JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

UNIVERSITY EXAMINATION 2012/2013

1ST YEAR 1ST SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE (ACTUARIAL SCIENCE)

(REGULAR)

COURSE CODE: SAS 201

TITLE: SAMPLE SURVEYS

DATE: 15/8/13 **TIME: 2.00-4.00 PM**

DURATION: 2 HOURS

INSTRUCTIONS

1. **This paper contains SIX (6) questions**
2. **Answer question 1 (Compulsory) and ANY other 2 Questions**
3. **Write all answers in the booklet provided**

JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

BACHELOR OF SCIENCE – ACTUARIAL SCIENCE
YEAR TWO SEMESTER ONE
SAS 201: SAMPLE SURVEY

QUESTION ONE (COMPULSORY 30MKS)

I. Define the following terms as used in sample survey:
 a. Sample survey (2mks)
 b. Sampling unit (2mks)
 c. Sampling frame (2mks)

II. State and explain **THREE** properties of estimators (6mks)

III. If random samples of size three and drawn without replacement from the population consisting of four numbers 4, 5, 5, 7. Find sample mean $\overline{X}$ for each sample and make sampling distribution of $\overline{X}$. Calculate the mean and standard deviation of this sampling distribution. Compare your calculations with population parameters (5mks)

Solution: We have population values 4, 5, 5, 7, population size N=4 and sample size n=3. Thus, the number of possible samples which can be drawn without replacement is $\binom{N}{n} = \binom{4}{3}$ =4

Sample No	Sample Values	Sample Mean ($\bar{X}$)
1	4, 5, 5	14/3
2	4, 5, 7	16/3
3	4, 5, 7	16/3
4	5, 5, 7	17/3

The sampling distribution of the sample mean $\bar{X}$ and its mean and standard deviation are:

$\bar{X}$	freq	$p(\bar{X})$	$\bar{X}p(\bar{X})$	$\bar{X}^2p(\bar{X})$
14/3	1	1/4	14/12	196/36
16/3	2	2/4	32/12	512/36
17/3	1	1/4	17/12	289/36
Total	4	1	63/12	997/36

$\mu_{\bar{X}} = \sum \bar{X}p(\bar{X}) = \frac{63}{12} = 5.25;\quad \sigma_{\bar{X}} = \sqrt{\sum \bar{X}^2p(\bar{X}) - [\bar{X}p(\bar{X})]^2} = \sqrt{\frac{997}{36} - \left(\frac{63}{12}\right)^2} = 0.3632$

The mean and standard deviation of the population are:

X	4	5	5	7	$\sum X \;=21$
X^2	16	25	25	49	$\sum X^2$=115

$$\mu = \frac{\sum X}{N} = \frac{21}{4} = 5.25; \quad \sigma \quad = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2} = \sqrt{\frac{115}{4} - \left(\frac{21}{4}\right)^2} = 1.0897$$

$$\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1.0897}{\sqrt{3}}\sqrt{\frac{4-3}{4-1}} = 0.3632$$

Hence, $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

IV. A corporation desires to estimate the total number of worker-hours lost for a given month because of accidents among all employees. Because laborers, technicians and administrators have different accident rates, the researcher decides to use stratified random sampling, with each group forming a separate stratum. Summary statistics are as follows:

	Stratum		
	I	II	III
N_i	132	92	27
σ_i^2	36	25	9
$N_i\sigma_i$	792	460	81
n_i	18	10	2
$\bar{y}_i$	8.83	6.7	4.5
s_i^2	81.56	50.46	24.50

Determine Neyman allocation for a sample of n=30 employees. (4mks)

$$\sum N_i \sigma_i = 1333; \; n_i = n \frac{N_i \sigma_i}{\sum N_i \sigma_i}$$

$$n_1 = 30\left(\frac{792}{1333}\right) = 17.82 \approx 18$$

$$n_2 = 30\left(\frac{460}{1333}\right) = 10.35 \approx 10$$

$$n_3 = 30\left(\frac{81}{1333}\right) = 1.82 \approx 2$$

V. National income from manufacturing industries is to be estimated for 1989 from a sample of 6 of the 19 industry categories that reported figures early for that year. Incomes from all 19 industries are known for 1980 and the total is $674 billion. From the data provided, find a difference estimator of the 1989 total income, and place a bound on the error of estimation. (4mks)

VI. A study was conducted to measure the body temperatures of a sample of female and male students at JOOUST. The summary statistics for these data are as follows:

	n	MEAN	MEDIAN	STDEV	SEMEAN
Male	10	97.880	97.850	0.555	0.176
Female	10	98.520	98.400	0.527	0.167

Find:

a) An approximate 95% confidence interval for the mean body temperature of men in the population (2mks)

b) An approximate 95% confidence interval for the difference in mean body temperature between women and men (3mks)

(a) An approximate 95% confidence interval for the mean body temperature of men in the population sampled here is

$$\bar{y} \pm 2\frac{s}{\sqrt{n}} = 97.88 \pm 2(0.176) = 97.88 \pm 0.35$$

The resulting interval is (97.53, 98.23).

(b) An approximate 95% confidence interval for the difference in mean body temperature between women and men is

$$(\bar{y}_1 - \bar{y}_2) \pm 2\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} =$$

$$(98.52 - 97.88) \pm 2\sqrt{0.167^2 + 0.176^2} =$$

$$0.64 \pm 0.49$$

This interval does not overlap zero; therefore, there is evidence that men have a lower mean body temperature than women.

QUESTION TWO (20MKS)

a. Consider the following population elements $y_1 = 10, y_2 = 6, y_3 = 4, y_4 = 4$ where a sample of size 2 is to be drawn. Show that for a simple random sampling without replacement, the standard estimate of the population mean is unbiased. (10mks)

b. Prove that in simple random sampling without replacement $\text{var}(\bar{y}) = \frac{N-n}{N}\frac{S^2}{n}$ (10mks)

QUESTION THREE (20MKS)

a) A sample of 30 students is to be drawn from a population of 300 students belonging to two colleges A and B. The means and standard deviations of their marks are given below;

	Total number of students	$\bar{y}_i$	S_i
College A	200	30	10
College B	100	60	40

Use the information to confirm that Neyman's allocation scheme is a more efficient scheme when compared to proportional allocation (10mks)

b) The "spring-like" effect in a golf club could be determined by measuring the coefficient of restitution. Drivers are randomly selected from two club makers and the coefficient of restitution is measured. The summary statistics for these data are as follows:

	n	Mean	Median	Stdev	SeMean
Club 1	10	0.9788	0.9785	0.55	0.176
Club 2	10	0.9852	0.984	0.52	0.167

Is there sufficient evidence to conclude that the coefficient of restitution for Club 1 differs from that for Club 2? (5mks)

c) A simple random sample of 100 water meters within a community is monitored to estimate the average daily water consumption per household over a specified dry spell. The sample mean and sample variance are found to be $\overline{Y} = 12.5 and S^2 = 1252$
If we assume that there are N = 10,000 households within the community, estimate « , the true mean daily consumption, and place a bound on the error of estimation. (5mks)

QUESTION FOUR (20MKS)

a) All the farms in a country are stratified by farm size and mean number of hectares of wheat per farm in each stratum, with the following results.

Farm size (hectares)	No. of farms	Mean wheat (hectares)	standard deviation
0-20	368	2.7	2.1
21-40	425	8.1	3.6
41-60	389	12.1	3.9
61-80	316	16.9	5.1
81-100	174	20.8	6.1
101-120	98	25.2	6.5
121+	138	31.8	9.1

For a sample of 100 farms, compute the sizes in each stratum under stratified simple random Sampling with;

Proportional allocation (5mks)

Neyman allocation. (5mks)

b) Let $\overline{y}_{sys}$ be the estimate of $\overline{Y}$ from the systematic sample, then show that
$E(\overline{y}_{sys}) = \overline{Y}$ (10mks)

QUESTION FIVE (20MKS)

a) Signatures to a petition were collected on 676 sheets each sheet had enough space for 42 signatures, but on many sheets, a smaller number of signatures had been collected. The numbers of signatures per sheet were counted on a random sample of 50 sheets (about a 7% sample). The results are given in the table below.

y_i	42	41	36	32	29	27	23	19	16	15	14	11	10	7	6	5	4	3
f_i	23	4	1	1	1	2	1	1	2	2	1	1	1	1	3	2	1	1

Estimate the total number of signatures to the petition and the 80% confidence limits. (10mks)

b) Prove that proportional allocation is more efficient than Neyman's allocation (10mks)