Paper Reference(s)

## 6683/01

## Edexcel GCE

## Statistics S1

## Advanced Level

# Friday 18 January 2013 - Afternoon <br> Time: 1 hour 30 minutes 

Materials required for examination<br>Items included with question papers<br>Mathematical Formulae (Pink) 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 7 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. A teacher asked a random sample of 10 students to record the number of hours of television, $t$, they watched in the week before their mock exam. She then calculated their grade, $g$, in their mock exam. The results are summarised as follows.

$$
\sum t=258 \quad \sum t^{2}=8702 \quad \sum g=63.6 \quad \mathrm{~S}_{g g}=7.864 \quad \sum g t=1550.2
$$

(a) Find $\mathrm{S}_{t t}$ and $\mathrm{S}_{g t}$.
(b) Calculate, to 3 significant figures, the product moment correlation coefficient between $t$ and $g$.

The teacher also recorded the number of hours of revision, $v$, these 10 students completed during the week before their mock exam. The correlation coefficient between $t$ and $v$ was -0.753 .
(c) Describe, giving a reason, the nature of the correlation you would expect to find between $v$ and $g$.
2. The discrete random variable $X$ can take only the values 1,2 and 3 . For these values the cumulative distribution function is defined by

$$
\mathrm{F}(x)=\frac{x^{3}+k}{40}, \quad x=1,2,3 .
$$

(a) Show that $k=13$.
(b) Find the probability distribution of $X$.

Given that $\operatorname{Var}(X)=\frac{259}{320}$,
(c) find the exact value of $\operatorname{Var}(4 X-5)$.
3. A biologist is comparing the intervals ( $m$ seconds) between the mating calls of a certain species of tree frog and the surrounding temperature $\left(t^{\circ} \mathrm{C}\right)$. The following results were obtained.

| $t{ }^{\circ} \mathrm{C}$ | 8 | 13 | 14 | 15 | 15 | 20 | 25 | 30 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $m$ secs | 6.5 | 4.5 | 6 | 5 | 4 | 3 | 2 | 1 |

(You may use $\sum t m=469.5, \quad S_{t t}=354, \quad S_{m m}=25.5$ )
(a) Show that $\mathrm{S}_{t \mathrm{~m}}=-90.5$.
(b) Find the equation of the regression line of $m$ on $t$ giving your answer in the form $m=a+b t$.
(c) Use your regression line to estimate the time interval between mating calls when the surrounding temperature is $10^{\circ} \mathrm{C}$.
(d) Comment on the reliability of this estimate, giving a reason for your answer.
4. The length of time, $L$ hours, that a phone will work before it needs charging is normally distributed with a mean of 100 hours and a standard deviation of 15 hours.
(a) Find $\mathrm{P}(L>127)$.
(b) Find the value of $d$ such that $\mathrm{P}(L<d)=0.10$.

Alice is about to go on a 6 hour journey. Given that it is 127 hours since Alice last charged her phone,
(c) find the probability that her phone will not need charging before her journey is completed.
5. A survey of 100 households gave the following results for weekly income $£ y$.

| Income $y(£)$ | Mid-point | Frequency $f$ |
| :---: | :---: | :---: |
| $0 \leq y<200$ | 100 | 12 |
| $200 \leq y<240$ | 220 | 28 |
| $240 \leq y<320$ | 280 | 22 |
| $320 \leq y<400$ | 360 | 18 |
| $400 \leq y<600$ | 500 | 12 |
| $600 \leq y<800$ | 700 | 8 |

(You may use $\sum f y^{2}=12452$ 800)
A histogram was drawn and the class $200 \leq y<240$ was represented by a rectangle of width 2 cm and height 7 cm .
(a) Calculate the width and the height of the rectangle representing the class $320 \leq y<400$.
(b) Use linear interpolation to estimate the median weekly income to the nearest pound.
(c) Estimate the mean and the standard deviation of the weekly income for these data.

One measure of skewness is $\frac{3(\text { mean }- \text { median) }}{\text { standard deviation }}$.
(d) Use this measure to calculate the skewness for these data and describe its value.

Katie suggests using the random variable $X$ which has a normal distribution with mean 320 and standard deviation 150 to model the weekly income for these data.
(e) Find $\mathrm{P}(240<X<400)$.
(f) With reference to your calculations in parts ( $d$ ) and $(e)$ and the data in the table, comment on Katie's suggestion.
6. A fair blue die has faces numbered $1,1,3,3,5$ and 5 . The random variable $B$ represents the score when the blue die is rolled.
(a) Write down the probability distribution for $B$.
(b) State the name of this probability distribution.
(c) Write down the value of $\mathrm{E}(B)$.

A second die is red and the random variable $R$ represents the score when the red die is rolled.
The probability distribution of $R$ is

| $r$ | 2 | 4 | 6 |
| :---: | :---: | :---: | :---: |
| $\mathrm{P}(R=r)$ | $\frac{2}{3}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

(d) Find $\mathrm{E}(\mathrm{R})$.
(e) Find $\operatorname{Var}(R)$.

Tom invites Avisha to play a game with these dice.
Tom spins a fair coin with one side labelled 2 and the other side labelled 5. When Avisha sees the number showing on the coin she then chooses one of the dice and rolls it. If the number showing on the die is greater than the number showing on the coin, Avisha wins, otherwise Tom wins.

Avisha chooses the die which gives her the best chance of winning each time Tom spins the coin.
(f) Find the probability that Avisha wins the game, stating clearly which die she should use in each case.
7. Given that

$$
\mathrm{P}(A)=0.35, \mathrm{P}(B)=0.45 \text { and } \mathrm{P}(A \cap B)=0.13 \text {, }
$$

find
(a) $\mathrm{P}(A \cup B)$,
(b) $\mathrm{P}\left(A^{\prime} \mid B^{\prime}\right)$.

The event $C$ has $\mathrm{P}(C)=0.20$.
The events $A$ and $C$ are mutually exclusive and the events $B$ and $C$ are independent.
(c) Find $\mathrm{P}(B \cap C)$.
(d) Draw a Venn diagram to illustrate the events $A, B$ and $C$ and the probabilities for each region.
(e) Find $\mathrm{P}\left([B \cup C]^{\prime}\right)$.

## J anuary 2013

## 6683 Statistics S1

Mark Scheme

\begin{tabular}{|c|c|c|}
\hline \begin{tabular}{l}
Question \\
Number
\end{tabular} \& Scheme \& Marks \\
\hline 1.80 \& \begin{tabular}{l}
\[
\begin{aligned}
\& \left(\mathrm{S}_{t t}\right)=8702-\frac{258^{2}}{10} \text { or }\left(\mathrm{S}_{g t}\right)=1550.2-\frac{258 \times 63.6}{10} \\
\& \left(\mathrm{~S}_{t t}=\right) 2045.6, \quad\left(\mathrm{~S}_{g t}=\right)-90.68 \quad \text { awrt }(2046), \quad \text { awrt }-90.7 \\
\& r=\frac{-90.68}{\sqrt{2045.6 \times 7.864}}=-0.714956 \ldots \quad \text { awrt }-0.715
\end{aligned}
\] \\
Positive \\
e.g. high \(v\) corresponds to low \(t\) and low \(t\) corresponds to high \(g\) so expect high \(v\) to corresponds to high \(g\) or expect more revision to result in a better grade
\end{tabular} \& \begin{tabular}{ll} 
M1 \& \\
A1, A1 \& \\
M1 A1 \& \\
\& (3) \\
B1 \\
B1 \& \\
B1 \& \\
\hline
\end{tabular} \\
\hline \& Notes \& \\
\hline (a)
(b)

(c) \& \multicolumn{2}{|l|}{```
M1 for at least one correct expression
$1^{\text {st }} \mathrm{A} 1$ for $\mathrm{S}_{t t}=$ awrt 2046 (Condone $S_{x x}=\ldots$ or even $S_{y y}=\ldots$ )
$2^{\text {nd }} \mathrm{A} 1$ for $\mathrm{S}_{g t}=$ awrt $-90.7 \quad$ (Condone $S_{x y}=\ldots$ )
On epen take care that $1^{\text {st }} \mathrm{A} 1$ is for $\mathrm{S}_{t t}$ and $2^{\text {nd }} \mathrm{A} 1$ is for $\mathrm{S}_{g t}$
M1 for attempt at correct formula.
Must have their $\mathrm{S}_{t t}, \mathrm{~S}_{\mathrm{g} t}$ and given $\mathrm{S}_{g g}$ in the correct places. Condone missing "-"
Award M1A0 for awrt -0.71 with no expression seen
M0 for $\frac{1550.2}{\sqrt{8702 \times 7.864}}$
Correct answer only is $2 / 2$
$1^{\text {st }}$ B1 for saying "positive". Condone dubious arguments here. Ignore mention of skew.
$2^{\text {nd }} \mathrm{B} 1$ for suitable reason that mentions at least $v$ and $g$ and supports positive correlation.
e.g. "the less revision done the lower the grade" is B1
"should do better with more revision" is B0 since does not mention grades
"both coefficients are similar" or two sketches of negative correlation with labelled axes is B1 since
$v, t$ and $g$ are implied
Allow use of letters $v$ and $g$
Allow equivalent terms e.g. "study" instead of "revision" or "score" instead of "grade"

```} \\
\hline
\end{tabular}





```

