

Question 1

$$\text{Average SAT score} = \frac{\sum x}{n} = 1082.625$$

Students who received the same score: 2 students, with a score of 1210.

$$\begin{aligned}\text{Range} &= \text{maximum score} - \text{minimum score} \\ &= 1300 - 825 \\ &= 475\end{aligned}$$

How much do scores vary from the average SAT score?

$$\text{std variance} = \sqrt{\frac{1}{n} \sum (x - \bar{x})^2}$$

$$= 161.41$$

$$\text{variance} = 161.41^2 =$$

Question 2.

What type of data do experiments gather?

- (i) Quantitative data: Data expressed in numbers
- (ii) Qualitative data: Data for description without measurement.
- (iii) Primary data: Data obtained directly from individuals.
- (iv) Secondary data: Data that has been collected by another researcher.

Question 3.

How is direct observation different from secondary data collection?

Observation is an example of primary data, since the researcher collected the data themselves whereas secondary data is used when the researcher did not collect the data personally.

Question 4

Year	Men	women
2004	713	573
2005	722	585
2006	743	600
2007	766	614
2008	798	638

Justify why the data in the table is cross-sectional. Give examples that support your answer.

The data is cross-section because:

- (i) The ~~data~~ ^{study} was took place at a single point of time every year. eg in the year 2004 the median weekly earnings is shown.
- (ii) It does not involve manipulating variables. The information that was present in a population was recorded.
- (iii) Numerous characteristics: It has data for both men and women.
- (iv) It can be used to look at the prevailing characteristics in the population.

Question 5.

Discrete data: Information that can only take certain values.

Continuous data: Data that can take any value. eg Height, weight.

Question 6.

Pareto chart: It is a type of chart that contains both bars and line graphs, where individual values are represented in descending order by bars and cumulative total is represented by the line.

It is used in business to highlight where defects originate from in a operational process.

Question 7

Which statistical measurement is most affected by outliers in a dataset? Mean.

Reasons:

This is because some observations lie too far away from the mean. It includes all values in a distribution and the outlier can increase or decrease the mean value.

Example

Question 8. * $X = 5, 12, 24, 28, 30$
 $\bar{X} = 19.8$

Question 8.

Amount (\$) Percentage (%)

less than 25000 39

25001 - 100,000 14

100,001 - 250,000 9

250001 - 500,000 7

Not sure 24

Bar

Answer: Pareto chart.

The data is ordinal, and it the last row of the table is larger than preceeding row.

Question 9

1264, 1264, 1296, 1104, 953, 889, 891, 846, 730, 791, 877, 923, 919, 976, 1020, 1057, 1036

$$\text{Mean} = \bar{X} = \frac{\sum X}{n} = \frac{15632}{16} = 977$$

$$\text{Standard deviation} = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} = 158.2$$

$$C.V = \frac{977}{158.2} \times 100 = 61.7$$

Ans: a. 977, 158.2, 16.1

Question 10

Why can't you depend on using the range of a large data set?

Ans

Range ^{describes} ~~represents~~ how well the central tendency describes the data. If the range is large, the central tendency is not as representative as it would be if the range would be small.

Question 11

Number of minutes per
customer

	frequency.	X	fx	$(x - \bar{x})^2$
0 to under 10	6	5	30	829.44
10 to under 20	22	15	330	353.44
20 to under 30	10	25	250	77.44
30 to under 40	18	35	630	1.44
40 to under 50	28 28	45	1260	125.44
50 to under 60	16	55	880	449.44
			3380	1836.64

$$\text{Average wait time} = \bar{x} = \frac{\sum fx}{\sum f} = \frac{3380}{100} = 33.8 \text{ mins.}$$

$$\begin{aligned} \text{Standard deviation} &= \sqrt{\frac{\sum (x - \bar{x})^2}{n}} = \sqrt{\frac{1836.64}{100}} = \sqrt{18.3664} \\ &= 4.29 \text{ mins.} \end{aligned}$$