FINAL EXAMINATION
FIRST SEMESTER SESSION 2011/2012

COURSE CODE / NAME : SQQS3083 SAMPLING TECHNIQUES FOR DECISION MAKING
DATE : 9 JANUARY 2012 (MONDAY)
TIME : 8.30-11.00 PM (2 ½ HOURS)
VENUE : DTSO

INSTRUCTION :
1. This examination paper contains FIVE (5) questions in ELEVEN (11) printed pages, excluding the cover page.
2. Answer ALL QUESTIONS on the space provided
3. List of Formulae are in page 10 to 11.
4. You are NOT ALLOWED to take the examination paper out of the hall.
5. State 4 DECIMAL POINTS in all the ways of your calculation.
6. Total mark is EIGHTY (80).

MATRIC NO : ____________________________ (with word) ____________________________ (with number)
IDENTIFICATION CARD NO. :
LECTURER : ____________________________
GROUP : __________________ TABLE NO. : __________

DO NOT OPEN THIS EXAMINATION PAPER UNTIL INSTRUCTED

CONFIDENTIAL
QUESTION 1 (6 MARKS)

a) Explain why non-response error may bias the results of a survey. (2 marks)

b) Describe ONE (1) technique to reduce errors in a survey. (2 marks)

c) Compare between stratified and cluster sampling. (2 marks)
QUESTION 2 (16 MARKS)

For the total of the population, \( N = \{0, 1, 2, 3\} \), we wish to select the sample size, \( n = 3 \) with equal probability; assume that the sampling is without replication and without replacement.

List all possible outcomes of simple random sampling that can be selected from the population.

Then estimate the variance of each sample mean, the mean for each sample and the total population for each sample

(16 marks)
QUESTION 3 (30 MARKS)

A corporation is interested in estimating the total earnings from sales of color television sets at the end of a 3 months period. The total earnings figures are available for all districts within the corporation for the corresponding 3 months period of the previous year. A simple random sample of 13 district offices is selected from the 123 offices within the corporation and shown in TABLE 1.

Given \( r_x = 128200 \) and \( \sum_{i=1}^{13} (y_i - r_x) = 155875.8286 \).

<table>
<thead>
<tr>
<th>Office</th>
<th>3 months data from previous year</th>
<th>3 months data from current year</th>
<th>Office</th>
<th>3 months data from previous year</th>
<th>3 months data from current year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>550</td>
<td>610</td>
<td>8</td>
<td>1200</td>
<td>1440</td>
</tr>
<tr>
<td>2</td>
<td>720</td>
<td>780</td>
<td>9</td>
<td>1350</td>
<td>1570</td>
</tr>
<tr>
<td>3</td>
<td>1500</td>
<td>1600</td>
<td>10</td>
<td>1750</td>
<td>2210</td>
</tr>
<tr>
<td>4</td>
<td>1020</td>
<td>1030</td>
<td>11</td>
<td>670</td>
<td>980</td>
</tr>
<tr>
<td>5</td>
<td>620</td>
<td>600</td>
<td>12</td>
<td>729</td>
<td>865</td>
</tr>
<tr>
<td>6</td>
<td>980</td>
<td>1050</td>
<td>13</td>
<td>1530</td>
<td>1710</td>
</tr>
<tr>
<td>7</td>
<td>928</td>
<td>977</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a) Using a ratio estimator, estimate \( r_y \) and place a bound on the error of estimation.

(7 marks)
b) Estimate the mean earnings for offices within the corporation and place a bound on the error of estimation.

(7 marks)

c) Using a regression estimator:

i. Find the regression line of population mean. Given the following information:

\[ \sum_{i=1}^{13} (y_i - \bar{y})(x_i - \bar{x}) = 2202930.692 \]

\[ \sum_{i=1}^{13} (x_i - \bar{x})^2 = 1846508.923 \]

(7 marks)
ii. Estimate the mean earnings $\mu_y$ and place a bound on the error of estimation. Given the following information:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = 2778516.769$$

(6 marks)

iii. Compare your answer in (b) and (c)(ii), are there any advantages to use the regression estimator here? Explain.

(3 marks)
QUESTION 4 (18 MARKS)

a) What is a systematic sampling? (1 mark)

b) The management of a particular company is interested in estimating the proportion of employees favoring a new investment policy. A systematic sample is obtained from employees leaving the building at the end of a particular workday.

<table>
<thead>
<tr>
<th>Employee sampled</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>195</td>
<td>0</td>
</tr>
<tr>
<td>198</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ \sum_{i=1}^{200} y_i = 133 \]

Based on the data in TABLE 2:

i. Find the sampling interval, \( k \). (1 mark)

ii. Calculate the population size, \( N \). (1 mark)
iii. Find the estimation of proportion in favor of the new policy and place a bound on the error of estimation.  

(6 marks)

c) The quality control section of an industrial firm uses systematic sampling to estimate the average amount of fill in 12-ounce cans coming off an assembly line. The data in TABLE 3 represent a 1-in-50 systematic sample of the production in 1 day.

<table>
<thead>
<tr>
<th>TABLE 3</th>
</tr>
</thead>
</table>
| \[ \sum_{i=1}^{36} y_i = 429.96 \]  
| \[ \sum_{i=1}^{36} (y_i - \bar{y})^2 = 0.2204 \]  
| \[ N = 1800 \] |

i. Estimate the average amount of fill in 12-ounce cans and place a bound on the error of estimation.  

(6 marks)

ii. Determine the sample size required to estimate \( \mu \) to within 0.03 units if the variance for population is the same as it estimator.  

(3 marks)
QUESTION 5 (10 MARKS)

a) Block statistics report the number of housing units, the number of residents, and the number of rooms within housing units for a random sample of eight blocks selected from a large city (N = 1200) and the data shown in TABLE 4

<table>
<thead>
<tr>
<th>Block</th>
<th>Housing units</th>
<th>Residents</th>
<th>Rooms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>40</td>
<td>58</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>39</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>12</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>52</td>
<td>98</td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>37</td>
<td>74</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>33</td>
<td>57</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>41</td>
<td>76</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>14</td>
<td>48</td>
</tr>
</tbody>
</table>

Based on the data in TABLE 4,

i. Calculate the average cluster size for the number of residents per housing unit for the population.

(2 marks)
ii. Estimate the average number of residents per housing units, and place a bound on the error of estimation. Given \( \sum_{i=1}^{8} (y_i - \bar{y}_m)^2 = 334.9076 \).

(8 marks)
APPENDIX: List of Formulae

1. Simple Random Sampling:

\[ (1- \alpha)\% \text{ confidence interval of differences of } x \text{ and } y = (\bar{x} - \bar{y}) \pm z_{\alpha} \left( \frac{\sigma_x + \sigma_y}{2} \right) \]

2. Stratified Random Sampling:

\[ \bar{y}_m = \frac{1}{N} \sum_{i=1}^{n} N_i \bar{y}_i \rightarrow \text{Estimation of the population mean, } \mu. \]

3. Ratio and regression Estimator:

i. \[ r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \rightarrow \text{estimation of the population ratio.} \]

ii. \[ \hat{\mu}_x = \bar{x} \rightarrow \text{Mean estimation} \]

iii. \[ s_r^2 = \frac{\sum_{i=1}^{n} (y_i - rx_i)^2}{n-1} \rightarrow \text{Standard deviation of ratio estimation.} \]

iv. \[ \hat{\sigma}_r^2 = \left( \frac{N-n}{nN} \left( \frac{1}{\hat{\mu}_x^2} \right) s_r^2 \right) \rightarrow \text{Estimated variance of } r. \]

v. \[ B = 2\hat{\sigma}_r \rightarrow \text{Bound on the error of estimation.} \]

vi. \[ b = \frac{\sum_{i=1}^{n} [(y_i - \bar{y})(x_i - \bar{x})]}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \rightarrow \text{slope of the regression estimator of population mean, } \mu \]

vii. \[ \hat{\mu}_{st} = \bar{y} + b(\mu_x - \bar{x}) \rightarrow \text{Regression estimator of a population mean, } \mu \]

viii. \[ \hat{\sigma}_{\mu_{st}}^2 = \left( \frac{N-n}{Nn} \right) \left( \frac{1}{n-2} \right) \left[ \sum_{i=1}^{n} (y_i - \bar{y})^2 - b^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \right] \rightarrow \text{Estimated variance of regression estimator of a population mean, } \mu \]

10
SQQS3083 Sampling Techniques For Decision Making

ix. $\mu_{xa} = \frac{\tau_{xa}}{N} \Rightarrow$ Estimator for the mean for strata $A$

x. $\mu_{y_{bc}} = \left(\frac{N_A}{N} \left(\bar{y}_A - \bar{x}_A\right)\right) (\mu_{xa}) + \left(\frac{N_b}{N} \left(\bar{y}_B - \bar{x}_B\right)\right) (\mu_{xb}) \Rightarrow$ separate ratio estimator for the population mean, $\mu$

xi. $\mu_s = \frac{\tau_{xa} + \tau_{xb}}{N} \Rightarrow$ Total mean for strata

xii. $\hat{\mu}_{y_{bc}} = \frac{\bar{y}_h}{\bar{x}_{st}} (\mu_s) \Rightarrow$ Combine ratio estimator for the population mean, $\mu$

4. Systematic Sampling:

i. $\hat{p} = \frac{\sum_{i=1}^{n} y_i}{n} \Rightarrow$ Estimator for the population proportion

ii. $\sigma_{\hat{p}}^2 = \frac{P(1-P)}{n-1} \left(\frac{N-n}{N}\right) \Rightarrow$ estimated variance of the population proportion.

iii. $\bar{y}_p = \frac{\sum_{i=1}^{n} y_i}{n} \Rightarrow$ Estimator for the population mean, $\mu$

iv. $\hat{\sigma}_{\bar{y}_p}^2 = \frac{s^2}{n} \left(\frac{N-n}{N}\right) \Rightarrow$ Estimated variance of population mean.

v. Sample size estimation:

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

5. Cluster Sampling:

i. $\bar{M} = \frac{\sum_{i=1}^{n} m_i}{N} \Rightarrow$ the average cluster size for the population

ii. $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} \Rightarrow$ Estimator of the population mean, $\mu$

iii. Estimated variance of population mean.

$$\sigma_{\bar{y}}^2 = \left(\frac{N-n}{NnM^2}\right)s_r^2$$

$$; s_r^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y}m_i)^2}{n-1}.$$