

Chapter 1 Exam Review

MDM4U

Jensen

SOLUTIONS

Section 1.1 - Visual Displays of Data

1. The number of goals by Jaromir Jagr in each of his 19 NHL seasons is recorded below

~~21~~, ~~32~~, ~~34~~, ~~32~~, ~~32~~, ~~62~~, ~~41~~, ~~38~~, ~~44~~, ~~42~~, ~~52~~, ~~31~~, ~~38~~, ~~31~~, ~~54~~, ~~30~~, ~~25~~, ~~19~~, ~~16~~

a) Construct a stem-and-leaf plot to display the data

Stem	Leaf
1	6, 9
2	5, 7
3	0, 1, 1, 2, 2, 2, 4, 5, 6
4	2, 4, 7
5	2, 4
6	2

b) Determine the percent of seasons where greater than 35 goals were scored.

$$\frac{7}{19} = 36.8\%$$

b) Find the median of the data (Q_2)

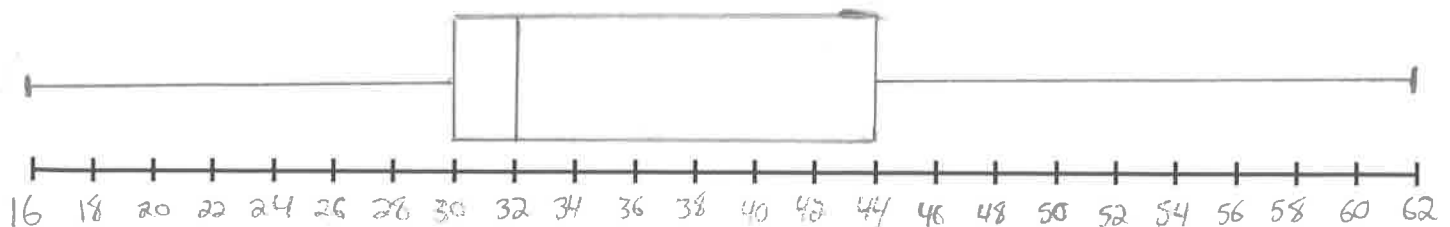
32

c) Find the median of the lower half of the data Q_1 and the median of the upper half of the data Q_3

$$Q_1 = 30$$

$$Q_3 = 44$$

d) Create a box and whisker plot for the data.



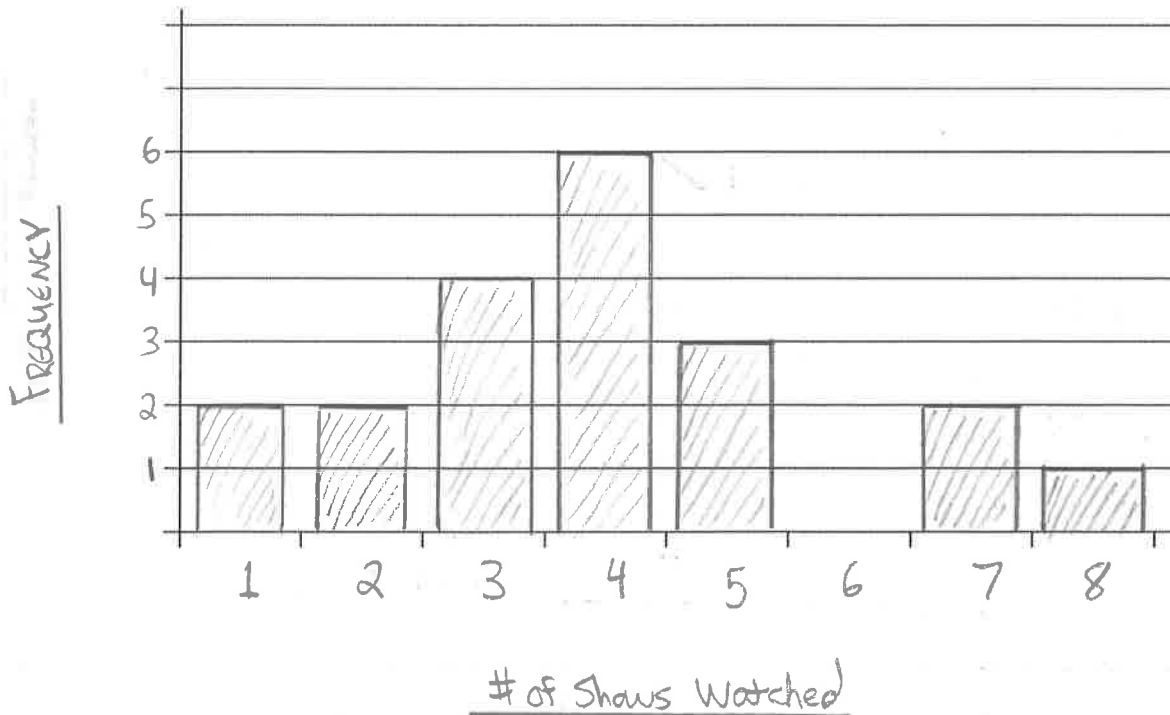
2. A group of teenagers were asked how many different television shows they watch on a regular basis. The results are as follows:

2, 4, 3, 5, 4, 4, 1, 7, 8, 1, 5, 3, 4, 3, 3, 4, 2, 5, 4, 7

a) Organize the data in a frequency table

# of shows watched	Tally	Frequency
1		2
2		2
3		4
4		6
5		3
6		0
7		2
8		1

b) Make a bar graph of the data



3. The heights of the 2013 Toronto Raptors (in centimeters) are listed below:

201, 183, 191, 211, 201, 201, 203, 213, 206, 206, 183, 208, 198, 198, 211

a) Determine the range of the data.

$$\text{Range} = 213 - 183 = 30$$

b) Determine an appropriate bin width that will divide the data into 7 intervals.

Round range
up to 35.

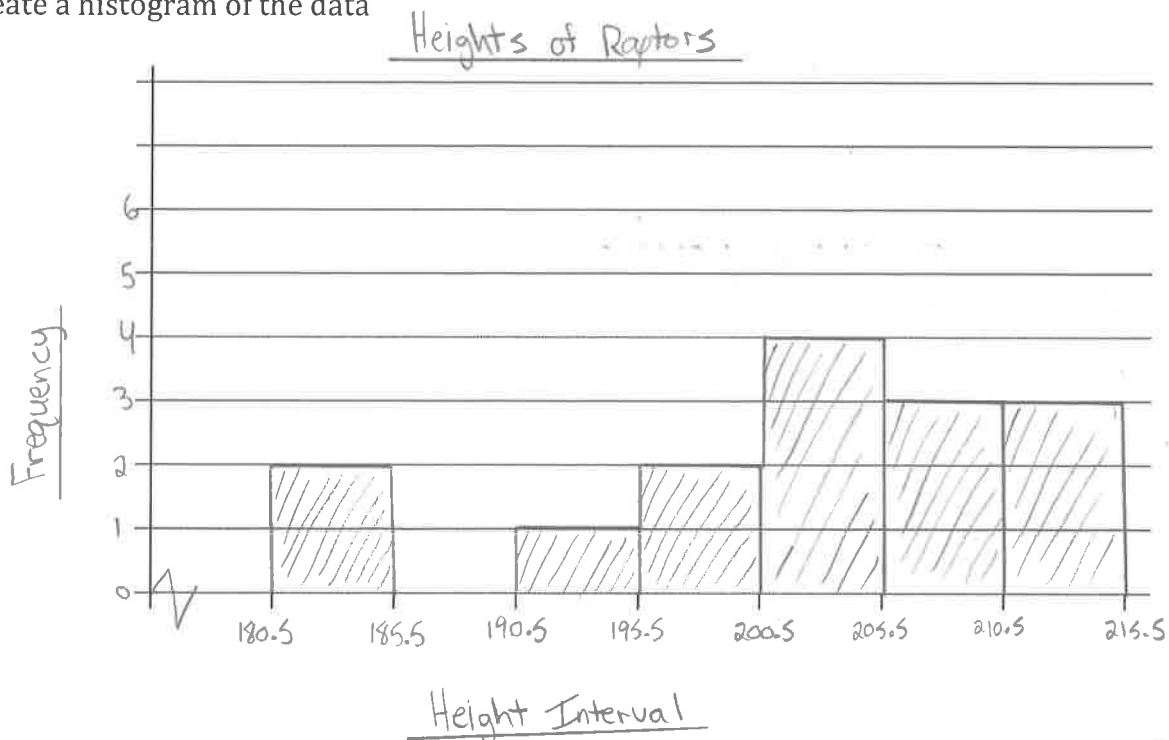
$$\text{Bin width} = \frac{35}{7} = 5$$

c) Create a frequency table for the data

$$\begin{aligned} \text{Starting Point} &= 183 - \frac{5}{2} \\ &= 183 - 2.5 \\ &= 180.5 \end{aligned}$$

Height Interval	Frequency
180.5 - 185.5	2
185.5 - 190.5	0
190.5 - 195.5	1
195.5 - 200.5	2
200.5 - 205.5	4
205.5 - 210.5	3
210.5 - 215.5	3

d) Create a histogram of the data

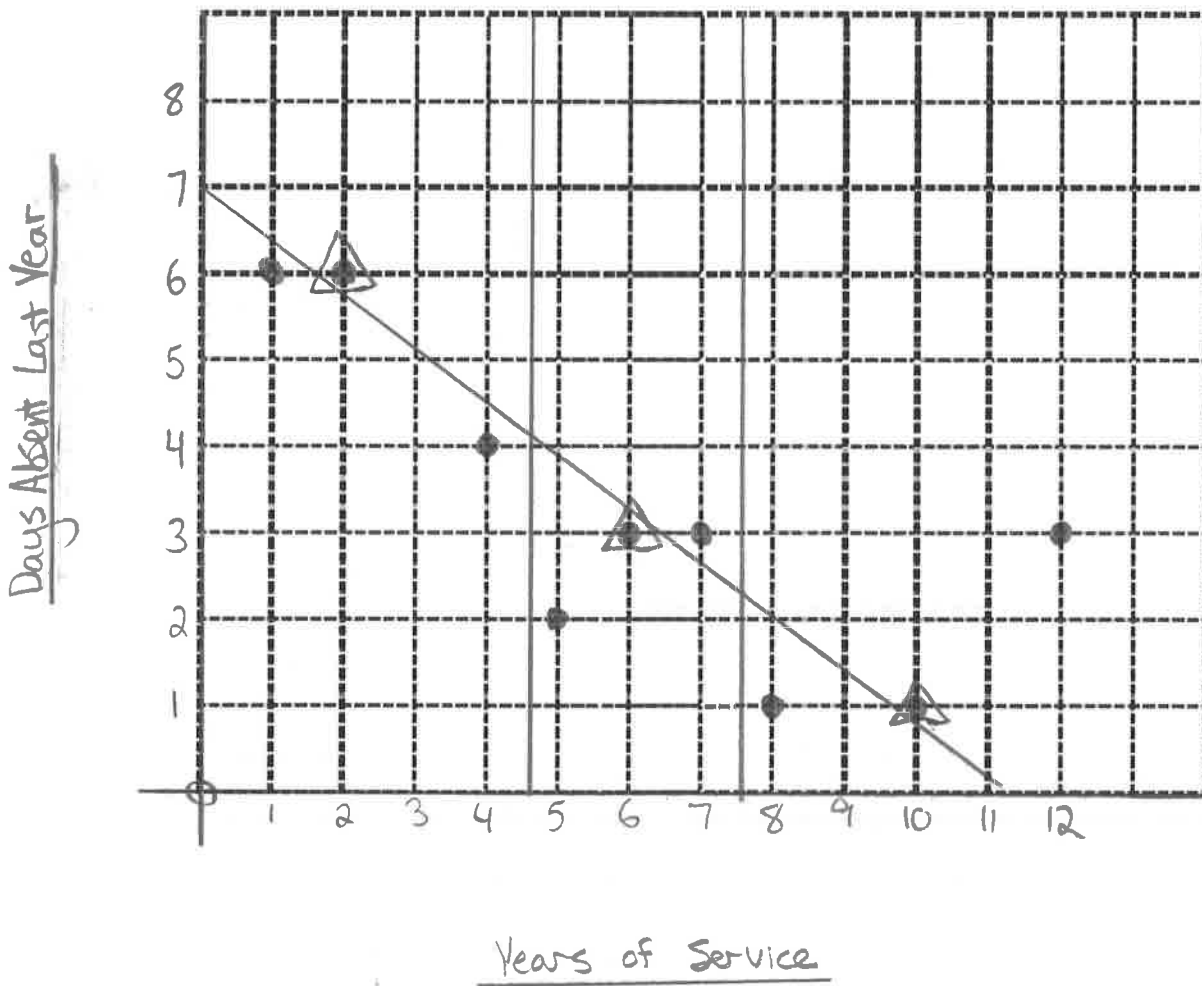


4. A company studied whether there was a relationship between its employees' years of service and number of days absent from work. The data for nine randomly selected employees are shown below.

x Years of Service	5	2	7	6	4	8	1	10	12
y Days Absent Last Year	2	6	3	3	4	1	6	1	3

a) Create a scatter plot for the information in the above table.

b) Construct a median-median line for the above data.



5. Two variables have a correlation coefficient of $r = 0.9$. This indicates

- a. a strong positive correlation
- b. a weak positive correlation
- c. a strong negative correlation
- d. a weak negative correlation

6. If two variables have no correlation, their correlation coefficient would have a value of

- a. +1
- b. -1
- c. 100
- d. 0

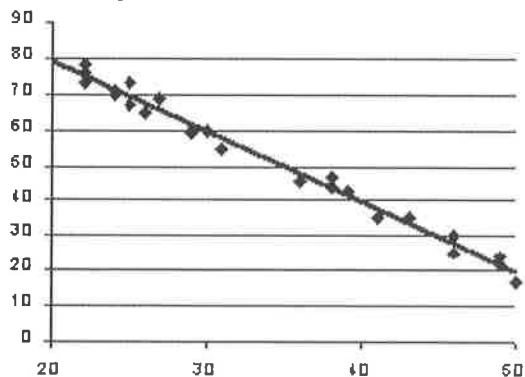
7. Two variables have a coefficient of determination of 0.64. The correlation coefficient could be

- a. -0.64
- b. 0.41
- c. -0.8
- d. 0.36

8. A relationship in which all data values lie on the regression line has a correlation coefficient of

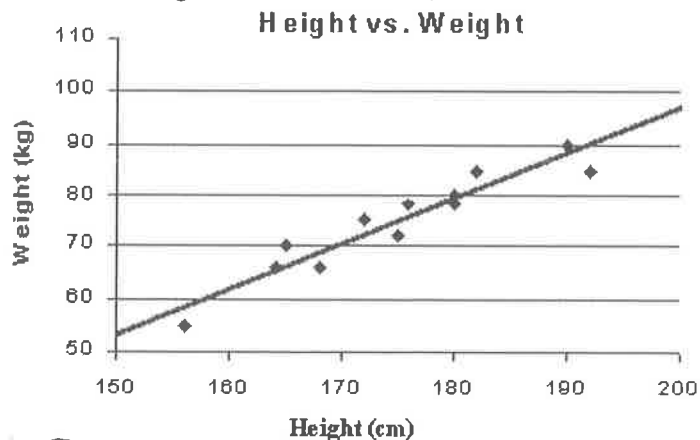
- a. 1
- b. 0
- c. -1
- d. +1 or -1

9. The regression line shown would have a correlation coefficient closest to



- a. +1
- b. 0.5
- c. -1
- d. 0

10. For the regression line shown, the coefficient of determination would be closest to



- a. +1
- b. 0
- c. -1
- d. 0.25

11. The residuals for a set of data represent the

- a. differences between consecutive x -values
- b. vertical differences between data points and the line of best fit
- c. data points that lie below the line of best fit
- d. data points that do not lie on the line of best fit

12. If a set of data has a very strong correlation, the residual values will be

- a. very large
- b. positive
- c. negative
- d. very small

13. A coefficient of determination, $r^2 = 0.75$, indicates that

- a. 75% of the data lie on the regression line
- b. the slope of the regression line is 0.75
- c. 75% of the variance in y is a result of the variance in x
- d. the data have a strong positive correlation

14. Which of the following statements is not true about variables having a correlation coefficient of 1?

- a. all data points lie on the line of best fit
- b. data points with larger x -values also have larger y -values
- c. there is a cause-effect relationship between the variables
- d. all of the residual values are zero

15. Which of the following is an example of a negative correlation?

- a. amount of studying and mark on a test
- b. temperature and number of kids at a pool
- c. a person's arm length and leg length
- d. number of people and slices of pizza per person

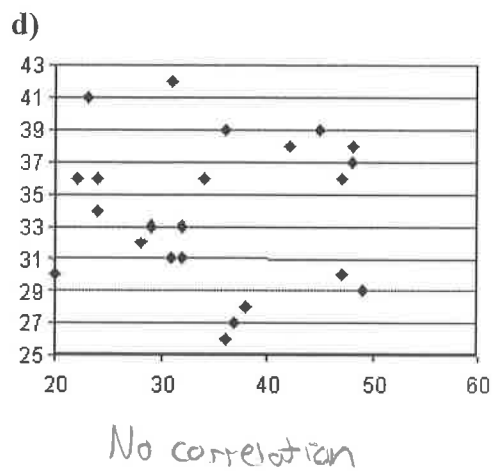
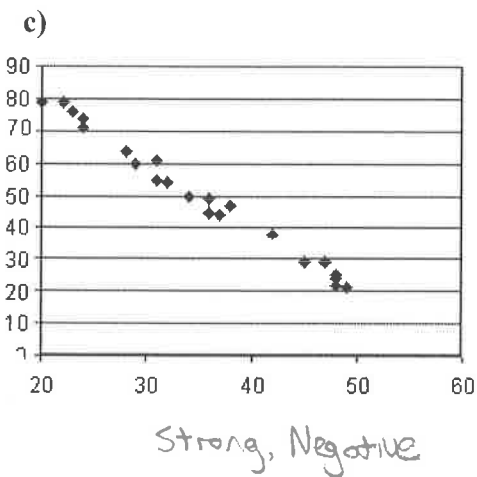
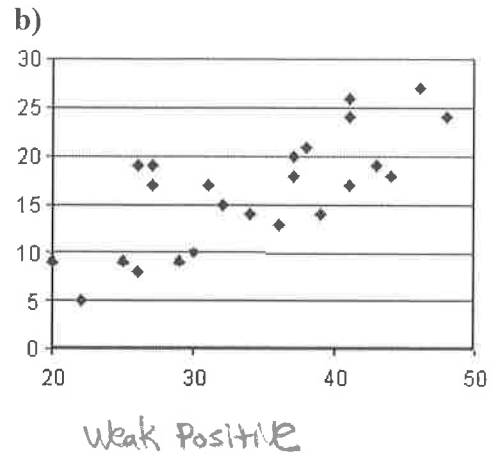
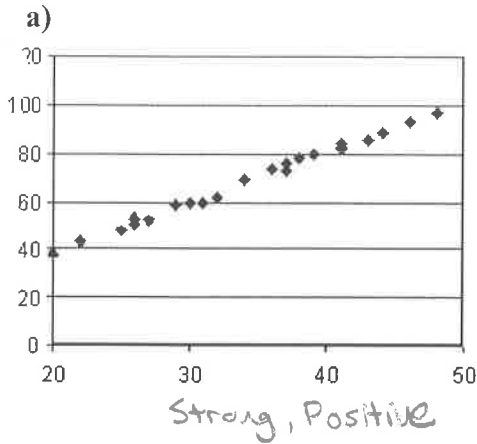
16. A set of data having small residual values means that

- a. the correlation coefficient is close to 0
- b. there is a positive correlation
- c. there is a negative correlation
- d. there is a strong correlation

17. A positive residual value means that the data point lies:

- a) close to the line of best fit
- b) above the line of best fit
- c) on the line of best fit

18. Describe the type of correlation represented by each of the following scatter plots.



19. A number from 0 to 1 that gives the relative strength of the relationship between two variables. It represents the percentage of variation in y due to the variation in x.

- a) coefficient of determination
- b) residual
- c) coefficient of correlation
- d) median value

20. What type of correlation is represented when the coefficient of correlation is -0.96 ?

- a) weak positive
- b) strong positive
- c) weak negative
- d) strong negative

21. What type of correlation is represented when the coefficient of ~~determination~~^{correlation} is 0.21 ?

- a) weak negative
- b) no correlation
- c) strong positive
- d) weak positive

22. This table shows the data for the full-time employees of a small company.

Age (year)	33	25	19	44	50	54	38	29
Annual Income (in thousands)	33	31	18	52	56	60	44	35

a) Use the TI-83 calculator to create a scatter plot and line of best fit. State the equation of the line of best fit, to 3 decimal places.

$$y = 1.150x - 0.864$$

b) Use the line of best fit to calculate the expected annual income for someone 45 years old.

$$y = 1.150(45) - 0.864$$

$$y = 50.886$$

c) State the correlation coefficient. Explain what it represents in this particular situation.

$$r = 0.98$$

Strong Positive Correlation between age and income.

d) State the coefficient of determination. Explain what it represents in this particular situation

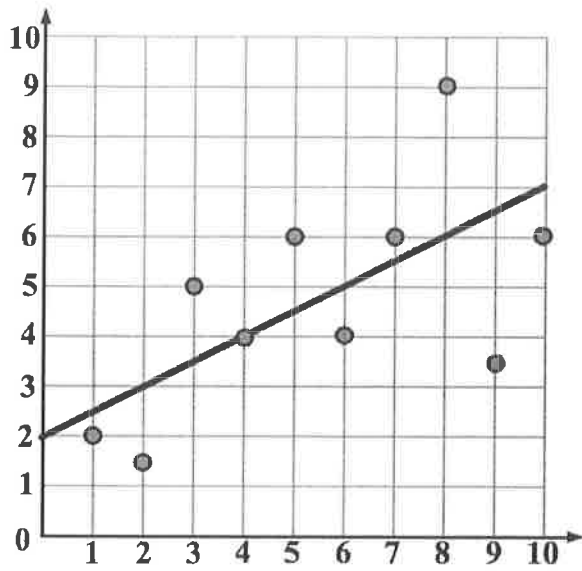
$$r^2 = 0.965$$

96.5% of the variation in income is due to the variation in age.

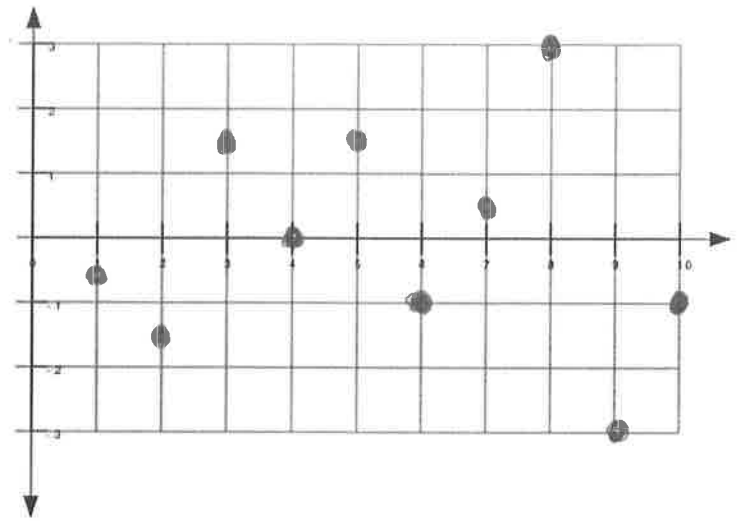
e) Calculate the residual for each point using the TI-83 calculator. List all the residuals in the table below.

Age (year)	33	25	19	44	50	54	38	29
Annual Income (in thousands)	33	31	18	52	56	60	44	35
Residuals	-4.099	3.1044	-2.993	2.2472	-0.6551	-1.257	1.1494	2.5028

23. Plot the residuals for the following scatter plot

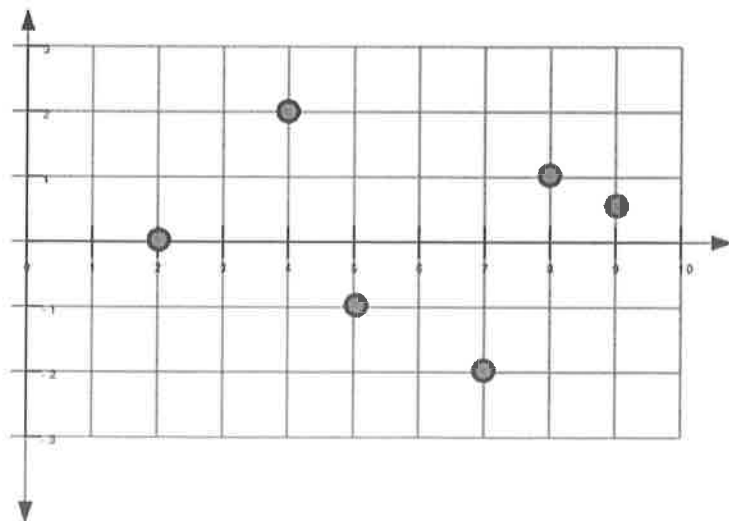


Residual Plot

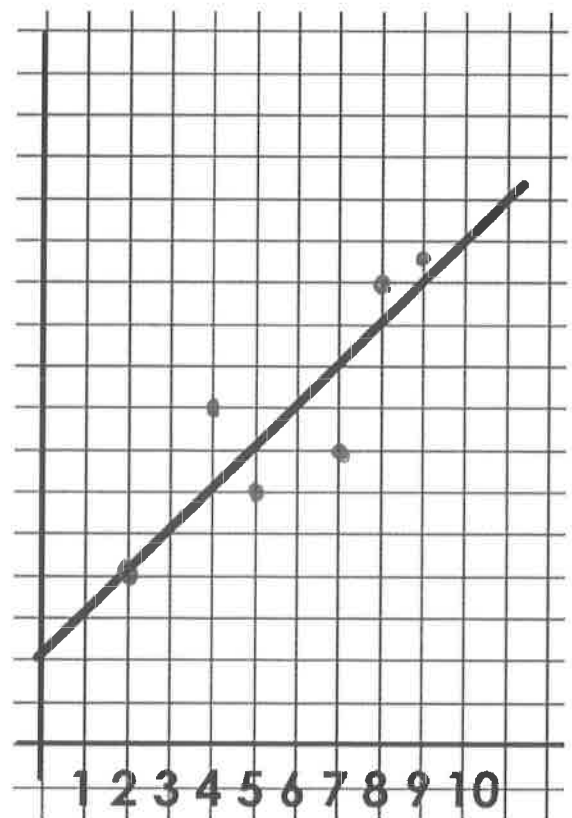


24. Sketch a graph that corresponds to the following residual plot. Label your axis. (3 marks)

Residual Plot

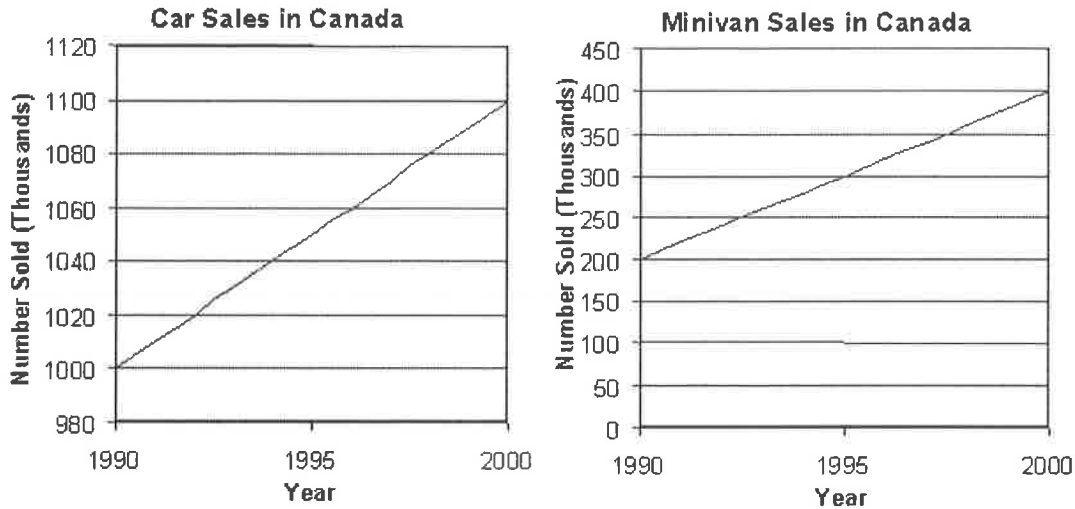


Scatter Plot



25. The following graphs represent the number of cars and minivans sold in Canada over time. Discuss how the caption misrepresents the data.

Car Sales Outpace Van Sales!



- Car sales have increased by 100,000 in the last 10 years.
- Minivan sales have increased by 200,000 in the last 10 years.
- The scale on the car graph makes it appear as though it is increasing at a higher rate; but minivan sales are increasing at a higher rate.