Edexcel GCE
Statistics S3
Advanced/Advanced Subsidiary
Thursday 22 May 2014 – Morning
Time: 1 hour 30 minutes

Materials required for examination
Mathematical Formulae (Pink)

Items included with question papers
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 or symbolic differentiation/integration, or have retrievable mathematical formulae stored in them.

Instructions to Candidates
In the boxes above, write your centre number, candidate number, your surname, initials and signature.
Check that you have the correct question paper.
Answer ALL the questions.
You must write your answer to each question in the space following the question.
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.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. (2).
There are 8 questions in this question paper. The total mark for this paper is 75.
There are 28 pages in this question paper. Any blank pages are indicated.

Advice to Candidates
You must ensure that your answers to parts of questions are clearly labelled.
You should show sufficient working to make your methods clear to the Examiner.
Answers without working may not gain full credit.
1. (a) Explain what you understand by a random sample from a finite population. (1)

(b) Give an example of a situation when it is not possible to take a random sample. (1)

A college lecturer specialising in shoe design wants to change the way in which she organises practical work.

She decides to gather ideas from her 75 students.

She plans to give a questionnaire to a random sample of 8 of these students.

(c) (i) Describe the sampling frame that she should use.

(ii) Explain in detail how she should use a table of random numbers to obtain her sample. (3)
Question 1 continued

___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________

(Total 5 marks)
2. The weights of pears in an orchard are assumed to have unknown mean $\mu$ and unknown standard deviation $\sigma$.

A random sample of 20 pears is taken and their weights recorded.

The sample is represented by $X_1, X_2, \ldots, X_{20}$. State whether or not the following are statistics. Give reasons for your answers.

(a) (i) $\frac{X_1 + 3X_{20}}{2}$

(ii) $\sum_{i=1}^{20} (X_i - \mu)$

(iii) $\sum_{i=1}^{20} \left( \frac{X_i - \mu}{\sigma} \right)$

(b) Find the mean and variance of $\frac{3X_1 - X_{20}}{2}$
Question 2 continued

___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________

(Total 7 marks)
3. A number of males and females were asked to rate their happiness under the headings “not happy”, “fairly happy” and “very happy”.

The results are shown in the table below

<table>
<thead>
<tr>
<th>Gender</th>
<th>Not happy</th>
<th>Fairly happy</th>
<th>Very happy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>9</td>
<td>43</td>
<td>34</td>
<td>86</td>
</tr>
<tr>
<td>Male</td>
<td>13</td>
<td>25</td>
<td>16</td>
<td>54</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>68</td>
<td>50</td>
<td>140</td>
</tr>
</tbody>
</table>

Stating your hypotheses, test at the 5% level of significance, whether or not there is evidence of an association between happiness and gender. Show your working clearly.

(10)
Question 3 continued
Question 3 continued
4. The random variable \( A \) is defined as

\[
A = B + 4C - 3D
\]

where \( B, C \) and \( D \) are independent random variables with

\[
B \sim N(6, 2^2) \quad C \sim N(7, 3^2) \quad D \sim N(4, 1.5^2)
\]

Find \( P(A < 45) \)
Question 4 continued

___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________

Q4 (Total 6 marks)
5. A research station is doing some work on the germination of a new variety of genetically modified wheat.

They planted 120 rows containing 7 seeds in each row.

The number of seeds germinating in each row was recorded. The results are as follows

<table>
<thead>
<tr>
<th>Number of seeds germinating in each row</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed number of rows</td>
<td>2</td>
<td>6</td>
<td>11</td>
<td>19</td>
<td>25</td>
<td>32</td>
<td>16</td>
<td>9</td>
</tr>
</tbody>
</table>

(a) Write down two reasons why a binomial distribution may be a suitable model.

(b) Show that the probability of a randomly selected seed from this sample germinating is 0.6

The research station used a binomial distribution with probability 0.6 of a seed germinating. The expected frequencies were calculated to 2 decimal places. The results are as follows

<table>
<thead>
<tr>
<th>Number of seeds germinating in each row</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected number of rows</td>
<td>0.20</td>
<td>2.06</td>
<td>s</td>
<td>23.22</td>
<td>t</td>
<td>31.35</td>
<td>15.68</td>
<td>3.36</td>
</tr>
</tbody>
</table>

(c) Find the value of s and the value of t.

(d) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the data can be modelled by a binomial distribution.
Question 5 continued
Question 5 continued

___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
Question 5 continued
6. A random sample \( X_1, X_2, \ldots, X_n \) is taken from a population with mean \( \mu \).

(a) Show that \( \bar{X} = \frac{1}{n} (X_1 + X_2 + \ldots + X_n) \) is an unbiased estimator of the population mean \( \mu \). (1)

A company produces small jars of coffee.

Five jars of coffee were taken at random and weighed.

The weights, in grams, were as follows

\[
\begin{align*}
197 & \quad 203 & \quad 205 & \quad 201 & \quad 195
\end{align*}
\]

(b) Calculate unbiased estimates of the population mean and variance of the weights of the jars produced by the company. (3)

It is known from previous results that the weights are normally distributed with standard deviation 4.8 g.

The manager is going to take a second random sample. He wishes to ensure that there is at least a 95% probability that the estimate of the population mean is within 1.25 g of its true value.

(c) Find the minimum sample size required. (4)
Question 6 continued
Question 6 continued

___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________
___________________________________________________________________________

Q6 (Total 8 marks)
7. A machine fills packets with \( X \) grams of powder where \( X \) is normally distributed with mean \( \mu \). Each packet is supposed to contain 1 kg of powder.

To comply with regulations, the weight of powder in a randomly selected packet should be such that \( P(X < \mu - 30) = 0.0005 \)

(a) Show that this requires the standard deviation to be 9.117 g to 3 decimal places. \( \text{ (3) } \)

A random sample of 10 packets is selected from the machine. The weight, in grams, of powder in each packet is as follows

\[
999.8 \quad 991.6 \quad 1000.3 \quad 1006.1 \quad 1008.2 \quad 997.0 \quad 993.2 \quad 1000.0 \quad 997.1 \quad 1002.1
\]

(b) Assuming that the standard deviation of the population is 9.117 g, test, at the 1% significance level, whether or not the machine is delivering packets with mean weight of less than 1 kg. State your hypotheses clearly. \( \text{ (7) } \)
Question 7 continued
Question 7 continued
8. The heights, in metres, and weights, in kilograms, of a random sample of 9 men are shown
in the table below

<table>
<thead>
<tr>
<th>Man</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height ((x))</td>
<td>1.68</td>
<td>1.74</td>
<td>1.75</td>
<td>1.76</td>
<td>1.78</td>
<td>1.82</td>
<td>1.84</td>
<td>1.88</td>
<td>1.98</td>
</tr>
<tr>
<td>Weight ((y))</td>
<td>75</td>
<td>76</td>
<td>100</td>
<td>77</td>
<td>90</td>
<td>95</td>
<td>110</td>
<td>96</td>
<td>120</td>
</tr>
</tbody>
</table>

(a) Given that \(S_{xx} = 0.0632\), \(S_{yy} = 1957.5556\) and \(S_{xy} = 9.3433\) calculate, to 3 decimal
places, the product moment correlation coefficient between height and weight for
these men.

(2)

(b) Use your value of the product moment correlation coefficient to test whether or not
there is evidence of a positive correlation between the height and weight of men. Use
a 5% significance level. State your hypotheses clearly.

(4)

Peter does not know the heights or weights of the 9 men. He is given photographs of them
and asked to put them in order of increasing weight. He puts them in the order

\[ A \ C \ E \ B \ G \ D \ I \ F \ H \]

(c) Find, to 3 decimal places, Spearman’s rank correlation coefficient between Peter’s
order and the actual order.

(6)

(d) Use your value of Spearman’s rank correlation coefficient to test for evidence of
Peter’s ability to correctly order men, by their weight, from their photographs. Use a
5% significance level and state your hypotheses clearly.

(4)
Question 8 continued
Question 8 continued