

STATISTICS

SUBJECT 4041

PAPER 2

GENERAL COMMENT

There was no improvement in both the candidature and the general standard. Good scripts were very few. The performance was poor. It was pleasing once again to note that almost all candidates used the sales given when answering the graph questions. Teachers need to stress on the importance of following instructions regarding the degree of accuracy and scales given when drawing graphs.

In Section B, questions 7 and 8 were very unpopular. In this section, many candidates failed to attempt the full quota of questions. This seemed to be because they did not know how to tackle some questions rather than through lack of time.

COMMENTS ON INDIVIDUAL QUESTIONS

QUESTION 1

This question was answered well by all candidates and most of them scored full marks.

QUESTION 2

There were many poor attempts at using the given assumed mean to find the actual mean. A surprising number of candidates attempted to calculate the mean using $\frac{\sum x}{n}$. Other mistakes included use of grouped data whereby the frequencies would be wrong. Those who attempted to find the deviations first, used $(\sum \bar{x} - xi)$ instead of $\sum((xi - \bar{x}))$. Such candidates got a negative $\frac{\sum d}{n}$ thereby producing a wrong mean.

Answer:

$$\text{Mean} = 20 + \frac{2}{14} = 20,1$$

QUESTION 3

Most candidates had some understanding of ideas in (i) and (ii) part (a) but were not always able to explain their thoughts clearly. This question would have been much better answered had candidates read it properly. Credit was given wherever possible for sensible statements. A number of candidates ignored the last part "when used in relation to statistical surveys". Part (b) was answered well but the majority scored 2 marks instead of 3.

Answers:

- (a) (i) The population is the ground of all items of interest for a particular statistical study.
- (ii) A census is a statistical survey in which every item in the population is measured or analyzed ... (or any equivalent).

QUESTION 4

Part (a) of this question was well done. Candidates who attempted this question scored full marks on this part. Part (b) (i) was poorly done. Candidates did not realize that two attempts to pass meant failing first then passing at 2nd attempt. Part (b) (ii) was fairly attempted but part (c) was again poorly done. Many of the candidates did not know how to use the information in their tree diagram though they had produced it correctly.

Answer:

- (a) 0,6 0,35 0,2
- (b) (i) $0,6 \times 0,65 = 0,39$
(ii) $1 - P(\text{fails three times}) = 0,958$
- (c) $(0,4 \times 0,6 \times 0,6) \times 3 = 0,432$

QUESTION 5

Generally this question was well done by most candidates. They really seemed prepared for this question. However, there were some few candidates who confused death rates per thousand with the number of deaths. In part (ii), some few candidates did not understand the significance of first multiplying the death rate per thousand by the standard population for each group then adding them to come up with the required standardized death rate. It is also stressed that the candidates should know the difference between the standardized death rate and the crude death rate.

Answers:

- (a) $X = 300$ $Y = 20\ 000$ $Z = 15$
- (b) (i) 6,35 per thousand
(ii) 7,95 per thousand

QUESTION 6

Variations answers were given to part (a) (i) and (ii). The question was not specific. In part (i) some candidates used 93, others 96 or 221 as the denominator, in the calculation of the percentage. Full marks were obtained in (b) (i) and (ii) by many students.

Answers:

(a) (i) $\frac{41}{93} \times 100 = 44\%$

(ii) 30%

(b) (i) 80

(ii) 40

QUESTION 7

This was an unpopular question. The question was attempted by very few candidates who earned only part marks on parts (a) (i) and (ii). Parts (c) and (d)(iii) proved to be the most difficult for most candidates who attempted this question. The candidates failed to apply the mathematical concepts. They could not solve the quadratic equation to come up with the possible values of y . The concepts demanded in this part (c) are not common in statistics but in mathematics. If candidates had read and understood part (d)(iii) well, they could have scored two marks. Many candidates did not understand how to tackle it.

Answers:

(a) (i) $x = 3$ (ii) median = 7

(b) Median is not affected by extreme values.

(c) (i) $S_2 = \frac{\sum x^2}{n} - (\bar{x})^2 \Leftrightarrow 24,4 = \frac{846+y^2}{5} - \left(\frac{2916+108y+y^2}{25}\right)$

$$\Leftrightarrow \begin{aligned} y^2 - 27y + 176 &= 0 \\ y &= 11 \text{ or } 16 \end{aligned}$$

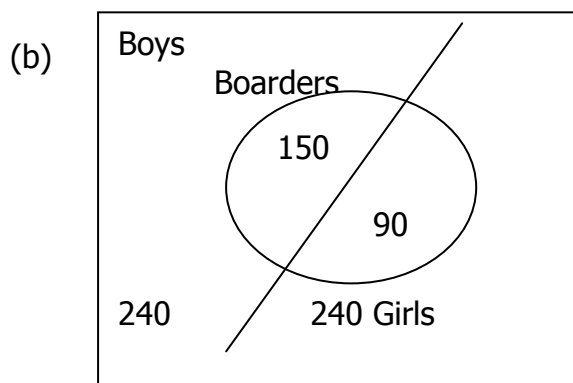
(d) (i) $\bar{x} = 14, \delta = 4$ (ii) $\bar{x} = 24, \delta = 8$ (iii) $\bar{x} = 37, \delta = 12$

QUESTION 8

This question was the least popular in Section B and was poorly answered by all candidates who attempted it. Many candidates could not calculate the required probabilities. Also most of them tried to write down an answer without showing any evidence of calculation. These answers were never correct. Lack of paying attention to the requirement of the question culminated in coming up with a correct figure to the wrong question. The Venn diagrams were wrongly filled in with figures written in parts they do not belong to.

Answers:

- (a) (i) 390 boys (ii) $240 - 150 = 90$ girl boarders



- (c) (i) $\frac{1}{3}$

(ii) $\frac{3}{8}$

(d) $0,043 \Rightarrow \frac{150}{720} \times \frac{149}{719}$

- (e) (i)

M	10	E	A
S			
40	20	30	
20			10

(ii) $\frac{30}{120} = \frac{1}{4}$

(iii) $\frac{20}{120} = \frac{1}{6}$

QUESTION 9

This was the only very popular question in Section B. There were many good attempts at this question and a number of candidate obtained maximum marks. Almost all candidates who attempted this question were able to fill in the required information in the table. The frequencies and the frequency densities were correct. They sued the correct scale to draw a histogram. Candidates also did very well in (c) (i) and (ii) and came up with correct sectional percentage bar charts.

Answers

(a)

DISTANCE	FREQUENCY	FREQUENCY DENSITY
$0 < x \leq 20$	3	0,15
$20 < x \leq 30$	3	0,3
$30 < x \leq 40$	5	0,5
$40 < x \leq 60$	17	0,85
$60 < x \leq 100$	10	0,25
$100 < x \leq 120$	2	0,1

(b) Histogram

(c) Hospital A	(ii) Sectional percentage bar chart
Men - 49%	
Women - 36%	
Children - 15%	

QUESTION 10

Another unpopular question badly answered. It seems candidates now had no choice. Anyway, part (a) comprising of 2 marks only was well done. Candidates managed to complete the table illustrating the possibility space.

Some of the candidates were not able to complete the table of probabilities correctly. Instead of writing the probabilities in the $P(X=x)$ row they wrote whole figure. Candidates used three terms instead of only two in the calculation of the expected gain or loss. In part (b) (ii) many candidates lost marks through vague statements. Part (c) seemed very difficult for almost all candidates. When calculating the expected value of prize in box candidates used the prizes given in the table and took the UCB yet they were supposed to find the midpoints.

Again candidates did not realize that they were to find probabilities first, so they just used 9, 3, 2, and 1 as is in the question. Candidates were not able to explain why the contestant should take cash. It is stressed that candidates be well versed with expectation, where there is a gain or a loss and be able to explain and interpret.

Answers:

(a)

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

<i>Total Score</i>	$X \leq 3$	$3 < x \leq 8$	<i>Over 8</i>
$P(X=x)$	$\frac{3}{36}$	$\frac{23}{36}$	$\frac{10}{36}$

(b) (i) Expectation = $\frac{10}{36} \times 18000 + \frac{3}{36} \times 60000 - 10000 = -0$

(ii) The game is fair because the amount he can expect to earn from the prizes is the same as what he has to pay per turn. (any equivalent).

(c) (i) Expected value of prize in box =

$$= \frac{9}{15} \times 5000 + \frac{3}{15} \times 30\,000 + \frac{2}{15} \times 75\,000 + \frac{1}{15} \times 5\,000\,000 = \$352\,333,33$$

(ii) He should take cash because his expected prize is less than \$500 000.

(iii) Once a box has been opened the probability associated with each prize changes.

QUESTION 11

Many attempts at this question, which was fairly popular, were spoiled by lack of suitable labels or even any labels at all. Some of the candidates failed to use the give scales.

Candidates followed the clear instruction to plot the original sales data and the centred moving averages. Some candidates did not position the Trend line correctly. Others could not explain when it is necessary to centre the moving averages. Here they confused the centre moving averages and the moving averages.

All candidates did not understand the phrase "derivation from the trend line" and used the trend line to quote the 4th quarter of 2006 figure which was wrong because it would not give an estimate of the sales of the sales but an estimate of the next centred moving averages, though read from a wrong point.

Answers:

$$w = 28$$

$$x = 142$$

$$y = 38,75$$

$$z = 40,5$$

- (iv) The trend line
- (v) Moving averages are centred when we have an even number of divisions in one cycle.
- (vi) 2006 4th quarters MA = 48 000 litres.

$$\begin{aligned}\text{Predicted sales} &= \left[48 + \left(\frac{10+11+12}{3} \right) \right] \text{ 000 litres.} \\ &= 48\ 000 + 11\ 000 = 59\ 000 \text{ litres.}\end{aligned}$$