## S1 January 2007

**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)**

**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)**

---

**3.** The random variable $X$ has probability function

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

(a) Construct a table giving the probability distribution of $X$.

**(3)**

Find

(b) $P(2 < X \leqslant 5)$,

**(2)**

(c) the exact value of $E(X)$.

**(2)**

(d) Show that $Var(X) = 1.97$ to 3 significant figures.

**(4)**

(e) Find $Var(2 - 3X)$.

**(2)**

---

**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)**

**[contd]**

(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)**

---

**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)**

---

**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)**

---

**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)**

---

**TOTAL FOR PAPER: 75 MARKS**

edexcel

# January 2007
# 6683 Statistics S1
# Mark Scheme

| Question number | Scheme | Marks |
|---|---|---|
| 1. (a) | (£) 17 Just **17** | B1 (1) |
| (b) | $\sum t = 212$ and $\sum m = 61$ (Accept as totals under each column in qu.) | B1, B1 |
| | $S_{tm} = 2485 - \frac{61\times 212}{10}$ , $= 1191.8$ awrt **1190** or 119 (3sf) | M1, A1 |
| | $S_{tt} = 983.6$ (awrt 984) and $S_{mm} = 1728.9$ (awrt 1730) (or 98.4 and 173) | A1, A1 (6) |
| (c) | $r = \frac{1191.8}{\sqrt{983.6\times 1728.9}}$ | M1, A1f.t. |
| | $= 0.913922\ldots$ awrt **0.914** | A1 (3) |
| (d) | 0.914 (Must be the same as (c) or awrt 0.914) | B1f.t. ($\|r\| < 1$) |
| | e.g. linear transformation, coding does not affect coefficient (or recalculate) | dB1 (2) |
| (e) | 0.914 suggests longer spent shopping the more spent. (Idea more time, more spent) | B1 |
| | 0.178 different amounts spent for same time. | B1 (2) |
| (f) | e.g. might spend short time buying 1 expensive item OR might spend a long time checking for bargains, talking, buying lots of cheap items. | B1g (1) |
| | | **15 marks** |

(b) M1 for one correct formula seen, f.t. their $\sum t, \sum m$ [Use 1st A1 for 1 correct, 2nd A1 for 2 etc]

(c) M1 for attempt at correct formula, $\frac{2485}{\sqrt{2101\times 5478}}$ scores M1A0A0

A1ft f.t. their values for $S_{tt}$ etc from (b) but don't give for $S_{tt}$ = 5478 etc (see above)

Answer only (awrt 0.914) scores 3/3, 0.913 (i.e. truncation) can score M1A1ft by implication.

(d) 2nd B1 dependent on 1st B1 Accept $\sum m = 261, \sum m^2 = 8541, \sum tm = 6725 \rightarrow 0.914$

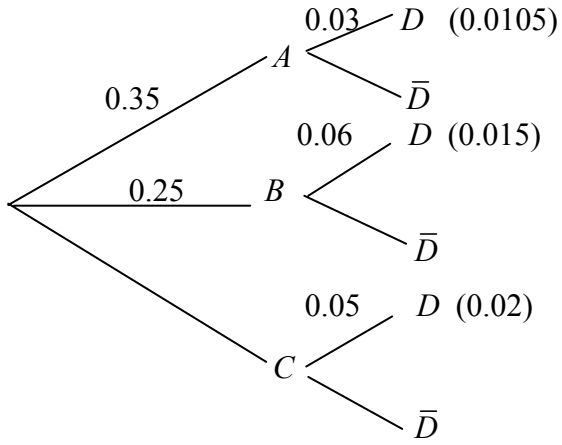
(e) One mark for a sensible comment relating to each coefficient

For 0.178 allow "little or no link between time and amount spent". Must be in context.

Just saying 0.914 is strong +ve correlation between amount spent and time shopping and 0.178 is weak correlation …scores B0B0.

(f) B1g for a sensible, practical suggestion showing that other factors might affect the amount spent.

E.g. different day (weekend vs weekday) or time of day (time spent queuing if busy)

| Question Number | Scheme | Marks |
|---|---|---|
| 2. (a) |  Correct tree shape<br>*A, B* and *C* and 0.35 and 0.25<br>*D* (x3) and 0.03, 0.06, 0.05<br>(May be implied by seeing $P(A \cap D)$ etc at the ends) | M1<br>A1<br>A1 (3) |
| (b)(i) | $P(A \cap D) = 0.35 \times 0.03$, = **0.0105** or $\frac{21}{2000}$ | M1, A1 |
| | P(*C*) = 0.4 (anywhere) | B1 |
| (ii) | P(*D*) = (i) + 0.25x 0.06 + (0.4x0.05) | M1 |
| | = **0.0455** or $\frac{91}{2000}$ | A1 (5) |
| (c) | $P(C \mid D) = \frac{P(C \cap D)}{P(D)}$, = $\frac{0.4 \times 0.05}{\text{(ii)}}$ | M1, A1ft |
| | = 0.43956… or $\frac{40}{91}$ **0.44** or awrt **0.440** | A1 (3) |
| | [Correct answers only score full marks in each part] | **11 marks** |

| | |
|---|---|
| (a) | M1 for tree diagram, 3 branches and then two from each. At least one probability attempted. |
| (b) | 1st M1 for 0.35x0.03. Allow for equivalent from their tree diagram. |
| | B1 for P(*C*) = 0.4, can be in correct place on tree diagram or implied by 0.4x0.05 in P(*D*). |
| | 2nd M1 for all 3 cases attempted and some correct probabilities seen, including +. Can ft their tree. |
| | Condone poor use of notation if correct calculations seen. E.g. $P(C \mid D)$ for $P(C \cap D)$. |
| (c) | M1 for attempting correct ratio of probabilities. There must be an attempt to substitute some values in a correct formula. If no correct formula and ration not correct ft score M0. |
| | Writing P(*D*\|*C*) and attempting to find this is M0. |
| | Writing P(*D*\|*C*) but calculating correct ratio – ignore notation and mark ratios. |
| | A1ft must have their 0.4 x0.05 divided by their (ii). |
| | If ratio is incorrect ft (0/3) unless correct formula seen and part of ratio is correct then M1. |

| Question Number | Scheme | Marks |
|---|---|---|
| 3. (a) | N.B. Part (a) doesn't have to be in a table, could be a list P($X = 1$) = …etc<br>$x$: 1, 2, 3, 4, 5, 6<br>P($X = x$): $\frac{1}{36}$, $\frac{3}{36}$, $\frac{5}{36}$, $\frac{7}{36}$, $\frac{9}{36}$, $\frac{11}{36}$<br>0.0278, 0.0833, 0.139, 0.194, 0.25, 0.306 (Accept awrt 3 s.f) | B1, B1, B1 (3) |
| (b) | P(3) + P(4) + P(5) =, $\frac{21}{36}$ or $\frac{7}{12}$ or awrt 0.583 | M1, A1 (2) |
| (c) | $E(X) = \frac{1}{36} + 2\times\frac{3}{36} + \ldots, = \frac{161}{36}$ or $4.47\dot{2}$ or $4\frac{17}{36}$ | M1, A1 (2) |
| (d) | $E(X^2) = \frac{1}{36} + 2^2\times\frac{3}{36} + \ldots, = \frac{791}{36}$ or full expression or $21\frac{35}{36}$ or awrt 21.97 | M1, A1 |
| | $\text{Var}(X) = \frac{791}{36} - \left(\frac{161}{36}\right)^2$ , = **1.9714… *** | M1, A1c.s.o. (4) |
| (e) | Var(2 – 3$X$) = 9×1.97 or $(-3)^2\times1.97$, = 17.73 awrt **17.7** or $\frac{2555}{144}$ | M1, A1 (2) |
| | | **13 marks** |

(a) 1st B1 for $x = 1, \ldots 6$ and at least one correct probability N.B. $\frac{3}{36} = \frac{1}{12}$ and $\frac{9}{36} = \frac{1}{4}$

2nd B1 for at least 3 correct probabilities

3rd B1 for a fully correct probability distribution.

(b) M1 for attempt to add the correct three probabilities, ft their probability distribution

(c) M1 for a correct attempt at E($X$). Minimum is as printed. Exact answer only scores M1A1.

[Division by 6 at any point scores M0, no ISW. Non-exact answers with no working score M0.]

(d) 1st M1 for a correct attempt at $E(X^2)$. Minimum as printed. $\frac{791}{36}$ or awrt 21.97 scores M1A1.

2nd M1 for their $E(X^2) - \left(\text{their } E(X)\right)^2$.

2nd A1 cso needs awrt 1.97 and $\frac{791}{36} - \left(\frac{161}{36}\right)^2$ or $\frac{2555}{1296}$ or any fully correct expression seen.

Can accept at least 4 sf for both. i.e. 21.97 for $\frac{791}{36}$, 4.472 for $\frac{161}{36}$, 20.00 for $\left(\frac{161}{36}\right)^2$.

(e) M1 for correct use of Var($aX + b$) formula or a full method.

NB $-3^2\times1.97$ followed by awrt 17.7 scores M1A1 BUT $-3^2\times1.97$ alone, or followed by – 17.7, scores M0A0.

| Question Number | Scheme | Marks |
|---|---|---|
| 4. (a) | Positive skew (both bits) | B1 (1) |
| (b) | $19.5 + \frac{(60-29)}{43}\times 10$, $= 26.7093\ldots$ awrt **26.7** | M1, A1 (2) |
| | (N.B. Use of 60.5 gives 26.825… so allow awrt 26.8) | |
| (c) | $\mu = \frac{3550}{120} = 29.5833...$ or $29\frac{7}{12}$ awrt **29.6** | B1 |
| | $\sigma^2 = \frac{138020}{120} - \mu^2$ or $\sigma = \sqrt{\frac{138020}{120} - \mu^2}$ | M1 |
| | $\sigma = 16.5829...$ or $(s = 16.652...)$ awrt **16.6** (or $s$ = 16.7) | A1 (3) |
| (d) | $\frac{3(29.6-26.7)}{16.6}$ | M1A1ft |
| | $= 0.52\ldots$ awrt **0.520** (or with $s$ awrt 0.518) | A1 (3) |
| | (N.B. 60.5 in (b) ...awrt 0.499[or with $s$ awrt 0.497]) | |
| (e) | $0.520 > 0$ correct statement about their (d) being >0 or < 0 | B1ft |
| | So it is consistent with (a) ft their (d) | dB1ft (2) |
| (f) | Use Median | B1 |
| | Since the data is skewed or less affected by outliers/extreme values | dB1 (2) |
| (g) | If the data are symmetrical or skewness is zero or normal/uniform distribution | B1 (1) |
| | ("mean =median" or "no outliers" or "evenly distributed" all score B0) | **14 marks** |

| | |
|---|---|
| (b) | M1 for (19.5 or 20) + $\frac{(60-29)}{43}\times 10$ or better. Allow 60.5 giving awrt 26.8 for M1A1<br>Allow their $0.5n$ [or $0.5(n+1)$] instead of 60 [or 60.5] for M1. |
| (c) | M1 for a correct expression for $\sigma, \sigma^2, s$ or $s^2$. NB $\sigma^2 = 274.99$ and $s^2 = 277.30$<br>Condone poor notation if answer is awrt16.6 (or 16.7 for $s$) |
| (d) | M1 for attempt to use this formula using their values to any accuracy. Condone missing 3.<br>1st A1ft for using their values to at least 3sf. Must have the 3.<br>2nd A1 for using accurate enough values to get awrt 0.520 (or 0.518 if using $s$)<br>NB Using only 3 sf gives 0.524 and scores M1A1A0 |
| (e) | 1st B1 for saying or implying correct sign for their (d). B1g and B1ft. Ignore "correlation" if seen.<br>2nd B1 for a comment about consistency with their (d) and (a) being positive skew, ft their (d) only<br>This is dependent on 1st B1: so if (d)>0, they say yes, if (d)<0 they say no. |
| (f) | 2nd B1 is dependent upon choosing median. |

| Question Number | Scheme | Marks |
|---|---|---|
| 5. (a) | Time is a <u>continuous</u> variable <u>or</u> data is in a <u>grouped</u> frequency table | B1 (1) |
| (b) | Area is proportional to frequency <u>or</u> $A \propto f$ or $A = kf$ | B1 (1) |
| (c) | $3.6 \times 2 = 0.8 \times 9$<br>1 child represented by 0.8 | M1<br>dM1<br>A1 cso (3) |
| (d) | (Total) = $\frac{24}{0.8}$, = **<u>30</u>** | M1, A1 (2) |
| | | **7 marks** |

(b) 1st B1 for one of these correct statements.

"Area proportional to frequency density" or "Area = frequency" is B0

(c) 1st M1 for a correct combination of any 2 of the 4 numbers: 3.6, 2, 0.8 and 9

e.g. $3.6 \times 2$ or $\frac{3.6}{0.8}$ or $\frac{0.8}{2}$ etc BUT e.g. $\frac{3.6}{2}$ is M0

2nd M1 dependent on 1st M1 and for a correct combination of 3 numbers leading to 4th.

May be in separate stages but must see all 4 numbers

A1cso for fully correct solution. Both Ms scored, no false working seen and <u>comment required.</u>

(d) M1 for $\frac{24}{0.8}$ seen or implied.

| Question Number | Scheme | Marks |
|---|---|---|
| 6. (a) | Used to simplify <u>or</u> represent a real world problem<br>Cheaper <u>or</u> quicker <u>or</u> easier (than the real situation) <u>or</u> more easily modified<br>To improve understanding of the real world problem<br>Used to predict outcomes from a real world problem (idea of predictions) | (any two lines)<br>B1<br>B1 (2) |
| (b) | (3 or 4) Model used to make predictions. (Idea of predicted values based on the model)<br>(4 or 3) (Experimental) data collected<br>(7) Model is refined. | B1<br>B1<br>B1 (3)<br>**5 marks** |

(a) 1st B1 For one line

2nd B1 For a second line

Be generous for 1st B1 but stricter for B1B1

(b) 1st & 2nd B1 These two points can be interchanged.

Idea of values from (experimental) data and predicted values based on the model.

1st B1 for predicted values from model e.g. "model used to gain suitable data"

2nd B1 for data collected. Idea of experimental data but "experiment" needn't be explicitly seen

3rd B1 This should be stage 7. Idea of refinement or revision or adjustment