

Statistics-S1 - 2007-January

Question 1

As part of a statistics project, Gill collected data relating to the length of time, to the nearest minute, spent by shoppers in a supermarket and the amount of money they spent. Her data for a random sample of 10 shoppers are summarised in the table below, where t represents time and $£m$ the amount spent over £20.

t (minutes)	$£m$
15	-3
23	17
5	-19
16	4
30	12
6	-9
32	27
23	6
35	20
27	6

- (a) Write down the actual amount spent by the shopper who was in the supermarket for 15 minutes.

(1)

- (b) Calculate S_{tt} , S_{mm} and S_{tm} .

(You may use $\Sigma t^2 = 5478$ $\Sigma m^2 = 2101$ $\Sigma tm = 2485$)

(6)

- (c) Calculate the value of the product moment correlation coefficient between t and m .

(3)

- (d) Write down the value of the product moment correlation coefficient between t and the actual amount spent. Give a reason to justify your value.

(2)

On another day Gill collected similar data. For these data the product moment correlation coefficient was 0.178

- (e) Give an interpretation to both of these coefficients.

(2)

- (f) Suggest a practical reason why these two values are so different.

(1)

Question 2

In a factory, machines A , B and C are all producing metal rods of the same length. Machine A produces 35% of the rods, machine B produces 25% and the rest are produced by machine C . Of their production of rods, machines A , B and C produce 3%, 6% and 5% defective rods respectively.

- (a) Draw a tree diagram to represent this information. (3)
- (b) Find the probability that a randomly selected rod is
- (i) produced by machine A and is defective,
 - (ii) is defective. (5)
- (c) Given that a randomly selected rod is defective, find the probability that it was produced by machine C . (3)
-

Question 3

The random variable X has probability function

$$P(X = x) = \frac{(2x-1)}{36} \quad x = 1, 2, 3, 4, 5, 6.$$

- (a) Construct a table giving the probability distribution of X . (3)
- Find
- (b) $P(2 < X \leq 5)$, (2)
- (c) the exact value of $E(X)$. (2)
- (d) Show that $\text{Var}(X) = 1.97$ to 3 significant figures. (4)
- (e) Find $\text{Var}(2 - 3X)$. (2)
-

Question 4

Summarised below are the distances, to the nearest mile, travelled to work by a random sample of 120 commuters.

Distance (to the nearest mile)	Number of commuters
0–9	10
10–19	19
20–29	43
30–39	25
40–49	8
50–59	6
60–69	5
70–79	3
80–89	1

For this distribution,

- (a) describe its shape, (1)

- (b) use linear interpolation to estimate its median. (2)

The mid-point of each class was represented by x and its corresponding frequency by f giving

$$\Sigma fx = 3550 \quad \text{and} \quad \Sigma fx^2 = 138020$$

- (c) Estimate the mean and the standard deviation of this distribution. (3)

One coefficient of skewness is given by

$$\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}.$$

- (d) Evaluate this coefficient for this distribution. (3)

- (e) State whether or not the value of your coefficient is consistent with your description in part (a). Justify your answer. (2)

- (f) State, with a reason, whether you should use the mean or the median to represent the data in this distribution. (2)

- (g) State the circumstance under which it would not matter whether you used the mean or the median to represent a set of data.

(1)

Question 5

A teacher recorded, to the nearest hour, the time spent watching television during a particular week by each child in a random sample. The times were summarised in a grouped frequency table and represented by a histogram.

One of the classes in the grouped frequency distribution was 20–29 and its associated frequency was 9. On the histogram the height of the rectangle representing that class was 3.6 cm and the width was 2 cm.

- (a) Give a reason to support the use of a histogram to represent these data.

(1)

- (b) Write down the underlying feature associated with each of the bars in a histogram.

(1)

- (c) Show that on this histogram each child was represented by 0.8 cm^2 .

(3)

The total area under the histogram was 24 cm^2 .

- (d) Find the total number of children in the group.

(2)

Question 6

(a) Give two reasons to justify the use of statistical models.

(2)

It has been suggested that there are 7 stages involved in creating a statistical model. They are summarised below, with stages 3, 4 and 7 missing.

Stage 1. The recognition of a real-world problem.

Stage 2. A statistical model is devised.

Stage 3.

Stage 4.

Stage 5. Comparisons are made against the devised model.

Stage 6. Statistical concepts are used to test how well the model describes the real-world problem.

Stage 7.

(b) Write down the missing stages.

(3)

Question 7

The measure of intelligence, IQ, of a group of students is assumed to be Normally distributed with mean 100 and standard deviation 15.

(a) Find the probability that a student selected at random has an IQ less than 91.

(4)

The probability that a randomly selected student has an IQ of at least $100 + k$ is 0.2090.

(b) Find, to the nearest integer, the value of k .

(6)