Paper Reference(s)

6683/01

Edexcel GCE

Solutions

Statistics S1

Advanced/Advanced Subsidiary

Monday 16 January 2006 – Morning

Time: 1 hour 30 minutes

Materials required for examination	**Items included with question papers**
Mathematical Formulae (Green or Lilac)	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.
The marks for individual questions and the parts of questions are shown in round brackets: e.g. **(2)**.
There are 7 questions on this paper. 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.

N20908A

1. Over a period of time, the number of people x leaving a hotel each morning was recorded. These data are summarised in the stem and leaf diagram below.

Number leaving	3 | 2 means 32							Totals
2	7	9	9					(3)
3	2	2	3	5	6			(5)
4	0	1	4	8	9			(5)
5	2	3	3	6	6	6	8	(7)
6	0	1	4	5				(4)
7	2	3						(2)
8	1							(1)

For these data,

(*a*) write down the mode,

(1)

(*b*) find the values of the three quartiles.

(3)

a) Mode = 56

b) $\frac{27}{4} = 6.75$ $Q_1 = 7^{th}$ item

$\frac{27+1}{2} = 14$ $Q_2 = 14^{th}$ item

$\frac{3 \times 27}{4} = 20.25$ $Q_3 = 21^{st}$ item

$Q_1 = 35$ $Q_2 = 52$ $Q_3 = 60$

Given that $\Sigma x = 1335$ and $\Sigma x^2 = 71\,801$, find

(*c*) the mean and the standard deviation of these data.

(4)

$\bar{x} = \frac{\Sigma x}{n} = \frac{1335}{27} = 49.\dot{4}$

$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2} = \sqrt{\frac{71801}{27} - \left(\frac{1335}{27}\right)^2} = 14.6$

One measure of skewness is found using

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}.$$

(*d*) Evaluate this measure to show that these data are negatively skewed.

(2)

(*e*) Give two other reasons why these data are negatively skewed.

(4)

---

d)

$$\frac{49.44 - 56}{14.6} = -0.449 \Rightarrow \text{negative skew}$$

e)

$$Q_2 - Q_1 > Q_3 - Q_2$$

$$17 > 8$$

Also mean < median

$$49.\dot{4} < 52$$

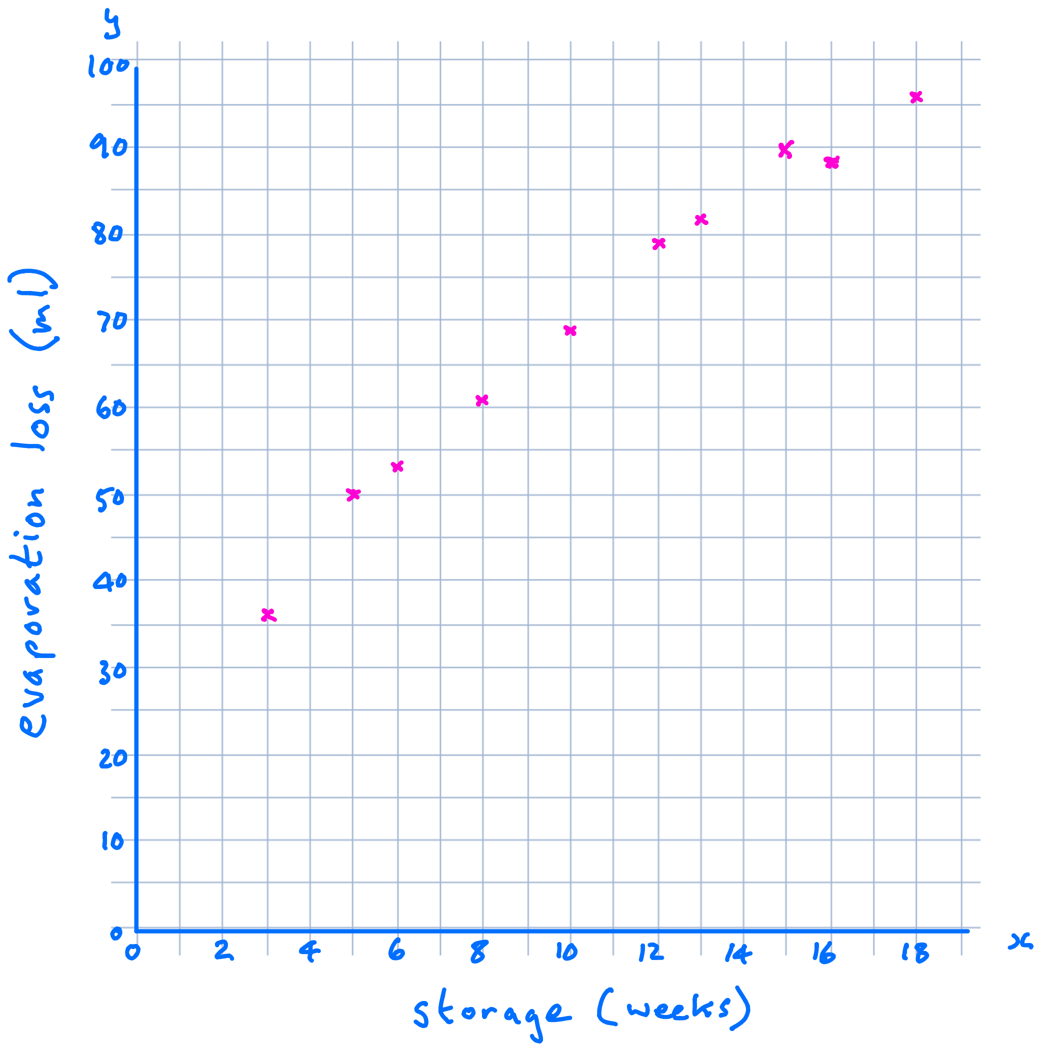
---

3. A manufacturer stores drums of chemicals. During storage, evaporation takes place. A random sample of 10 drums was taken and the time in storage, x weeks, and the evaporation loss, y ml, are shown in the table below.

x	3	5	6	8	10	12	13	15	16	18
y	36	50	53	61	69	79	82	90	88	96

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

(3)



(*b*) Give a reason to support fitting a regression model of the form $y = a + bx$ to these data. **(1)**

(*c*) Find, to 2 decimal places, the value of a and the value of b.

(You may use $\Sigma x^2 = 1352$, $\Sigma y^2 = 53\,112$ and $\Sigma xy = 8354$.)

No longer on syllabus but can do with calculator **(7)**

(*d*) Give an interpretation of the value of b. **(1)**

(*e*) Using your model, predict the amount of evaporation that would take place after

(i) 19 weeks,

(ii) 35 weeks. **(2)**

(*f*) Comment, with a reason, on the reliability of each of your predictions. **(4)**

---

b) Data appear to have an approximate linear relationship

c) By calc $a = 29.021$

$b = 3.904$

d) b is the rate of evaporation loss in ml per week

e) $y = 29.021 + 3.904x$

i) $x = 19$ $y = 29.021 + 3.904 \times 19 = 103$ ml

ii) $x = 35$ $y = 29.021 + 3.904 \times 35 = 166$ ml

f) Answer for 19 weeks likely to be reliable since only just outside the range of values of x used in model.

35 weeks is well outside the range of values for x. Extrapolation means this answer is likely to be unreliable. No evidence that linear relationship is maintained beyond the range of values of x used in model.

4. A bag contains 9 blue balls and 3 red balls. A ball is selected at random from the bag and its colour is recorded. The ball is not replaced. A second ball is selected at random and its colour is recorded.

(*a*) Draw a tree diagram to represent the information.

(3)

Find the probability that

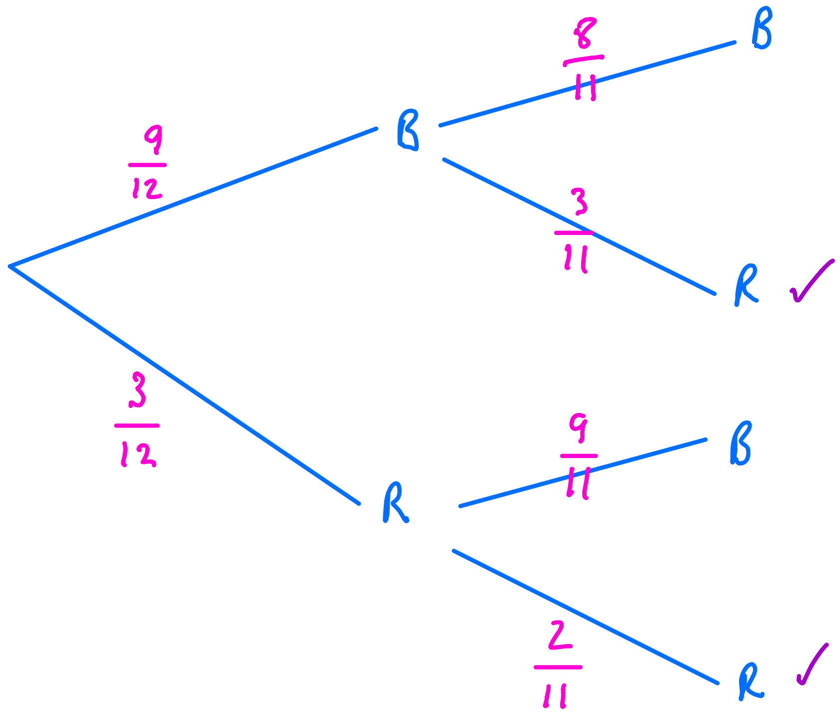
(*a*) the second ball selected is red,

(2)

(*b*) both balls selected are red, given that the second ball selected is red.

(2)

a)



$\frac{9}{12} \times \frac{3}{11}$

$\frac{3}{12} \times \frac{2}{11}$

a) $P(\text{2nd is Red}) = \frac{9}{12} \times \frac{3}{11} + \frac{3}{12} \times \frac{2}{11} = \frac{1}{4}$

b) $P(\text{Both Red} \mid \text{2nd Red}) = \dfrac{P(\text{Both Red} \cap \text{2nd Red})}{P(\text{2nd Red})}$

$= \dfrac{\frac{3}{12} \times \frac{2}{11}}{\frac{1}{4}}$

$= \dfrac{2}{11}$

---

5. (*a*) Write down two reasons for using statistical models.

(2)

(*b*) Give an example of a random variable that could be modelled by

(i) a normal distribution,

(ii) a discrete uniform distribution.

(2)

5. (a)	To simplify a real world problem To improve understanding / describe / analyse a real world problem Quicker and cheaper than using real thing To predict possible future outcomes Refine model / change parameters possible Any 2	B1B1	(2)
(b)	(i) e.g.s height, weight (ii) score on a face after tossing a fair die	B1B1	(2)
		Total 4 marks	

6. For the events A and B,

$$P(A \cap B') = 0.32, \; P(A' \cap B) = 0.11 \text{ and } P(A \cup B) = 0.65.$$

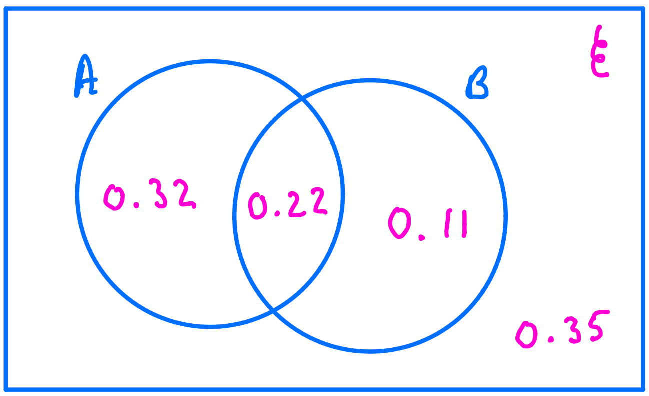
(*a*) Draw a Venn diagram to illustrate the complete sample space for the events A and B. **(3)**

(*b*) Write down the value of P(A) and the value of P(B). **(3)**

(*c*) Find $P(A \mid B')$. **(2)**

(*d*) Determine whether or not A and B are independent. **(3)**

a)



b) $P(A) = 0.54$

$P(B) = 0.33$

c) $P(A \mid B')$

$= \dfrac{P(A \cap B')}{P(B')}$

$= \dfrac{0.32}{0.67}$

$= \dfrac{32}{67}$

$= 0.4776$

d) $P(A \cap B) = 0.22$

$P(A) \times P(B) = 0.54 \times 0.33$

$= 0.1782$

$P(A \cap B) \neq P(A) \times P(B)$

$\therefore$ A and B are not independent

7. The heights of a group of athletes are modelled by a normal distribution with mean 180 cm and a standard deviation 5.2 cm. The weights of this group of athletes are modelled by a normal distribution with mean 85 kg and standard deviation 7.1 kg.

Find the probability that a randomly chosen athlete

(*a*) is taller than 188 cm,

(3)

(*b*) weighs less than 97 kg.

(2)

(*c*) Assuming that for these athletes height and weight are independent, find the probability that a randomly chosen athlete is taller than 188 cm and weighs more than 97 kg.

(3)

(*d*) Comment on the assumption that height and weight are independent.

(1)

---

a) $H \sim N(180, 5.2^2)$

By calc $P(H > 188) = 0.0619679$

$= 0.0620$

---

b) $W \sim N(85, 7.1^2)$

By calc $P(W < 97) = 0.9544995$

$= 0.9545$

---

c) $P(H > 188 \cap W > 97)$

$= 0.0619679 \times (1 - 0.9544995)$

$= 0.00282$

---

d) In general, small people weigh less than tall people, so height and weight will not be independent

---