
UNIVERSITI SAINS MALAYSIA

Second Semester Examination
Academic Session 2007/2008

April 2008

MSG 368 – Sample Survey and Sampling Technique
[Tinjauan Sampel dan Teknik Pensampelan]

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of FOURTEEN pages of printed material before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi EMPAT BELAS muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]

Instructions: Answer all four [4] questions.

[Arahan: Jawab semua empat [4] soalan.]

1. (a) The Waterworks Department in Selangor wanted to estimate the average daily water consumption per household (in gallons) over the next five years. To do this, the department monitored 100 water meters randomly selected in Shah Alam. The sample mean $\bar{y} = 12.5$ and sample variance $s^2 = 1252$ were found. If we assume that there will be $N = 10,000$ households in Shah Alam in five years' time,

- (i) estimate the population mean μ
- (ii) find a 95% confidence interval for μ
- (iii) estimate the true total daily water consumption
- (iv) how large a sample of houses is required to estimate μ to within ± 2 gallons?

(b) Define the meaning of the following terms:

- (i) Observation unit
- (ii) Elementary unit
- (iii) Sampling frame
- (iv) Sampling unit

(c) (i) For the following survey, describe the target population, sampling frame, sampling unit and observation unit

A survey is conducted to find the average weight of cows in a region. A list of all farms is available for the region and 50 farms are selected at random. Then the weight of each cow at the 50 selected farms is recorded.

(ii) Decide if the following sampling method produce a simple random sample of students from a class of 30 students. If not, explain why.

If the class consists of 15 boys and 15 girls, assign each of the boys to a number from 1 to 15, and each of the girls to a number from 16 to 30. Then use a random number table to select three numbers from 1 to 15 and three numbers from 16 to 30. Select the students assigned those numbers as your sample.

[100 marks]

1. (a) *Jabatan Air Negeri Selangor hendak menganggar purata kegunaan air harian (dalam gelen) setiap rumah dalam tempoh lima tahun akan datang. Untuk mencapai tujuan ini Jabatan Air telah menyalurkan 100 buah meter air yang dipilih secara rawak di Shah Alam. Min sampel $\bar{y} = 12.5$ dan varians sampel $s^2 = 1252$ telah didapati. Sekiranya dianggapkan bahawa terdapat $N = 10,000$ buah rumah di Shah Alam dalam masa lima tahun lagi,*

- (i) *anggarkan min populasi μ*
- (ii) *kira selang keyakinan 95% bagi μ*
- (iii) *anggarkan jumlah air yang sebenar yang digunakan setiap hari*
- (iv) *berapa besar saiz sampel perumahan yang diperlukan untuk menganggarkan μ antara ± 2 gelen dengan 95% keyakinan?*

(b) *Terangkan maksud sebutan-sebutan yang berikut:*

- (i) *Unit peninjauan*
- (ii) *Unit awalan*
- (iii) *Kerangka pensampelan*
- (iv) *Unit pensampelan*

(c) (i) *Untuk tinjauan yang berikut, huraikan populasi sasaran, kerangka pensampelan, unit pensampelan dan unit peninjauan.*

Satu tinjauan dikendalikan untuk mencari purata berat lembu di suatu kawasan. Satu senarai kesemua ladang-ladang boleh diperolehi untuk kawasan tersebut dan 50 ladang telah dipilih secara rawak. Seterusnya berat bagi setiap lembu dari 50 ladang yang terpilih direkodkan.

(ii) *Tentukan jika kaedah pensampelan yang berikut menghasilkan sampel rawak ringkas bagi pelajar-pelajar yang terdiri daripada 30 orang. Jika tidak, terangkan kenapa.*

Jika sebuah kelas terdiri daripada 15 pelajar lelaki dan 15 pelajar perempuan. Setiap pelajar lelaki ditetapkan dengan nombor dari 1 hingga 15 dan setiap pelajar perempuan ditetapkan dengan nombor 16 hingga 30. Kemudian gunakan jadual nombor rawak untuk memilih 3 nombor dari 1 hingga 15 dan 3 nombor dari 16 hingga 30. Pilih pelajar-pelajar yang diperuntukkan nombor tersebut sebagai sampel anda.

[100 markah]

2. (a) The USM Students Association will be conducting a survey to obtain students' opinions on the increment of the study fees. This increment will be implemented by the university in the near future. You are assigned to conduct this survey. A simple random sample will be taken from the population.

Determine the sample size formula to estimate the population proportion p with a bound on the error of estimation B .

- (b) A sampler proposes to take a stratified random sample. He expected that his field costs will be of the form $\sum c_n n_n$. His advance estimates of relevant quantities for the two strata are as follows:

Stratum	W_h	s_h	c_h
1	0.4	10	RM4
2	0.6	20	RM9

Assume that $N = 200$

- (i) Find the values of n_1/n and n_2/n that minimize the total field cost for a given value of $V(\bar{y}_{st})$.
- (ii) Find the sample size required, under this optimum allocation, to make $V(\bar{y}_{st}) = 1$.
- (iii) How much will the total field cost be?
- (c) The personnel manager of a corporation wants to estimate, for one year, the total number of days used for sick leave among all 50 plants under his supervision. The 50 plants are divided into 22 "small" plants and 28 "large" plants. From past experience, the manager figures that the small plants may use from 0 to 90 days of sick leave, whereas the large plants may use from 15 to 200 days of sick leave. If it is desired to estimate the total to within 100 days:
- (i) find the appropriate allocation sample size.
- (ii) find the appropriate allocation of the sample to the two strata.

[100 marks]

3. (a) An experienced farmer makes an eye estimate of the weight of peaches x_i on each tree in an orchard of $N = 200$ trees. He finds a total weight of $\tau_x = 11600$ lb. The peaches are picked and weighed on a simple random sample of 10 trees, with the following results:

2. (a) Persatuan Mahasiswa USM mahu menjalankan suatu tinjauan untuk mendapat pandangan para pelajar tentang kenaikan yuran pengajian yang akan dilaksanakan oleh universiti dalam masa terdekat. Anda telah ditugaskan untuk menjalankan tinjauan ini. Satu sampel rawak ringkas akan diambil dari populasi tersebut. Tentukan formula menentukan saiz sampel untuk menganggar kadaran populasi p dengan batas ralat penganggaran B .
- (b) Suatu contoh pensampelan yang dipertimbangkan ialah dengan melaksanakan sampel rawak berstrata. Beliau menjangkakan bahawa perbelanjaan kajiannya akan berbentuk $\sum c_n n_n$. Terlebih dahulu anggaran beliau terhadap kuantiti yang berkaitan untuk dua strata adalah seperti yang berikut:

Stratum	W_h	s_h	c_h
1	0.4	10	RM4
2	0.6	20	RM9

Andai $N = 200$

- (i) Cari nilai bagi n_1/n dan n_2/n yang meminimumkan jumlah perbelanjaan kajian dengan diberikan nilai $V(\bar{y}_{st})$.
- (ii) Cari sampel saiz yang diperlukan dengan menggunakan peruntukan optimum agar $V(\bar{y}_{st}) = 1$.
- (iii) Berapakah jumlah perbelanjaan kajian?
- (c) Seorang pengurus personel dari sebuah koperasi ingin menganggar untuk selama setahun, jumlah bilangan hari yang digunakan bagi cuti sakit di antara kesemua 50 kilang yang di bawah pengawasannya. Lima puluh buah kilang tersebut dibahagikan kepada 22 buah kilang "kecil" dan 28 buah kilang "besar". Dari pengalaman yang lalu, pengurus dapat membayangkan bahawa kilang yang kecil akan menggunakan dari 0 hingga 90 hari cuti sakit, sementara kilang yang besar akan menggunakan dari 15 hingga 200 hari cuti sakit. Jika diinginkan untuk menganggar jumlah supaya berada dalam lingkungan 100 hari:
- (i) cari peruntukan yang bersesuaian untuk saiz sampel.
- (ii) cari peruntukan yang bersesuaian untuk sampel bagi kedua-dua strata.

[100 markah]

3. (a) Seorang petani yang berpengalaman membuat anggaran bagi berat buah peach x_i bagi setiap pohon di sebuah ladang dengan $N = 200$ pohon. Dia mendapati bahawa jumlah berat $\tau_x = 11600$ lb. Buah-buahan peach dipetik dan dicatat beratnya daripada 10 pohon yang dipilih secara rawak ringkas. Berikut adalah keputusannya:

	Tree Number										Total
	1	2	3	4	5	6	7	8	9	10	
Actual weight (y_i)	61	42	50	58	67	45	39	57	71	53	543
Estimate weight (x_i)	59	47	52	60	67	48	44	58	76	58	569

- (i) As an estimate of the total actual weight τ_y , we take
- $$\hat{\tau}_{yD} = N[\mu_x + (\bar{y} - \bar{x})]$$
- Compute the estimate and find the bound on the error of estimation.
- (ii) Compute the relative efficiency of the ratio estimation to the difference estimation. Comment on your results.
- (b) State under what conditions should you employ a ratio estimator of a population total rather than an estimator of the form $N\bar{y}$.
- (c) Suppose that there are only 3 firms in Balik Pulau, as shown in the table below:

Balik Pulau Population		
Firms	Sales, x	Employees, y
A	13	9
B	12	8
C	9	6

- (i) Calculate μ_x and μ_y .
- (ii) Show the possible outcomes of a simple random sample without replacement of two firms, the probabilities of these outcomes and the associated values of the ratio estimator of μ_y .
- [100 marks]
4. (a) Suppose that the elementary school 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 the purpose of estimating the number of school children in the city who are color-blind (as measured by a standard test), and the accompanying data are obtained from this sample.

Bilangan Pohon

	1	2	3	4	5	6	7	8	9	10	Jumlah
Berat Sebenar (y_i)	61	42	50	58	67	45	39	57	71	53	543
Anggaran Berat (x_i)	59	47	52	60	67	48	44	58	76	58	569

- (i) Sebagai anggaran bagi jumlah τ_y , kita gunakan

$$\hat{\tau}_{yD} = N[\mu_x + (\bar{y} - \bar{x})]$$

Kira anggaran tersebut dan cari batas ralat penganggaran.

- (ii) Kirakan kecekapan relatif penganggaran nisbah kepada penganggaran beza. Komen keputusan anda.
- (b) Nyatakan dalam keadaan manakah anda menggunakan penganggar nisbah bagi jumlah populasi daripada penganggar yang dalam bentuk Ny .
- (c) Katakan hanya terdapat 3 firma di Balik Pulau, seperti yang dicatatkan di jadual berikut:

Populasi Balik Pulau		
Firma	Jualan, x	Pekerja, y
A	13	9
B	12	8
C	9	6

- (i) Hitung μ_x dan μ_y .
- (ii) Tunjukkan kesudahan yang mungkin dari sampel rawak ringkas tanpa pengembalian bagi 2 firma, kebarangkalian daripada kesudahan tersebut dan nilai-nilai bagi penganggar nisbah, μ_y yang bersepadan

[100 markah]

4. (a) Katakan sekolah tadika di sebuah bandar dikumpulkan kepada 30 sekolah daerah, dengan setiap sekolah daerah mengandungi 4 buah sekolah. Andaikan sampel berkelompok peringkat pertama bagi 3 buah sekolah daerah yang dipilih untuk tujuan menganggar bilangan kanak-kanak sekolah di bandar yang rabun warna (yang disukat melalui satu ujian piawai), dan data yang berikut diperoleh daripada sampel tersebut:

Sample School District	School	Number of children	Number 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) Estimate and obtain a 95% confidence interval for the total number of color-blind children.
- (ii) Estimate the proportion of all children who are color-blind and place a bound on the error of estimation. Explain your result.
- (b) Foresters are interested in determining the mean timber volume per acre for 500 one-acre plots ($N = 500$). A 1-in-30 systematic sample is conducted. Use the data presented in the table below

Plot sampled	Volume (in board feet)
5	7030
35	6720
65	6850
95	7210
125	7150
155	7370
185	7000
215	6930
245	6570
275	6910
305	7380
335	7540
365	6720
395	6900
425	7200
455	7100
485	6860

<i>Sampel Sekolah Daerah</i>	<i>Sekolah</i>	<i>Bilangan kanak-kanak</i>	<i>Bilangan kanak-kanak yang rabun warna</i>
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) *Anggar dan dapatkan selang keyakinan 95% bagi jumlah bilangan kanak-kanak yang rabun warna.*
- (ii) *Anggarkan kadaran semua kanak-kanak yang rabun warna dan cari batas ralat penganggarannya.*
- (b) *Pegawai-pegawai yang menjaga hutan amat berminat untuk menentukan purata balak seekar bagi 500 plot satu-ekar. Satu sampel bersistematik 1-dalam-30 dilaksanakan. Dengan menggunakan data dalam jadual di bawah :*

<i>Plot yang disampelkan</i>	<i>Isipadu (dalam keluasan meter)</i>
5	7030
35	6720
65	6850
95	7210
125	7150
155	7370
185	7000
215	6930
245	6570
275	6910
305	7380
335	7540
365	6720
395	6900
425	7200
455	7100
485	6860

- (i) Estimate the average timber volume per plot and place a bound on the error of estimation.
 - (ii) What sample size is needed to estimate τ with a bound on the error of estimation approximately equal to 20000 volume per plot?
- (c) Explain the following terms and give one example:
- (i) Two-stage cluster sampling
 - (ii) Primary unit sampling

[100 marks]

- (i) *Anggarkan purata isipadu balak per plot dan cari batas ralat penganggarannya.*
- (ii) *Apakah saiz sampel yang diperlukan untuk menganggar τ , dengan batas ralat penganggaran menghampiri 20000 isipadu per plot?*
- (c) *Jelaskan maksud sebutan-sebutan yang berikut dan berikan satu contoh:*
- (i) *Pensampelan berkelompok peringkat kedua*
- (ii) *Pensampelan unit utama*

[100 markah]

Appendix

Sample	Sample Variance
$\frac{\sum_{i=1}^n y_i}{n}$	$\frac{s^2}{n} \left(\frac{N-n}{N} \right), s^2 = \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n-1}$
$N\bar{y}$	$N^2 \frac{s^2}{n} \left(\frac{N-n}{N} \right)$
$\frac{a}{n}$	$\frac{\hat{p}(1-\hat{p})}{n-1} \left(\frac{N-n}{N} \right)$
$\frac{\sum_{i=1}^L N_i \bar{y}_i}{N}$	$\sum_{i=1}^L \frac{N_i^2}{N^2} \left(\frac{N_i - n_i}{N_i} \right) \frac{s_i^2}{n_i}$
$\frac{\sum_{i=1}^L N_i \hat{p}_i}{N}$	$\sum_{i=1}^L \frac{N_i^2}{N^2} \left(\frac{N_i - n_i}{N_i} \right) \frac{\hat{p}_i (1 - \hat{p}_i)}{n_i - 1}$
$\frac{\bar{y}}{\bar{x}}$	$\left(\frac{N-n}{nN} \right) \left(\frac{1}{\mu_x^2} \right) \left(\frac{\sum_{i=1}^n (y_i - rx_i)^2}{n-1} \right)$
$\bar{y} + b(\mu_x - \bar{x})$, $b = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$	$\left(\frac{N-n}{Nn} \right) \left(\frac{1}{n-2} \right) \left(\sum_{i=1}^n (y_i - \bar{y})^2 - b^2 \sum_{i=1}^n (x_i - \bar{x})^2 \right)$
$\mu_x + \bar{d}$	$\left(\frac{N-n}{Nn} \right) \left(\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1} \right)$
$\left(\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i} \right)$	$\left(\frac{N-n}{Nn\bar{M}^2} \right) \left(\frac{\sum_{i=1}^n (y_i - \bar{y}m_i)^2}{n-1} \right)$

Sample	Sample Variance
$\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i}$	$\left(\frac{N-n}{Nn-\bar{M}^2} \right) \frac{\sum_{i=1}^n (a_i - \hat{p}m_i)^2}{n-1}$
$M\bar{y}$	$M^2 \left(\frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (y_i - \bar{y}m_i)^2}{n-1}$
$\hat{\mu} = \frac{1}{n\bar{M}} \sum_{i=1}^n M_i \bar{y}_i$	$\left(\frac{N-n}{N} \right) \left(\frac{1}{n\bar{M}^2} \right) S_b^2 + \frac{1}{nN\bar{M}^2} \sum_{i=1}^n M_i^2 \left(\frac{M_i - m_i}{M_i} \right) \left(\frac{S_i^2}{m_i} \right)$ <p>where</p> $S_b^2 = \frac{\sum_{i=1}^n (M_i \bar{y}_i - \bar{M} \hat{\mu})^2}{n-1}$ $S_i^2 = \frac{\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2}{m_i - 1}, i = 1, 2, \dots, n$
$\hat{\mu}_r = \frac{\sum_{i=1}^n M_i \bar{y}_i}{\sum_{i=1}^n M_i}$	$\left(\frac{N-n}{N} \right) \left(\frac{1}{n\bar{M}^2} \right) S_r^2 + \frac{1}{nN\bar{M}^2} \sum_{i=1}^n M_i^2 \left(\frac{M_i - m_i}{M_i} \right) \left(\frac{S_i^2}{m_i} \right)$ <p>where $S_r^2 = \frac{\sum_{i=1}^n (M_i \bar{y}_i - \hat{\mu}_r M_i)^2}{n-1}$</p>
$\hat{p} = \frac{\sum_{i=1}^n M_i \hat{p}_i}{\sum_{i=1}^n M_i}$	$\left(\frac{N-n}{N} \right) \left(\frac{1}{n\bar{M}^2} \right) S_r^2 + \frac{1}{nN\bar{M}^2} \sum_{i=1}^n M_i^2 \left(\frac{M_i - m_i}{M_i} \right) \left(\frac{\hat{p}_i \hat{q}_i}{m_i - 1} \right)$ <p>where $S_r^2 = \frac{\sum_{i=1}^n (M_i \hat{p}_i - \hat{p} M_i)^2}{n-1}$</p>

Sample size	
$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} ; D = \frac{B^2}{4} ; D = \frac{B^2}{4N^2}$	
$n = \frac{\sum_{i=1}^L \frac{N_i^2 \sigma_i^2}{w_i}}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2} ; w_i = \frac{n_i}{n}$	
$n = \frac{\left(\sum_{k=1}^L N_k \sigma_k / \sqrt{C_k} \right) \left(\sum_{i=1}^L N_i \sigma_i \sqrt{C_i} \right)}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2}$ $n_i = \frac{n N_i \sigma_i / \sqrt{C_i}}{\sum_{i=1}^L N_i \sigma_i / \sqrt{C_i}}$	
$n = \frac{\left(\sum_{i=1}^L N_i \sigma_i \right)^2}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2} ; n_i = n \left(\frac{N_i \sigma_i}{\sum_{i=1}^L N_i \sigma_i} \right)$	Neyman Allocation
$n = \frac{\sum_{i=1}^L N_i \sigma_i^2}{ND + \frac{1}{N} \sum_{i=1}^L N_i \sigma_i^2} ; n_i = n \left(\frac{N_i}{\sum_{i=1}^L N_i} \right)$	Proportioanal Allocation
$n = \frac{\sum_{i=1}^L N_i^2 p_i q_i / w_i}{N^2 D + \sum_{i=1}^L N_i p_i q_i} ; n_i = n \left(\frac{N_i \sqrt{p_i q_i / c_i}}{\sum_{i=1}^L N_i \sqrt{p_i q_i / c_i}} \right)$	
$n = \frac{N\sigma^2}{ND + \sigma^2} ; D = \frac{B^2 \mu x^2}{4} ; D = \frac{B^2}{4} ; D = \frac{B^2}{4N^2}$	
$n = \frac{N\sigma_r^2}{ND + \sigma_r^2} ; D = \frac{B^2 (\bar{M})^2}{4} ; D = \frac{B^2}{4N^2}$	