Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2014/2015

COURSE NAME	:	SURVEY AND SAMPLING METHODS
COURSE CODE	:	BWB 21103
PROGRAMME	:	2 BWQ
EXAMINATION DATE	:	DECEMBER 2014/JANUARY 2015
DURATION	:	3 HOURS
INSTRUCTION	:	ANSWER **ALL** QUESTIONS

THIS QUESTION PAPER CONSISTS OF **EIGHT (8)** PAGES

Q1 (a) Define the sampling frame and elementary unit.

(2 marks)

(b) Differentiate the concepts of observational study and experimental study.

(4 marks)

(c) Differentiate the concepts of validity and reliability.

(4 marks)

(d) As part of a marketing program, a city block containing 4 households was selected and a sample of 3 households was sampled as follows. Jeremiah, the research assistant, identified the households and numbered each household from 1 to 4. He then want to list all combinations of the 4 households 2 at a time. These are:

1 and 2 1 and 3 1 and 4 2 and 3 2 and 4 3 and 4

Unfortunately, Jeremiah got this job because he was related to the Department Chairman, and he was very careless. He forgot the combination 3 and 4. He chose a random number between 1 and 5 (it turned out to be 4) which corresponded to the combination 2 and 3. Thus households 2 and 3 were sampled. The variable of interest was out-of-pocket medical expenses incurred by the household (**Table Q1 (d)**).

Table Q1 (d) : Out-Of-Pocket Medical Expenses Incurred by the Household

Household	Expenses ($)
1	345.00
2	126.00
3	492.00
4	962.00

(i) Compute the value of population mean.

(1 mark)

(ii) Based on Jeremiah's sampling procedure, what is the expected value of mean and variance of the estimated mean out-of-pocket medical expenses?

(4 marks)

(iii) Calculate the value of bias and mean square error (MSE).

(4 marks)

(iv) Is the estimate derived from this sampling procedure unbias?

(1 mark)

Q2 (a) Define a simple random sampling with replacement and without replacement.

(2 marks)

(b) Give two methods on how to take a simple random sample.

(2 marks)

(c) List all possible simple random samples of size $n=3$ that can be selected from the population $\{0, 1, 2, 3, 4\}$.

(4 marks)

(d) A community within a city contains 3000 households and 10,000 persons. For purposes of planning a community satellite to the local health department, it is desired to estimate the total number of physician visits made during a calendar year by members of the community. For this information to be useful, it should be accurate to within 10% of the true value. A small pilot survey of 10 households, conducted for purposes of gathering preliminary information, yielded the accompanying data on physician visits made during the previous calendar year (**Table Q2 (d)**). Use 99.7% confidence level of estimated.

Table Q2 (d) : Total Number Of Physician Visits During Previous Year

Household	Total number of physician visits
1	12
2	27
3	16
4	17
5	1
6	21
7	34
8	12
9	24
10	30

(i) Using these data as preliminary information, identify the sample size needed to meet the specifications of the survey.

(10 marks)

(ii) If one person want to take 250 sample size, examine whether it is enough compare to the result of sample size form Question **Q2 (d)(i)**? Why?

(2 marks)

Q3 (a) Estimated variances are large when the sampling ratio coincides with a periodicity in the frame. (True/False)

(2 marks)

(b) State the advantage of using repeated systematic sampling.

(2 marks)

(c) Suppose that a study is planned of the level of the pesticide dieldrin, which is believed to be a carcinogen, in a 7.5 mile stretch of a particular river. To assure representativeness, a map of the river is divided into 36 zones, (see the **Figure Q3(c)** on page 8) and a 1 in 4 systematic sample of these zones is to be selected. Water samples will be drawn by taking a boat out to the geographic center of the designated zone, and drawing a grab sample of water from a depth of several centimeters below the surface level. The levels of dieldrin, in micrograms per liter, for each of these zones are shown in parentheses.

(i) Using systematic sampling of 1 in 4, show the list of sample of zones that should be selected.

(2 marks)

(ii) Calculate the 90% confidence interval for the average level of dieldrin in this stretch of the river.

(12 marks)

(iii) What advantages can you identify for this method of sampling the river over simple random sampling?

(2 marks)

Q4 (a) What is strata?

(2 marks)

(b) Why we used stratified sampling?

(2 marks)

(c) A sample survey of households in a community containing 1500 households is to be conducted for the purpose of determining the total number of persons over 18 years of age in the community who have one or more permanent teeth (other than third molars) missing. Since this variable is thought to be correlated with age and income, the strata shown in the accompanying **Table Q4(c)** are formed by using available population data. A stratified random sample of 100 families is to be taken.

Table Q4(c) : Population Data

Variable	Stratum			
	1	2	3	4
Age				
Mean	30	32	25	27
Standard deviation	15	15	10	10
Annual family income (x $1000)				
Mean	15	7	15	8
Standard deviation	5	3	3	2
No. of families	300	500	100	600

(i) Compute the number of families to be taken from each stratum if proportional allocation is used.

(4 marks)

(ii) Calculate the number of families to be taken from each stratum if optimal allocation is used based on annual family income.

(6 marks)

(iii) Calculate the number of families to be taken from each stratum if optimal allocation is used based on age.

(6 marks)

Q5 (a) Why we used cluster sampling?

(2 marks)

(b) Suppose that the elementary schools in a city are grouped into 30 school districts, with each school district containing four schools. Suppose that a simple one-stage cluster sample of three school districts is taken for purposes of estimating the number of school children in the city who are color-blind (as measured by a standard test), and that the accompanying data are obtained from this sample (**Table Q5(b)**).

Table Q5(b) : The Number Of School Children In The City Who Are Color-Blind

Sample School District	School	No. Of Children	No. Of Color-Blind Children
1	1	130	2
	2	150	3
	3	160	3
	4	120	5
2	1	110	2
	2	120	4
	3	100	0
	4	120	1
3	1	89	4
	2	130	2
	3	100	0
	4	150	2

(i) Calculate and obtain a 95% confidence interval for the total number of color-blind children.

(9 marks)

(ii) Calculate and obtain a 95% confidence interval for the proportion of all children who are color-blind.

(9 marks)

- END OF QUESTION –

FINAL EXAMINATION

SEMESTER/SESSION: SEM I/2014/2015 PROGRAMME : 2 BWQ
COURSE NAME : SURVEY AND SAMPLING METHODS COURSE CODE: BWB21103

Chapter 2: $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, $B(\hat{d}) = E(\hat{d}) - d$, $E(\bar{x}) = \sum_{i=1}^{C} \bar{x}_i \pi_i$

$Var(\bar{x}) = \sum_{i=1}^{C} [(\bar{x}_i - E(\bar{x}]^2 \pi_i$, $MSE(\hat{d}) = Var(\hat{d}) + B^2(\hat{d})$

Chapter 3: $z_{0.025} = 1.96$, $z_{0.05} = 1.6449$, $z_{0.0015} = 3$

$s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$, $n \geq \frac{z^2 N V_x^2}{z^2 V_x^2 + (N-1)\varepsilon^2}$, $\hat{\sigma}_x^2 = \left(\frac{N-1}{N}\right) s_x^2$, $V_x^2 = \frac{\hat{\sigma}_x^2}{\bar{x}^2}$

Chapter 4: $\hat{\sigma}_x^2 = \left(\frac{N-1}{N}\right) s_x^2$, $\widehat{Var}(\bar{x}) = \left(\frac{\hat{\sigma}_x^2}{n}\right)\left(\frac{N-n}{N-1}\right)$, $\bar{x} \pm z_{1-\left(\frac{\alpha}{2}\right)} \sqrt{\widehat{Var}(\bar{x})}$,

Chapter 5: $n_h = N_h \times \frac{n}{N}$, $n_h = \left(\frac{N_h \sigma_{hx}}{\sum_{h=1}^{L} N_h \sigma_{hx}}\right)(n)$

Chapter 6: $x'_{clu} = \left(\frac{M}{m}\right) x$, $\widehat{SE}(x'_{clu}) = \left(\frac{M}{\sqrt{m}}\right) \hat{\sigma}_{1x} \sqrt{\frac{M-m}{M-1}}$,

$\hat{\sigma}_{1x} = \sqrt{\left[\frac{\sum_{i=1}^{m} (x_i - \bar{x}_{clu})^2}{m-1}\right]} \sqrt{\frac{M-1}{M}}$,

$x'_{clu} - 1.96 \times \widehat{SE}(x'_{clu}) \leq X \leq x'_{clu} + 1.96 \times \widehat{SE}(x'_{clu})$,

$r_{clu} = \left(\frac{x}{y}\right)$,

$\widehat{SE}(r_{clu}) =$

$$r_{clu} \sqrt{\frac{[\widehat{SE}(\bar{x}_{clu}^2)]^2}{\bar{x}_{clu}^2} + \frac{[\widehat{SE}(\bar{y}_{clu}^2)]^2}{\bar{y}_{clu}^2} - \frac{2}{m} \times \left(\frac{M-m}{M}\right) \times \left(\frac{1}{\bar{x}_{clu}\bar{y}_{clu}}\right) \times \left[\frac{\sum_{i-1}^{m} (x_i - \bar{x}_{clu})(y_i - \bar{y}_{clu})}{m-1}\right]}$$

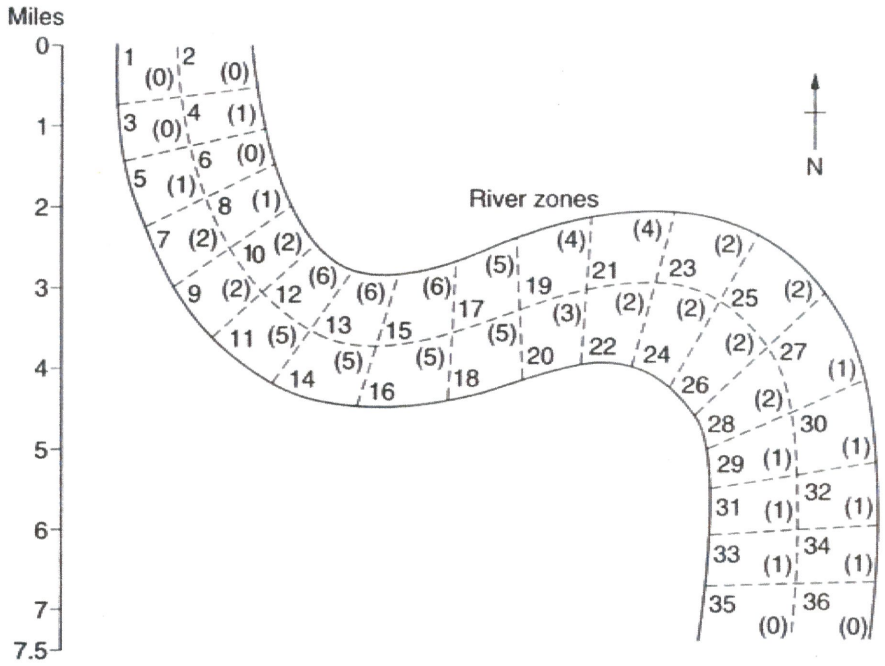
Miles
0
1
2
3
4
5
6
7
7.5
River zones
N
1 (0) 2 (0)
3 (0) 4 (1)
5 (1) 6 (0)
7 (2) 8 (1)
9 (2) 10 (2)
11 (5) 12 (6)
13 (5) 14
15 (5) 16
17 (6) 18
19 (5) 20
21 (4) 22
23 (4) 24
25 (2) 26
27 (2) 28
29 (2) 30
31 (1) 32
33 (1) 34
35 (0) 36
(1)
(1)
(1)
(0)

Figure Q3(c)