CONFIDENTIAL

Universiti Tun Hussein Onn Malaysia

# UNIVERSITI TUN HUSSEIN ONN MALAYSIA

## FINAL EXAMINATION
## SEMESTER II
## SESSION 2017/2018

| | | |
|---|---|---|
| COURSE NAME | : | INTRODUCTION TO STATISTICAL ANALYSIS |
| COURSE CODE | : | BWJ 10703 |
| PROGRAMME CODE | : | BWW |
| EXAMINATION DATE | : | JUNE / JULY 2018 |
| DURATION | : | 3 HOURS |
| INSTRUCTION | : | ANSWERS **ALL** QUESTIONS |

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

CONFIDENTIAL

**Q1** The following data represent the length of life, in seconds, of 30 fruit flies subject to a new spray in a controlled laboratory experiment

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 11 | 7 | 10 | 6 | 8 | 7 | 5 | 9 | 12 |
| 9 | 6 | 5 | 12 | 5 | 11 | 9 | 8 | 6 | 5 |
| 12 | 9 | 10 | 8 | 9 | 6 | 10 | 7 | 11 | 8 |

(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 time, number of fruit flies and cumulative frequency.

(2 marks)

(c) Calculate the median.

(3 marks)

(d) Construct the table of probability distribution function.

(2 marks)

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

(4 marks)

(f) Compute the expectation and variance.

(5 marks)

(g) Calculate the probability that the length of life of fruit flies only six or twelve seconds.

(3 marks)

(h) Find the percentage of fruit flies that have length of life not more than seven seconds.

(4 marks)



**Q2** (a) Biologists doing studies in a particular environment often tag and release subjects in order to estimate the size of a population or the prevalence of certain features in the population. 10 animals of a certain population thought to be extinct (or near extinction) are caught, tagged and released in a certain region. After a period of time, a random sample of 15 of this type of animal is selected in the region. What is the probability that 5 of those selected are tagged if there are 25 animals of this type in the region?

(3 marks)

(b) A research scientist reports that mice will live an average of 40 months when their diets are sharply restricted and the enriched with vitamins and proteins. Assuming that the lifetimes of such mice are normally distributed with a standard deviation of 5 months, find the probability that a given mouse will live 37 months but less than 49 months

(5 marks)

(c) The probability that a patient recovers from a delicate heart operation is 0.9. Of the next 100 patients having this operation,

(i) calculate the probability that between 5 and 16 inclusive not survive.
(4 marks)

(ii) obtain the number of patients that fewer than 12 not survive.
(5 marks)

(d) The average number of field mice per acre in a wheat field is estimated to be 12. Find the probability that

(i) not less than 7 field mice are found on a given acre.
(3 marks)

(ii) not more than 25 field mice on the next two acres inspected.
(5 marks)

**Q3** The chemical benzene is highly toxic to humans. However, it is used in the manufacture of any medicine dyes, leather and coverings. Government regulations dictate that for any production process involving benzene, the water in the output of the process must not exceed 7950 parts per million (ppm) of benzene. For a particular process of concern, the water sample was collected by a manufacturer 36 times randomly and the sample average was 7960 ppm. It is known from the historical data that the standard deviation is 100 ppm. Assume that the distribution of benzene concentration is normal.

(a) Formulate the appropriate null and alternative hypotheses.
(2 marks)

(b) Identify either the hypothesis tests a left-tailed, right-tailed or two-tailed test.
(1 mark)

(c) Determine an appropriate statistical test for this problem. Give your reason.
(3 marks)



(d) Use result in part **Q3 (a)** to construct the rejection region if the test using a 0.05 level of significance.
(4 marks)

(e) Based on the sample data and result in part **Q3 (d)**, conducts hypotheses test whether that water sample fulfills the government regulation at 5% level significant.
(7 marks)

(f) Conducts hypotheses test if a new water sample that collected 36 times randomly with the sample average 7980 ppm fulfills the government regulation at 5% level significant.
(7 marks)

(g) Interpret your result in part **Q3 (e)** and **Q3 (f)**.
(1 mark)

**Q4** Anderson et al. (1990) performed a study on the effects of oat bran on serum cholesterol for males with high or borderline high values of serum cholesterol. High values of serum cholesterol are greater than or equal to 240 mg/dL (6.20 mmol/L). We wish to use the data from the study to determine whether or not there is a linear relation between body mass index and serum cholesterol. The body mass index is defined as weight (in kilograms) divided by the square of height (in meters). The data are shown in **Table Q4**.

**Table Q4** : Body Mass Index and Serum Cholesterol

| Serum Cholesterol | Body Mass Index |
|---|---|
| 7.29 | 29.0 |
| 8.43 | 21.6 |
| 5.43 | 27.2 |
| 6.96 | 25.2 |
| 6.65 | 25.1 |
| 8.20 | 27.9 |
| 5.92 | 31.9 |
| 8.04 | 26.3 |
| 7.96 | 21.8 |
| 5.77 | 24.8 |
| 6.23 | 24.5 |
| 6.26 | 23.5 |
| 6.21 | 24.8 |
| 5.92 | 24.4 |

(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 mass index and serum cholesterol.

(3 marks)

(c) Fit the simple linear regression model to the data.



(10 marks)

(d) Estimate the body mass index when the serum cholesterol equal to 6.2.

(2 marks)

(e) Compute a correlation coefficient to determine if there is relationship between body mass index and serum cholesterol or not. Give your interpretation.

(5 marks)

(f) Find the value of coefficient of determination to determine if there is relationship between body mass index and serum cholesterol or not. Interpret your result.

(3 marks)

**- END OF QUESTIONS -**

## FINAL EXAMINATION

SEMESTER / SESSION : SEM II / 2017/2018
COURSE NAME : INTRODUCTION TO STATISTICAL ANALYSIS
PROGRAMME CODE : BWW
COURSE CODE : BWJ 10703

## FORMULA

$$\sum_{i-1}^{n} \Pr(X_i) = 1$$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

$$E(X) = \sum_{i=1}^{n} x_i \times \Pr(X = x_i)$$

$$Var(X) = E[(x-\mu)^2] = E(x^2) - [E(x)]^2$$

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

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

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

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim Z_\alpha$$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_\alpha(v)$$

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim T_\alpha(v)$$

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

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

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

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

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

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

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

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

$$\Pr(X = r) = \frac{e^{-\mu}\mu^r}{r!}$$

$$Z = \frac{x-\mu}{\sigma} \sim Z_\alpha$$

$$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim Z_\alpha$$

$$T = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim T_\alpha(v = n-1)$$

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

$$\bar{x} = \frac{1}{n}\sum_{i-1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

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

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$r^2 = \frac{(S_{xy})^2}{S_{xx}S_{yy}} = R^2$$