Write your name here

| Surname | Other names |
|---|---|
| | |

**Pearson Edexcel**
**International**
**Advanced Level**

Centre Number | | | | | |

Candidate Number | | | | |

# Statistics S1

## Advanced/Advanced Subsidiary

Thursday 15 January 2015 – Afternoon
**Time: 1 hour 30 minutes**

Paper Reference
**WST01/01**

**You must have:**
Mathematical Formulae and Statistical Tables (Blue)

Total Marks

**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 formulae stored in them.**

## Instructions

- Use **black** ink or ball-point pen.
- If pencil is used for diagrams/sketches/graphs it must be dark (HB or B). Coloured pencils and highlighter pens must not be used.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions and ensure that your answers to parts of questions are clearly labelled.
- Answer the questions in the spaces provided – *there may be more space than you need.*
- You should show sufficient working to make your methods clear. Answers without working may not gain full credit.
- Values from the statistical tables should be quoted in full. When a calculator is used, the answer should be given to an appropriate degree of accuracy.

## Information

- The total mark for this paper is 75.
- The marks for **each** question are shown in brackets – *use this as a guide as to how much time to spend on each question.*

## Advice

- Read each question carefully before you start to answer it.
- Try to answer every question.
- Check your answers if you have time at the end.

*Turn over* ▶

P45064A



5/5/5/5/

P 4 5 0 6 4 A 0 1 2 4

PEARSON

Leave blank

**1.** The discrete random variable $X$ has probability function $p(x)$ and cumulative distribution function $F(x)$ given in the table below.

| $x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $p(x)$ | 0.10 | $a$ | 0.28 | $c$ | 0.24 |
| $F(x)$ | 0.10 | 0.26 | $b$ | 0.76 | $d$ |

(a) Write down the value of $d$

**(1)**

(b) Find the values of $a$, $b$ and $c$

**(3)**

(c) Write down the value of $P(X > 4)$

**(1)**

Two independent observations, $X_1$ and $X_2$, are taken from the distribution of $X$.

(d) Find the probability that $X_1$ and $X_2$ are both odd.

**(2)**

Given that $X_1$ and $X_2$ are both odd,

(e) find the probability that the sum of $X_1$ and $X_2$ is 6
Give your answer to 3 significant figures.

**(3)**

Leave blank

**Question 1 continued**

**(Total 10 marks)**

**Q1**

Leave blank

**2.** A sports teacher recorded the number of press-ups done by his students in two minutes. He recorded this information for a Year 7 class and for a Year 11 class.

The back-to-back stem and leaf diagram shows this information.

| Totals | Year 7 class | | Year 11 class | Totals |
|---|---|---|---|---|
| (6) | 8 7 6 5 5 4 | 1 | | |
| (10) | 9 7 7 6 5 4 4 4 2 2 | 2 | 0 5 6 9 | (4) |
| (7) | 8 7 5 4 3 3 0 | 3 | 3 4 5 8 8 | (5) |
| (5) | 9 9 7 2 2 | 4 | 0 5 6 7 9 | (5) |
| (3) | 8 4 0 | 5 | 0 3 5 5 6 6 7 7 7 9 9 | (11) |
| | | 6 | 0 3 3 3 3 4 8 | (7) |

Key: 2|4|0 means 42 press-ups for a Year 7 student and 40 press-ups for a Year 11 student

(a) Find the median number of press-ups for each class.

**(2)**

For the Year 11 class, the lower quartile is 38 and the upper quartile is 59

(b) Find the lower quartile and the upper quartile for the Year 7 class.

**(2)**

(c) Use the medians and quartiles to describe the skewness of each of the two distributions.

**(3)**

(d) Give two reasons why the normal distribution should not be used to model the number of press-ups done by the Year 11 class.

**(2)**

Leave blank

**Question 2 continued**

**Q2**

**(Total 9 marks)**

Leave blank

**3.** The table shows the price of a bottle of milk, $m$ pence, and the price of a loaf of bread, $b$ pence, for 8 different years.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $m$ | 29 | 29 | 35 | 39 | 41 | 43 | 44 | 46 |
| $b$ | 75 | 83 | 91 | 121 | 120 | 126 | 119 | 126 |

(You may use $S_{bb} = 3083.875$ and $S_{mm} = 305.5$)

(a) Find the exact value of $\sum bm$

**(1)**

(b) Find $S_{bm}$

**(3)**

(c) Calculate the product moment correlation coefficient between $b$ and $m$

**(2)**

(d) Interpret the value of the correlation coefficient.

**(1)**

A ninth year is added to the data set. In this year the price of the bottle of milk is 46 pence and the price of a loaf of bread is 175 pence.

(e) Without further calculation, state whether the value of the product moment correlation coefficient will increase, decrease or stay the same when all nine years are used. Give a reason for your answer.

**(2)**

Leave blank

**Question 3 continued**

Leave blank

**Question 3 continued**

Leave blank

**Question 3 continued**

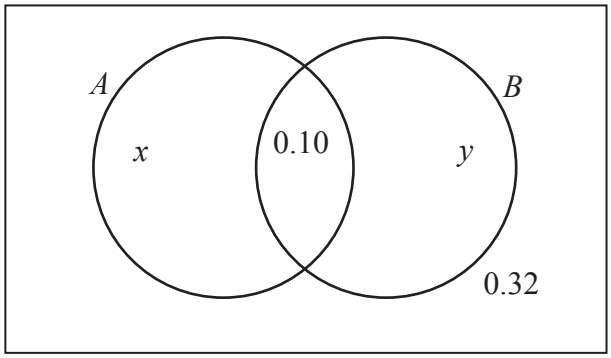
**Q3**

**(Total 9 marks)**

Leave blank

**4.** Events $A$ and $B$ are shown in the Venn diagram below

where $x$, $y$, 0.10 and 0.32 are probabilities.




(a) Find an expression in terms of $x$ for

(i) P($A$)

(ii) P($B \mid A$)

**(3)**

(b) Find an expression in terms of $x$ and $y$ for P($A \cup B$)

**(1)**

Given that P($A$) = 2P($B$)

(c) find the value of $x$ and the value of $y$

**(5)**

Leave blank

**Question 4 continued**

Leave blank

**Question 4 continued**

Leave blank

**Question 4 continued**

**Q4**

**(Total 9 marks)**

Leave blank

**5.** The resting heart rate, $h$ beats per minute (bpm), and average length of daily exercise, $t$ minutes, of a random sample of 8 teachers are shown in the table below.

| $t$ | 20 | 35 | 40 | 25 | 45 | 70 | 75 | 90 |
|---|---|---|---|---|---|---|---|---|
| $h$ | 88 | 85 | 77 | 75 | 71 | 66 | 60 | 54 |

(a) State, with a reason, which variable is the response variable.

**(2)**

The equation of the least squares regression line of $h$ on $t$ is

$$h = 93.5 - 0.43t$$

(b) Give an interpretation of the gradient of this regression line.

**(1)**

(c) Find the value of $\bar{t}$ and the value of $\bar{h}$

**(2)**

(d) Show that the point $(\bar{t}, \bar{h})$ lies on the regression line.

**(1)**

(e) Estimate the resting heart rate of a teacher with an average length of daily exercise of 1 hour.

**(1)**

(f) Comment, giving a reason, on the reliability of the estimate in part (e).

**(2)**

The resting heart rate of teachers is assumed to be normally distributed with mean 73 bpm and standard deviation 8 bpm.

The middle 95% of resting heart rates of teachers lies between $a$ and $b$

(g) Find the value of $a$ and the value of $b$.

**(4)**

Leave blank

**Question 5 continued**

Leave blank

**Question 5 continued**

Leave blank

**Question 5 continued**

**(Total 13 marks)**

**Q5**

Leave blank

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

$$P(X = x) = \frac{x^2}{k} \qquad x = 1, 2, 3, 4$$

(a) Show that $k = 30$

**(2)**

(b) Find $P(X \neq 4)$

**(2)**

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

**(2)**

(d) Find the exact value of $\text{Var}(X)$

**(4)**

Given that $Y = 3X - 1$

(e) find $E(Y^2)$

**(4)**

Leave blank

**Question 6 continued**

Leave blank

**Question 6 continued**

Leave blank

**Question 6 continued**

**Q6**

**(Total 14 marks)**

Leave blank

**7.** The birth weights, $W$ grams, of a particular breed of kitten are assumed to be normally distributed with mean 99 g and standard deviation 3.6 g

(a) Find $P(W > 92)$

**(3)**

(b) Find, to one decimal place, the value of $k$ such that $P(W < k) = 3P(W > k)$

**(4)**

(c) Write down the name given to the value of $k$.

**(1)**

For a different breed of kitten, the birth weights are assumed to be normally distributed with mean 120 g

Given that the 20th percentile for this breed of kitten is 116 g

(d) find the standard deviation of the birth weight of this breed of kitten.

**(3)**

Leave blank

**Question 7 continued**

Leave blank

**Question 7 continued**

**Q7**

**(Total 11 marks)**

**TOTAL FOR PAPER: 75 MARKS**

**END**