

IEB Statistics practice — Memorandum

Every answer below was confirmed by an independent verifier at the moment the question was made.

The least-squares line of best fit

1

(a) Answer: $y = 98.357 - 4.028x$

1. $A = 98.357$

2. $B = -4.028$

(b) Answer: $r = -0.99$

1. $r = -0.99$

(c) Answer: a strong negative correlation

1. $r = -0.99$: $r < 0$ and $|r|$ is at least 0.8

(d) Answer: $y = 78.22$

1. $y = 98.357 - 4.028(5)$

2. $y = 78.22$

(e) Answer: Yes, reasonably: x lies within the data (interpolation)

1. the data runs from $x = 0$ to $x = 14$; $x = 5$ lies inside it

The estimated mean of grouped data

2

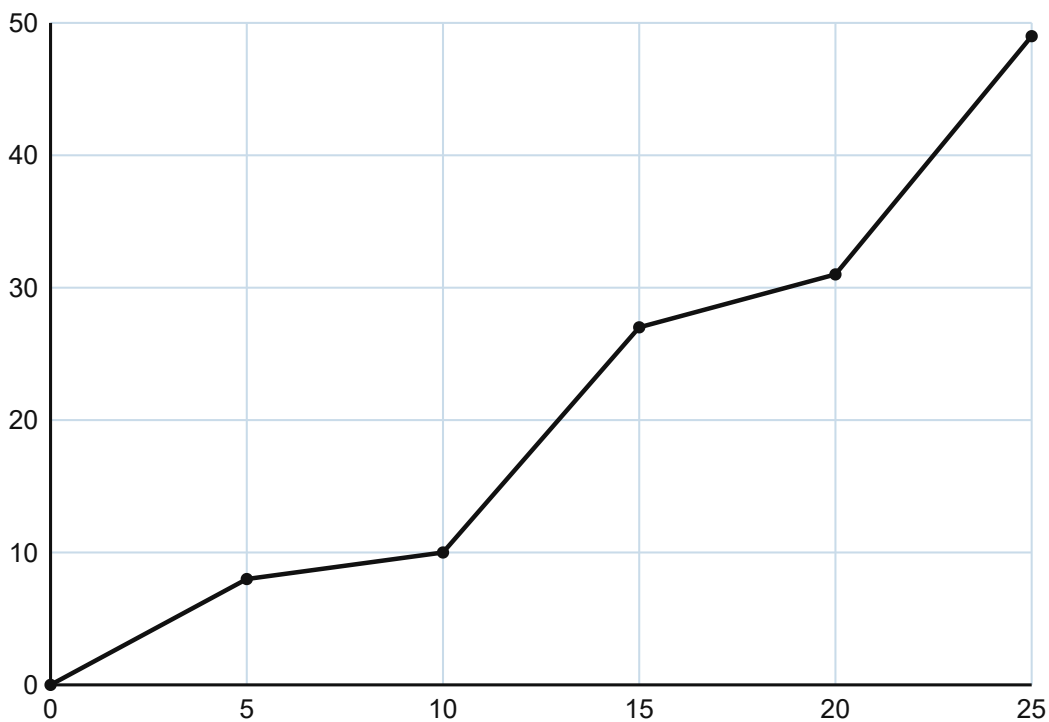
(a) Answer: 14.74

1. midpoints: 2.5, 7.5, 12.5, 17.5, 22.5
2. $\Sigma f \cdot x = 722.5$ and $n = 49$
3. mean = $\frac{722.5}{49} = 14.74$

(b) Answer: $20 \leq x < 25$

1. the highest frequency is 18

(c) Answer: See the ogive below.



1. cumulative frequencies: 8, 10, 27, 31, 49
2. plot each against the upper limit of its class, starting at (0; 0)

(d) Answer: about 14.3

1. the median is at a cumulative frequency of $\frac{49}{2} = 24.5$
2. read across to the ogive and down: about 14.3

(e) Answer: about 10.9

1. Q1 at 12.25: about 10.7; Q3 at 36.75: about 21.6

2. $IQR = 21.6 - 10.7 = 10.9$

(f) Answer: Skewed to the right (positively)

1. mean 14.74 and median about 14.3

2. mean $>$ median

(a) Answer: $x = 5$

1. $\frac{515 + 45x}{35 + x} = 18.5$

2. $515 + 45x = 18.5(35 + x)$

3. $x = 5$

(b) Answer: $0 \leq x < 10$

1. the highest frequency is 15

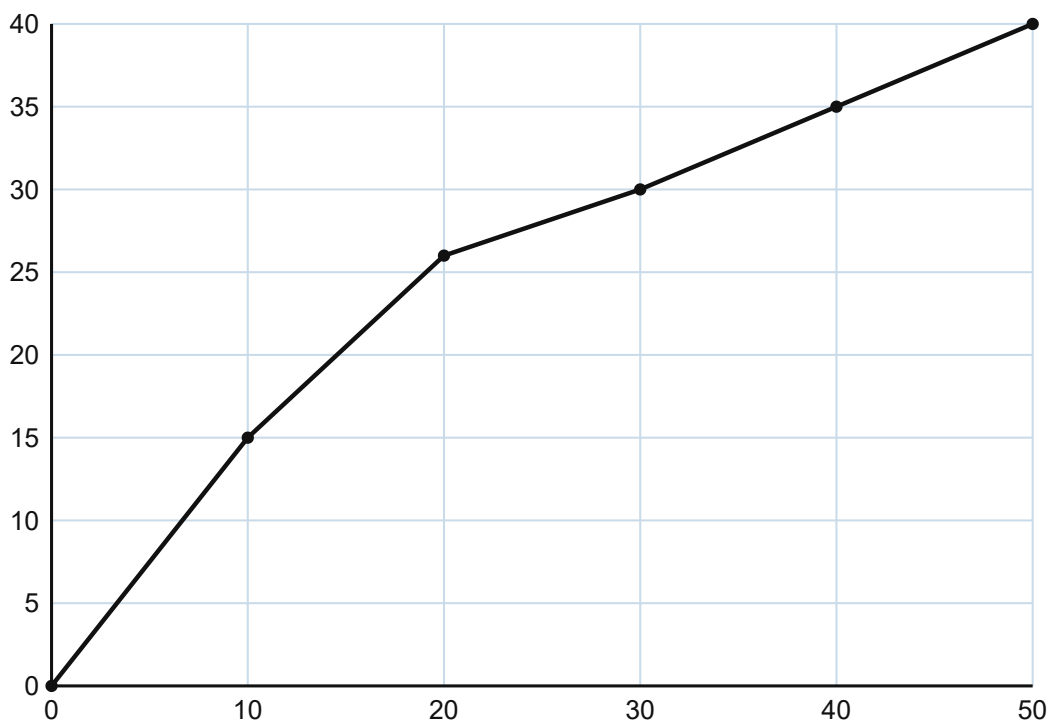
(c) Answer: $10 \leq x < 20$

1. $n = 40$: the median lies between the 20th and 21st values

2. cumulative frequencies: 15, 26, 30, 35, 40

3. $10 \leq x < 20$

(d) Answer: See the ogive below.



1. cumulative frequencies: 15, 26, 30, 35, 40

2. plot each against the upper limit of its class, starting at (0; 0)

Scatter plots

4

(a) Answer: (30; 71)

1. (30; 71) lies far from the trend of the other points
-

(b) Answer: $y = -55.747 + 5.604x$

1. $A = -55.747$
 2. $B = 5.604$
-

(c) Answer: $r = 1.00$

1. $r = 1.00$
-

(d) Answer: y increases by about 5.604

1. $B = 5.604$ is the change in y for each 1 added to x
-

(e) Answer: Weaker

1. with the outlier $r = 0.94$; without it $r = 1.00$
2. $|r|$ is smaller with the outlier: it weakened the correlation

Mean, median and mode

5

(a) Answer: 19.25

1. $\Sigma x = 231$ and $n = 12$

2. $\text{mean} = \frac{231}{12} = 19.25$

(b) Answer: $\sigma = 6.19$

1. $\sigma^2 = \Sigma(x - 19.25)^2 / 12 = 38.35$

2. $\sigma = 6.19$

(c) Answer: 9

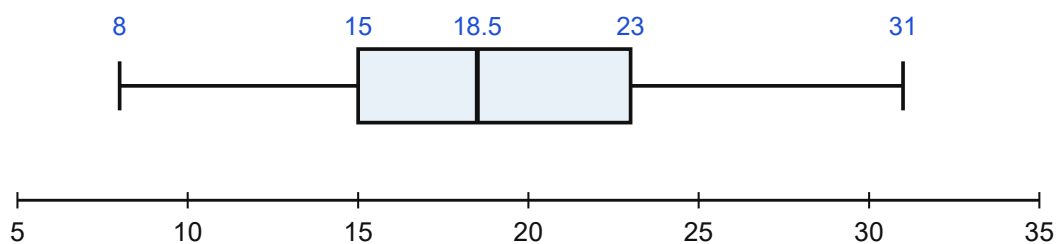
1. between 13.06 and 25.44

2. 9 values

(d) Answer: 8; 15; 18.5; 23; 31

1. minimum 8; Q1 15; median 18.5; Q3 23; maximum 31

(e) Answer: See the plot below.



1. the box from Q1 to Q3 with the median marked; whiskers to the minimum and maximum

(f) Answer: No: $31 \leq 35$

1. $\text{IQR} = 23 - 15 = 8$

2. $\text{Q3} + 1.5 \times \text{IQR} = 23 + 12 = 35$

3. $31 \leq 35$

(g) Answer: The mean increases by 10; the standard deviation stays the same.

1. every value, and so their mean, moves by the same amount
2. the distances from the mean do not change

6

(a) Answer: $x = 19$

1. $n = 7$: the median is the 4th value, $x + 10$
2. $x + 10 = 29$, so $x = 19$

(b) Answer: 28.14

1. the data: 5; 23; 24; 29; 37; 38; 41
2. mean = $\frac{197}{7} = 28.14$

(c) Answer: 11.47

1. $\sigma = 11.47$ (calculator)

(d) Answer: 1

1. mean + $\sigma = 28.14 + 11.47 = 39.61$
2. 1 value(s) above it

(e) Answer: The mean and the standard deviation; the median stays the same

1. the total grows by 10, so the mean grows
2. the top value moves further from the rest, so σ grows
3. the middle value does not move