# Edexcel GCE

# Statistics S1 Bronze Level B1

# (Question Paper)

Paper Reference(s)

# 6683/01

# Edexcel GCE

## Statistics S1
## Bronze Level B1

## 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 |
|---|---|---|---|---|---|
| 74 | 68 | 62 | 55 | 48 | 43 |

**1.** A random sample of 50 salmon was caught by a scientist. He recorded the length $l$ cm and weight $w$ kg of each salmon.

The following summary statistics were calculated from these data.

$$\sum l = 4027 \qquad \sum l^2 = 327\ 754.5 \qquad \sum w = 357.1 \qquad \sum lw = 29\ 330.5 \qquad S_{ww} = 289.6$$

(*a*)  Find $S_{ll}$ and $S_{lw}$ .

**(3)**

(*b*)  Calculate, to 3 significant figures, the product moment correlation coefficient between $l$ and $w$.

**(2)**

(*c*)  Give an interpretation of your coefficient.

**(1)**

**January 2011**

---

**2.** The 19 employees of a company take an aptitude test. The scores out of 40 are illustrated in the stem and leaf diagram below.

```
          2 | 6 means a score of 26

     0 | 7                    (1)
     1 | 88                   (2)
     2 | 4468                 (4)
     3 | 2333459              (7)
     4 | 00000                (5)
```

Find

(*a*)  the median score,

**(1)**

(*b*)  the interquartile range.

**(3)**

The company director decides that any employees whose scores are so low that they are outliers will undergo retraining.

An outlier is an observation whose value is less than the lower quartile minus 1.0 times the interquartile range.

(*c*)  Explain why there is only one employee who will undergo retraining.

**(2)**

(*d*)  Draw a box plot to illustrate the employees' scores.

**(3)**

**January 2010**

---

2

**3.** The discrete random variable $X$ can take only the values 2, 3, 4 or 6. For these values the probability distribution function is given by

| $x$ | 2 | 3 | 4 | 6 |
|---|---|---|---|---|
| P($X = x$) | $\dfrac{5}{21}$ | $\dfrac{2k}{21}$ | $\dfrac{7}{21}$ | $\dfrac{k}{21}$ |

where $k$ is a positive integer.

(*a*) Show that $k = 3$.

(2)

Find

(*b*) F(3),

(1)

(*c*) E($X$),

(2)

(*d*) E($X^2$),

(2)

(*e*) Var ($7X - 5$).

(4)

**January 2012**

3

**4.** In a study of how students use their mobile telephones, the phone usage of a random sample of 11 students was examined for a particular week.

The total length of calls, $y$ minutes, for the 11 students were

$$17, 23, 35, 36, 51, 53, 54, 55, 60, 77, 110$$

(*a*) Find the median and quartiles for these data.

**(3)**

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(*b*) Show that 110 is the only outlier.

**(2)**

(*c*) Draw a box plot for these data indicating clearly the position of the outlier.

**(3)**

The value of 110 is omitted.

(*d*) Show that $S_{yy}$ for the remaining 10 students is 2966.9

**(3)**

These 10 students were each asked how many text messages, $x$, they sent in the same week. The values of $S_{xx}$ and $S_{xy}$ for these 10 students are $S_{xx} = 3463.6$ and $S_{xy} = -18.3$.

(*e*) Calculate the product moment correlation coefficient between the number of text messages sent and the total length of calls for these 10 students.

**(2)**

A parent believes that a student who sends a large number of text messages will spend fewer minutes on calls.

(*f*) Comment on this belief in the light of your calculation in part (*e*).

**(1)**

**January 2009**

**5.** A person's blood group is determined by whether or not it contains any of 3 substances $A$, $B$ and $C$.

A doctor surveyed 300 patients' blood and produced the table below.

| Blood contains | No. of Patients |
|---|---|
| only $C$ | 100 |
| $A$ and $C$ but not $B$ | 100 |
| only $A$ | 30 |
| $B$ and $C$ but not $A$ | 25 |
| only $B$ | 12 |
| $A$, $B$ and $C$ | 10 |
| $A$ and $B$ but not $C$ | 3 |

(*a*)  Draw a Venn diagram to represent this information.

**(4)**

(*b*)  Find the probability that a randomly chosen patient's blood contains substance $C$.

**(2)**

Harry is one of the patients. Given that his blood contains substance $A$,

(*c*)  find the probability that his blood contains all 3 substances.

**(2)**

Patients whose blood contains none of these substances are called universal blood donors.

(*d*)  Find the probability that a randomly chosen patient is a universal blood donor.

**(2)**

**May 2008**

**6.** The following shows the results of a survey on the types of exercise taken by a group of 100 people.

> 65 run
> 48 swim
> 60 cycle
> 40 run and swim
> 30 swim and cycle
> 35 run and cycle
> 25 do all three

(*a*) Draw a Venn Diagram to represent these data.

**(4)**

Find the probability that a randomly selected person from the survey

(*b*) takes none of these types of exercise,

**(2)**

(*c*) swims but does not run,

**(2)**

(*d*) takes at least two of these types of exercise.

**(2)**

Jason is one of the above group.

Given that Jason runs,

(*e*) find the probability that he swims but does not cycle.

**(3)**

**January 2012**

**7.** A teacher took a random sample of 8 children from a class. For each child the teacher recorded the length of their left foot, $f$ cm, and their height, $h$ cm. The results are given in the table below.

| $f$ | 23 | 26 | 23 | 22 | 27 | 24 | 20 | 21 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $h$ | 135 | 144 | 134 | 136 | 140 | 134 | 130 | 132 |

(You may use $\sum f = 186$    $\sum h = 1085$    $S_{ff} = 39.5$    $S_{hh} = 139.875$    $\sum fh = 25\,291$)

(a) Calculate $S_{fh}$.

**(2)**

(b) Find the equation of the regression line of $h$ on f in the form $h = a + bf$.
Give the value of $a$ and the value of $b$ correct to 3 significant figures.

**(5)**

(c) Use your equation to estimate the height of a child with a left foot length of 25 cm.

**(2)**

(d) Comment on the reliability of your estimate in part (c), giving a reason for your answer.

**(2)**

The left foot length of the teacher is 25 cm.

(e) Give a reason why the equation in part (b) should not be used to estimate the teacher's height.

**(1)**

**May 2011**

**TOTAL FOR PAPER: 75 MARKS**

**END**