DEGREE EXAMINATION

ST1505 Understanding Data

Wednesday 30 May 2007   (3 pm to 5 pm)

Only calculators approved by the Department of Mathematical Sciences may be used in this examination. Calculator memories must be clear at the start of the examination.

Marks may be deducted for answers that do not show clearly how the solution is reached.

Attempt ALL THREE of the questions in SECTION A and TWO of the questions in SECTION B. Each question in section A is worth 10 marks and each question in section B is worth 20 marks.

SECTION A

1. A gardener picks some baby aubergines from his greenhouse. The weights of the aubergines in grammes are given in the table below:

$$62 \quad 82 \quad 67 \quad 68 \quad 78 \quad 74 \quad 59 \quad 77 \quad 76 \quad 80$$
$$61 \quad 75 \quad 57 \quad 74 \quad 64 \quad 73 \quad 72 \quad 84 \quad 68 \quad 76.$$

 (i) Draw a stem and leaf plot of the data choosing a suitable bin width and leaf unit.

 (ii) State whether the sample mean is greater than, about equal to or less than the sample median. Justify your answer without doing any calculations.

 (iii) Find the sample median and the first and third quartiles.

 (iv) Another gardener picks twenty baby aubergines from his greenhouse. The weights have a similar range as the first gardener. What sort of plot would be used to compare the quartiles of the weights of the aubergines from the two gardeners?

2. **(a)** Examine the following statement and the conclusion drawn from it. Assuming the statement is true present an argument against the validity of the conclusion.

  A survey of pensioners found that the mean weight of men aged 65 was 69.1 kilogrammes whereas the corresponding figure for men aged 75 was 66.9.

  Therefore men lose weight as they grow older.

(b) In the entrance hall of a railway station there are ten public telephones. At any given time the probability that a telphone is not working is 0.1 independently of the other telephones.

Find the probability that, at a given time,

 (i) all telephones are working;

 (ii) at least 9 of the telephones are working.

A passenger arrives at the entrance hall and finds that none of the telephones are being used. She tries each telephone in turn until she finds one that is working. What is the probability that she will need to try exactly two telephones?

3. **(a)** Explain what is meant by *stratified* sampling and *cluster* sampling. Describe a statistical investigation for which stratified sampling would be appropriate.

(b) A sample of 100 observations from a (continuous) variable X was taken. MINITAB produced the frequency histogram:
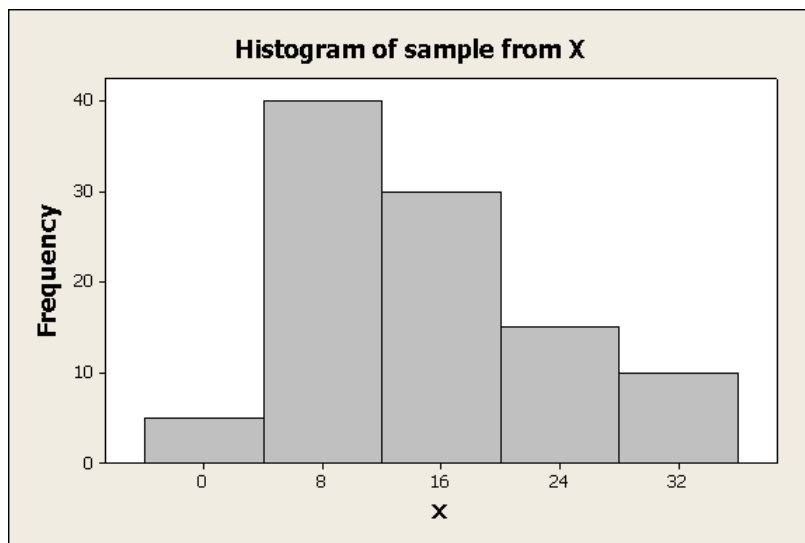
Using information from the histogram, estimate the sample mean and the sample variance. (The 'heights' of the bars are 5, 40, 30, 15 and 10.)

SECTION B

4. **(a)** Assume that the resting heart rate of women in a population is normally distributed with mean 71 beats per minute and standard deviation 12 beats per minute.

Find the probability that a randomly selected woman has a resting heart rate:

   (i) greater than 80 beats per minute;

   (ii) between 68 and 80 beats per minute;

   (iii) less than 68 or greater than 80 beats per minute.

A random sample of 25 women is taken from the population and their resting heart rates are measured. What is the probability that the sample mean heart rate is less than 69 beats per minute?

(b) A random sample of the resting heart rate of 100 women was taken from a different population, with

$$\sum_{i=1}^{100} x_i = 6774 \cdot 8 \quad \text{and} \quad \sum_{i=1}^{100} x_i^2 = 472283.$$

Calculate a 95% confidence interval for the mean heart rate of the population.

5. **(a)** Of the adult population of a town, 52% are women. 80% of the men in the town are on the electoral register and 75% of the women are on the electoral register. A person is chosen at random from the adult population of the town. What is the probability that the person is a man given that the person is not on the electoral register.

(b) Surgeons can use two methods for performing a certain operation, method A and method B say. They want to compare the time taken to do the operation by the two different methods. The times (in minutes) of a random sample of 15 operations using method A and a random sample of 20 operations using method B are taken. The sample mean and sample variance of the times for the two methods are given in the table below:

	Sample Mean	Sample Variance	n
Method A	44	10.2	15
Method B	47	11.9	20

(i) Check that it is reasonable to pool variances. Find the pooled variance.

(ii) Is there evidence to conclude that the mean time of operations is different for the two methods?

6. The table below summarises the results of a survey of men at a large company which investigated the association between blood pressure (classified as high or low) and occupation.

	Occupation		
Blood Pressure	Sales	Management	Professional
Low	34	38	17
High	63	51	49

(i) What proportion of the salesmen had high blood pressure?

(ii) Using a suitable (cluster) bar chart, compare the levels of blood pressure for the three types of occupation. Comment briefly on any differences of the blood pressure between the occupations indicated by your chart.

(iii) A chi-squared test of independence between blood pressure and type of work was performed using MINITAB. Edited output from MINITAB follows:

```
Expected counts are printed below observed counts

      Sales  Manage   Prof  Total
  1     34      38     17     89
      34.26    (a)   23.31

  2     63      51     49    163
      62.74   57.57  42.69

Total   97      89     66    252

Chi-Sq = 0.002 + (b) + 1.708 + 0.001 + 0.749 +0.933 = (c),   DF = (d)
```

State the null and alternative hypotheses of the test.

Write down the formula used to calculate the expected counts.

Evaluate the missing values (a), (b), (c) and (d) ((b) is the contribution to Chi-Sq from the cell 'low' and 'management').

Is there any evidence of an association between blood pressure and the type of work? Report the results of this test briefly in a non-technical way.