208

UNIVERSITY OF SWAZILAND

SUPPLEMENTARY EXAMINATION PAPER 2005

TITLE OF PAPER	: SAMPLE SURVEY THEORY
COURSE CODE	: ST 306 *(NEW PROGRAMME)*
TIME ALLOWED	: TWO (2) HOURS
REQUIREMENTS	: CALCULATOR AND STATISTICAL TABLES FORMULA SHEET ATTACHED
INSTRUCTIONS	: ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

Question 1

A society has two grades of membership (Grade I and Grade II) and a worldwide membership of about 7000 individuals. About 20% of the members fall into Grade II, and about 75% of all members are concentrated into three geographical areas (A, B, C), the rest being spread throughout the rest of the world.

The society publishes a journal which is sent by post to all members, and it wishes to carry out a survey to discover if members find the journal useful, and what aspects of their subject they most wish to see covered in the journal. The society is particularly anxious to discover whether members of both grades find the journal useful.

The society keeps its membership list as a computer file, with one record for each member. The records are stored in alphabetical order of members' names, though the secretary is assured that separate lists could be provided for the different grades of membership.

Consider five possible methods for selecting a sample of members to receive, by post, a questionnaire for this purpose:
simple random sampling,
stratified random sampling,
cluster sampling,
systematic sampling.

For each method, discuss
(i) whether it would be possible to use this method of sampling for this purpose,
(ii) whether the method would be a good one to choose for the purpose.

(5 marks for each method)

Question 2

A wholesale food distributor in a large city wants to assess the demand for a new product based on mean monthly sales. He plans to sell this product in a sample of stores he services. He only services four large chains in the city. Hence, for administrative convenience, he decides to use stratified random sampling with each chain as a stratum. A stratified random sample of 20 stores yields the following sales figures after a month:

Stratum (Chain)			
1	2	3	4
$N_1 = 24$ $n_1 = 4$ $\bar{y}_1 = 99.3$ $s_1 = 9.00$	$N_1 = 36$ $n_1 = 6$ $\bar{y}_1 = 100.0$ $s_1 = 7.46$	$N_1 = 30$ $n_1 = 5$ $\bar{y}_1 = 98.0$ $s_1 = 6.28$	$N_1 = 30$ $n_1 = 5$ $\bar{y}_1 = 100.0$ $s_1 = 10.61$
94 90 103 110	91 99 93 105 111 101	108 96 100 93 93	92 110 94 91 113

a) Explain what is meant by *stratification with proportional allocation,* and verify that this has been used to construct the above stratum sample sizes. (4)

b) Write down an expression for the unbiased estimator of the population mean in terms of the stratum means and derive its standard error. (5)

c) Estimate the mean monthly sales and obtain an estimate of the standard error of your estimator. Construct an approximate 95% confidence interval for the population mean. (3)

d) Suppose instead that a simple random sample of 20 stores from the population of 120 stores had been selected, with the same responses as given in the table. Had this been done, the estimate for the population standard deviation would have been 7.75. Construct an approximate 95% confidence interval for the population mean. (2)

e) Compare the efficiencies of your estimators in parts (c) and (d). Suggest why stratified random sampling gives a less precise estimate of the population mean than simple random sampling in this case. (3)

f) Advise the wholesaler on how to select appropriate strata for a stratified random sample when the objective of stratification is to produce estimators with small variance.

(3)

Question 3

A region has 3510 farms which cluster naturally into 90 different "villages". In any village, x denotes the total number of farms and y the total number of cattle. A simple random sample of 15 villages (clusters) was selected, giving the data shown in the table.

Village (cluster) i	Number of farms x_i	Number of cattle y_i	Mean number of cattle per farm $\bar{z}_i = y_i/x_i$
1	35	418	11.94
2	25	402	16.08
3	48	360	7.50
4	30	394	13.13
5	70	515	7.36
6	55	910	16.55
7	66	600	9.09
8	18	316	17.56
9	30	288	9.60
10	32	350	10.94
11	64	784	12.25
12	24	290	12.08
13	48	795	16.56
14	40	478	11.95
15	82	906	11.05
Total	667	7806	

$\Sigma x_i y_i = 393716$; $\Sigma x_i^2 = 34883$; $\Sigma y_i^2 = 4759890$; $\Sigma \bar{z}_i = 183.64$; $\Sigma \bar{z}_i^2 = 2395.8018$

Estimate the mean number of cattle per farm in the region as a whole and its 95% confidence interval. (20)

Question 4

The wholesale price for oranges in large shipments is based on the sugar content of the load. The exact sugar content cannot be determined prior to the purchase and extraction of the juice from the load. The data below show the sugar content in grams (y) and weight in grams (x) of a random sample of ten oranges from a consignment with total weight 820 kg.

Sugar content (gm) y	9.5	13.6	11.4	9.1	15.0	12.3	8.6	9.5	10.5	11.4
Weight (gm) x	182	218	195	191	227	209	177	186	190	200

$\Sigma xy = 22194.80$, $\Sigma y^2 = 1268.69$, $\Sigma x^2 = 392389$.

a) Explain why a ratio estimator is appropriate for these data. (5)
b) Estimate the total sugar content for the oranges and give a standard error of your estimate, giving reasons for your choice of estimator. (15)

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

$$\hat{\mu}_{srs} = \bar{y} \qquad \hat{V}(\hat{\mu})_{srs} = \frac{s^2}{n}\left(\frac{N-n}{N}\right)$$

$$\hat{\tau}_{srs} = N\hat{\mu}_{srs} \qquad \hat{V}(\hat{\tau})_{srs} = N^2\hat{V}(\hat{\mu})_{srs}$$

$$\hat{p}_{srs} = \sum_{i=1}^{n} \frac{y_i}{n} \qquad \hat{V}(\hat{p})_{srs} = \frac{\hat{p}(1-\hat{p})}{(n-1)}\left(\frac{N-n}{N}\right)$$

$$\hat{\tau}_{pps} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\pi_i}\right) \qquad \hat{V}(\hat{\tau})_{pps} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{\pi_i} - \hat{\tau}_{pps}\right)^2$$

$$\hat{\mu}_{pps} = \frac{1}{N}\hat{\tau}_{pps} \qquad \hat{V}(\hat{\mu})_{pps} = \frac{1}{N^2}\hat{V}(\hat{\tau})_{pps}$$

$$\hat{\mu}_{sys} = \sum_{i=1}^{n} \frac{y_i}{n} \qquad \hat{V}(\hat{\mu})_{sys} = \frac{s^2}{n}\left(\frac{N-n}{N}\right)$$

$$\hat{\tau}_{sys} = N\hat{\mu}_{sys} \qquad \hat{V}(\hat{\tau})_{sys} = N^2\hat{V}(\hat{\mu})_{sys}$$

$$\hat{p}_{sys} = \sum_{i=1}^{n} \frac{y_i}{n} \qquad \hat{V}(\hat{p})_{sys} = \frac{\hat{p}(1-\hat{p})}{(n-1)}\left(\frac{N-n}{N}\right)$$

$$\hat{\mu}_{rsys} = \sum_{i=1}^{ns} \frac{\hat{\mu}_i}{ns} \qquad \hat{V}(\hat{\mu})_{rsys} = \left(\frac{N-n}{N}\right)\sum_{i=1}^{ns} \frac{(\hat{\mu}_i - \hat{\mu}_{rsys})^2}{ns(ns-1)}$$

$$\hat{\tau}_{rsys} = N\hat{\mu}_{rsys} \qquad \hat{V}(\hat{\tau})_{rsys} = N^2\hat{V}(\hat{\mu})_{rsys}$$

$$\hat{\mu}_{str} = \frac{1}{N}\sum_{i=1}^{L} N_i\bar{y}_i \qquad \hat{V}(\hat{\mu})_{str} = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2\left(\frac{N_i-n_i}{N_i}\right)\frac{s_i^2}{n_i}$$

$$\hat{\tau}_{str} = N\hat{\mu}_{str} \qquad \hat{V}(\hat{\tau})_{str} = N^2\hat{V}(\hat{\mu})_{str}$$

$$\hat{p}_{str} = \frac{1}{N}\sum_{i=1}^{L} N_i\hat{p}_i \qquad \hat{V}(\hat{p})_{str} = \frac{1}{N^2}\sum_{i=1}^{L} N_i^2\left(\frac{N_i-n_i}{N_i}\right)\left(\frac{\hat{p}_i(1-\hat{p}_i)}{n_i-1}\right)$$

$$\hat{\mu}_{pstr} = \sum_{i=1}^{L} w_i\bar{y}_i \qquad \hat{V}(\hat{\mu})_{pstr} = \frac{1}{n}\left(\frac{N-n}{N}\right)\sum_{i=1}^{L} w_i s_i^2 + \frac{1}{n^2}\sum_{i=1}^{L}(1-w_i)s_i^2$$

$r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$ $\hat{V}(r) = (\frac{N-n}{N})(\frac{1}{n\mu_x^2})\frac{\sum_{i=1}^{n}(y_i - rx_i)^2}{(n-1)}$

$\hat{\rho} = \frac{\hat{cov}(x,y)}{s_x s_y}$ $\hat{V}(r) = \frac{1-(n/N)}{n}(\frac{1}{\mu_x^2})(s_y^2 + r^2 s_x^2 - 2r\hat{\rho}s_x s_y)$

$\hat{\tau}_{ratio} = r\tau_x$ $\hat{V}(\hat{\tau})_{ratio} = \tau_x^2 \hat{V}(r)$

$\hat{\mu}_{ratio} = r\mu_x$ $\hat{V}(\hat{\mu})_{ratio} = \mu_x^2 \hat{V}(r)$

$Y_i = \beta_0 + \beta_1(X_i) + \varepsilon_i$ $\sum_{i=1}^{n}(y_i - rx_i)^2 = \sum_{i=1}^{n} y_i^2 + r^2 \sum_{i=1}^{n} x_i^2 - 2r \sum_{i=1}^{n} y_i x_i$

$b_0 = \bar{y} - b_1\bar{x}$ $b_1 = \hat{\rho}(s_y/s_x)$

$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ $\hat{\rho} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$

$\hat{\mu}_{reg} = \bar{y} + b_1(\mu_x - \bar{x})$ $\hat{V}(\hat{\mu})_{reg} = (\frac{N-n}{N})\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n(n-1)}$

$\hat{y}_i = b_0 + b_1(x_i)$ $\hat{V}(\hat{\mu})_{reg} \approx (\frac{N-n}{N})\frac{MSE}{n}$

$\hat{\mu}_{diff} = \bar{y} + (\mu_x - \bar{x})$ $\hat{V}(\hat{\mu})_{diff} = (\frac{N-n}{N})\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n(n-1)}$

$\sum_{i=1}^{n}(d_i - \bar{d})^2 = \sum_{i=1}^{n} d_i^2 - n\bar{d}^2$ $\hat{RE}(\frac{E1}{E2}) = \frac{\hat{V}(E2)}{\hat{V}(E1)}$

$\hat{\mu}_{cts1} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$ $\hat{V}(\hat{\mu})_{cts1} = (\frac{N-n}{N})\frac{\sum_{i=1}^{n}(y_i - \bar{y}m_i)^2}{nM^2(n-1)}$

$\hat{V}(\hat{\mu})_{cts1} = (\frac{N-n}{N})(\frac{1}{nM^2})(s_y^2 + \hat{\mu}_{cts1}^2 s_m^2 - 2\hat{\mu}_{cts1}\hat{\rho}s_y s_m)$

$\hat{\tau}_{cts1(1)} = M\hat{\mu}_{cts1}$ $\hat{V}(\hat{\tau})_{cts1(1)} = M^2\hat{V}(\hat{\mu})_{cts1}$

$$\hat{\tau}_{cts1(2)} = N\bar{y}_t = N\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) \qquad \hat{V}(\hat{\tau})_{cts1(2)} = \left(\frac{N-n}{N}\right)\left(\frac{N^2}{n}\right)\frac{\sum_{i=1}^{n}(y_i-\bar{y}_t)^2}{(n-1)}$$

$$\bar{m} = \frac{\sum_{i=1}^{n} m_i}{n} \qquad \sum_{i=1}^{n}(y_i - \bar{y}m_i)^2 = \sum_{i=1}^{n} y_i^2 + \bar{y}^2\sum_{i=1}^{n} m_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i m_i$$

$$\hat{p}_{cts1} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i} \qquad \hat{V}(\hat{p})_{cts1} = \left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)\frac{\sum_{i=1}^{n}(a_i-\hat{p}m_i)^2}{(n-1)}$$

$$\Pi_i = \frac{m_i}{M} \qquad \hat{V}(\hat{p})_{cts1} = \left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)(s_a^2 + \hat{p}^2 s_m^2 - 2\hat{p}\hat{\rho}s_a s_m)$$

$$\hat{\tau}_{cts1,pps} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\Pi_i} \qquad \sum_{i=1}^{n}(a_i - \hat{p}m_i)^2 = \sum_{i=1}^{n} a_i^2 + \hat{p}^2\sum_{i=1}^{n} m_i^2 - 2\hat{p}\sum_{i=1}^{n} a_i m_i$$

$$\hat{\tau}_{cts1,pps} = \frac{M}{n}\sum_{i=1}^{n}\bar{y}_i \qquad \hat{V}(\hat{\tau})_{cts1,pps} = \frac{M^2}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i - \hat{\tau})^2$$

$$\hat{\mu}_{cts1,pps} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i \qquad \hat{V}(\hat{\mu})_{cts1,pps} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i - \hat{\mu})^2$$

$$\hat{\mu}_{cts2} = \left(\frac{N}{M}\right)\frac{\sum_{i=1}^{n} M_i\bar{y}_i}{n} \qquad \hat{V}(\hat{\mu})_{cts2} = \left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)s_b^2 + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2\left(\frac{M_i-m_i}{M_i}\right)\left(\frac{s_i^2}{m_i}\right)$$

$$s_b^2 = \frac{\sum_{i=1}^{n}(M_i\bar{y}_i - \bar{M}\hat{\mu})^2}{n-1} \qquad s_i^2 = \frac{\sum_{i=1}^{m_i}(y_{ij}-\bar{y}_i)^2}{m_i-1}$$

$$\hat{\tau}_{cts2} = M\hat{\mu}_{cts2} \qquad \hat{V}(\hat{\tau})_{cts2} = M^2\hat{V}(\hat{\mu})_{cts2}$$

$$\hat{\mu}_{cts2,ratio} = \frac{\sum_{i=1}^{n} M_i\bar{y}_i}{\sum_{i=1}^{n} M_i} \qquad \hat{V}(\hat{\mu})_{cts2,ratio} = \left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)s_r^2 + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2\left(\frac{M_i-m_i}{M_i}\right)\left(\frac{s_i^2}{m_i}\right)$$

$$s_r^2 = \frac{\sum_{i=1}^{n} M_i^2(\bar{y}_i - \hat{\mu}_{cts2,r})^2}{n-1} \qquad s_r^2 = \frac{\sum_{i=1}^{n} M_i^2(\hat{p}_i - \hat{p}_{cts2,r})^2}{n-1}$$

$$\hat{p}_{cts2,ratio} = \frac{\sum_{i=1}^{n} M_i\hat{p}_i}{\sum_{i=1}^{n} M_i} \qquad \hat{V}(\hat{p})_{cts2,ratio} = \left(\frac{N-n}{N}\right)\left(\frac{1}{n\bar{M}^2}\right)s_r^2 + \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2\left(\frac{M_i-m_i}{M_i}\right)\left(\frac{\hat{p}_i(1-\hat{p}_i)}{m_i-1}\right)$$

$\hat{\mu}_{cts2,pps} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$ $\hat{V}(\hat{\mu})_{cts2,pps} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i - \hat{\mu}_{cts2,pps})^2$

$\hat{\tau}_{cts2,pps} = M\hat{\mu}_{cts2,pps}$ $\hat{V}(\hat{\tau}) = M^2\hat{V}(\hat{\mu})_{cts2,pps}$

n for μ (SRS): $n = \frac{N\sigma^2}{(N-1)(B^2/4)+\sigma^2}$

n for τ (SRS): $n = \frac{N\sigma^2}{(N-1)(B^2/4N^2)+\sigma^2}$

n for p (SRS): $n = \frac{Np(1-p)}{(N-1)(B^2/4)+p(1-p)}$

n for μ (SYS): $n = \frac{N\sigma^2}{(N-1)(B^2/4)+\sigma^2}$

n for p (SYS): $n = \frac{Np(1-p)}{(N-1)(B^2/4)+p(1-p)}$

$k \leq \frac{N}{n}$ $k' = k(ns)$

n for μ (STR): $n = \frac{\sum_{i=1}^{L} N_i^2(\sigma_i^2/w_i)}{N^2(B^2/4)+\sum_{i=1}^{L} N_i\sigma_i^2}$

n for τ (STR): $n = \frac{\sum_{i=1}^{L} N_i^2(\sigma_i^2/w_i)}{N^2(B^2/4N^2)+\sum_{i=1}^{L} N_i\sigma_i^2}$

Allocations for STR μ:

$n_i = n\left(\frac{N_i\sigma_i/\sqrt{c_i}}{\sum_{k=1}^{L} N_k\sigma_k/\sqrt{c_k}}\right)$ $n = \frac{\left(\sum_{k=1}^{L} N_k\sigma_k/\sqrt{c_k}\right)\left(\sum_{i=1}^{L} N_i\sigma_i\sqrt{c_i}\right)}{N^2(B^2/4)+\sum_{i=1}^{L} N_i\sigma_i^2}$

$n_i = n\left(\frac{N_i\sigma_i}{\sum_{k=1}^{L} N_k\sigma_i}\right)$ $n = \frac{\left(\sum_{i=1}^{L} N_i\sigma_i\right)^2}{N^2(B^2/4)+\sum_{i=1}^{L} N_i\sigma_i^2}$

$n_i = n\left(\frac{N_i}{N}\right)$ $n = \frac{\sum_{i=1}^{L} N_i\sigma_i^2}{N^2(B^2/4)+(1/N)\sum_{i=1}^{L} N_i\sigma_i^2}$

Allocations for STR τ: change $N^2(B^2/4)$ to $N^2(B^2/4N^2)$

Allocations for STR p:

$$n_i = n\left(\frac{N_i\sqrt{p_i(1-p_i)/c_i}}{\sum_{k=1}^{L} N_k\sqrt{p_k(1-p_k)/c_k}}\right)$$

$$n = \frac{\sum_{i=1}^{L} N_i^2 p_i(1-p_i)/w_i}{N^2(B^2/4)+\sum_{i=1}^{L} N_i p_i(1-p_i)}$$

n for μ (ratio): $n = \frac{N\sigma^2}{N(B^2/4)+\sigma^2}$

n for τ (ratio): $n = \frac{N\sigma^2}{N(B^2/4N^2)+\sigma^2}$

n for μ (CTS1): $n = \frac{N\sigma_c^2}{N(B^2M^2/4)+\sigma_c^2}$

n for τ (CTS1(1)): $n = \frac{N\sigma_c^2}{N(B^2/4N^2)+\sigma_c^2}$

n for τ (CTS1(2)): $n = \frac{N\sigma_t^2}{N(B^2/4N^2)+\sigma_t^2}$

$s_c^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y}m_i)^2}{(n-1)}$ with $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$ $\quad$ $s_t^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y}_t)^2}{(n-1)}$ with $\bar{y}_t = \frac{\sum_{i=1}^{n} y_i}{n}$

n for p (CTS1): $n = \frac{N\sigma_c^2}{N(B^2M^2/4)+\sigma_c^2}$

$$s_c^2 = \frac{\sum_{i=1}^{n}(a_i-\hat{p}m_i)^2}{(n-1)}$$