 2018
Put all working out and code into a single pdf file.

1 Question 1

A

There was a limited edition set of toys, where only 6 were produced and by hand. Their weights are:
$W = [2.05, 1.8, 2.2, 2.15, 1.75, 1.95]$
a) Would it be correct to say we are computing the sample mean or the population mean? (1point)
b) Compute this mean by hand, or type it out but show the working. (1point)
c) Would you compute the sample variance or the population variance? (1point)
d) Compute the variance by hand or type it out but show the working. (1point)
e) What are the median and mode values? (1point)
f) Was there any skewness in the production process in terms of the weights of these toys? (2points)

B

Someone goes fishing and catches a bucket of fish of the same species. Their weights are:
$W_{fish} = [1.25, 2, 1.75, 1, 1.3, 1.8, 1.7, 1.2, 1.5]$ Looking at pictures of fish, the expected weight of the fish in the image of a webpage is stated to be 1.65.
a) What does the degrees of freedom refer to in the Student-T distribution? (1point)
b) When would we use the student-T test over the Z score? (2points)
c) What is the t-test equation? (1point)
d) Using the t-test can you test the weights of the fish caught support or reject that these fish belong to the same species seen in the catalogue? (4points)

C

We observe from our campus the temperature and count the number of squirrels. Our obesrvations are
$T = [52, 52, 50, 54, 50, 52, 54, 80, 80]$ $Sq = [8, 10, 6, 9, 6, 12, 12, 1, 0]$
a) What is the covariance of these vectors? (1point)
b) What is the covariance matrix? (1point)
c) What is the correlation coefficient? (1point)
d) Find the coefficient of determination? (1point)
e) Can you make any statement or conclusion given these numbers as to the relationship of these paired observations? (1point)
f) If we try to fit a straight line model to these variables; What is the best choice for the dependent and independent variable and why? What is the slope β_1 ?

What is the intercept value β_0 ? (show your working) (4points)
g) Which aspect of the data is not being captured and does this affect the methodology of the straight line model used? (2points)
h) What is a good measure to assess the quality of the straight line fitted above? Use this measure. (1point)

D

We have a marketing manager who believes that for 1 every dollar put into advertising, 2 dollars are returned. This manager is very confident about the hypothesis. You observe the advert spending budget and sales return:
$B = [10, 20, 40, 80, 160, 320]$
$S = [0, 10, 90, 150, 300, 600]$
a) Assuming a straightline model, use a test statistic to support or reject the hypothesis that the return from the budget is 2. (6points)

E

A student looks their horoscope rating every morning before taking an exam. The student keeps a log of the horoscope ratings and the exam performance. This is the log:
$H = [3, 4, 9, 10, 6, 7]$
$E = [80, 90, 75, 95, 85, 85]$
Test if least squares regression produces a significant F statistic:
$F = \frac{MSRegression}{MSResidual} = \frac{{}^{(SSY-SSE)}/_{k}}{{}^{SSE}/_{(n-k-1)}}$
a) What conclusion can you arrive at? (4points)

Question 2

A

Here is the data on some countries:

Country	Roadlength	Population	GDP (PPP)
USA	5486610	324,532,000	57,294
India	5472144	1,293,057,000	6,664
Russia	139600	144,554,993	25,185
Italy	487700	60,674,003	36,191
Vietnam	199567	92,700,000	6,377
Peru	139295	31,151,643	13,018
Ireland	96155	4,757,976	69,375

Using SAS or Python, analyze this data in any meaningful way you see fit to produce some insight. Possibilities include correlation coefficients, F statistics, the slopes, the number of dependent variables, or even finding more data on each country to include in the table that will add to the regression. You can choose the dependent variable among these 3. It is possible to go through the results of straightline regression and interpret the resuls with figures. (6 points)

B

You have 2 CSV files with the prices of Bitcoin (BTC). a) Find the mean and standard error for the prices in each file (4 points)
b) Using the mean of the 2 year prices, compare it to that of the data in the 2018-01-24 to 2018-02-08 with a one sample T-test. What conclusion can you arrive at? (3 points)