THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

This exam paper must not be removed from the venue

Venue ____________________

Seat Number ________

Student Number |__|__|__|__|__|__|__|__|

Family Name _____________________

First Name _____________________

School of Civil Engineering

EXAMINATION

Semester One Final Examinations, 2017

CIVL2530-1 Probability, Statistics and Scientific Computing (Probability and Statistics)

This paper is for St Lucia Campus students.

Examination Duration: 120 minutes

Reading Time: 10 minutes

Exam Conditions:

This is a Central Examination

This is a Closed Book Examination - no materials permitted

During reading time - write only on the rough paper provided

This examination paper will be released to the Library

Materials Permitted In The Exam Venue:

(No electronic aids are permitted e.g. laptops, phones)

Calculators - Casio FX82 series or UQ approved (labelled)

Materials To Be Supplied To Students:

Formulae and tables booklet

Instructions To Students:

Additional exam materials (eg. answer booklets, rough paper) will be provided upon request.

Answer all questions on the exam paper.

For Examiner Use Only

Question	Mark
1a	
1b	
1c	
1d	
2a	
2b	
2c	
2d	
3a	
3b	
3c	
3d	
4a	
4b	
4c	
4d	

Total ________

Instructions

The exam consists of 4 questions, 1-4. Each question has four items, a-d.

Within each question:

Item (a) carries a weight of 8 marks.

Item (b) carries a weight of 7 marks.

Item (c) carries a weight of 6 marks.

Item (d) carries a weight of 4 marks.

The total marks in the exam are 100.

Answer ALL questions in the spaces provided.

If more space is required, use the back of the PREVIOUS page.

Show all your working and include sketches where appropriate.

Work written in the formulae and tables booklet will NOT be marked.

Question 1:

The reading on a sensor is a continuous random variable X with density function

$$f(x) = \begin{cases} K, & -2 \le x \le 2, \\ 0, & \text{otherwise}, \end{cases}$$

where K is a positive constant.

(a) What is the variance of X as a numerical value?

(b) Sketch an illustration of the cumulative distribution function of X. Make sure to label the axes and include values on the axes.

(c) Assume that 4 such independent sensors are operating. Let N denote the number of sensors that have a positive reading. Write an expression for the probability mass function of the random variable N.

(d) You measure this system of sensors, over 10 days, obtaining i.i.d. observations $N_1, \ldots, N_{10}$ each distributed as N in (c). The total number of sensors with a positive reading is $S = \sum_{i=1}^{10} N_i$. What is the variance of the random variable S?

Question 2:

A pair of measurements, denoted (X, Y) are distributed as a bivariate Normal random variable. It is known that $E(X) = E(Y) = 0$ and $\text{var}(X) = \text{var}(Y) = 4$. Further, $\text{cov}(X, Y) = 2$.

(a) Find the correlation coefficient between X and Y.

(b) Let $U = \text{sign}(X \times Y)$ where

$$\text{sign}(x) = \begin{cases} -1, & x < 0, \\ 0, & x = 0, \\ +1, & x > 0. \end{cases}$$

Hence the support of the discrete random variable U is $\{-1, +1\}$. Does it hold that,

$$P(U = 1) \;>\; P(U = -1),$$

or is it the other way around? Use your answer to (a) to explain your answer.

(c) The measurements, X and Y are summed such that $W = X + Y$. Determine the probability that $|W| \leq 5$, showing your working.

(d) Assume two independent pairs of measurements are made, (X_1, Y_1) and (X_2, Y_2), with each pair having the same bivariate distribution as (X, Y) above. Denote,

$$M = \frac{X_1 + X_2}{2} + \frac{Y_1 + Y_2}{2}$$

What is the variance of M? Show your working.

Question 3:

An engineer is designing a kayak that holds a single **person** together with a **bag**. It is known that kayak users have a mean (person) weight of $\mu_1 = 78$ and a standard deviation of $\sigma_1 = 12.3$ (all units in kilograms). Denote the (unknown) population mean and population standard deviation of bag weights, by μ_2 and σ_2 respectively.

The engineer measures bags of 16 randomly selected users. A sample mean of $\overline{x} = 25.2$ and a sample standard deviation of $s = 4.2$ are obtained.

(a) Find a 99% confidence interval for μ_2. State your assumptions on the population of bags.

(b) Previous manufacturers have assumed a bag weight of 30. However the data collected indicate a potentially lower weight. Hence, the engineer wishes to claim that bag weight assumptions were too conservative. She decides to test the hypothesis, $H_0 : \mu_2 = 30$ vs. $H_1 : \mu_2 < 30$, setting $\alpha = 10\%$. What does the engineer conclude? Carry out the hypothesis test.

(c) The engineer wishes to assess the power of the hypothesis test in (b) if it holds that $\mu_2 = 23$ and $\sigma_2 = 4$. For this she produces the following code but is missing numerical values in “AAAA” and “BBBB”. What values should she use? Explain your answer.

```
using Distributions
function tStatisticUnderH1()
    data = [rand(Normal(23,4)) for _ in 1:16]
    xBar = mean(data)
    s = std(data)
    tStatistic = (xBar - AAAA)/(s/4)
    return tStatistic
end
sum([tStatisticUnderH1() < BBBB for _ in 1:10^6])/10^6
```

(d) At this point, the engineer assumes that $\mu_2 = 25.2$ and $\sigma_2 = 4.2$. She further assumes that person weights are independent of bag weights and that all weights are Normally distributed. To what weight capacity (person + bag) should the engineer design the Kayak to accommodate 99% of the users? Show your working.

Question 4:

An engineer is comparing the density of two types of materials, A and B. She measures $n_1 = 16$ density observations of material A and $n_2 = 26$ density observations of material B (all measurements in g/cm^3). Denoting A by 1 and B by 2, the sample means are,

$$\overline{x}_1 = 5.81, \qquad \overline{x}_2 = 6.5,$$

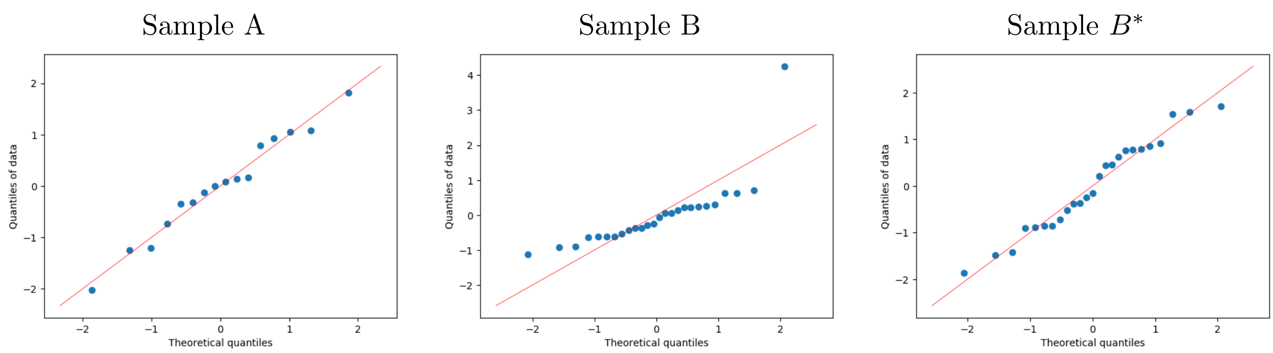
and the sample standard deviations are,

$$s_1 = 2.39, \qquad s_2 = 3.11.$$

Further, the maximal observations for each sample are,

$$\max_1 = 10.15, \qquad \max_2 = 19.6.$$

(a) The engineer produces Normal probability plots of samples A, B and B^* where B^* excludes the maximal observation (has $n_2 = 25$).



The vertical axis gives "quantiles of the data" and the horizontal axis gives "theoretical quantiles". Comment on the normality of the samples, based on the plots.

(b) The engineer became aware of equipment failure associated with the maximal observation of sample B and hence decides to use sample B^*. Calculate $\overline{x}_2$ and s_2 for B^*.

(c) Carry out a hypothesis test for $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Use samples A and B^* and assume that the population variances are equal. State your conclusion with $\alpha = 0.05$.

(d) Below is computer output of the test above with an **unequal** variance assumption. Under this assumption, you wish to carry out an hypothesis test for $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 < \mu_2$. Denote the student t density (pdf) with v degrees of freedom as $f(x\,;\,v)$ (v may be non-integer). Use $f(\cdot\,;\,\cdot)$ and numerical values from the output below to write an expression for the p-value. (You may round numerical values for brevity).

```
Two sample t-test (unequal variance)
------------------------------------
Population details:
    parameter of interest:   Mean difference
    value under h_0:         0
    point estimate:          -0.16522552415180058
    95% confidence interval: (-1.5683388857073237,1.2378878374037225)

Test summary:
    outcome with 95% confidence: fail to reject h_0
    two-sided p-value:           0.8099938625942766 (not significant)

Details:
    number of observations:   [16,25]
    t-statistic:              -0.24311622813636102
    degrees of freedom:       23.85344603495639
    empirical standard error: 0.6796153651212766
```