CONFIDENTIAL

Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER II
SESSION 2016/2017

COURSE NAME	:	INTRODUCTION TO STATISTICAL ANALYSIS
COURSE CODE	:	BWJ 10703
PROGRAMME CODE	:	BWW
EXAMINATION DATE	:	JUNE 2017
DURATION	:	3 HOURS
INSTRUCTION	:	ANSWERS **FIVE** (5) QUESTIONS ONLY

TERBUKA

THIS QUESTION PAPER CONSISTS OF **FIVE (5)** PAGES

CONFIDENTIAL

Q1 In a study of survival times from diagnosis (in months) for 20 patients with acute lymphoblastic leukemia presented as the following.

Male patients :	6	5	2	2	2	4	1	6	8	4
Female patients :	5	1	7	3	2	1	2	7	5	6

(a) Classify the given data either qualitative, quantitative, discrete, continuous, ordinal or nominal. Your answer can be more than one.

(2 marks)

(b) Construct a table of frequency distribution by obtaining the number of month, number of patient and cumulative frequency.

(4 marks)

(c) Calculate the sample mean and median.

(5 marks)

(d) Let Pr(M) be the probability of selecting male patients and Pr(F) be the probability of selecting female patients.

(i) Are Pr(M) and Pr(F) mutually exclusive? Give your reason.

(2 marks)

(ii) What is the probability that the selected patient is survive from acute lymphoblastic leukemia in 2 months?

(3 marks)

(iii) If a male patients are selected at random, calculate the percentage that he survives from acute lymphoblastic leukemia less than 3 months.

(4 marks)

Q2 (a) Assume that the number of sightings for bald eagles (*Haliaeetus leucocephalus*) at a site over a 4-week period is Poisson distributed (temporally random) with a mean 10. What is the probability of having not more than 15 sightings over a 6-week period?

(6 marks)

(b) A researcher knows that the variance in height for a large population of 20-year-old farmed pine trees is 33 m^2. The researcher finds that the sample mean of 20-year-old pine tree heights from a sample of 55 is 76 m. Find out the number of pine tree that its height is 65 m but less than 81 m.

(8 marks)

(c) "Rapid HIV tests" provide results for the presence/absence of HIV antibodies in a matter of minutes. One such test, which uses oral fluids, has a probability of 0.08 of producing a false-positive result. Given that 100 people, who are antibody free, are tested using this method, calculate the percentage that more than 5 but less than 15 will receive a false positive result? Assume that test results for different individuals are independent.

(6 marks)

Q3 (a) Assume that leaf biomass (in grams) from the plant *Salix Artica* follows the probability density function

$$f(x)=\begin{cases}2(x+1)^{-3} & 0<x<10\\ 0 & \text{otherwise}\end{cases}$$

Calculate the probability of leaf being between 3g and 5g.

(5 marks)

(b) Bliss and R.A. Fisher (1953) examined female European red mite counts (*Panonychus ulmi*) on Mcintosh apple trees [*Malus domestica* (Mcintosh)]. Counts of the mites on 150 leaves are shown as **Table Q3(b)** below.

Table Q3 (b)

Mites per leaf	0	1	2	3	4	5	6	7	8
Leaves observed	70	38	17	10	9	3	2	1	0

(i) Construct the table of probability distribution function.

(3 marks)

(ii) Is the random variable discrete or continuous? Then, prove it.

(4 marks)

(iii) Calculate the expected number of mites per leaf.

(3 marks)

(iv) Compute the standard deviation of the number of mites per leaf.

(5 marks)

Q4 It is claimed that a new treatment is more effective than the standard treatment for prolonging the lives of terminal cancer patients. The standard treatment has been in use for a long time, and from records in medical journals, the mean survival period is known to be 4.2 years. The new treatment is administered to 80 patients and their duration of survival recorded. The sample mean and the standard deviation are found to be 4.5 and 1.1 years, respectively.

(a) State the appropriate null and alternative hypotheses.

(4 marks)

(b) Is the hypothesis test a left-tailed, right-tailed or two-tailed test?

(1 marks)

(c) Construct the rejection region if the test using a 0.05 level of significance.

(5 marks)

(d) Is the claim supported by these results?

(10 marks)

Q5 The following observations of body weight (kg) and food consumption (calories/day) were collected in a study on obesity in adolescent girls as in **Table Q5**.

Table Q5

Food Consumption (calories/day)	Body Weight (kg)
2680	60.4
3280	81.1
3890	94.9
3170	86.4
3390	90.3
2670	60.4
2770	77.8
3330	85.0
2710	71.6
2600	64.6
2880	75.1
3430	89.6
3160	84.4
3330	93.0
2360	61.3
3030	74.9

(a) Indicate which variable is to be independent variable and which is to be the dependent variable.

(2 marks)

(b) Calculate the sample mean for body weight and food consumption.

(3 marks)

(c) Compute a correlation coefficient to determine if there is relationship between body weight and food consumption or not. Interpret your result.

(8 marks)

(d) Obtain the estimated regression model. Interpret your result.

(5 marks)

(e) Estimate for the weight of a girl whose daily consumption was 3000 calories.

(2 marks)

- END OF QUESTIONS -

FINAL EXAMINATION

SEMESTER / SESSION : SEM II / 2016/2017
COURSE NAME : INTRODUCTION TO STATISTICAL ANALYSIS
PROGRAMME CODE : BWW
COURSE CODE : BWJ 10703

FORMULA

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

$$\Pr(A \cap B) = \Pr(B) \times \Pr(A \mid B)$$

$$\Pr(X = r) = \binom{n}{r} p^r (1-p)^{n-r} = {}^nC_r p^r q^{n-r}$$

$$\Pr(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} = \frac{{}^rC_x \, {}^{N-r}C_{n-x}}{{}^NC_n}$$

$$\tilde{x} = L_B + \left(\frac{\frac{(\Sigma f_i)+1}{2} - F_B}{f_m}\right) \times C$$

$$\hat{x} = L_B + \left(\frac{\Delta_B}{\Delta_B + \Delta_A}\right) \times C$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} f_i \; x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} f_i \; x_i\right)^2$$

$$\hat{\beta}_1 = \frac{\sum xy - \frac{\Sigma x \Sigma y}{n}}{\sum x^2 - \frac{(\Sigma x)^2}{n}} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2]}\sqrt{[n\Sigma y^2 - (\Sigma y)^2]}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$