UNIVERSITI SAINS MALAYSIA

First Semester Examination
2018/2019 Academic Session

December 2018/January 2019

MSG368 - Sample Survey And Sampling Technique
(Tinjauan Sample Dan Teknik Persampelan)

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of SIXTEEN (16) pages of printed material before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi ENAM BELAS (16) muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]

Instructions: Answer **EIGHT (8)** questions.

*[**Arahan**: Jawab **LAPAN (8)** soalan.]*

In the event of any discrepancies, the English version shall be used.

[Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi Bahasa Inggeris hendaklah diguna pakai.]

...2/-

Question 1

(a) List 3 types of non response which may arise in an interview survey conducted at the homes of selected interviewees.

(b) Suggest 3 methods how non response might be reduced in a mail survey by appropriate follow-up procedures.

(c) Explain why non response is a problem in social survey.

[8 marks]

Soalan 1

(a) Senaraikan 3 jenis tanpa respon yang mungkin wujud dalam tinjauan dilakukan di rumah mereka yang terpilih untuk ditemuduga.

(b) Cadangkan 3 kaedah bagaimana tanpa respon boleh dikurangkan dalam suatu tinjauan mel dengan mengikut prosedur susulan yang bersesuaian.

(c) Jelaskan mengapa tanpa respon adalah suatu masalah dalam tinjauan sosial.

[8 markah]

Question 2

A research center wishes to estimate the unemployment rate of a certain district with 25,100 people in the labour force. Since the unemployment rate varies across different age groups, the population were stratified into three age groups and the results are given below:

Age Group	Proportion in labour force	Sample size	Number of unemployed persons
Below 26	15%	24	6
26 – 40	50%	60	3
Above 40	35%	40	5

(a) Calculate the number of people in the labour force for each age group.

(b) Estimate the overall unemployment rate in the district and determine an estimate of the standard error for this estimator.

(c) Suppose that this study serves as a pilot study and the costs of sampling people in different age groups are the same. Estimate the total sample size and the strata sample sizes such that the error of the estimation is 0.02 at the 95% level of confidence.

[15 marks]

Soalan 2

Sebuah pusat penyelidikan ingin menganggarkan kadar pengangguran di sebuah daerah tertentu dengan 25,100 orang dalam tenaga buruh. Oleh kerana kadar pengangguran yang berbeza-beza mengikut kumpulan umur yang berbeza, populasi telah distratifikasikan kepada tiga kumpulan umur dan keputusan adalah seperti di bawah:

Kumpulan Umur	***Kadaran dalam tenaga buruh***	***Saiz sampel***	***Bilangan pengangguran***
Below 26	*15%*	*24*	*6*
26 – 40	*50%*	*60*	*3*
Above 40	*35%*	*40*	*5*

(a) Kirakan jumlah orang dalam tenaga buruh untuk setiap kumpulan umur.

(b) Anggarkan kadar pengangguran keseluruhan di daerah dan tentukan anggaran ralat piawai untuk penganggar ini.

(c) Andaikan bahawa kajian ini berfungsi sebagai kajian perintis dan kos untuk pensampelan dalam kumpulan umur yang berbeza adalah sama. Anggarkan jumlah saiz sampel dan saiz sample strata supaya ralat penganggaran adalah 0.02 pada tahap keyakinan 95%.

[15 markah]

Question 3

A household survey is designed to estimate the proportion P of families who own their flat. It is known that the value of P should lie between 20% to 30%.

(a) With simple random sampling, how large is the value of the sample size which is necessary to estimate P with a margin of error not exceeding 3% ?

(b) Without any knowledge of the possible values of P, can you provide a conservative estimate of n necessary to estimate P with a bound on the error of estimation not exceeding 3%?

(c) If instead of the margin of error that is given but it is the margin of standard error that is given to be 1.5%, what is the new conservative estimate for the sample size?

Assume the population size to be very large.

[10 marks]

Soalan 3

Suatu tinjauan isi rumah dirancang untuk menganggarkan perkadaran P keluarga yang memiliki rumah pangsa. Adalah diketahui bahawa nilai P harus terletak di antara 20% hingga 30%.

(a) Dengan pensampelan rawak mudah, berapa besar nilai saiz sampel yang diperlukan untuk menganggar P dengan batas ralat penganggaran tidak melebihi 3%?

(b) Tanpa sebarang pengetahuan tentang nilai-nilai yang mungkin, adakah anda dapat memberikan anggaran konservatif n yang diperlukan untuk menganggar P dengan batas ralat penganggaran tidak melebihi 3%?

(c) Jika bukannya batasan ralat yang diberi tetapi ia adalah batasan ralat piawai yang diberi iaitu 1.5%, apakah anggaran konservatif yang baharu untuk saiz sampel?

Anggap saiz populasi adalah sangat besar.

[10 markah]

Question 4

A company provides its salesmen with cars for work related travelling. The company operates in 12 branches with 810 cars in all and a simple random sampling of 4 branches is selected. For each selected branch, the number of cars and the total distance (in thousands of miles) travelled last year by cars are reported in the following table:

Branch	Number of cars	Total mileage
2	60	1459.2
5	110	3036.0
8	20	568.2
9	50	1277.5
Total	240	6340.9

(a) State the type of sampling design used.

(b) Estimate the average number of miles travelled by a company car in the past year and place a bound on the error of estimation.

(c) How large a sample should be taken in a future survey in order to estimate the total number of miles with a bound of 85 miles on the error of estimation?

[12 marks]

Soalan 4

Sebuah syarikat menyediakan kepada jurujualnya kereta agar dapat digunakan dalam perjalanan urusan pekerjaan. Syarikat ini beroperasi di 12 cawangan dengan 810 buah kereta kesemuanya dan suatu pensampelan rawak mudah telah memilih 4 cawangan.
Bagi setiap cawangan yang terpilih, bilangan kereta dan jumlah jarak (dalam ribuan batu) yang dilalui tahun lepas oleh kereta telah dilaporkan dalam jadual berikut:

Cawangan	*Bilangan kereta*	*Jumlah perbatuan*
2	*60*	*1459.2*
5	*110*	*3036.0*
8	*20*	*568.2*
9	*50*	*1277.5*
Total	*240*	*6340.9*

(a) Nyatakan jenis reka bentuk pensampelan yang digunakan.

(b) Anggarkan purata bilangan batu yang dilalui oleh sebuah kereta syarikat pada tahun lepas dan tentukan batas ralat penganggarannya.

(c) Berapakah besar sampel yang perlu diambil dalam tinjauan pada masa hadapan untuk menganggarkan jumlah batu dengan batas sebanyak 85 batu bagi ralat penganggaran?

[12 markah]

Question 5

A fashion shop wishes to estimate the percentage increase in monthly sales after the launch of a promotion campaign. A simple random sample of 10 branches is drawn from 256 branches and the monthly sales in thousand ringgit and related information are given below:

Branch	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Monthly sales after, y	65	109	60	124	128	104	65	61	49	56
Monthly sales before, x	52	100	60	128	104	98	48	64	96	48

$$\sum x_i = 798, \sum y_i = 821, \sum x_i^2 = 71,028, \sum y_i^2 = 75,765, \sum x_i y_i = 71,672$$

The shops records show that the total monthly sales for all branches before the launch of the campaign was 20,500 in thousand ringgit.

(a) Estimate the percentage increase in monthly sales after the launch of the promotion campaign and its standard error.

(b) Estimate the average monthly sales after the launch of the promotion campaign and calculate its standard error by using:

(i) the simple random sample estimator.

(ii) the ratio estimator.

(iii) the difference estimator

(c) Given that the correlation coefficient $\rho_{y,x} = 0.7854$, compare the above three estimators in part (b) and explain your results.

[20 marks]

Soalan 5

Sebuah kedai fesyen ingin menganggarkan peningkatan peratusan jualan bulanan selepas pelancaran kempen promosi. Satu sampel rawak 10 cawangan diambil dari 256 cawangan dan penjualan bulanan dalam ribuan ringgit dan maklumat yang berkaitan diberikan di bawah:

Cawangan	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Selepas jualan bulanan, y	*65*	*109*	*60*	*124*	*128*	*104*	*65*	*61*	*49*	*56*
Sebelum jualan bulanan, x	*52*	*100*	*60*	*128*	*104*	*98*	*48*	*64*	*96*	*48*

$$\sum x_i = 798, \sum y_i = 821, \sum x_i^2 = 71,028, \sum y_i^2 = 75,765, \sum x_i y_i = 71,672$$

Rekod-rekod daripada kedai menunjukkan bahawa jumlah jualan bulanan untuk semua cawangan sebelum pelancaran itu adalah 20,500 dalam ribuan ringgit.

(a) *Anggarkan peningkatan peratusan jualan bulanan selepas pelancaran kempen promosi dan juga anggarkan ralat piawainya.*

(b) *Anggarkan purata jualan bulanan selepas pelancaran kempen promosi serta kirakan ralat piawainya dengan menggunakan*

(i) *penganggar sampel rawak mudah.*

(ii) *penganggar nisbah.*

(iii) *penganggar perbezaan.*

(c) *Diberikan pekali korelasi* $\rho_{y,x} = 0.7854$ *bandingkan ketiga-tiga penganggar di bahagian (b) dan jelaskan jawapan anda.*

[20 markah]

Question 6

A videotape hire company has shops in each of 5 regions : three regions have 12 shops each, the others just 8 shops each. To estimate the total number of video films hired from the company in a particular week, the sales manager phones 12 shops chosen by picking 3 regions at random and then making a random choice of 3, 5 and 4 shops from the chosen regions.

The results were as follows:

First region : 260, 296, 182 (12 shops)
Second region : 156, 261, 130, 302, 241 (8 shops)
Third region : 196, 356, 368, 284 (12 shops)

(a) State the type of sampling design used.

(b) Estimate the total number of video films hired from the company.

(c) Place a bound on the error of estimation.

[12 marks]

Soalan 6

Sebuah syarikat yang menyewa video mempunyai beberapa buah kedai di setiap 5 kawasan: tiga kawasan mempunyai 12 buah kedai di setiap kawasan, sementara selainnya hanya 8 buah kedai di setiap kawasan. Untuk menganggarkan jumlah bilangan filem video yang disewa daripada syarikat dalam minggu yang tertentu, pengurus jualan menelefon 12 buah kedai yang dipilih dengan memilih 3 kawasan secara rawak dan kemudiannya membuat pilihan rawak 3, 5 dan 4 buah kedai dari kawasan yang terpilih.

Keputusannya adalah seperti yang berikut:

Kawasan pertama : 260, 296, 182 (12 kedai)
Kawasan kedua : 156, 261, 130, 302, 241 (8 kedai)
Kawasan ketiga : 196, 356, 368, 284 (12 kedai)

(a) Nyatakan jenis reka bentuk pensampelan yang digunakan.

(b) Anggarkan jumlah filem video yang disewa daripada syarikat.

(c) Tentukan batas ralat penganggarannya.

[12 markah]

Question 7

The editor of a local daily newspaper is interested in estimating the average number of misprints in the daily over a year of 365 days. All the editions of the daily corresponding to 365 days were labeled from 1 to 365. A systematic sample of 28 editions of the paper was selected. The selected editions were then carefully examined and misprints were counted. These are given in the following table along with the serial numbers of the selected editions of the paper.

Edition Number	Number of misprints	Edition Number	Number of misprints	Edition Number	Number of misprints
73	3	190	9	307	11
86	11	203	4	320	9
99	4	213	8	333	6
112	2	229	6	346	10
125	8	242	5	359	8
138	0	255	7	7	1
151	6	268	16	20	5
164	13	281	10	33	0
177	5	294	8	46	3
				59	7

(a) What are the possibilities for the first random start?

(b) Identify the sampling interval, k for this design.

(c) Estimate the average number of misprints in a daily edition.

(d) Determine an appropriate 90% confidence interval for the estimated total number of misprints in a daily edition.

[13 marks]

Soalan 7

Penyunting sebuah suratkhabar harian tempatan berminat untuk menganggarkan bilangan purata salah cetak dalam suratkhabar selama setahun yang mempunyai 365 hari. Semua edisi harian yang sepadan dengan 365 hari ditandakan dari 1 hingga 365. Satu sampel sistematik daripada 28 edisi suratkhabar telah dipilih. Edisi-edisi yang dipilih kemudiannya diperiksa dengan teliti dan salah cetak telah dikira. Ini diberikan dalam jadual yang berikut bersama –sama nombor siri suratkhabar yang terpilih.

Edition Number	***Number of misprints***	***Edition Number***	***Number of misprints***	***Edition Number***	***Number of misprints***
73	*3*	*190*	*9*	*307*	*11*
86	*11*	*203*	*4*	*320*	*9*
99	*4*	*213*	*8*	*333*	*6*
112	*2*	*229*	*6*	*346*	*10*
125	*8*	*242*	*5*	*359*	*8*
138	*0*	*255*	*7*	*7*	*1*
151	*6*	*268*	*16*	*20*	*5*
164	*13*	*281*	*10*	*33*	*0*
177	*5*	*294*	*8*	*46*	*3*
				59	*7*

(a) *Apakah kemungkinan-kemungkinan bagi permulaan rawak yang pertama?*

(b) *Kenal pastikan selang pensampelan, k bagi reka bentuk ini.*

(c) *Anggarkan purata bilangan salah cetak dalam edisi harian.*

(d) *Tentukan selang keyakinan 90% yang sesuai untuk anggaran jumlah salah cetak dalam edisi harian*

[13 markah]

Question 8

(a) In simple random sampling without replacement where n units are selected from N units, show that the probability that the ith unit is selected on the kth draw is $1/N$.

(b) An estimator of the variance of the regression estimator of μ is

$$\hat{V}(\hat{\mu}_{yL}) = \frac{N-n}{Nn}\frac{\sum_{i=1}^{n}(y_i - a - bx_i)^2}{n-2}$$

Show that

$$\hat{V}(\hat{\mu}_{yL}) = \frac{N-n}{Nn}\frac{n-1}{n-2}S_y{}^2(1-\hat{\rho}^2)$$

[10 marks]

Soalan 8

(a) Dalam pensampelan rawak ringkas tanpa penggantian dengan n unit dipilih daripada N unit, tunjukkan bahawa kebarangkalian bahawa unit ke-i dipilih pada cabutan ke-k ialah $1/N$.

(b) Suatu penganggar bagi varians penganggar regresi μ *ialah*

$$\hat{V}(\hat{\mu}_{yL}) = \frac{N-n}{Nn}\frac{\sum_{i=1}^{n}(y_i - a - bx_i)^2}{n-2}$$

Tunjukkan bahawa

$$\hat{V}(\hat{\mu}_{yL}) = \frac{N-n}{Nn}\frac{n-1}{n-2}S_y{}^2(1-\hat{\rho}^2)$$

[10 markah]

Appendix

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

Sample	Sample Variance
$\left(\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}\right)$	$\left(\frac{N-n}{Nn\bar{M}^2}\right)\left(\frac{\sum_{i=1}^{n}\left(y_i-\bar{y}m_i\right)^2}{n-1}\right)$
$\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}$	$\left(\frac{N-n}{Nn\bar{M}^2}\right)\frac{\sum_{i=1}^{n}\left(a_i-\hat{p}m_i\right)^2}{n-1}$
$M\bar{y}$	$M^2\left(\frac{N-n}{Nn\bar{M}^2}\right)\frac{\sum_{i=1}^{n}\left(y_i-\bar{y}m_i\right)^2}{n-1}$
$\frac{N}{n}\sum_{i=1}^{n} y_i$	$N^2\left(\frac{N-n}{Nn}\right)S_t^2$ with $S_t^2=\sum_{i=1}^{n}\frac{(y_i-\bar{y}_t)^2}{n-1}$
$\hat{\mu}=\frac{1}{n\bar{M}}\sum_{i=1}^{n} M_i\bar{y}_i$	$\left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)S_b^2+\frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2\left(\frac{M_i-m_i}{M_i}\right)\left(\frac{S_i^2}{m_i}\right)$ with $S_b^2=\frac{\sum_{i=1}^{n}\left(M_i\bar{y}_i-\bar{M}\hat{\mu}\right)^2}{n-1}$ $S_i^2=\frac{\sum_{j=1}^{m_i}\left(y_{ij}-\bar{y}_i\right)^2}{m_i-1}$ $\left(\frac{N-n}{N}\right)\frac{MSB}{mn}+\left(1-\frac{m}{\bar{M}}\right)\left(\frac{1}{N}\right)\frac{MSW}{m}$

Sample	Sample Variance
$\hat{\mu}_r = \frac{\sum_{i=1}^{n} M_i \bar{y}_i}{\sum_{i=1}^{n} M_i}$	$\left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)S_r^2 + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2\left(\frac{M_i - m_i}{M_i}\right)\left(\frac{S_i^2}{m_i}\right)$ With $S_r^2 = \frac{\sum_{i=1}^{n_i}\left(M_i\bar{y}_i - \hat{\mu}_r M_i\right)^2}{n-1}$
$\hat{p} = \frac{\sum_{i=1}^{n} M_i \hat{p}_i}{\sum_{i=1}^{n} M_i}$	$\left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)S_r^2 + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2\left(\frac{M_i - m_i}{M_i}\right)\left(\frac{\hat{p}_i\hat{q}_i}{m_i - 1}\right)$ with $S_r^2 = \frac{\sum_{i=1}^{n_i}\left(M_i\hat{p}_i - \hat{p}M_i\right)^2}{n-1}$

Sample Size

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2} \quad ; \quad D = \frac{B^2}{4} \quad ; \quad D = \frac{B^2}{4N^2}$$

$$n = \frac{\sum_{i=1}^{L} \frac{N_i \sigma_i^{\,2}}{w_i}}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^{\,2}} \quad ; \qquad w_i = \frac{n_i}{n}$$

$$n = \frac{\left(\sum_{k=1}^{L} N_k \sigma_k \big/ \sqrt{C_k}\right)\left(\sum_{i=1}^{L} N_i \sigma_i \sqrt{C_i}\right)}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^{\,2}} \quad , \quad n = \frac{(C - C_o)\sum_{i=1}^{L} N_i \sigma_i / \sqrt{C_i}}{\sum_{i=1}^{L} N_i \sigma_i \sqrt{C_i}}$$

$$n_i = \frac{n\, N_i \sigma_i \big/ \sqrt{C_i}}{\sum_{i=1}^{L} N_i \sigma_i \Big/ \sqrt{C_i}}$$

Optimal Allocation

$$n = \frac{\left(\sum_{i=1}^{L} N_i \sigma_i\right)^2}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^{\,2}} \quad ; \quad n_i = n\left(\frac{N_i \sigma_i}{\sum_{i=1}^{L} N_i \sigma_i}\right)$$

Neyman Allocation

$$n = \frac{N\sum_{i=1}^{L} N_i \sigma_i^{\,2}}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^{\,2}} \quad ; \quad n_i = n\left(\frac{N_i}{\sum_{i=1}^{L} N_i}\right)$$

Proportional Allocation

$$n = \frac{\sum_{i=1}^{L} N_i^2 p_i q_i / w_i}{N^2 D + \sum_{i=1}^{L} N_i p_i q_i} \quad ; \quad n_i = n\left(\frac{N_i \sqrt{p_i q_i / c_i}}{\sum_{i=1}^{L} N_i \sqrt{p_i q_i / c_i}}\right)$$

...16/-

$n = \frac{N\sigma^2}{ND + \sigma^2}$; $D = \frac{B^2\mu_x^2}{4}$; $D = \frac{B^2}{4}$; $D = \frac{B^2}{4N^2}$
$n = \frac{N\sigma_r^{\,2}}{ND + \sigma_r^{\,2}}$; $D = \frac{B^2\overline{M}^2}{4}$; $D = \frac{B^2}{4N^2}$

Intra class correlation coefficient
$\rho_w = \frac{2}{\sigma^2 nk(n-1)} \sum_{i=1}^{k} \sum_{j<u}^{n} (y_{ij} - \mu)(y_{iu} - \mu)$

- oooOOooo -