UTHM
Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
SEMESTER I
SESSION 2016/2017

TERBUKA

COURSE NAME	:	SURVEY AND SAMPLING METHODS
COURSE CODE	:	BWB 21103
PROGRAMME CODE	:	BWQ
EXAMINATION DATE	:	DECEMBER 2016 / JANUARY 2017
DURATION	:	3 HOURS
INSTRUCTION	:	ANSWER ALL QUESTIONS

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

Q1   (a)   Define the concept of probability sampling and give **THREE (3)** examples of probability sampling techniques.

(4 marks)

(b)   As part of an AIDS education program, 120 intravenous drug users seronegative for HIV (Human Immunodeficiency Virus) in the first screening were given instructions on sterilizing their needles with bleach and practicing "safe sex". One year after the program's inception, a sample of 30 of these subjects was taken by numbering the participants from 1 to 120 and taking all subjects whose numbers are divisible by 4 (e.g., 4, 8, 12, etc.).

(i)   What are the sampling frame in this sample design?

(1 mark)

(ii)   What are the sampling units in this sample design?

(1 mark)

(iii)   Does each person in the population have an equal chance of being selected in the sample?

(1 mark)

(iv)   What kind of sampling technique used in this study?

(1 mark)

(c) For the purpose of studying sampling distribution and sampling technique, we assume that all population values are known and shown in **Table Q1**.

Table Q1 : Number of Element in Population

Population	Number of Element in Population
1	100
2	105
3	157
4	131
5	190

(i) Calculate the value of mean population.

(2 marks)

(ii) For the first sampling plan, we choose the population of 2, 3, and 5 to be our sample. Calculate the value of mean sample.

(2 marks)

(iii) For the second sampling plan, select all possible sample if sample of two populations are selected. List down all possible samples with their mean values. Then, calculate the expected value of mean.

(3 marks)

(iv) From your analysis, which sampling plan has greater validity? Your analysis can be done by comparing the result obtained in part (i), (ii) and (iii). Verify your result by calculating the bias value.

(5 marks)

CONFIDENTIAL

Q2 (a) Define a simple random sampling.

(2 marks)

(b) A community in the San Francisco Bay area consists of 30 families with their weekly net income (RM) data is shown in **Table Q2**.

Table Q2: Weekly Net Income of 30 Families

Family Number	Family Size	Weekly Net Income (RM)	Family Number	Family Size	Weekly Net Income (RM)
1	2	372	16	5	450
2	3	372	17	3	414
3	3	522	18	4	498
4	5	390	19	2	510
5	4	348	20	4	438
6	7	552	21	2	396
7	2	528	22	5	348
8	4	574	23	3	462
9	2	498	24	4	414
10	5	372	25	7	390
11	3	378	26	3	462
12	6	372	27	3	414
13	4	360	28	6	570
14	4	450	29	2	462
15	2	540	30	2	414

(i) A researcher, named Steve, wishes to do a survey to estimate the total weekly net income by taking 10 families as a sample using simple random sampling technique. He choose family number 1, 4, 6, 10, 12, 14, 18, 21, 25 and 30 randomly to be a sample. Compute the value of estimated total weekly net income and estimated standard error from the sample.

(5 marks)

TERBUKA

CONFIDENTIAL

(ii) The second researcher, Alice wish to do the same survey as Steve but she takes more size sample which is 20 families. The sample contains family number 1, 2, 3, 5, 8, 10, 11, 13, 14, 16, 17, 19, 20, 21, 23, 24, 26, 27, 28, and 29. Compute the value of estimated total weekly net income and estimated standard error from the sample.

(5 marks)

(iii) By comparing result obtained in part (i) and (ii), which sample do you think is more reliable? Give your reasons.

(3 marks)

(iv) Based on data samples taken by Steve and Alice, which samples have larger proportion of family size 3 or more?

(5 marks)

Q3 (a) **Table Q3(a)** shows the summary of result from four types of sampling.

Table Q3(a): Summary of Sampling

Sample Design	Estimated Varience of mean	Intraclass Correlation Coefficient
Systematic sampling design 1	64.6	0.125
Systematic sampling design 2	137.8	0.865
Systematic sampling design 3	14.1	- 0.381
Simple random sampling	40.3	-

(i) By analyzing the result, categorize the sampling design into unorder list, monotonically ordered list, and periodicity in list.

(3 marks)

(ii) Which sampling technique has high reliability of estimated? Give your reason.

(2 marks)

TERBUKA

(iii)   Which sampling design is categorized into homogeneity and heterogeneity?

(2 marks)

(b)   The number of beds in each twelve hospitals is shown in the **Table Q3(b)**. Use the information provided to answer the following questions.

Table Q3(b): Number of Beds in Twelve Hospitals

Hospital	Number of Beds
1	824
2	312
3	329
4	648
5	358
6	252
7	256
8	263
9	138
10	150
11	189
12	201

(i)   Using systematic sampling of 1 in 4 with the initial random number of 2, show the list of sample of hospitals that should be selected and calculate the average number of beds.

(3 marks)

(ii)   Calculate the 95% confidence interval for the average number of beds. Interpret your result.

(10 marks)

Q4   (a)   Describe **FOUR (4)** advantages of using stratified sampling compared to simple random sampling.

(4 marks)

(b)   A marathon was conducted in a large city on September 3, 1989. Based on the entry application, **Table Q4** was obtained. It is desired to take a sample of approximately 500 persons from this list for a purpose of estimating the average number of miles run per week in preparation for this marathon.

Table Q4: Population Data

Variable	Stratum		
	1	**2**	**3**
Age Group			
Mean number of marathons completed	1.9	2.3	3.1
Standard deviation	0.6	0.8	0.7
Gender			
Mean number of marathons completed	1.8	2.1	2.8
Standard deviation	0.5	0.7	0.9
No. of persons	2300	1478	978

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

(6 marks)

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

(5 marks)

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

(5 marks)

TERBUKA

CONFIDENTIAL

Q5 (a) Describe **TWO (2)** advantages of using cluster sampling.

(2 marks)

(b) A survey which uses a simple one-stage cluster sampling design was conducted in a large city in China. The clusters in this instance were entities known as "neighbourhood group", which are essentially groups of contiguous households. The city comprises nine districts which, in this design, comprise the strata. Within each district, a simple random sample of two neighbourhood groups was selected, and all individuals within each neighbourhood group were interviewed concerning their overall health status. **Table Q5(b)** represents a sample data from the survey on the number of individuals over 30 years of age who were edentulous. Strata 1, 2, 4 and 6 each contain 200 neighbourhood groups, strata 3 contains 175 neighbourhood groups and strata 5 contains 150 neighbourhood groups.

Table Q5(b): The Number of Edentulous Persons

District	Neighbourhood Group	No. of Persons > 30 years old	No. of Edentulous Persons
1	1	28	7
	2	35	9
2	1	29	12
	2	43	26
3	1	61	19
	2	48	12
4	1	15	10
	2	39	28
5	1	21	9
	2	46	15
6	1	12	0
	2	25	4

(i) Construct 95% confidence interval for the total number of edentulous persons over 30 years of age. Interpret your result.

(10 marks)

(ii) Construct 95% confidence interval for the mean number of edentulous persons over 30 years of age. Interpret your result.

(8 marks)

- END OF QUESTIONS -

TERBUKA

CONFIDENTIAL