



## STATISTICS

### Checklist

Make sure you:

- Understand how Statistics has been divided into P1 & P2 in the new CAPS curriculum.  
P1: Probability & Counting Techniques  
P2: Data Handling
- Understand the 5 *sections* of Data Handling
  - Working with grouped data – Estimate for the mean, working in intervals etc.
  - Calculating the Standard Deviation
  - Calculation of 5 number summary and sketching / interpreting the box & whisker
  - Working with cumulative frequency & sketching the ogive - reading values from the ogive sketch
  - Determining (and sketching) the *Line of best fit* – (including correlation -  $r$ )

### Exam Questions

#### Question 1

Gr11 – Grouped data: mean & std dev, cum freq, ogive, reading off graph

This frequency table gives the weights of the first 55 'millennium' babies born in a city hospital.

Weight (kg)	Frequency $f$	
$1.0 \leq w < 1.5$	2	
$1.5 \leq w < 2.0$	7	
$2.0 \leq w < 2.5$	10	
$2.5 \leq w < 3.0$	14	
$3.0 \leq w < 3.5$	9	
$3.5 \leq w < 4.0$	8	
$4.0 \leq w < 4.5$	5	
$4.5 \leq w$	0	
	Total = 55	

- Construct a cumulative frequency table for the data.
- Draw a cumulative frequency curve.



- c) Estimate, from your graph, the median and quartiles for the weights of the babies.
- d) Calculate an estimate for the MEAN weight and the Standard Deviation of the babies, rounded to 2 decimal places .

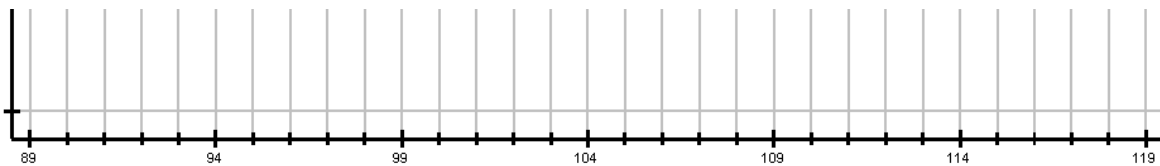
**Question 2**

*5 Number summary, IQR, Box & Whisker, Skewness*

Consider the IQ's of twenty-seven 15-year-old girls:

90 ; 90 ; 90 ; 90 ; 92 ; 93 ; 94 ; 96 ; 99 ; 100 ; 100 ; 100 ; 100 ; 100 ; 100 ; 101 ; 101 ; 101 ; 102 ; 103 ; 103 ; 105 ; 106 ; 106 ; 110 ; 113 ; 116

- a) Give the five number summary:
- b) Use your answer to the question above to draw a box and whisker plot on the axes provided.



- c) Calculate the IQR.
- d) Is the data positively or negatively skewed? Justify your answer.





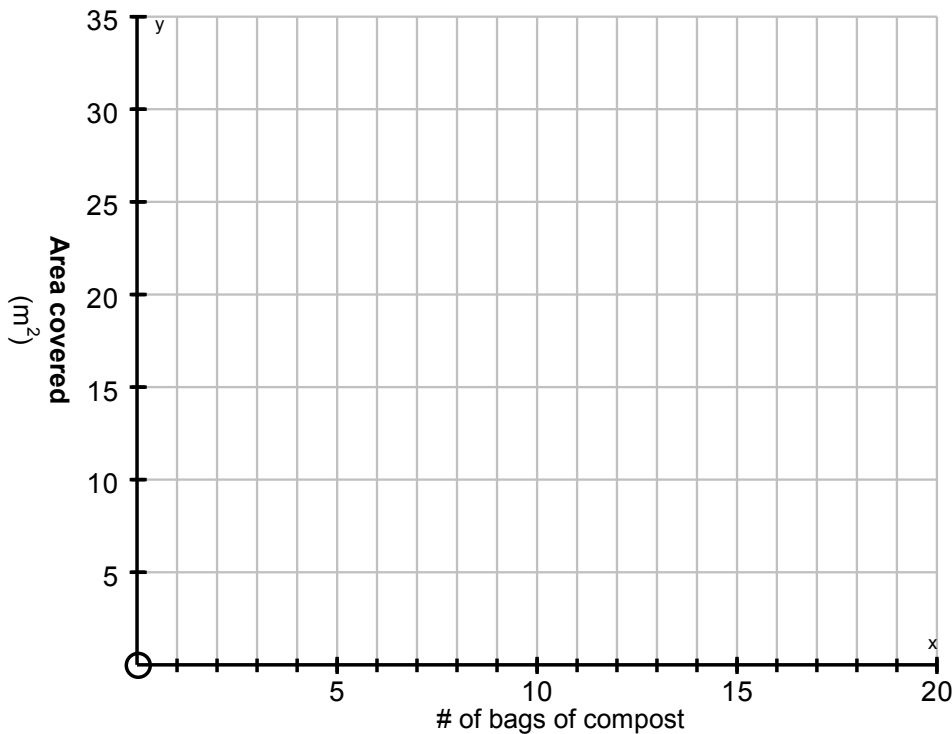
### Question 3

*Bivariate data – line of best fit and  $r$*

The following data is a summary of the number of bags of compost ( $x$ ) used to cover an area of ground ( $y$ ) in square metres.

$x$	15	6	10	18	9	7	14	11	5	8
$y$	35	18	24	32	24	20	32	29	14	22

- a) Draw a scatter plot of  $x$  and  $y$ .



- b) Find the equation of the best fit line (rounded to 2 decimal places).  
c) Estimate the area of ground (in  $m^2$ ) covered by 16 bags.  
d) Draw in the line of best fit, showing 2 significant points.  
e) Determine the value of  $r$  and interpret the significance of this value.

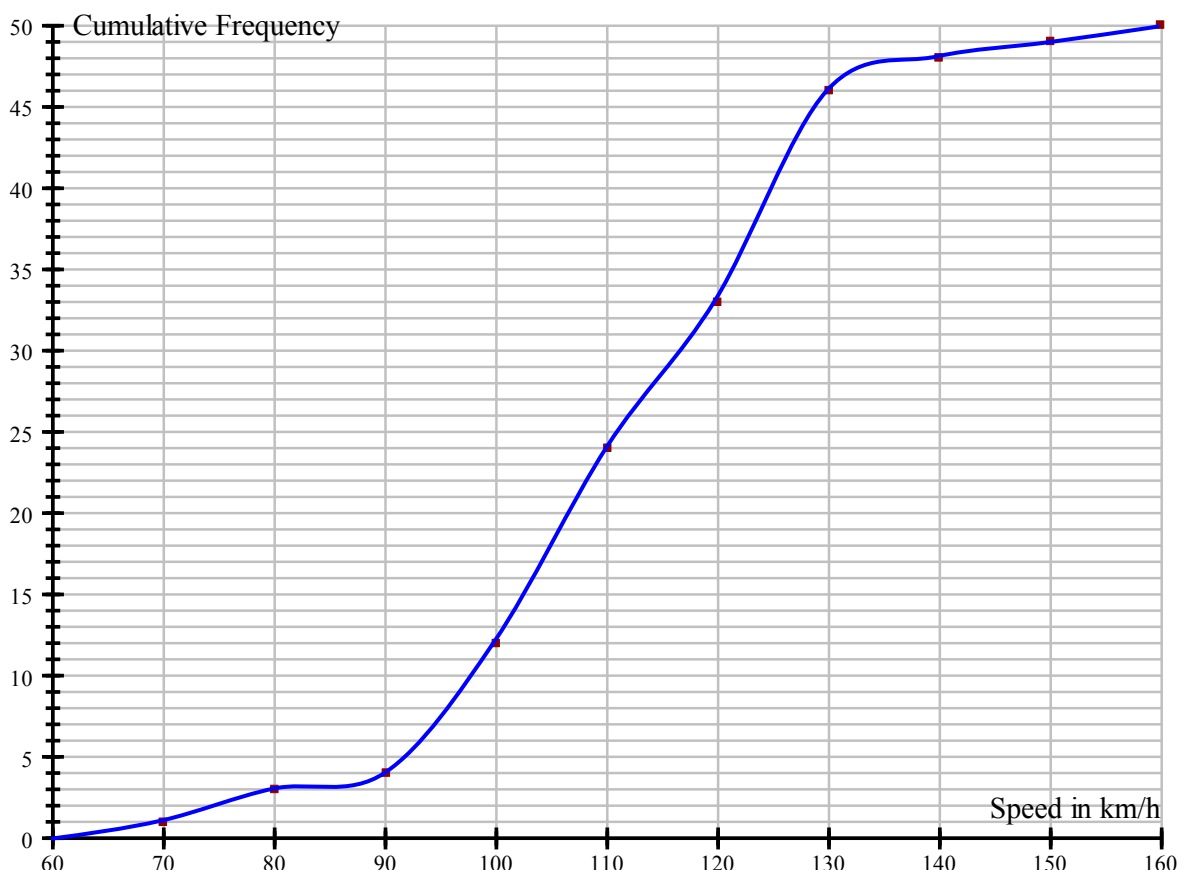


**Question 4**

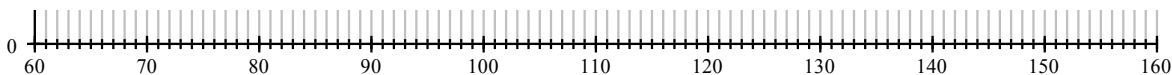
*IEB Exemplar Question*

The speeds of 50 motorists were recorded on the N3 between Durban and Johannesburg. The speed limit on this particular stretch of road is 120km/h.

The ogive curve, showing the relationship between the speeds of the cars versus the cumulative frequency, is shown below.



- a) How many cars were travelling within the legal speed of 120km/h? (1)
- b) What was the median speed? (1)
- c) Above what speed were the fastest 25% of the drivers driving and at which quartile is this represented? (2)
- d) If the lowest speed recorded was 62km/h and the fastest was 158km/h, draw a box and whisker plot of the data from the ogive plot using the axes provided. (3)





- e) Complete the table, in the answer booklet, using the information from the given ogive curve. (2)

Speed	Frequency
$60 < x \leq 70$	1
$70 < x \leq 80$	2
$80 < x \leq 90$	
$90 < x \leq 100$	8
$100 < x \leq 110$	
$110 < x \leq 120$	
$120 < x \leq 130$	13
$130 < x \leq 140$	2
$140 < x \leq 150$	
$150 < x \leq 160$	1

- f) Using the values calculated above, calculate an estimate for mean speed and the standard deviation, rounded to 1 decimal digit. (2)





## SOLUTIONS TO STATISTICS

### Question 1

a)

Weight (kg)	Frequency $f$	Cumulative Frequency
$1.0 \leq w < 1.5$	2	2
$1.5 \leq w < 2.0$	7	9
$2.0 \leq w < 2.5$	10	19
$2.5 \leq w < 3.0$	14	33
$3.0 \leq w < 3.5$	9	42
$3.5 \leq w < 4.0$	8	50
$4.0 \leq w < 4.5$	5	55
$4.5 \leq w$	0	55
	Total = 55	

b)



c)  $Q_1 = 2,2 \text{ kg}$ ;  $Q_2 = 2,8 \text{ kg}$ ;  $Q_3 = 3,4 \text{ kg}$

d) Est. Mean = 2,84 (2 d.p.)

$$\bar{x} = \frac{2 \times 1,25 + 7 \times 1,75 + 10 \times 2,25 + 14 \times 2,75 + 9 \times 3,25 + 8 \times 3,75 + 5 \times 4,25}{55} = 2,84$$

Std Deviation:  $\sigma = 0,80$

### Question 2

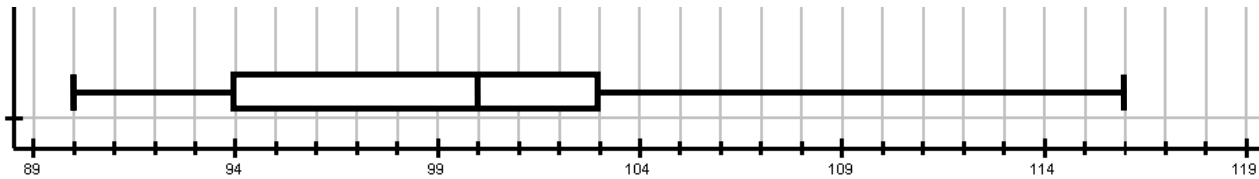
$n = 27$ .

90 ; 90 ; 90 ; 90 ; 92 ; 93 ; **94** ; 96 ; 99 ; 100 ; 100 ; 100 ; 100 ; **100** ; 100 ; 101 ; 101 ; 101 ; 102 103 ; **103** ; 105 ; 106 ; 106 ; 110 ; 113 ; 116



a) Min: 90;  $Q_1 = 94$ ;  $Q_2 = 100$ ;  $Q_3 = 103$ ; Max: 116

b)

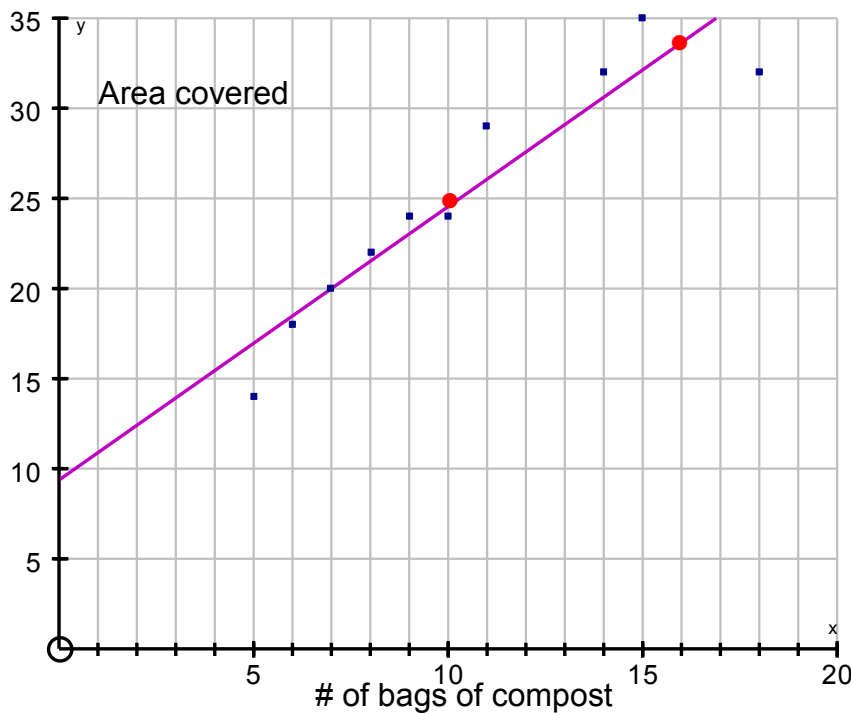


c)  $IQR = Q_3 - Q_1 = 103 - 94 = 9$

d) The data is positively skewed since the greatest distance from the median is to the right 'whisker'.

**Question 3**

a) Scatter plot of  $x$  and  $y$ .



b) Equation of the best fit line:  $\hat{y} = 9,37 + 1,52x$

c) Substitute  $x = 16$ :  $\hat{y} = 9,37 + 1,52(16) = 33,69\text{m}^2$

d) Line of best fit: Use points  $(\bar{x}; \bar{y}) = (10,3; 25)$  and any other point.

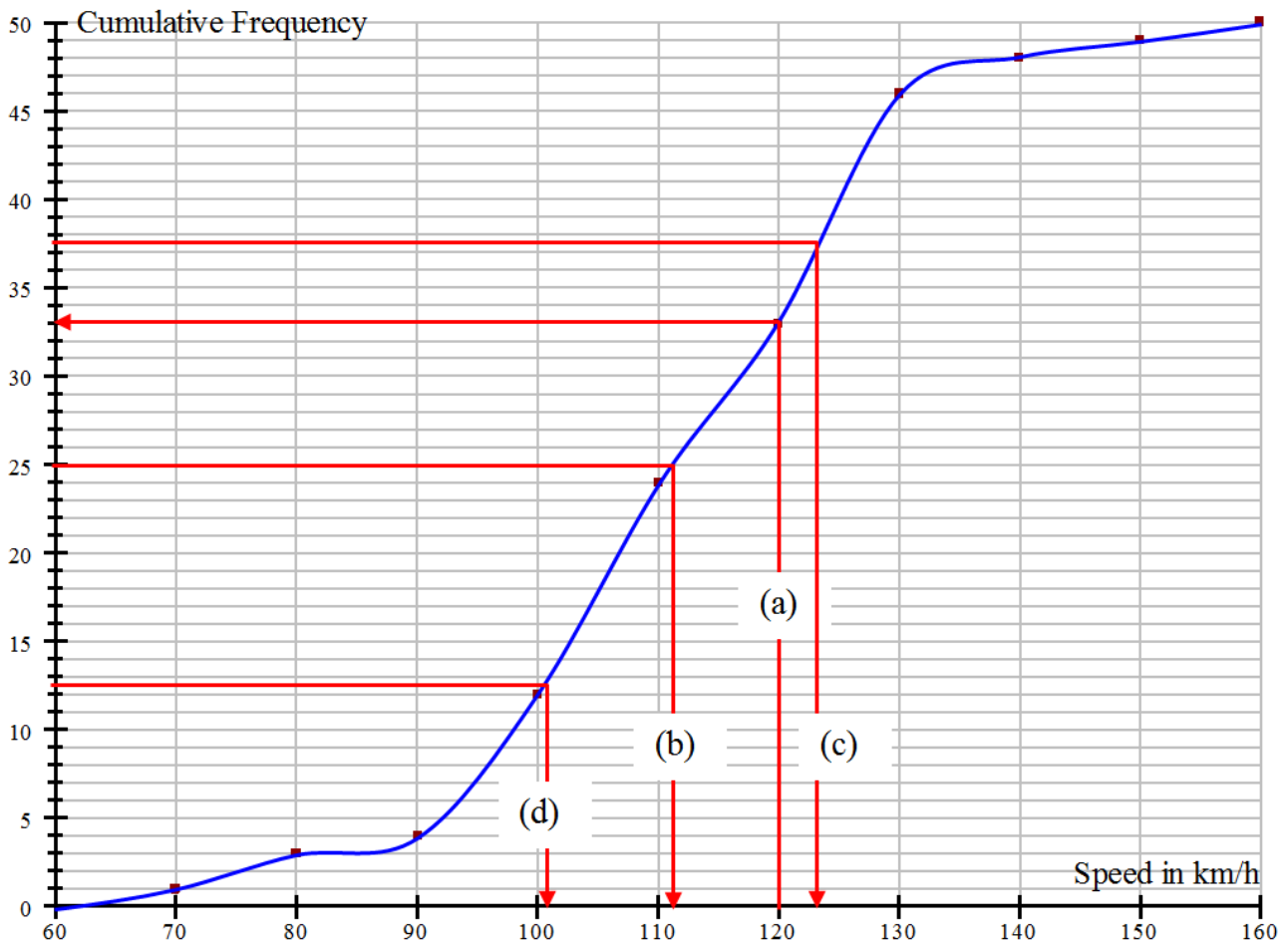
Using  $x = 16 \rightarrow (16; 33,69)$

e)  $r = 0,9371$

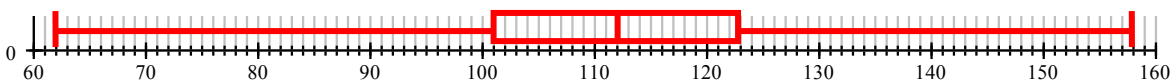
This means that there is a very strong, positive linear correlation between the given variables.



Question 4



- a) 33 cars within speed limit of 120 km/h
- b) median speed =  $\pm 112$  km/h (Read off  $\pm 25$  cars on y-axis)
- c) Fastest 25%  $\rightarrow$  Upper Quartile ( $Q_3$ )  
Read from  $\pm 0,75 \times 50 = 37,5$  on y-axis: i.e.  $\pm 123$  km/h
- d) Find ( $Q_1$ ) from curve:  $\pm 101$  km/h  
Min: 62; Max 158  
 $Q_1 = 101; Q_2 = 112; Q_3 = 123$







e)

Speed	Frequency
$60 < x \leq 70$	1
$70 < x \leq 80$	2
$80 < x \leq 90$	1
$90 < x \leq 100$	8
$100 < x \leq 110$	12
$110 < x \leq 120$	9
$120 < x \leq 130$	13
$130 < x \leq 140$	2
$140 < x \leq 150$	1
$150 < x \leq 160$	1

f)

$$\bar{x} = 111 \text{ km/h}$$

$$\sigma = 17,2 \text{ km/h}$$