

B. Sc. Examination by course unit 2009

MTH5119 Sampling, Surveys and Simulation

Duration: 2 hours

Date and time: 27 April 2009, 1000h–1200h

Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.

You may attempt as many questions as you wish and all questions carry equal marks. Except for the award of a bare pass, only the best **FOUR** questions answered will be counted.

Calculators **ARE** permitted in this examination. The unauthorized use of material stored in pre-programmable memory constitutes an examination offence. Please state on your answer book the name and type of machine used.

Statistical functions provided by the calculator may be used provided that you state clearly where you have used them.

The New Cambridge Statistical Tables are provided.

Complete all rough workings in the answer book and cross through any work which is not to be assessed.

Candidates should note that the Examination and Assessment Regulations state that possession of unauthorized materials by any candidate who is under examination conditions is an assessment offence. Please check your pockets now for any notes that you may have forgotten that are in your possession. If you have any, then please raise your hand and give them to an invigilator now.

Exam papers must not be removed from the examination room.

Examiner(s): R. A. Sugden and H. Grossmann

Question 1 (a) Using the random string of digits

756260135

as economically as possible, draw a simple random sample of size 3 from a population size 7 by Waterman's algorithm. [11]

(b) If the units labelled $1, \dots, 7$ have responses 2.5, 3.2, 2.8, 3.3, 2.9, 3.0, 2.3 respectively, estimate the population *total* T from your selected sample in (a), together with an appropriate variance estimate. [12]

(c) Does the nominal 95% confidence interval for T include the true value? [2]

Question 2 A simple random sample of size n is drawn from a finite population of size N .

(a) Give the appropriate variance estimator for the unbiased estimator of the population mean when the sample is poststratified into poststratum sample sizes n_h , means \bar{y}_h , and variances s_h^2 , and the poststratum population sizes are N_h , $h = 1, \dots, H$. [5]

(b) A simple random sample of size 27 from a population size 100 is poststratified according to two factors A and B and the respective population size, sample size, sample mean and sample variance are given below in each cell of the following table (all real numbers are given to 3 s.f.):

	$B = 1$	$B = 2$	all B
$A = 1$	40,10	20,7	60,17
	3.60	4.20	3.85
	0.645	0.732	0.985
$A = 2$	10,4	30,6	40,10
	3.90	4.50	4.26
	0.708	0.832	1.13
all A	50,14	50,13	100,27
	3.69	4.34	4.00
	0.854	0.904	1.25

By calculating the variance estimator of (a) in each case, indicate which of the following four poststratifications you would use for estimating the population mean and say why: (i) by both A and B ($H = 4$), (ii) by A only ($H = 2$), (iii) by B only ($H = 2$), (iv) by neither A nor B ($H = 1$)? [20]

Question 3 A simple random sample of size 2 is to be drawn from a population size 4 and it is desired to estimate the population mean of a variable Y . A suitable size measure X has values 2,3,3,3 and the corresponding values of Y are 3,7,8,9.

(a) Find the sampling distribution of the ratio estimator. [11]

(b) From your sampling distribution find the exact bias and variance to 5 s.f. [8]

(c) Hence find the approximation error in the large sample formula for the variance given that $S_Y^2 = 6.91\bar{6}$, $S_X^2 = 0.25$, and $S_{XY} = 1.25$. [6]

Question 4 (a) Using the random string

756260135

as economically as possible, draw a PPS sample of size 2 with replacement from a population with cluster sizes 2,6,4,3,3,2. [8]

(b) Given that the total responses for each cluster are 3,9,6,4,6,4 respectively, use your drawn sample to unbiasedly estimate the population mean response per individual R . [7]

(c) Use your sample to calculate an appropriate variance estimate, and a nominal 95% confidence interval for R . [10]

Question 5 (a) Show that if U is a uniform random variable on the interval $(0, 1)$ then $X = -\frac{1}{\lambda} \log U$ has an exponential distribution with parameter $\lambda > 0$. [8]

(b) Using the random string

756260135,

simulate two values from an exponential distribution mean 2, using 4 s.f. throughout. [7]

(c) Hence, using the Box-Müller transformation method, generate the values of a pair of independent standard normal random variables. [10]

Question 6 A sample survey of students in the U.K. is to be undertaken in order to find out how much time is spent on reading, replying to and sending email. Describe some of the main problems with non-sampling error which the survey researcher must consider, writing your description as an essay of not more than 300 words. [25]

End of Paper