SULIT

POLITEKNIK MALAYSIA

BAHAGIAN PEPERIKSAAN DAN PENILAIAN
JABATAN PENDIDIKAN POLITEKNIK
KEMENTERIAN PENDIDIKAN TINGGI

JABATAN KEJURUTERAAN AWAM

PEPERIKSAAN AKHIR

SESI DISEMBER 2015

DCC3132: STATISTICS

TARIKH : 14 APRIL 2016
MASA : 2.30 PM – 4.30 PM (2 JAM)

Kertas ini mengandungi **LAPAN BELAS (18)** 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) Define the following terms:

Berikan definisi terma berikut:

i) Statistics

Statistik

[3 marks]

[*3 markah*]

ii) Population

Populasi

[2 marks]

[*2 markah*]

CLO1
C2

b) Identify whether the statements below are qualitative or quantitative.

Kenalpasti samada kenyataan berikut merupakan pembolehubah kuantitatif atau kualitatif.

i) Health condition of village residents.

Taraf kesihatan penduduk-penduduk di sebuah kampung.

ii) The candidates which is listed in one of the election.

Calon yang diundi atas kertas undi yang terdapat dalam satu kotak mengundi, semasa pilihanraya dijalankan.

iii) The qualification of SPM students to enter the higher education level.
Kelayakan pelajar-pelajar lepasan SPM untuk memasuki institusi-institusi pengajian tinggi.

iv) Time taken by the novelist to write short story in one of the famous magazine.
Masa yang diambil oleh seorang pengarang terkenal untuk menulis sebuah cerpen dalam sebuah majalah terkenal.

v) Brands of cellular phones displayed in a telecommunication store.
Jenama telefon bimbit yang dipamerkan di kedai telekomunikasi.

[10 marks]
[*10 markah*]

CLO1
C3

c) Choose whether the following statements are nominal data or ordinal scale.
Pilih sama ada penyataan berikut adalah data nominal atau skala ordinal.

i) Numbers on the jerseys of basketball players.
Nombor pada jersey pemain bola keranjang.

ii) Marital status of respondents.
Status perkahwinan responden.

iii) Academic qualification of respondents.
Kelayakan akademik responden.

iv) IQ score of students.
Kedudukan markah IQ pelajar.

v) Social class of residents
Kelas social penduduk.

[10 marks]
[*10 markah*]

QUESTION 2
SOALAN 2

CLO1
C1

(a) Sampling techniques
Teknik pensampelan

i) Define non-probability sampling technique.
Definisikan maksud teknik pensampelan bukan kebarangkalian.

[2 marks]
[2 markah]

ii) List **THREE (3)** types of the non-probability sampling technique.
*Senaraikan **TIGA (3)**jenis teknik pensampelan bukan kebarangkalian.*

[3 marks]
[3 markah]

CLO1
C2

(b) There are several methods of collecting data and each has its own advantages and disadvantages.
Terdapat beberapa kaedah untuk mengumpulkan data dan setiap kaedah mempunyai kebaikan dan keburukan tersendiri.

i) Explain Mail (or postal) Questionnaire Method.
Terangkan Kaedah Soalselidik Secara Pos.

[4 marks]
[4 markah]

ii) Describe **THREE (3)** advantages and disadvantage of the mail (or postal) Questionnaire Method.
*Terangkan **TIGA (3)** kebaikan dan keburukan bagi kaedah Soalselidik Secara Pos.*

[6 marks]
[6 markah]

CLO1
C3

(c) Choose the types of sampling techniques used in each of the following situation:
Pilih teknik pensampelan yang digunakan didalam situasi yang berikut:

i) A survey on social media will be conducted in the state of Perak. The population is divided into rural and urban areas. About five percent of the urban and rural population will be selected randomly.
Sebuah kajian mengenai media sosial dijalankan di negeri Perak. Populasi negeri tersebut dibahagikan kepada kawasan bandar dan pedalaman. Sebanyak lima peratus dari kawasan bandar dan pedalaman dipilih secara rawak.

ii) The researcher had a total population of 100 individuals and need 12 subjects. He selected the fifth person as his starting number and chose 8 intervals for each pick.
Pengkaji telah mendapat jumlah populasi seramai 100 orang dan memerlukan 12 orang sebagai subjek. Beliau memilih orang yang kelima sebagai permulaan dan memilih sela 8 bagi setiap pemilihan.

iii) Random numbers generated by a computer for the selection of serial number of lorry engine for testing.
Nombor rawak dijana menggunakan komputer untuk pemilihan nombor siri enjin lori untuk diuji.

iv) An initial group of respondents for house price survey was selected. After being interviewed, these respondents were asked to identify others who belong to the target population of interest. This procedure was applied until 200 respondent obtained.
Satu kumpulan permulaan responden telah dipilih untuk kajian harga rumah. Selepas ditemubual,responden diminta untuk mengenalpasti responden lain yang termasuk didalam sasaran populasi. Prosedur ini diteruskan sehingga mendapat 200 responden.

v) A research was conducted on given quotas based on demographics. The researcher was told to interview 100 subjects with 50 males and 50 females between the ages of 18 to 21, employed and living at Melaka.
Satu kajian telah dijalankan berdasarkan kuota demografik yang diberikan. Pengkaji telah diminta untuk menemubual 100 subjek dengan 50 lelaki dan 50 perempuan berumur diantara 18 hingga 21 dan bekerja serta tinggal di Melaka.

[10 marks]
[10 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) Mean of data is a set of numerical values which is the arithmetic average of the data values in the set. List **FIVE (5)** characteristics of Mean.

*Min bagi data adalah satu set nilai berangka bagi purata aritmetik dalam set data. Senaraikan **LIMA (5)** ciri-ciri daripada min.*

[5 marks]

[*5 markah*]

CLO2 C3

(b) The following set of data represent billions of dollars owned by 10 hypothetical wealthy men. Calculate the mean, median and mode these top 10 rich men.

Set data berikut mewakili jumlah wang dalam billion dollar yang dimiliki oleh 10 orang terkaya. Kirakan min , median dan mod bagi 10 orang kaya tersebut.

12.6, 13.7, 18.0, 18.0, 18.0, 20.0, 20.0, 41.2, 48.0, 60.0

[10 marks]

[*10 markah*]

CLO2 C4

(c) The following data represent the costs per load (in cents) of selling used item by an Environmental Organization. Determine the mean value.

Data menunjukkan kos per muatan (dalam sen) penjualan item terpakai oleh Pertubuhan Alam Sekitar. Tentukan nilai min.

Class limits	Frequency
13-19	3
20-26	7
27-33	12
34-40	4
41-47	6
48-54	2
55-61	0
62-68	3

[10 marks]

[*10 markah*]

QUESTION 2

SOALAN 2

CLO2
C3

(a) Given the recorded data of the height of patients obtained from the monthly medical check-up tabulated in **Table 2a**.

Diberi data rekod ketinggian pesakit semasa mendapatkan rawatan pemeriksaan bulanan seperti dalam ***Jadual 2a****.*

Table 2a / *Jadual 2a*

Height / *Ketinggian* (cm)		
173	180	150
173	155	155
178	175	158
160	170	157
175	179	178
163	185	179
165	184	183
166	182	177
169	170	165
188	166	166

(i) Complete the data using frequency distribution.

Lengkapkan data menggunakan jadual kekerapan

[10 marks]

[*10 markah*]

(ii) Sketch the histogram of the distribution.

Lakarkan taburan histogram

[5 marks]

[*5 markah*]

CLO2
C4

(b) **Table 2b** shows demand by market segment for both natural and synthetic latex in Malaysia. Illustrate this information in the form of percentage using a pie chart.

Jadual 2b *menunjukkan keperluan penggunaan bagi kedua-dua produk getah asli dan sintetik di Malaysia. Lukiskan maklumat ini dengan menggunakan carta pai dalam bentuk peratus.*

Table 2b / *Jadual 2b*

Global latex demand by market segment (applications) *Permintaan getah global mengikut segmen pasaran (permohonan)*	
Market segment / *segmen pasaran*	Demand / *permintaan* (1000 tons)
Paints and coatings / *Cat dan salutan*	2170
Paper and paperboard / *Kertas dan papan kertas*	1960
Adhesives and sealants / *Pelekat dan pengedap*	1600
Carpets / *permaidani*	705
Medical / *perubatan*	334
Non-wovens / *Bukan wovens*	246
Textiles / *Tekstil*	788
Other applications / *permohonan lain*	600

[10 marks]

[*10 markah*]

QUESTION 3

SOALAN 3

CLO2
C3

(a) There are 20 students in Geomatics Unit (8 males and 12 females), 10 students in Architecture Unit (7 males and 3 females) and 12 students in Civil Engineering Unit (7 males and 5 females). If one of students is selected randomly, calculate the probability of:

Terdapat 20 orang pelajar Unit Geomatik (8 lelaki dan 12 perempuan), 10 orang pelajar Unit Seni Bina (7 lelaki dan 3 perempuan) dan 12 orang pelajar Unit Kejuruteraan Awam (7 lelaki dan 5 perempuan) Jika seorang pelajar tersebut dipilih secara rawak, kira kebarangkalian bagi:

i. Architecture Unit or female

Unit Seni Bina atau perempuan.

[5 marks]

[*5 markah*]

ii. Geomatics Unit or male

Unit Geomatik atau lelaki.

[5 marks]

[*5 markah*]

iii. Non Civil Engineering Unit

Bukan Unit Kejuruteraan Awam.

[5 marks]

[*5 markah*]

CLO2
C4

(b) i. The advertising director for television show has 7 advertisements to be used on the program. If he selects one of them for the opening of the show, one for the middle of the show and one for the ending of the show, calculate possible ways how it can be accomplished.

Pengarah pengiklanan persembahan tv mempunyai 7 iklan untuk sesuatu program. Jika dia memilih salah satu iklan pada permulaan persembahan, satu di pertengahan persembahan dan satu lagi di hujung persembahan, kirakan bilangan cara berkemungkinan untuk disiapkan.

[5 marks]

[*5 markah*]

ii. An editor of a publishing company has received 8 books to review. He decides that he can use 3 reviews in his Publishing Company. Determine different ways that these 3 reviews can be selected.

Editor Percetakan telah menerima 6 buah buku untuk disemak. Dia membuat keputusan untuk menggunakan 3 semakan untuk syarikat percetakannya. Tentukan bilangan cara yang berbeza yang boleh dipilih.

[5 marks]

[*5 markah*]

QUESTION 4

SOALAN 4

CLO2
C3

(a) The director of a Green Technology Association wants to determine whether there is any type of relationship between the amount of a contribution from selling reuse items (in Ringgit) and the years of selling it. The data are as follows in the table below:

Pengarah Persatuan Teknologi Hijau ingin mengenalpasti hubungkait antara sumbangan dari jualan barangan terpakai (dalam Ringgit) dengan jumlah tahun penjualan. Data adalah seperti jadual di bawah:

Year (x) *Tahun (x)*	1	5	3	10	7	6
Contribution (y) *Sumbangan (y)*	500	100	300	50	75	80

Calculate the value of correlation and develop an equation of the regression line.

Hence determine the y value when x = 4 years.

Kirakan nilai korelasi dan bina persamaan bagi garis regresi tersebut. Seterusnya dapatkan nilai y apabila x = 4 tahun.

[15 marks]

[15 markah]

CLO2
C4

(b) The survey of number of absences and final grade for 7 student is shown in the table below. Determine value for correlation coefficient from the data below.

Kajian bilangan ketidakhadiran dan gred akhir bagi 7 orang pelajar adalah seperti jadual di bawah. Tentukan nilai pekali korelasi daripada data di bawah.

Number of absences, X ***Bilangan ketidakhadiran, X***	**Final grade (%),Y** ***Gred akhir (%),Y***
6	82
2	86
15	43
9	74
12	58
5	90
8	78

[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: $k_1 \cdot k_2 \cdot k_3 \cdots k_n$

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

Sample size for a proportion: $n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$

where $\hat{p} = \frac{X}{n}$ and $\hat{q} = 1 - \hat{p}$

Confidence interval for variance:

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$

Confidence interval for standard deviation:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\text{right}}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\text{left}}}}$$

Chapter 8 Hypothesis Testing

z test: $z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ for any value n. If $n < 30$, population must be normally distributed.

$z = \frac{\overline{X} - \mu}{s/\sqrt{n}}$ for σ unknown and $n \geq 30$

t test: $t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$ for $n < 30$ (d.f. $= n - 1$)

z test for proportions: $z = \frac{\hat{p} - p}{\sqrt{pq/n}}$

Chi-square test for a single variance: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

(d.f. $= n - 1$)

Chapter 9 Testing the Difference between Two Means, Two Variances, and Two Proportions

z test for comparing two means (independent samples):

$$z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Formula for the confidence interval for difference of two means (large samples):

$$(\overline{X}_1 - \overline{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Note: s_1^2 and s_2^2 can be used when $n_1 \geq 30$ and $n_2 \geq 30$.

F test for comparing two variances: $F = \frac{s_1^2}{s_2^2}$

where s_1^2 is the larger variance and
d.f.N. $= n_1 - 1$, d.f.D. $= n_2 - 1$

t test for comparing two means (independent samples, variances not equal):

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

(d.f. = the smaller of $n_1 - 1$ or $n_2 - 1$)

Formula for the confidence interval for difference of two means (small independent samples, variance unequal):

$$(\overline{X}_1 - \overline{X}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(d.f. = smaller of $n_1 - 1$ and $n_2 - 1$)

t test for comparing two means (independent samples, variances equal):

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

(d.f. $= n_1 + n_2 - 2$)

Formula for the confidence interval for difference of two means (small independent samples, variances equal):

$$(\overline{X}_1 - \overline{X}_2) - t_{\alpha/2}\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + t_{\alpha/2}\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

and d.f. $= n_1 + n_2 - 2$.

t test for comparing two means for dependent samples:

$$t = \frac{\overline{D} - \mu_D}{s_D/\sqrt{n}} \quad \text{where} \quad \overline{D} = \frac{\Sigma D}{n} \quad \text{and}$$

$$s_D = \sqrt{\frac{\Sigma D^2 - [(\Sigma D)^2/n]}{n-1}} \qquad (\text{d.f.} = n - 1)$$

Formula for confidence interval for the mean of the difference for dependent samples:

$$\overline{D} - t_{\alpha/2}\frac{S_D}{\sqrt{n}} < \mu_D < \overline{D} + t_{\alpha/2}\frac{S_D}{\sqrt{n}}$$

(d.f. $= n - 1$)

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

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

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:

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

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

Formula for the multiple correlation coefficient:

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

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 $\bar{X}_{GM} = \frac{\Sigma X}{N}$

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

d.f.D. $= N - k$ where $k =$ number of groups

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

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

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

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

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

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: $z = \frac{(X + 0.5) - (n/2)}{\sqrt{n}/2}$

where n = sample size (greater than or equal to 26)

X = smaller number of + or − signs

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

where

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

$\sigma_R = \sqrt{\frac{n_1n_2(n_1 + n_2 + 1)}{12}}$

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: $z = \frac{w_s - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$

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.