S2 Sampling and Estimations from Samples Past Qu.s

1) June 2004 no.2

A random sample of 5 people is to be drawn from 10 people whose surnames are Ali, Budd, Cook, Dost, Evans, Fox, Grant, Hall, Ip, Jenks. Use the following extract from a table of random numbers, starting at the beginning, to obtain the sample, and write down the names of the people chosen. Make your method clear.

$$079\ 831\ 130\ 709\ 938\ 423\ 756\ 281\ 787\ 118$$ [4]

2) Jan 2005 no.2

The editor of a local newspaper is attempting to determine what proportion of the adults in the area served by the newspaper is interested in environmental matters. One issue of the newspaper therefore contains a questionnaire which readers are invited to complete and return.

(i) Give two reasons why the results obtained may be biased. [2]

(ii) Describe briefly an unbiased method of obtaining the information. [3]

3) June 2006 no.4

(i) Explain briefly what is meant by a random sample. [1]

Random numbers are used to select, with replacement, a sample of size n from a population numbered $000, 001, 002, \ldots, 799$.

(ii) If $n = 6$, find the probability that exactly 4 of the selected sample have numbers less than 500. [3]

(iii) If $n = 60$, use a suitable approximation to calculate the probability that at least 40 of the selected sample have numbers less than 500. [6]

4) Jan 2007 no.2

A school has 900 pupils. For a survey, Jan obtains a list of all the pupils, numbered 1 to 900 in alphabetical order. She then selects a sample by the following method. Two fair dice, one red and one green, are thrown, and the number in the list of the first pupil in the sample is determined by the following table.

		Score on green dice					
		1	2	3	4	5	6
Score on red dice	1, 2 or 3	1	2	3	4	5	6
	4, 5 or 6	7	8	9	10	11	12

For example, if the scores on the red and green dice are 5 and 2 respectively, then the first member of the sample is the pupil numbered 8 in the list.

Starting with this first number, every 12th number on the list is then used, so that if the first pupil selected is numbered 8, the others will be numbered 20, 32, 44, ...

(i) State the size of the sample. [1]

(ii) Explain briefly whether the following statements are true.

(a) Each pupil in the school has an equal probability of being in the sample. [1]

(b) The pupils in the sample are selected independently of one another. [1]

(iii) Give a reason why the number of the first pupil in the sample should not be obtained simply by adding together the scores on the two dice. Justify your answer. [2]

5) June 2007 no.1

A random sample of observations of a random variable X is summarised by

$$n = 100, \quad \Sigma x = 4830.0, \quad \Sigma x^2 = 249\,509.16.$$

(i) Obtain unbiased estimates of the mean and variance of X. [4]

(ii) The sample mean of 100 observations of X is denoted by $\bar{X}$. Explain whether you would need any further information about the distribution of X in order to estimate $P(\bar{X} > 60)$. [You should not attempt to carry out the calculation.] [2]

6) Jan 2007 no.4

A set of observations of a random variable W can be summarised as follows:

$$n = 14, \quad \Sigma w = 100.8, \quad \Sigma w^2 = 938.70.$$

(i) Calculate an unbiased estimate of the variance of W. [4]

(ii) The mean of 70 observations of W is denoted by $\bar{W}$. State the approximate distribution of $\bar{W}$, including unbiased estimate(s) of any parameter(s). [3]

7) Jan 2006 no.6

Alex obtained the actual waist measurements, w inches, of a random sample of 50 pairs of jeans, each of which was labelled as having a 32-inch waist. The results are summarised by

$$n = 50, \quad \Sigma w = 1615.0, \quad \Sigma w^2 = 52\,214.50.$$

Test, at the 0.1% significance level, whether this sample provides evidence that the mean waist measurement of jeans labelled as having 32-inch waists is in fact greater than 32 inches. State your hypotheses clearly. [10]

8) June 2004 no.8

A psychologist is testing the effect of background music on students' work. She knows that the average time for a randomly chosen student to complete a particular task in the absence of background music is 38.5 seconds. A sample of 50 students took the test with background music being played. The times taken, t seconds, are summarised as follows.

$$n = 50, \quad \Sigma t = 1967, \quad \Sigma t^2 = 77\,959.$$

(i) Test, at the 5% significance level, whether the presence of background music has an effect on the times taken by students to complete the task. State your hypotheses clearly. [10]

(ii) Give a reason why it is necessary to use the Central Limit Theorem in carrying out your test. [1]

8) June 2007 no.3

The proportion of adults in a large village who support a proposal to build a bypass is denoted by p. A random sample of size 20 is selected from the adults in the village, and the members of the sample are asked whether or not they support the proposal.

(i) Name the probability distribution that would be used in a hypothesis test for the value of p. [1]

(ii) State the properties of a random sample that explain why the distribution in part **(i)** is likely to be a good model. [2]

9) Jan 2008 no.2

A village has a population of 600 people. A sample of 12 people is obtained as follows. A list of all 600 people is obtained and a three-digit number, between 001 and 600 inclusive, is allocated to each name in alphabetical order. Twelve three-digit random numbers, between 001 and 600 inclusive, are obtained and the people whose names correspond to those numbers are chosen.

(i) Find the probability that all 12 of the numbers chosen are 500 or less. [3]

(ii) When the selection has been made, it is found that all of the numbers chosen are 500 or less. One of the people in the village says, "The sampling method must have been biased." Comment on this statement. [2]

10) Jan 2008 no.4

The random variable Y has the distribution $N(\mu, \sigma^2)$. The results of 40 independent observations of Y are summarised by

$$\Sigma y = 3296.0, \quad \Sigma y^2 = 286\,800.40.$$

(i) Calculate unbiased estimates of μ and σ^2. [4]

(ii) Use your answers to part (i) to estimate the probability that a single random observation of Y will be less than 60.0. [2]

(iii) Explain whether it is necessary to know that Y is normally distributed in answering part (i) of this question. [1]

11) June 2008 no.1

The head teacher of a school asks for volunteers from among the pupils to take part in a survey on political interests.

(i) Explain why a sample consisting of all the volunteers is unlikely to give a true picture of the political interests of all pupils in the school. [2]

(ii) Describe a better method of obtaining the sample. [3]

12) June 2008 no.3

In a factory the time, T minutes, taken by an employee to make a single item is a normally distributed random variable with mean 28.0. A new ventilation system is installed, after which the times taken to produce a random sample of 40 items are measured. The sample mean is 26.44 minutes and it is given that $\frac{\Sigma t^2}{40} - 26.44^2 = 37.05$. Test, at the 10% significance level, whether there is evidence of a change in the mean time taken to make an item. [8]

13) Jan 2009 no.2

The continuous random variable Y has the distribution $N(23.0, 5.0^2)$. The mean of n observations of Y is denoted by $\overline{Y}$. It is given that $P(\overline{Y} > 23.625) = 0.0228$. Find the value of n. **[4]**

14) Jan 2009 no.7

A motorist records the time taken, T minutes, to drive a particular stretch of road on each of 64 occasions. Her results are summarised by

$$\Sigma t = 876.8, \quad \Sigma t^2 = 12\,657.28.$$

(i) Test, at the 5% significance level, whether the mean time for the motorist to drive the stretch of road is greater than 13.1 minutes. **[11]**

(ii) Explain whether it is necessary to use the Central Limit Theorem in your test. **[1]**

15) June 2009 no.3

A survey is to be carried out to draw conclusions about the proportion p of residents of a town who support the building of a new supermarket. It is proposed to carry out the survey by interviewing a large number of people in the high street of the town, which attracts a large number of tourists.

(i) Give two different reasons why this proposed method is inappropriate. [2]

(ii) Suggest a good method of carrying out the survey. [3]

(iii) State two statistical properties of your survey method that would enable reliable conclusions about p to be drawn. [2]

16) June 2009 no.6

The continuous random variable R has the distribution $N(\mu, \sigma^2)$. The results of 100 observations of R are summarised by

$$\Sigma r = 3360.0, \quad \Sigma r^2 = 115\,782.84.$$

(i) Calculate an unbiased estimate of μ and an unbiased estimate of σ^2. [4]

(ii) The mean of 9 observations of R is denoted by $\overline{R}$. Calculate an estimate of $P(\overline{R} > 32.0)$. [4]

(iii) Explain whether you need to use the Central Limit Theorem in your answer to part **(ii)**. [2]

17) June 2009 no.7

The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{2}{9}x(3-x) & 0 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the variance of X. [5]

(ii) Show that the probability that a single observation of X lies between 0.0 and 0.5 is $\frac{2}{27}$. [2]

(iii) 108 observations of X are obtained. Using a suitable approximation, find the probability that at least 10 of the observations lie between 0.0 and 0.5. [6]

(iv) The mean of 108 observations of X is denoted by $\overline{X}$. Write down the approximate distribution of $\overline{X}$, giving the value(s) of any parameter(s). [3]

18) Jan 2010 no.1

The values of 5 independent observations from a population can be summarised by

$$\Sigma x = 75.8, \quad \Sigma x^2 = 1154.58.$$

Find unbiased estimates of the population mean and variance. [4]

19) Jan 2010 no.2

A college has 400 students. A journalist wants to carry out a survey about food preferences and she obtains a sample of 30 pupils from the college by the following method.

- Obtain a list of all the students.

- Number the students, with numbers running sequentially from 0 to 399.

- Select 30 random integers in the range 000 to 999 inclusive. If a random integer is in the range 0 to 399, then the student with that number is selected. If the number is greater than 399, then 400 is subtracted from the number (if necessary more than once) until an answer in the range 0 to 399 is selected, and the student with that number is selected.

(i) Explain why this method is unsatisfactory. [2]

(ii) Explain how it could be improved. [1]

20) June 2010 no.2

A university has a large number of students, of whom 35% are studying science subjects. A sample of 10 students is obtained by listing all the students, giving each a serial number and selecting by using random numbers.

(i) Find the probability that fewer than 3 of the sample are studying science subjects. [3]

(ii) It is required that, in selecting the sample, the same student is not selected twice. Explain whether this requirement invalidates your calculation in part (i). [2]

21) June 2010 no.7

A machine is designed to make paper with mean thickness 56.80 micrometres. The thicknesses, x micrometres, of a random sample of 300 sheets are summarised by

$$n = 300, \qquad \Sigma x = 17\,085.0, \qquad \Sigma x^2 = 973\,847.0.$$

Test, at the 10% significance level, whether the machine is producing paper of the designed thickness. [11]