Paper Reference(s)

# 6683/01
# Edexcel GCE
## Statistics S1
## Gold Level G1

# Time: 1 hour 30 minutes

| **Materials required for examination papers** | **Items included with question** |
|---|---|
| Mathematical Formulae (Green) | Nil |

**Candidates may use any calculator allowed by the regulations of the Joint Council for Qualifications. Calculators must not have the facility for symbolic algebra manipulation, differentiation and integration, or have retrievable mathematical formulas stored in them.**

## Instructions to Candidates

Write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, initials and signature.

## Information for Candidates

A booklet 'Mathematical Formulae and Statistical Tables' is provided.
Full marks may be obtained for answers to ALL questions.
There are 7 questions in this question paper. The total mark for this paper is 75.

## Advice to Candidates

You must ensure that your answers to parts of questions are clearly labelled.
You must show sufficient working to make your methods clear to the Examiner. Answers without working may gain no credit.

**Suggested grade boundaries for this paper:**

| A* | A | B | C | D | E |
|---|---|---|---|---|---|
| 61 | 54 | 47 | 40 | 34 | 28 |

**1.** Gary compared the total attendance, *x*, at home matches and the total number of goals, *y*, scored at home during a season for each of 12 football teams playing in a league. He correctly calculated:

$$S_{xx} = 1022500, \qquad S_{yy} = 130.9, \qquad S_{xy} = 8825.$$

(*a*) Calculate the product moment correlation coefficient for these data.

**(2)**

(*b*) Interpret the value of the correlation coefficient.

**(1)**

Helen was given the same data to analyse. In view of the large numbers involved she decided to divide the attendance figures by 100. She then calculated the product moment correlation coefficient between $\dfrac{x}{100}$ and *y*.

(*c*) Write down the value Helen should have obtained.

**(1)**

**May 2010**

**2.** An experiment consists of selecting a ball from a bag and spinning a coin. The bag contains 5 red balls and 7 blue balls. A ball is selected at random from the bag, its colour is noted and then the ball is returned to the bag.

When a red ball is selected, a biased coin with probability $\frac{2}{3}$ of landing heads is spun.

When a blue ball is selected a fair coin is spun.

(*a*) Copy and complete the tree diagram below to show the possible outcomes and associated probabilities.



**Ball**      **Coin**

Red — Heads / Tails

Blue — Heads / Tails
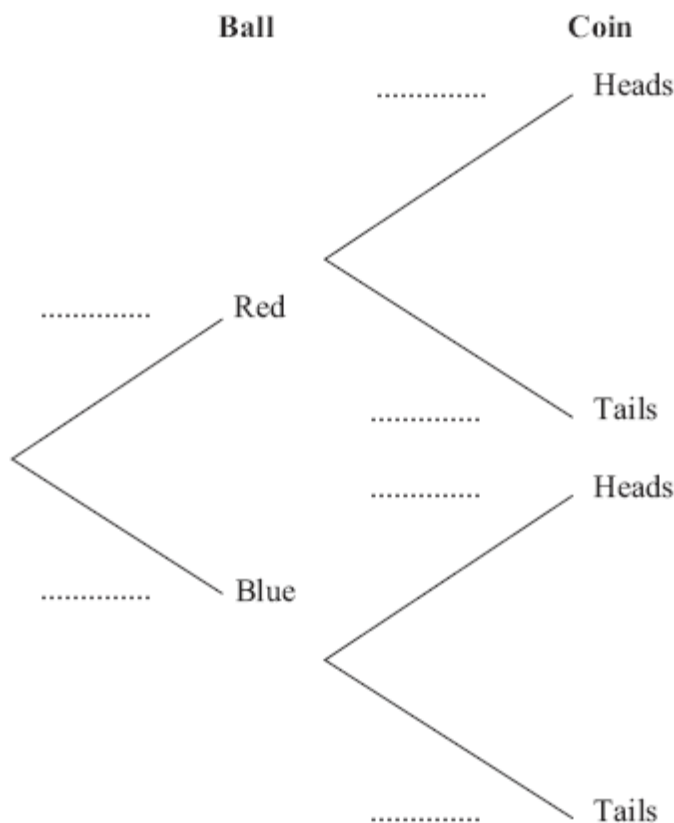
**(2)**

Shivani selects a ball and spins the appropriate coin.

(*b*) Find the probability that she obtains a head.

**(2)**

Given that Tom selected a ball at random and obtained a head when he spun the appropriate coin,

(*c*) find the probability that Tom selected a red ball.

**(3)**

Shivani and Tom each repeat this experiment.

(*d*) Find the probability that the colour of the ball Shivani selects is the same as the colour of the ball Tom selects.

**(3)**

**May 2010**

**3.** Past records show that the times, in seconds, taken to run 100 m by children at a school can be modelled by a normal distribution with a mean of 16.12 and a standard deviation of 1.60.

A child from the school is selected at random.

(*a*) Find the probability that this child runs 100 m in less than 15 s.

**(3)**

On sports day the school awards certificates to the fastest 30% of the children in the 100 m race.

(*b*) Estimate, to 2 decimal places, the slowest time taken to run 100 m for which a child will be awarded a certificate.

**(4)**

**May 2011**

**4.** A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

| Hours | 1–10 | 11–20 | 21–25 | 26–30 | 31–40 | 41–59 |
|---|---|---|---|---|---|---|
| Frequency | 6 | 15 | 11 | 13 | 8 | 3 |
| Mid-point | 5.5 | 15.5 | | 28 | | 50 |

(*a*) Find the mid-points of the 21−25 hour and 31−40 hour groups.

**(2)**

A histogram was drawn to represent these data. The 11−20 group was represented by a bar of width 4 cm and height 6 cm.

(*b*) Find the width and height of the 26−30 group.

**(3)**

(*c*) Estimate the mean and standard deviation of the time spent watching television by these students.

**(5)**

(*d*) Use linear interpolation to estimate the median length of time spent watching television by these students.

**(2)**

The teacher estimated the lower quartile and the upper quartile of the time spent watching television to be 15.8 and 29.3 respectively.

(*e*) State, giving a reason, the skewness of these data.

**(2)**

**May 2010**

**5.** The heights of an adult female population are normally distributed with mean 162 cm and standard deviation 7.5 cm.

(*a*) Find the probability that a randomly chosen adult female is taller than 150 cm.

**(3)**

Sarah is a young girl. She visits her doctor and is told that she is at the 60th percentile for height.

(*b*) Assuming that Sarah remains at the 60th percentile, estimate her height as an adult.

**(3)**

The heights of an adult male population are normally distributed with standard deviation 9.0 cm.

Given that 90% of adult males are taller than the mean height of adult females,

(*c*) find the mean height of an adult male.

**(4)**

**May 2012**

---

**6.** A packing plant fills bags with cement. The weight $X$ kg of a bag of cement can be modelled by a normal distribution with mean 50 kg and standard deviation 2 kg.

(*a*) Find P($X > 53$).

**(3)**

(*b*) Find the weight that is exceeded by 99% of the bags.

**(5)**

Three bags are selected at random.

(*c*) Find the probability that two weigh more than 53 kg and one weighs less than 53 kg.

**(4)**

**May 2008**

---

**7.**     The score $S$ when a spinner is spun has the following probability distribution.

| $s$ | 0 | 1 | 2 | 4 | 5 |
|---|---|---|---|---|---|
| $P(S = s)$ | 0.2 | 0.2 | 0.1 | 0.3 | 0.2 |

(*a*)  Find $E(S)$.

**(2)**

(*b*)  Show that $E(S^2) = 10.4$.

**(2)**

(*c*)  Hence find $Var(S)$.

**(2)**

(*d*)  Find

   (i)   $E(5S - 3)$,

   (ii)  $Var(5S - 3)$.

**(4)**

(*e*)  Find $P(5S - 3 > S + 3)$.

**(3)**

The spinner is spun twice.

The score from the first spin is $S_1$ and the score from the second spin is $S_2$.
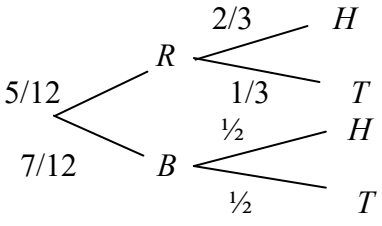
The random variables $S_1$ and $S_2$ are independent and the random variable $X = S_1 \times S_2$.

(*f*)   Show that $P(\{S_1 = 1\} \cap X < 5) = 0.16$.

**(2)**

(*g*)  Find $P(X < 5)$.

**(3)**

**May 2013 (R)**

**TOTAL FOR PAPER: 75 MARKS**

**END**

| Question Number | Scheme | Marks |
|---|---|---|
| **1. (a)** | $r = \dfrac{8825}{\sqrt{1022500 \times 130.9}}$, $\quad = $ awrt **0.763** | M1 A1 **(2)** |
| (b) | Teams with high attendance scored more goals (oe, statement in context) | B1 **(1)** |
| (c) | 0.76(3) | B1ft **(1)** **[4]** |
| **2. (a)** |  $\quad$ P($R$) and P($B$) $\quad$ 2$^{\text{nd}}$ set of probabilities | B1 B1 **(2)** |
| (b) | P($H$) = $\dfrac{5}{12} \times \dfrac{2}{3} + \dfrac{7}{12} \times \dfrac{1}{2}$, $= \dfrac{41}{72}$ or awrt 0.569 | M1 A1 **(2)** |
| (c) | P($R\mid H$) = $\dfrac{\frac{5}{12} \times \frac{2}{3}}{"\frac{41}{72}"}$, $= \dfrac{20}{41}$ or awrt 0.488 | M1 A1ft A1 **(3)** |
| (d) | $\left(\dfrac{5}{12}\right)^2 + \left(\dfrac{7}{12}\right)^2$ $= \dfrac{25}{144} + \dfrac{49}{144} = \dfrac{74}{144}$ or $\dfrac{37}{72}$ or awrt 0.514 | M1 A1ft A1 **(3)** **[10]** |
| **3. (a)** | $(z = \pm)\dfrac{15 - 16.12}{1.6} (= -0.70)$ $\quad$ P($Z$ < -0.70) = 1 - 0.7580 $\qquad = \underline{0.2420}$ $\qquad$ **(awrt 0.242)** | M1 M1 A1 **(3)** |
| (b) | [P($T < t$)=0.30 implies] $z = \dfrac{t - 16.12}{1.6} = -0.5244$ $\dfrac{t - 16.12}{1.6} = -0.5244 \Rightarrow t = 16.12 - 1.6 \times "0.5244"$ $t = $ awrt $\underline{15.28}$ (allow awrt 15.28/9) | M1 A1 M1 A1 **(4)** **[7]** |

| Question Number | Scheme | Marks |
|---|---|---|
| 4. (a) | 23,    35.5 | B1 B1 (2) |
| (b) | Width of 10 units is 4 cm so width of 5 units is **2 cm** | B1 |
| | Height = $2.6 \times 4 =$ **10.4 cm** | M1 A1 (3) |
| (c) | $\sum fx = 1316.5 \Rightarrow \bar{x} = \dfrac{1316.5}{56} =$   awrt **23.5** | M1 A1 |
| | $\sum fx^2 = 37378.25$  can be implied | B1 |
| | So $\sigma = \sqrt{\dfrac{37378.25}{56} - \bar{x}^2} =$  awrt **10.7**      allow $s = 10.8$ | M1 A1 (5) |
| (d) | $Q_2 = (20.5) + \dfrac{(28-21)}{11} \times 5 = 23.68\ldots$      **awrt 23.7 or 23.9** | M1 A1 (2) |
| (e) | $Q_3 - Q_2 = 5.6,\quad Q_2 - Q_1 = 7.9$    (or $\bar{x} < Q_2$) | M1 |
| | [7.9 >5.6 so ]    **negative skew** | A1 (2) **[14]** |
| 5. (a) | $[z =] \pm \left( \dfrac{150 - 162}{7.5} \right)$ | M1 |
| | $[z =] -1.6$ | A1 |
| | $\big[ P(F > 150) = P(Z > -1.6) = \big] = 0.9452(0071\ldots)$      **awrt 0.945** | A1 (3) |
| (b) | $z = \pm\, 0.2533$ (or better seen) | B1 |
| | $(\pm)\dfrac{s - 162}{7.5} = 0.2533\,(47\ldots)$ | M1 |
| | $s = 163.9$      **awrt 164** | A1 (3) |
| (c) | $z = \pm\, 1.2816$ (or better seen) | B1 |
| | $\dfrac{162 - \mu}{9} = -1.2815515\ldots$ | M1 A1 |
| | $\mu = 173.533\ldots$      **awrt 174** | A1 (4) **[10]** |

| Question Number | Scheme | Marks |
|---|---|---|
| **6.** (a) | $z = \dfrac{53 - 50}{2}$ | M1 |
| | $P(X > 53) = 1 - P(Z < 1.5)$ | B1 |
| | $= 1 - 0.9332$ | A1 |
| | | (3) |
| (b) | $P(X \leq x_0) = 0.01$ | M1 |
| | $\dfrac{x_0 - 50}{2} = -2.3263$ | M1 B1 |
| | $x_0 = 45.3474$         awrt 45.3 or 45.4 | M1 A1 |
| | | (5) |
| (c) | P(2 weigh more than 53kg and 1 less) $= 3 \times 0.0668^2 (1 - 0.0668)$ | B1 M1 A1ft |
| | $= 0.012492487..$      awrt 0.012 | A1 |
| | | (4) |
| | | **[12]** |
| **7.** (a) | $E(S) = 0 + 1 \times 0.2 + 2 \times 0.1 + 4 \times 0.3 + 5 \times 0.2 = [0.2 + 0.2 + 1.2 + 1.0]$ | M1 |
| | **2.6** | A1 |
| | | (2) |
| (b) | $E(S^2) = 0 + 1 \times 0.2 + 2^2 \times 0.1 + 4^2 \times 0.3 + 5^2 \times 0.2$   or   $0.2 + 0.4 + 4.8 + 5$ | M1 |
| | **10.4** (*) | A1cso |
| | | (2) |
| (c) | $Var(S) = 10.4 - \left("2.6"\right)^2$ | M1 |
| | **3.64**   or   $\frac{91}{25}$ (o.e.) | A1 |
| | | (2) |
| (d)(i) | $5E(S) - 3 = 5 \times "2.6" - 3$ ,       $= \underline{\mathbf{10}}$ | M1, A1 |
| (ii) | $5^2 Var(S) = 25 \times 3.64$,       $= \underline{\mathbf{91}}$ | M1, A1 |
| | | (4) |
| (e) | $5S - 3 > S + 3 \Rightarrow 4S > 6$   or   $S > 1.5$,     so need $P(S \geq 2)$ | M1, A1 |
| | $P(S \geq 2) = \underline{\mathbf{0.6}}$ | A1 |
| | | (3) |
| (f) | $P(S_1 = 1) \times P(S_2 \leq 4), = 0.2 \times 0.8 = 0.16$   (*) | M1, A1cso |
| | | (2) |
| (g) | $P(S_1 = 2) \times P(S_2 \leq 2) = 0.1 \times 0.5$      $= 0.05$ <br> $P(S_1 = 4) \times P(S_2 \leq 1) = 0.3 \times 0.4$      $= 0.12$        Full method – all cases listed | M1 |
| | $P(S_1 = 5) \times P(S_2 = 0) = 0.2 \times 0.2$      $= 0.04$ <br> $P(S_1 = 0) \times P(S_2 = \text{any value}) = 0.2 \times 1 = 0.20$      all correct products | A1 |
| | $= \underline{\mathbf{0.57}}$ | A1 |
| | | (3) |
| | | **[18]** |

## Examiner reports

### Question 1

Typically candidates successfully used the correct formula in order to calculate the product moment correlation coefficient in part (a). However, a number of candidates lost the accuracy mark by only giving a rounded answer to two decimal places. Providing an interpretation of their value of the correlation coefficient was less straightforward. Most frequently candidates made general remarks and described the correlation as positive without relating this to the context of the question. Of those who did attempt to provide an interpretation, many failed to appreciate that it was the attendance at the matches being compared to the total number of goals scored and not the number of home matches that were played.

Part (c) was answered well overall and correct answers were often justified by accompanying statements which indicated that linear coding does not affect the product moment correlation coefficient. Some candidates, however, seemed unaware of this fact and a common mistake was to divide their original product moment correlation coefficient by 100. In addition many candidates failed to recognise the significance of them being asked to write down their answer and chose to perform a full calculation in order to obtain the product moment correlation coefficient, which sometimes led to processing errors.

### Question 2

Overall there were very few errors made when candidates completed their tree diagrams. A small number of candidates repeated their probabilities of $\frac{2}{3}$ (for obtaining a head) and $\frac{1}{3}$ (for obtaining a tail) on the second branches for the fair coin. Occasionally the $\frac{5}{12}$ and $\frac{7}{12}$ probabilities were placed on the wrong branches and, in a few instances, quantities rather than probabilities were used. The vast majority of candidates were able to calculate the probability that Shivani selects a head correctly, or at least follow through the correct method from their tree diagrams, with few errors seen.

In contrast the quality of candidates' attempts at part (c) was extremely varied. Very few candidates quoted the correct formula despite it being given in the formula booklet, and of those who did, few realised that the numerator should be $\frac{5}{12} \times \frac{2}{3}$. The numerator was quite often seen as $\frac{5}{12}$ alone, and a number of candidates failed to recognise that their denominator should be their answer to part (b), leading in some cases to a repeated fraction in the numerator and denominator. P($H/R$) was sometimes calculated instead of P($R/H$).

The final part of the question was attempted fairly successfully overall. Indeed, many of the candidates who had erred in previous parts of the question were able to gain some credit, as most could identify at least one of $(\frac{5}{12})^2$ or $(\frac{7}{12})^2$. The special case pertaining to no replacement was occasionally seen.

### Question 3

Part (a) was tackled well by the majority of candidates although some were still unsure whether to give the answer as 0.7580 or 1 – 0.7580. There were a surprising number who mis-read the 16.12 value as 16.2 and lost the final accuracy mark.

In part (b) a number of candidates are still not using the table of percentage points to obtain the more accurate $z$ value of 0.5244 but, as usual this only incurred a one mark penalty. Those who standardised and equated to a $z$ value often failed to consider the sign of this value carefully and obtained the commonly seen answer of 16.96.

**Question 4**

Finding the midpoints of the given groups was predominantly carried out correctly with very few errors seen. In contrast, attempts at finding the width and height of the 26–30 group were extremely varied, with most candidates finding this particularly challenging, especially in finding the height. In the majority of cases, candidates obtained the wrong width and height, mostly with no clear strategy, although these did multiply together to make 20.8 in some cases. Calculation of the mean was carried out successfully on the whole, although there were some apparent misconceptions, with quite a few candidates merely summing the midpoints (without multiplying by the frequency) and dividing this by 56.

The standard deviation proved to be more problematic, with frequent mistakes in both the formula and in their calculations. Some candidates used the sum of the $f^2x$'s and others the sum of the $(fx)^2$ or the sum of the $fx$'s all squared in their formula. Very few candidates calculated $s$. Most candidates were able to use the correct interpolation technique to obtain the median, although many lost the accuracy mark through their use of 21 as the lower class boundary (which was relatively common) and /or 4 as the class width. Quite a few candidates worked with 28.5. A few candidates tried to apply the correct formula to the wrong class interval, however. Some candidates appeared to have a limited understanding of the class boundaries and failed to recognise the continuous nature of the data.

The majority of candidates were able to carry out a suitable test to determine the skewness of the data correctly. This mostly involved comparing $Q_3 - Q_2$ to $Q_2 - Q_1$ (with or without explicit substitutions), although the wrong conclusion was often drawn, either following on from a previous error in evaluating the median or from a lack of understanding of what their result was showing. A few students evaluated 3(mean – median)/standard deviation. Quite often the result of their test was described in words not figures, for example $Q_2$ is closer to $Q_3$ than $Q_1$. Some candidates merely attempted to describe the skewness without carrying out any test.

**Question 5**

Most candidates now can standardise with confidence but there is still some confusion when using the tables for negative $z$ values and a number of candidates are still not clear about the difference between a probability (area under the curve) and a $z$ value (the point on the horizontal axis). The notation is still not handled well and statements such as "$P\left(Z > \dfrac{150-162}{7.5}\right) = -1.6$" or "$0.4 = 0.2533$" are often seen as candidates stumble towards a correct answer.

In part (a) the majority gained the first 2 marks for standardising and arriving at a correct value for $z$ but many lost the final mark for subtracting 0.9452 from 1. A clearly labelled diagram may have helped them avoid this error.

In part (b) some were confused by the reference to 60th percentile and simply found 60% of 162. Whilst many could standardise correctly equations such as $\dfrac{s-162}{7.5} = 0.6$ were quite common. Those who did realise that $\dfrac{s-162}{7.5}$ should be set equal to a $z$ value often used 0.25 rather than the more accurate 0.2533 available from the table of percentage points.

Similar problems arose in part (c) and there was the added complication of obtaining compatible signs for a correct equation. Some candidates made double sign errors when solving $\frac{162 - \mu}{9} = 1.2816$ to arrive at an answer of 174 and this lost the accuracy marks as their answer did not follow from their equation. The use of $\pm$ 1.28 was condoned here if the candidates had already lost a mark for using 0.25 rather than 0.2533 in part (b).

## Question 6

Part (a) was answered with the highest degree of success with all but the weakest students not gaining 3 marks. Many candidates incorrectly interpreted the sign of the inequality in part (b) and went on to calculate a score above the mean. Many of those who arrived at an answer got circa 54 kg and failed to consider the reasonableness of their answer i.e. if the mean is 50kg, then 99% of the packets cannot weigh more than 54kg. Candidates often did not use the percentage points table quoting 2.33 instead of 2.3263 and this was reflected in the very small proportion of B marks awarded in part (b). Few candidates equated their standardised equation to a z-score with a consistent sign. Part (c) really sorted out those who really understood what was going on and who were hazily following the rules. Many failed to recognize that they only required the answer from part (a), a marked number specifically calculating the individual probabilities anew. The number who recognized that there was a factor of three involved was small although a significant number scored 2 marks by using $p^2(1 - p)$. Quite marked was the significant number who multiplied 0.0688 by two rather than squaring it. There was also a fair number who added the probabilities rather than multiplying.

## Question 7

Parts (a), (b), (c) and (d) were answered very well and most candidates scored well here. In part (e) most knew they had to solve the inequality and arrived at $S > 1.5$ but some misinterpreted this and gave $P(S > 1)$ rather than $P(S > 2)$. The final 2 parts caused problems for many candidates. A popular approach was to calculate a sample space showing all 25 products of $S_1$ and $S_2$ but they often thought that these 25 cases were all equally likely.

Unfortunately this false assumption would lead to the correct answer in part (f), but not of course the available marks, and many candidates never properly considered the correct distribution of $X$. Some candidates did list the 15 required cases and were able to obtain the correct probability but such cases were rare.

## Statistics for S1 Practice Paper Gold Level G1

| | | | | Mean score for students achieving grade: | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Qu | Max Score | Modal score | Mean % | ALL | A* | A | B | C | D | E | U |
| 1 | 4 | | 67 | 2.68 | 3.54 | 3.32 | 2.86 | 2.59 | 2.34 | 2.18 | 1.82 |
| 2 | 10 | | 64 | 6.42 | 9.23 | 8.55 | 7.12 | 6.18 | 5.28 | 4.63 | 3.50 |
| 3 | 7 | | 55 | 3.83 | 6.26 | 5.81 | 4.84 | 3.85 | 2.78 | 1.83 | 0.70 |
| 4 | 14 | | 60 | 8.44 | 12.76 | 11.75 | 9.80 | 8.16 | 6.60 | 5.13 | 3.45 |
| 5 | 10 | | 51 | 5.12 | 8.99 | 8.23 | 6.27 | 4.76 | 3.55 | 2.67 | 1.26 |
| 6 | 12 | | 45 | 5.34 | | 8.29 | 5.98 | 4.81 | 3.70 | 2.71 | 1.33 |
| 7 | 18 | | 71 | 12.73 | 15.76 | 14.35 | 12.75 | 11.34 | 10.36 | 8.91 | 6.58 |
| | 75 | | 59 | 44.56 | | 60.30 | 49.62 | 41.69 | 34.61 | 28.06 | 18.64 |