SULIT

BAHAGIAN PEPERIKSAAN DAN PENILAIAN
JABATAN PENDIDIKAN POLITEKNIK
KEMENTERIAN PENDIDIKAN TINGGI

JABATAN KEJURUTERAAN AWAM

PEPERIKSAAN AKHIR

SESI JUN 2015

DCC3132: STATISTICS

TARIKH : 04 NOVEMBER 2015
TEMPOH : 8.30AM – 10.30AM (2 JAM)

Kertas ini mengandungi **TIGA BELAS (13)** halaman bercetak.
Bahagian A: Struktur (2 soalan)
Bahagian B: Struktur (4 soalan)
Dokumen sokongan yang disertakan : Formula

JANGAN BUKA KERTAS SOALAN INI SEHINGGA DIARAHKAN

(CLO yang tertera hanya sebagai rujukan)

SECTION A : 50 MARKS
BAHAGIAN A : 50 MARKAH

INSTRUCTION:
This section consists of **TWO (2)** structured questions. Answer **ALL** questions.

ARAHAN:
*Bahagian ini mengandungi **DUA (2)** soalan berstruktur. Jawab **SEMUA** soalan.*

QUESTION 1

SOALAN 1

CLO1
C1

(a) i. Define statistics.
Takrifkan statistik.

[3 marks]
[*3 markah*]

ii. List **TWO (2)** types of statistic.

CLO1
C2

*Senaraikan **DUA (2)** jenis statistik.*

[2 marks]
[*2 markah*]

(b) There are **TWO(2)** types of variable. Explain both.
*Terdapat **DUA(2)** jenis pemboleh ubah. Terangkan kedua-dua pemboleh ubah tersebut.*

[10 marks]
[*10 markah*]

CLO1
C3

(c) Students are required to analyze the level of customers' satisfaction with the food at a cafeteria in Hostel A. Apply the Statistics Problem-Solving to analyze the statement above.
Pelajar dikehendaki untuk menganalisa tahap kepuasan pelanggan terhadap makanan yang disediakan di kafeteria Kamsis A. Aplikasikan kaedah Penyelesaian Masalah Statistik untuk menganalisa kajian ini.

[10 marks]
[*10 markah*]

QUESTION 2

SOALAN 2

CLO1 C1

(a) There are many methods used to collect or obtain data for statistical analysis. State **THREE (3)** of the most popular methods.

*Terdapat beberapa kaedah untuk mengumpul data dalam analisis statistik. Nyatakan **TIGA (3)** kaedah yang biasa digunakan.*

[5 marks]

[*5 markah*]

CLO1 C2

(b) There are two types of data classification while collecting a data. Explain briefly **TWO (2)** types of data classification.

*Terdapat dua jenis klasifikasi data semasa mengumpul data. Terangkan **DUA (2)** jenis pengkelasan data.*

[10 marks]

[*10 markah*]

CLO1 C3

(c) The engineer of the concrete precast factory wishes to estimate the average time (in minutes) for an operator to complete a certain job. The factory has 98 operators. 8 operators selected at random and their time to complete a particular job is recorded.

i) Recognize the variable of interest and also the type of variable used.

[5 marks]

ii) Choose and interpret what sampling method used in this research.

[5 marks]

Seorang jurutera di kilang pembuatan konkrit pratuang ingin mendapatkan masa purata(dalam minit) bagi operator menyelesaikan tugasan mereka, Kilang tersebut mempunyai 98 orang operator. 8 orang operator dipilih secara rawak dan masa menyiapkan kerja mereka direkodkan.

i) Kenalpasti pembolehubah dan jenis pembolehubah yang digunakan.

[5 markah]

ii) Pilih dan jelaskan kaedah persampelan yang digunakan dalam kajian ini.

[5 markah]

SECTION B : 50 MARKS
BAHAGIAN B : 50 MARKAH

INSTRUCTION:
This section consists of **FOUR (4)** structured questions. Answer **TWO (2)** questions only.

ARAHAN:
*Bahagian ini mengandungi **EMPAT (4)** soalan berstruktur. Jawab **DUA (2)** soalan sahaja.*

QUESTION 1

SOALAN 1

CLO2 C1

a) Identify the median for the following data.

Kenalpasti median bagi data di bawah.

11 14 4 7 21 18 22 16 19 15 7 8 19 8 9 4 18

[5 marks]

[*5 markah*]

CLO2 C3

b) The age distribution of insurance agents in an insurance company is as shown in **Table B1.**

*Taburan umur ejen insurans di sebuah syarikat insurans adalah seperti yang ditunjukkan dalam **Jadual B1**.*

Table B1/*Jadual B1* Age distribution of insurance agents

Age (year)	Number of agents
21-25	10
26-30	35
31-35	16
36-40	14
41-45	12
46-50	10
51-55	3

i) Calculate the mean age of the insurance agents.
Kirakan nilai min bagi umur ejen insurans.

[4 marks]
[*4 markah*]

ii) Draw a histogram for the above data.
Lukiskan histogram bagi data di atas.

[4 marks]
[*4 markah*]

iii) From the histogram, calculate the mode value and explain its meaning.
Daripada histogram tersebut, kira nilai mod dan berikan maksud bagi nilai mod tersebut.

[2 marks]
[*2 markah*]

CLO2
C4

c) The data below shows the number of vehicles that arrive at Jalan Duta toll booth during 16 intervals of the 10 minute-duration.

Data di bawah menunjukkan bilangan kenderaan yang tiba di plaza tol Jalan Duta yang diambil 16 kali selang 10 minit.

25 55 34 32 25 18 25 32

29 28 44 40 34 28 25 42

i) Calculate the mean, median and mode for the data given.
Kirakan min, median dan mod bagi data yang diberi.

[4 marks]
[*4 markah*]

ii) Analyze the first quartile and third quartile.
Berikan analisis kuartil pertama dan kuartil ketiga.

[4 marks]
[*4 markah*]

iii) Draw a box-and-whisker plot for the data given.
Lukiskan kotak-dan-rerambut bagi data yang diberi.

[2 marks]
[*2 markah*]

QUESTION 2

SOALAN 2

CLO2
C3

a) Approximately 20 000 diploma were awarded in the Polytechnic in 2009 to students majoring in five programmes : civil engineering, electrical engineering, mechanical engineering, food technology and information science & technology as shown in the pie chart below. Answer the questions given.

Hampir 20 000 diploma telah dianugerahkan di Politeknik pada tahun 2009 dalam 5 program utama: Kejuruteraan Awam, Kejuruteraan Elektrik, Kejuruteraan Mekanikal, Teknologi Makanan dan Sains dan Teknologi Maklumat seperti yang ditunjukkan dalam carta pai di bawah. Jawab soalan yang diberikan.

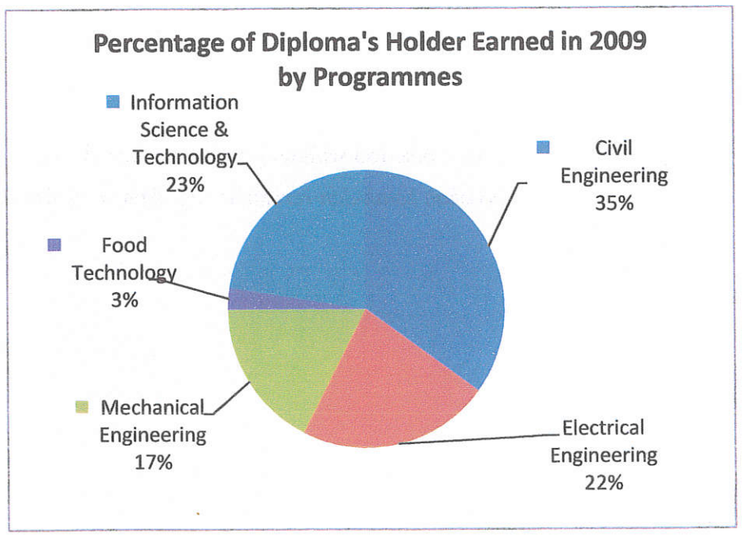

Figure 2a /*Rajah 2a*

i) Calculate the number of diploma's holders from each programme.
Kirakan bilangan pemegang diploma bagi setiap program.

[5 marks]
[*5 markah*]

ii) Calculate the total of diploma holders from the engineering field who have been awarded.
Kirakan bilangan penerima diploma dalam bidang kejuruteraan.

[2 marks]
[*2 markah*]

iii) Hence, illustrate a horizontal bar chart based on the number of diploma holder in 2009.
Seterusnya, lukis carta palang melintang berdasarkan bilangan penerima diploma untuk tahun 2009.

[8 marks]
[*8 markah*]

CLO2
C4

b) Each number in following **Table 2b** represents the weight of a new car in pounds.
*Setiap nombor dalam **Jadual 2b** berikut mewakili berat kereta baru dalam pound.*

Table 2b/*Jadual 2b*

2250	1760	2000	2100
1640	1820	2300	2210
2150	1930	2060	2350

i) Develop a frequency distribution table with the first class interval 1600-1799.
Bina jadual taburan kekerapan dengan selang kelas bermula dengan 1600-1799.

[4 marks]
[*4 markah*]

ii) Based on frequency distribution table in Question 2b (i), draw a graph of histogram.
Berdasarkan jadual taburan kekerapan dalam soalan 2b(i), lukiskan graf histogram.

[6 marks]
[*6 markah*]

QUESTION 3
SOALAN 3

CLO2
C3

a) Suppose there are 4 books on fairy tales, 5 novels and 3 plays. They have to be arranged so that the books on fairy tales are together, novels are together and plays are together, but we no longer require that they should be in a specific order. Calculate how many ways can this be done?

Katakan terdapat 4 buah buku mengenai cerita dongeng, 5 novel dan 3 drama. Mereka perlu diatur supaya buku cerita dongeng disusun bersama-sama , novel disusun bersama-sama dan buku drama disusun bersama-sama , tetapi kita tidak memerlukan buku-buku tersebut berada dalam aturan tertentu. Kirakan berapa carakah ia boleh dilakukan?

[15 marks]
[*15 markah*]

CLO2
C4

b) A committee of 5 persons is to be formed from 6 men and 4 women. Determine how many ways can it be done when:

Satu jawatankuasa yang mempunyai 5 orang ahli akan dibentuk daripada 6 lelaki dan 4 wanita. Tentukan berapakah cara ia boleh dilakukan apabila:

(i) At least 2 women are included.
Sekurang-kurangnya 2 wanita dimasukkan.

[5 marks]
[*5 markah*]

(ii) At most 2 women are included.
Paling banyak 2 wanita dimasukkan.

[5 marks]
[*5 markah*]

QUESTION 4
SOALAN 4

CLO2
C3

a) Calculate Pearson's correlation coefficient r for the following set of data.
Kira pekali korelasi Pearson r bagi set data berikut

x	3	5	8	10	13	15	18	20	28
y	30	35	41	50	51	60	65	66	70

[15 marks]
[*15 markah*]

CLO2
C4

b) Five students sat for both oral and written tests. The results are shown in the following table.
Lima pelajar menduduki kedua-dua ujian lisan dan bertulis. Keputusan ditunjukkan dalam jadual berikut.

Student / *Pelajar*	A	B	C	D	E
Written test score / *Skor ujian bertulis*	62	60	82	90	82
Oral test score / *Skor ujian lisan*	60	65	70	83	72

Determine Pearson's correlation coefficient.
Tentukan pekali korelasi Pearson ini.

[10 marks]
[*10 markah*]

SOALAN TAMAT

Formulas and Tables

Elementary Statistics: A Step-By-Step Approach
Bluman / Mayer, 1st Canadian Edition

Chapter 3 Data Description

Mean for individual data: $\overline{X} = \frac{\Sigma X}{n}$

Mean for grouped data: $\overline{X} = \frac{\Sigma f \cdot X_m}{n}$

Standard deviation for a sample:

$$s = \sqrt{\frac{\Sigma X^2 - [(\Sigma X)^2/n]}{n-1}}$$

Standard deviation for grouped data:

$$s = \sqrt{\frac{\Sigma f \cdot X_m^2 - [(\Sigma f \cdot X_m)^2/n]}{n-1}}$$

Range rule of thumb: $s \approx \frac{\text{range}}{4}$

Chapter 4 Probability and Counting Rules

Addition rule 1 (mutually exclusive events):

$$P(A \text{ or } B) = P(A) + P(B)$$

Addition rule 2 (events not mutually exclusive):

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Multiplication rule 1 (independent events):

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

Multiplication rule 2 (dependent events):

$$P(A \text{ and } B) = P(A) \cdot P(B \mid A)$$

Conditional probability: $P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$

Complementary events: $P(\overline{E}) = 1 - P(E)$

Fundamental counting rule: Total number of outcomes of a sequence when each event has a different number of possibilities: $k_1 \cdot k_2 \cdot k_3 \cdot \cdots \cdot k_n$

Permutation rule: Number of permutations of n objects taking r at a time is ${}_nP_r = \frac{n!}{(n-r)!}$

Combination rule: Number of combinations of r objects selected from n objects is ${}_nC_r = \frac{n!}{(n-r)!r!}$

Chapter 5 Discrete Probability Distributions

Mean for a probability distribution: $\mu = \Sigma[X \cdot P(X)]$
Variance and standard deviation for a probability distribution:

$$\sigma^2 = \Sigma[X^2 \cdot P(X)] - \mu^2$$
$$\sigma = \sqrt{\Sigma[X^2 \cdot P(X)] - \mu^2}$$

Expectation: $E(X) = \Sigma[X \cdot P(X)]$

Binomial probability: $P(X) = \frac{n!}{(n-X)!X!} \cdot p^X \cdot q^{n-X}$

Mean for binomial distribution: $\mu = n \cdot p$
Variance and standard deviation for the binomial distribution:
$\sigma^2 = n \cdot p \cdot q \qquad \sigma = \sqrt{n \cdot p \cdot q}$
Multinomial probability:

$$P(X) = \frac{n!}{X_1!X_2!X_3!\cdots X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdot p_3^{X_3} \cdots p_k^{X_k}$$

Poisson probability: $P(X; \lambda) = \frac{e^{-\lambda}\lambda^X}{X!}$ where

$X = 0, 1, 2, \ldots$

Hypergeometric probability: $P(X) = \frac{{}_aC_X \cdot {}_bC_{n-X}}{{}_{a+b}C_n}$

Chapter 6 The Normal Distribution

Standard score $z = \frac{X - \mu}{\sigma}$ or $\frac{X - \overline{X}}{s}$

Mean of sample means: $\mu_{\overline{X}} = \mu$

Standard error of the mean: $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

Central limit theorem formula: $z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$

Chapter 7 Confidence Intervals and Sample Size

z confidence interval for means:

$$\overline{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

t confidence interval for means:

$$\overline{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Sample size for means: $n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$ where E is the maximum error of estimate

Confidence interval for a proportion:

$$\hat{p} - (z_{\alpha/2})\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + (z_{\alpha/2})\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Formulas and Tables
Elementary Statistics: A Step-By-Step Approach
Bluman / Mayer, 1st Canadian Edition

z test for comparing two proportions:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$ $\quad \hat{p}_1 = \frac{X_1}{n_1}$

$\bar{q} = 1 - \bar{p}$ $\quad \hat{p}_2 = \frac{X_2}{n_2}$

Formula for the confidence interval for the difference of two proportions:

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

Chapter 10 Correlation and Regression

Correlation coefficient:

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

t test for correlation coefficient: $t = r\sqrt{\frac{n-2}{1-r^2}}$

(d.f. $= n - 2$)

The regression line equation: $y' = a + bx$

where $a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Coefficient of determination: $r^2 = \frac{\text{explained variation}}{\text{total variation}}$

Standard error of estimate:

$$s_{est} = \sqrt{\frac{\Sigma y^2 - a\Sigma y - b\Sigma xy}{n-2}}$$

Prediction interval for y:

$$y' - t_{\alpha/2}s_{est}\sqrt{1 + \frac{1}{n} + \frac{n(x - \overline{X})^2}{n\Sigma x^2 - (\Sigma x)^2}} < y < y' + t_{\alpha/2}s_{est}\sqrt{1 + \frac{1}{n} + \frac{n(x - \overline{X})^2}{n\Sigma x^2 - (\Sigma x)^2}}$$

(d.f. $= n - 2$)

Formula for the multiple correlation coefficient:

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

Formula for the F test for the multiple correlation coefficient:

$$F = \frac{R^2/k}{(1 - R^2)/(n - k - 1)}$$

(d.f.N. $= n - k$ and d.f.D. $= n - k - 1$)

Formula for the adjusted R^2:

$$R_{adj}^2 = 1 - \left[\frac{(1 - R^2)(n-1)}{n - k - 1}\right]$$

Chapter 11 Other Chi-Square Tests

Chi-square test for goodness-of-fit:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

(d.f. = no. of categories $- 1$)

Chi-square test for independence and homogeneity of proportions:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

[d.f. = (rows $- 1$)(col. $- 1$)]

Chapter 12 Analysis of Variance

ANOVA test: $F = \frac{s_B^2}{s_W^2}$ where $\overline{X}_{GM} = \frac{\Sigma X}{N}$

d.f.N. $= k - 1$ where $N = n_1 + n_2 + \cdots + n_k$
d.f.D. $= N - k$ where k = number of groups

$$s_B^2 = \frac{\Sigma n_i(\overline{X}_i - \overline{X}_{GM})^2}{k-1}$$

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

Scheffé test: $F_S = \frac{(\overline{X}_i - \overline{X}_j)^2}{s_W^2(1/n_i + 1/n_j)}$ and

$F' = (k - 1)(\text{C.V.})$

Tukey test: $q = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{s_W^2/n}}$

Formulas for two-way ANOVA:

$$MS_A = \frac{SS_A}{a-1} \qquad F_A = \frac{MS_A}{MS_W}$$

$$MS_B = \frac{SS_B}{b-1} \qquad F_B = \frac{MS_B}{MS_W}$$

$$MS_{A\times B} = \frac{SS_{A\times B}}{(a-1)(b-1)} \qquad F_{A\times B} = \frac{MS_{A\times B}}{MS_W}$$

$$MS_W = \frac{SS_W}{ab(n-1)}$$

Formulas and Tables

Elementary Statistics: A Step-By-Step Approach
Bluman / Mayer, 1st Canadian Edition

Chapter 13 Nonparametric Statistics

z test value in the sign test: $z = \dfrac{(X + 0.5) - (n/2)}{\sqrt{n}/2}$

where n = sample size (greater than or equal to 26)
X = smaller number of + or − signs

Wilcoxon rank sum test: $z = \dfrac{R - \mu_R}{\sigma_R}$

where

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

R = sum of the ranks for the smaller sample size (n_1)
n_1 = smaller of the sample sizes
n_2 = larger of the sample sizes
$n_1 \geq 10$ and $n_2 \geq 10$

Wilcoxon signed-rank test: $z = \dfrac{w_s - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}$

where

n = number of pairs where the difference is not 0
w_s = smaller sum in absolute value of the signed ranks

Kruskal-Wallis test:

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k}\right) - 3(N+1)$$

where

R_1 = sum of the ranks of sample 1
n_1 = size of sample 1
R_2 = sum of the ranks of sample 2
n_2 = size of sample 2
⋮
R_k = sum of the ranks of sample k
n_k = size of sample k
$N = n_1 + n_2 + \cdots + n_k$
k = number of samples

Spearman rank correlation coefficient:

$$r_S = 1 - \frac{6\,\Sigma d^2}{n(n^2 - 1)}$$

where

d = difference in the ranks
n = number of data pairs

Procedure Table

Solving Hypothesis-Testing Problems (Traditional Method)

STEP 1 State the hypotheses, and identify the claim.
STEP 2 Find the critical value(s) from the appropriate table in Appendix C.
STEP 3 Compute the test value.
STEP 4 Make the decision to reject or not reject the null hypothesis.
STEP 5 Summarize the results.

Procedure Table

Solving Hypothesis-Testing Problems (*P*-value Method)

STEP 1 State the hypotheses and identify the claim.
STEP 2 Compute the test value.
STEP 3 Find the *P*-value.
STEP 4 Make the decision.
STEP 5 Summarize the results.