

Paper Reference(s)

**6683**

**Edexcel GCE**

Solutions

**Statistics S1**

**Advanced Subsidiary**

**Friday 14 January 2005 – Morning**

**Time: 1 hour 30 minutes**

**Materials required for examination**

Mathematical Formulae (Lilac)

Graph Paper (ASG2)

**Items included with question papers**

Nil

**Candidates may use any calculator EXCEPT those with the facility for symbolic algebra, differentiation and/or integration. Thus candidates may NOT use calculators such as the Texas Instruments TI 89, TI 92, Casio CFX 9970G, Hewlett Packard HP 48G.**

**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 seven 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 gain no credit.

1. A company assembles drills using components from two sources. Goodbuy supplies 85% of the components and Amart supplies the rest. It is known that 3% of the components supplied by Goodbuy are faulty and 6% of those supplied by Amart are faulty.

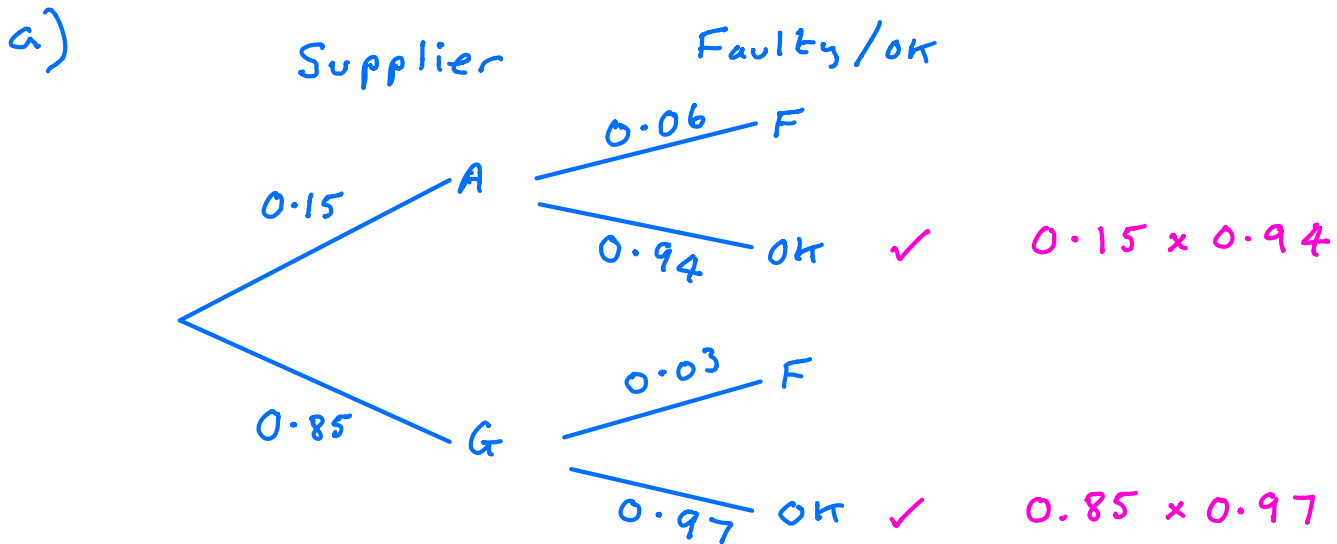
(a) Represent this information on a tree diagram.

(3)

An assembled drill is selected at random.

(b) Find the probability that it is not faulty.

(3)



b)

$$P(\text{ok}) = 0.15 \times 0.94 + 0.85 \times 0.97$$
$$= 0.9655$$

---

2. The number of caravans on Seaview caravan site on each night in August last year is summarised in the following stem and leaf diagram.

Caravans	1   0 means 10	Totals
1	0 5	(2)
2	1 2 4 8	(4)
3	0 3 3 4 7 8 8	(8)
4	1 3 5 8 8 8 9 9	(9)
5	3 6 6 7	(5)
6	2 3 4	(3)

- (a) Find the three quartiles of these data.

(3)

During the same month, the least number of caravans on Northcliffe caravan site was 31. The maximum number of caravans on this site on any night that month was 72. The three quartiles for this site were 38, 45 and 52 respectively.

- (b) On graph paper and using the same scale, draw box plots to represent the data for both caravan sites. You may assume that there are no outliers.

(6)

- (c) Compare and contrast these two box plots.

(3)

- (d) Give an interpretation to the upper quartiles of these two distributions.

(2)

a) 31 data items

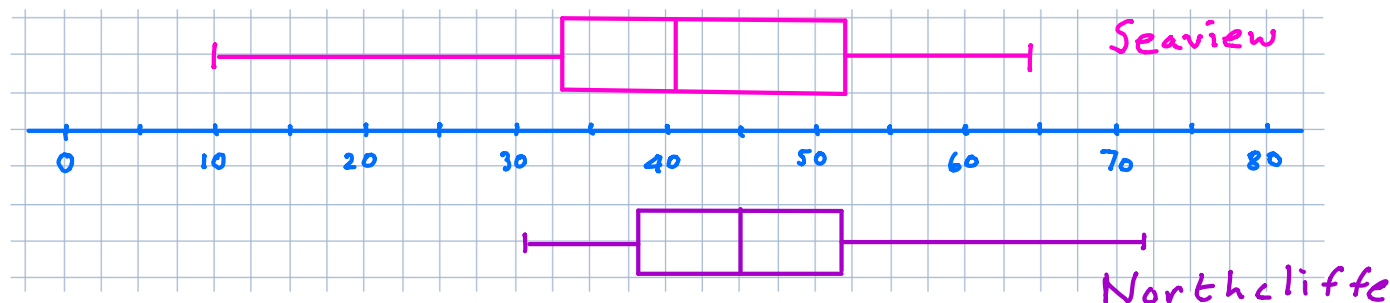
$$\frac{31}{4} = 7.75 \quad Q_1 = 8^{\text{th}}$$

$$\frac{31+1}{2} = 16 \quad Q_2 = 16^{\text{th}}$$

$$\frac{3 \times 31}{4} = 23.25 \quad Q_3 = 24^{\text{th}}$$

$$Q_1 = 33 \quad Q_2 = 41 \quad Q_3 = 52$$

b)



c) On average more caravans stayed at Northcliffe since the Northcliffe median was 45 and the Seaview median 41.

Seaview had a greater variation than Northcliffe as the IQR for Seaview was 19 and the IQR for Northcliffe was 14.

Seaview distribution shows negative skew whereas Northcliffe distribution shows positive skew.

---

d) For 75% of the nights both sites had a maximum of 52 caravans

---

3. The following table shows the height  $x$ , to the nearest cm, and the weight  $y$ , to the nearest kg, of a random sample of 12 students.

$x$	148	164	156	172	147	184	162	155	182	165	175	152
$y$	39	59	56	77	44	77	65	49	80	72	70	52

(a) On graph paper, draw a scatter diagram to represent these data.

(3)

(b) Write down, with a reason, whether the correlation coefficient between  $x$  and  $y$  is positive or negative.

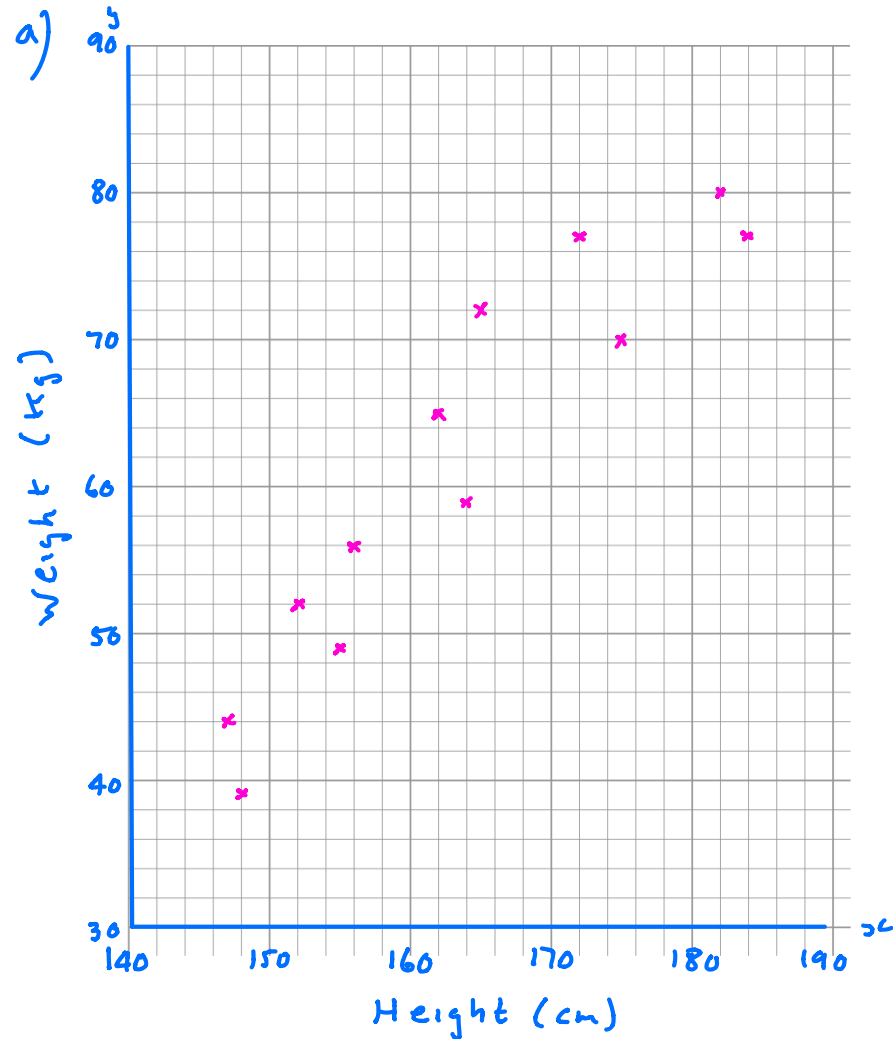
(2)

The data in the table can be summarised as follows.

$$\Sigma x = 1962, \quad \Sigma y = 740, \quad \Sigma y^2 = 47\,746, \quad \Sigma xy = 122\,783, \quad S_{xx} = 1745.$$

(c) Find  $S_{xy}$ .

(2)



b) positive  
since  $y$  increases  
as  $x$  increases

c)

$$\begin{aligned}
 S_{xy} &= \sum xy - n\bar{x}\bar{y} \\
 &= 122783 - \frac{1962 \times 740}{12} \\
 &= 1793
 \end{aligned}$$


---

The equation of the regression line of  $y$  on  $x$  is  $y = -106.331 + bx$ .

- (d) Find, to 3 decimal places, the value of  $b$ . (2)
- (e) Find, to 3 significant figures, the mean  $\bar{y}$  and the standard deviation  $s$  of the weights of this sample of students. (3)
- (f) Find the values of  $\bar{y} \pm 1.96s$ . (2)
- (g) Comment on whether or not you think that the weights of these students could be modelled by a normal distribution. (1)
-

d) By calculator  $b = 1.028$  to 3 d.p.

---

e) By calculator  $\bar{y} = 61.7$  kg to 3 s.f.  
 $s = 13.3$  kg to 3 s.f.

---

f)

$$\begin{aligned}\bar{y} + 1.96s &= 61.7 + 1.96 \times 13.3 = 87.8 \text{ kg} \\ \bar{y} - 1.96s &= 61.7 - 1.96 \times 13.3 = 35.6 \text{ kg}\end{aligned}$$

---

g) All weights between 35.6 and 87.8 kg  
so could be normal.

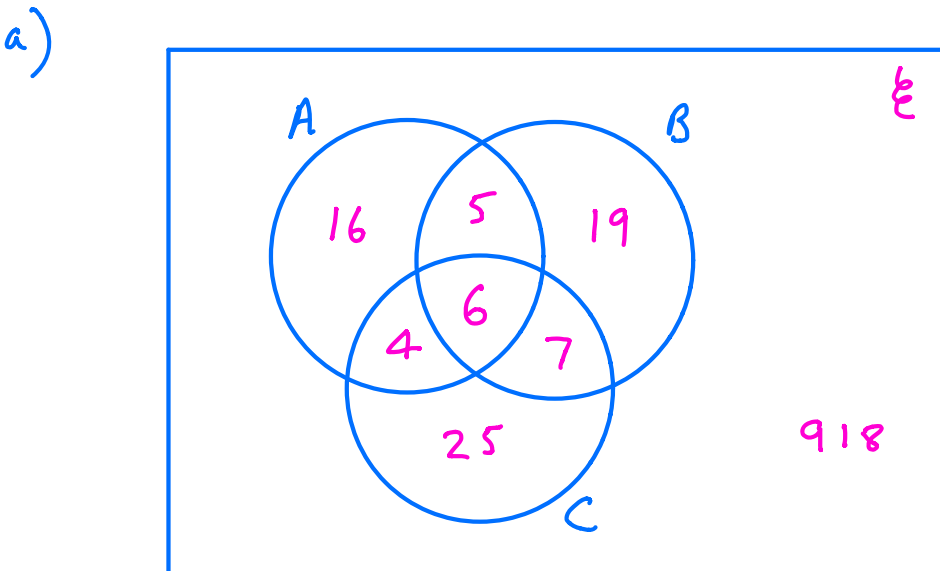
---

5. Articles made on a lathe are subject to three kinds of defect,  $A$ ,  $B$  or  $C$ . A sample of 1000 articles was inspected and the following results were obtained.

31 had a type  $A$  defect  
37 had a type  $B$  defect  
42 had a type  $C$  defect  
11 had both type  $A$  and type  $B$  defects  
13 had both type  $B$  and type  $C$  defects  
10 had both type  $A$  and type  $C$  defects  
6 had all three types of defect.

(a) Draw a Venn diagram to represent these data.

(6)



Find the probability that a randomly selected article from this sample had

(b) no defects, (1)

(c) no more than one of these defects. (2)

An article selected at random from this sample had only one defect.

(d) Find the probability that it was a type  $B$  defect. (2)

Two different articles were selected at random from this sample.

(e) Find the probability that both had type  $B$  defects. (2)

---

$$b) \quad P(\text{No defects}) = \frac{981}{1000}$$

$$c) \quad P(\text{No more than one defect}) = \frac{918 + 25 + 19 + 16}{1000} \\ = \frac{978}{1000}$$

---

$$d) \quad P(B \text{ defect} \mid \text{One defect}) = \frac{19}{16 + 19 + 25} = \frac{19}{60}$$

---

$$e) \quad P(\text{Two selected had } B \text{ defects}) \\ = \frac{37}{1006} \times \frac{36}{999} = \frac{1}{750} = 0.001\bar{3}$$

---

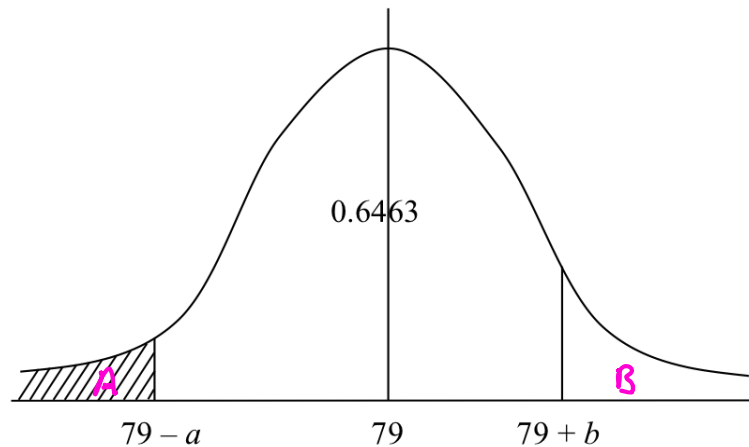
7. The random variable  $X$  is normally distributed with mean 79 and variance 144.

Find

(a)  $P(X < 70)$ , By calc = 0.2266 (3)

(b)  $P(64 < X < 96)$ . By calc = 0.8161 (3)

It is known that  $P(79 - a \leq X \leq 79 + b) = 0.6463$ . This information is shown in the figure below.



Given that  $P(X \geq 79 + b) = 2P(X \leq 79 - a)$ ,

(c) show that the area of the shaded region is 0.1179. (3)

(d) Find the value of  $b$ . (4)

c)  $A + B = 1 - 0.6463 = 0.3537$   
 $B = 2A \Rightarrow 3A = 0.3537$   
 $A = \frac{0.3537}{3} = 0.1179$

d)  $B = 2A = 0.2358$   
Area below B = 0.7642  
By calc  $79 + b = 87.6385$   
 $b = 8.6385$   
 $b \approx 8.64$