## Statistics S1

# Advanced/Advanced Subsidiary 

# Friday 5 June 2015 - Morning 

Time: 1 hour 30 minutes

Materials required for examination<br>Mathematical Formulae (Pink)<br>Items included with question papers<br>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

In the boxes on the answer book, write the name of the examining body (Edexcel), your centre number, candidate number, the unit title (Statistics S1), the paper reference (6683), your surname, other name and signature.
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 for Candidates

A booklet 'Mathematical Formulae and Statistical Tables’ is provided.
Full marks may be obtained for answers to ALL questions.
This paper has 6 questions.
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 not gain full credit.

1. Each of 60 students was asked to draw a $20^{\circ}$ angle without using a protractor. The size of each angle drawn was measured. The results are summarised in the box plot below.

(a) Find the range for these data.
(b) Find the interquartile range for these data.

The students were then asked to draw a $70^{\circ}$ angle.
The results are summarised in the table below.

| Angle, $\boldsymbol{a}$, (degrees) | Number of students |
| :---: | :---: |
| $55 \leq a<60$ | 6 |
| $60 \leq a<65$ | 15 |
| $65 \leq a<70$ | 13 |
| $70 \leq a<75$ | 11 |
| $75 \leq a<80$ | 8 |
| $80 \leq a<85$ | 7 |

(c) Use linear interpolation to estimate the size of the median angle drawn. Give your answer to 1 decimal place.
(d) Show that the lower quartile is $63^{\circ}$.

For these data, the upper quartile is $75^{\circ}$, the minimum is $55^{\circ}$ and the maximum is $84^{\circ}$.
An outlier is an observation that falls either more than $1.5 \times$ (interquartile range) above the upper quartile or more than $1.5 \times$ (interquartile range) below the lower quartile.
(e) (i) Show that there are no outliers for these data.
(ii) On graph paper, draw a box plot for these data.
(f) State which angle the students were more accurate at drawing. Give reasons for your answer.
2. An estate agent recorded the price per square metre, $p £ / \mathrm{m}^{2}$, for 7 two-bedroom houses. He then coded the data using the coding $q=\frac{p-a}{b}$, where $a$ and $b$ are positive constants. His results are shown in the table below.

| $p$ | 1840 | 1848 | 1830 | 1824 | 1819 | 1834 | 1850 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $q$ | 4.0 | 4.8 | 3.0 | 2.4 | 1.9 | 3.4 | 5.0 |

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

The estate agent also recorded the distance, $d \mathrm{~km}$, of each house from the nearest train station. The results are summarised below.

$$
\mathrm{S}_{d d}=1.02 \quad \mathrm{~S}_{q q}=8.22 \quad \mathrm{~S}_{d q}=-2.17
$$

(b) Calculate the product moment correlation coefficient between $d$ and $q$.
(c) Write down the value of the product moment correlation coefficient between $d$ and $p$.

The estate agent records the price and size of 2 additional two-bedroom houses, $H$ and $J$.

| House | Price (£) | Size $\left(\mathrm{m}^{2}\right)$ |
| :---: | :---: | :---: |
| $H$ | 156400 | 85 |
| $J$ | 172900 | 95 |

(d) Suggest which house is most likely to be closer to a train station. Justify your answer.
3. A college has 80 students in Year 12.

20 students study Biology.
28 students study Chemistry.
30 students study Physics.
7 students study both Biology and Chemistry.
11 students study both Chemistry and Physics.
5 students study both Physics and Biology.
3 students study all 3 of these subjects.
(a) Draw a Venn diagram to represent this information.

A Year 12 student at the college is selected at random.
(b) Find the probability that the student studies Chemistry but not Biology or Physics.
(c) Find the probability that the student studies Chemistry or Physics or both.

Given that the student studies Chemistry or Physics or both,
(d) find the probability that the student does not study Biology.
(e) Determine whether studying Biology and studying Chemistry are statistically independent.
4. Statistical models can provide a cheap and quick way to describe a real world situation.
(a) Give two other reasons why statistical models are used.

A scientist wants to develop a model to describe the relationship between the average daily temperature, $x^{\circ} \mathrm{C}$, and her household's daily energy consumption, $y \mathrm{kWh}$, in winter.

A random sample of the average daily temperature and her household's daily energy consumption are taken from 10 winter days and shown in the table.

| $x$ | -0.4 | -0.2 | 0.3 | 0.8 | 1.1 | 1.4 | 1.8 | 2.1 | 2.5 | 2.6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $y$ | 28 | 30 | 26 | 25 | 26 | 27 | 26 | 24 | 22 | 21 |

[You may use $\sum x^{2}=24.76 \quad \sum y=255 \quad \sum \sum x y=283.8 \quad \mathrm{~S}_{x x}=10.36$ ]
(b) Find $\mathrm{S}_{x y}$ for these data.
(c) Find the equation of the regression line of $y$ on $x$ in the form $y=a+b x$.

Give the value of $a$ and the value of $b$ to 3 significant figures.
(d) Give an interpretation of the value of $a$.
(e) Estimate her household's daily energy consumption when the average daily temperature is $2^{\circ} \mathrm{C}$.

The scientist wants to use the linear regression model to predict her household's energy consumption in the summer.
(f) Discuss the reliability of using this model to predict her household's energy consumption in the summer.
5. In a quiz, a team gains 10 points for every question it answers correctly and loses 5 points for every question it does not answer correctly. The probability of answering a question correctly is 0.6 for each question. One round of the quiz consists of 3 questions.

The discrete random variable $X$ represents the total number of points scored in one round. The table shows the incomplete probability distribution of $X$.

| $x$ | 30 | 15 | 0 | -15 |
| :--- | :---: | :---: | :---: | :---: |
| $\mathrm{P}(X=x)$ | 0.216 |  |  | 0.064 |

(a) Show that the probability of scoring 15 points in a round is 0.432 .
(b) Find the probability of scoring 0 points in a round.
(c) Find the probability of scoring a total of 30 points in 2 rounds.
(d) Find E( $X$ ).
(e) Find $\operatorname{Var}(X)$.

In a bonus round of 3 questions, a team gains 20 points for every question it answers correctly and loses 5 points for every question it does not answer correctly.
(f) Find the expected number of points scored in the bonus round.
6. The random variable $Z \sim N(0,1)$.
$A$ is the event $Z>1.1$
$B$ is the event $Z>-1.9$
$C$ is the event $-1.5<Z<1.5$
(a) Find
(i) $\mathrm{P}(A)$,
(ii) $\mathrm{P}(B)$,
(iii) $\mathrm{P}(C)$,
(iv) $\mathrm{P}(A \cup C)$.

The random variable $X$ has a normal distribution with mean 21 and standard deviation 5 .
(b) Find the value of $w$ such that $\mathrm{P}(X>w \mid X>28)=0.625$.


| Question | Scheme | Marks |
| :---: | :---: | :---: |
| 2. (a) | $\begin{array}{ll} \frac{1840-a}{b}=4.0 & \underline{1848-a} \\ a=\underline{\mathbf{1 8 0 0}} & b=\underline{\mathbf{1 0}} \end{array}$ | $\begin{aligned} & \text { M1 } \\ & \text { A1 } \end{aligned}$ |
| (b) | $r=\frac{-2.17}{\sqrt{1.02 \times 8.22}}=-0.749417343 \ldots . \quad \text { awrt }-\underline{\mathbf{0 . 7 4 9}}$ | M1A1 |
| (c) | -0.749 | (2) B1ft |
|  | House H: 156400/85 $=\left[£ 1840 / \mathrm{m}^{2}\right.$ or $\left.q=4\right]$ <br> House J: 172900/95 = [£1820/m $\mathrm{m}^{2}$ or $\left.q=2\right]$ | (1) M1 |
|  | Since ( $r=-0.749$,) there is negative correlation. or <br> The higher the price (per square metre), the lower the distance from the train station. | $\mathrm{dM} 1$ A1 |
|  |  | $\begin{array}{\|c} (3) \\ (8 \mathrm{marks}) \\ \hline \end{array}$ |
|  | Notes |  |
| (a) | M1 for setting up two suitable equations which could lead to $a$ and $b$ (may be implied by one correct answer) <br> A1 for $a=1800$ and $b=10$ ( $a=10$ and $b=1800$ is A0) Correct answer only is $2 / 2$ |  |
| (b) | M1 for a correct expression (condone missing -) <br> A1 for awrt -0.749 <br> ( -0.75 or awrt 0.749 with no working scores M1 A0). |  |
| (c) | B1ft for -0.749 or ft their answer to (b) to at least 2sf. Must be in the range | <'(b)'<1 |
| (d) | M1 for calculating price/square metre for both $H$ and $J$. Can be implied by sight of 1840 and 1820 (so OK if not labelled or mis-labelled) |  |
| $r>0$ | dM1 dependent on $1^{\text {st }}$ M1 for a statement that correlation is negative or a contextualised interpretation of the negative correlation. If $r>0$ allow equivalent statements about positive correlation A1 (dependent on both Ms) for House H is likely to be closer (No ft if $r>0$ | 0) |


| Question | Scheme | Marks |
| :---: | :---: | :---: |
| 3. (a) |  | B1 M1 A1 A1 B1 |
| (b) | $\frac{13 '}{80} \quad \underline{\text { or }} 0.1625$ | B1ft |
| (c) | $\frac{28+30-11}{80}$ or $\frac{2+3+4+8+13+17}{80}$ or $1-\frac{(11+22)}{80}=\frac{47}{80}$ or 0.5875 | M1 A1 |
| (d) | $\frac{" 17+8+13 "}{44 "}$ or $\frac{\frac{" 38 "}{80}}{\frac{877 " 1}{80}}$ or $1-\frac{" 2+3+4 "}{447 "}=\frac{38}{47}$ (condone awrt 0.809 ) | M1 A1cao |
| (e) | $\begin{aligned} & \mathrm{P}(B \mid C)=\frac{7}{28}, \mathrm{P}(B)=\frac{20}{80} \\ & \mathrm{P}(C \mid B)=\frac{7}{20}, \mathrm{P}(C)=\frac{28}{80} \\ & \mathrm{P}(B \cap C)=\frac{7}{80}, \mathrm{P}(B)=\frac{20}{80} \mathrm{P}(C)=\frac{28}{80} \end{aligned}$ <br> $\mathrm{P}(B \mid C)=\mathrm{P}(B), \mathrm{P}(C \mid B)=\mathrm{P}(C)$ these may be implied by correct conclusion $\mathrm{P}(B \cap C)=\mathrm{P}(B) \times \mathrm{P}(C)$ this approach requires the product to be seen So, they are independent. | (2) <br> M1 <br> M1 <br> A1 (3) <br> (13 marks) |
|  | Notes |  |
| (a) | B1 for 3 intersecting circles with 3 in the centre. Allow probs. or integers in diagram. M1 for some correct subtraction e.g. at least one of $2,4,8$ or for $B: 20-\operatorname{their}(2+3+4)$ etc A1 for 2,4 and 8 (ignore labels) <br> A1 for 11, 13 and 17 (must be in compatible regions with 2, 4, 8 if no labels) <br> B1 for correct labels and 22 and box (Do not treat "blank" as 0 so can't use 0 for ft in (c)) <br> M1 for a correct expression seen in (c) ( or ft their diagram). Correct ans M1A1 <br> M1 for denominator of 47 or ft their numerator from part (c) and numerator of 38 or their $(17+8+13)$ or (their 47$)-$ their $(2+3+4)$. Correct ans M1A1 <br> M1 for stating at least the required probs.\& labelled for a correct test (can ft their diagram) M1 for use of a correct test with $B$ and $C$ Must see product attempted for $\mathrm{P}(B \cap C)$ test. A1 for a correct test with all probabilities correct and a correct concluding statement. NB M0M1A0 should be possible but A1 requires both Ms |  |
|  |  |  |
| (e) |  |  |


| Question | Scheme | Marks |
| :---: | :---: | :---: |
| 4. (a) | To simplify (or represent) a real world problem (o.e.) <br> To improve understanding (o.e.) <br> To analyse a real world problem or can change variables/replicate easily (oe) To make predictions or find estimates (o.e.) | $\begin{array}{\|l\|l} \text { B1g } \\ \text { B1h } \end{array}$ |
| (b) |  | B1 |
|  | $S_{x y}=283.8-\frac{12 ' \times 255}{10}, \quad=-\underline{\mathbf{2 2 . 2}}$ | M1,A1cao |
| (c) | $\begin{aligned} & \left.b=\frac{'^{-22.2 '}}{10.36}=,-2.142857 \ldots . \quad \text { (A1 for awrt }-2.1\right) \\ & {[a=\bar{y}-b \bar{x} \quad \Rightarrow] a=\frac{255}{10}-{ }^{\prime} b \cdot \times \frac{" 12 "}{10}=28.07143} \end{aligned}$ | $\begin{aligned} & \text { M1A1 } \\ & \text { M1 } \end{aligned}$ |
|  | $y=28.1-2.14 x$ <br> [Condone: $y=28.1+-2.14 x]$ | A1 |
| (d) | (28 | B1 |
| (e) | $y=$ | $\begin{array}{ll}  & \text { (1) } \\ \text { M1 } \\ \text { A1 } \end{array}$ |
| (f) |  |  |
|  | The regression model is based on temperatures from the winter, so not reliable in the summer. <br> Stating it is reliable (whatever the reason) is B0B0 | B1 <br> dB1 <br> (2) <br> (14 marks) |
|  | Notes |  |
| (a) | Make sure reasons refer to models and not tests <br> $1^{\text {st }} \mathrm{B} 1 \mathrm{~g}$ (be fairly generous) for a sensible reason not using "quick", "cheap" or "describe" $2^{\text {nd }} \mathrm{B} 1 \mathrm{~h}$ (be slightly harder) for two convincing reasons (both based on the list above) Use professional judgement and mark as B0B0 or B1B0 or B1B1 do not use B0B1 |  |
| (b) | B1 for $\sum x=12$ (May be by the table) (Can be implied by 3060 seen or the M1 for attempt at correct formula (ft their $\sum x$ where $10<\sum x<14$ ) A1 for -22.2 only | next line) |
| (c) | M1 for a correct expression for $b$ (ft their $\mathrm{S}_{x y} \neq 283.8$ ) <br> A1 for awrt - 2.1 (allow -15/7) <br> M1 for a correct expression for $a$ and ft their 12 (allow use of a letter $b$ ) A1 for $y=28.1-2.14 x$ (awrt 28.1 and awrt -2.14) Must be $y$ and $x$ and | actions |
| (d) | B1 for a contextualised interpretation e.g. the amount of energy used when temperature is $0\left[{ }^{\circ} \mathrm{C}\right.$ ] or [28.1] kWh used when temp. is $0\left[{ }^{\circ} \mathrm{C}\right.$ ] [Can ft their 28.1]Need temp or ${ }^{\circ}$ sign [B0 for "value of $y$ when $x=0$ " since no context in words] |  |
| (e) | M1 for substituting $x=2$ into their equation |  |
| (f) | B1 for reasoning to suggest that temperatures are different in summer or the based only on data from the winter. Allow mention of extrapolation (o.e.) dB1 so not reliable. | model was |




