

NAME: Solutions.

There are 6 problems with total of 25 parts. Each PART is worth 2 points. You have 50 minutes. Good luck.

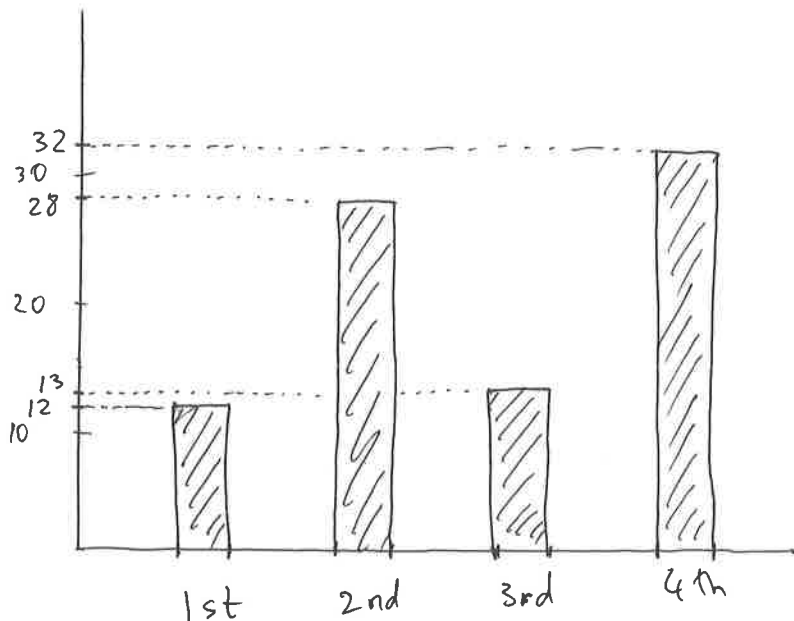
1. A high school competes in track events. Below are the total number of medals that the high school won.

1st Place	12
2nd Place	28
3rd Place	13
4th Place	32

a. What type of data does this data set represent – measurement or categorical?

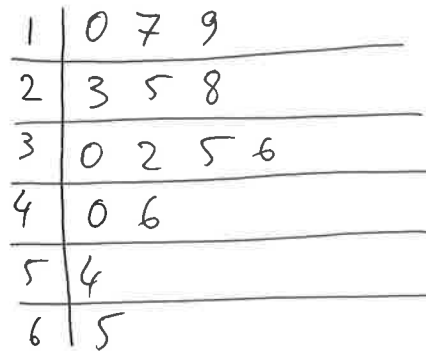
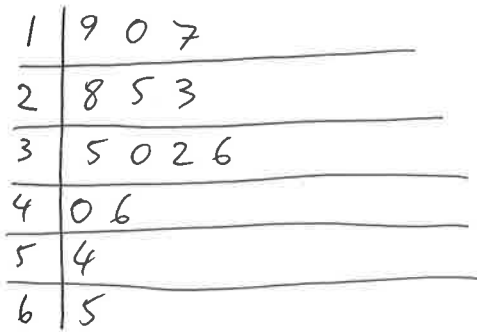
Categorical.

b. Make a bar graph for this data set.

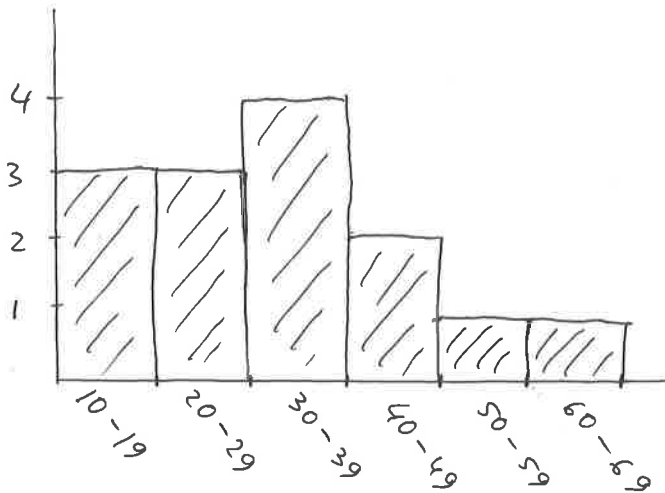


2. Consider the following data set: 28, 35, 19, 25, 30, 40, 32, 46, 54, 36, 23, 10, 65, 17.

a. Make a stem-and-leaf plot.



b. Make a histogram.



c. Rewrite the data set in increasing order and find the median.

10, 17, 19, 23, 25, 28, 30, 32, 35, 36, 40, 46, 54, 65.

$$\text{Median} = \frac{30 + 32}{2} = 31.$$

d. Find the 25th percentile.

Median of 10, 17, 19, 23, 25, 28, 30

$$25^{\text{th}} \text{ percentile} = 23.$$

e. Find the 75th percentile.

Median of 32, 35, 36, 40, 46, 54, 65

$$75^{\text{th}} \text{ percentile} = 40.$$

f. How do you detect outliers? I.e., what are upper and lower bounds for 'normal' data?

$$IQR = 40 - 23 = 17.$$

$$25^{\text{th}} \text{ percentile} - \frac{1}{2} \times IQR = 23 - \frac{1}{2} \cdot 17 = 23 - 8.5 = 14.5$$

$$75^{\text{th}} \text{ percentile} + \frac{1}{2} \times IQR = 40 + \frac{1}{2} \cdot 17 = 40 + 8.5 = 48.5$$

Numbers outside the interval $(14.5, 48.5)$ are outliers.

g. Find outliers, if any.

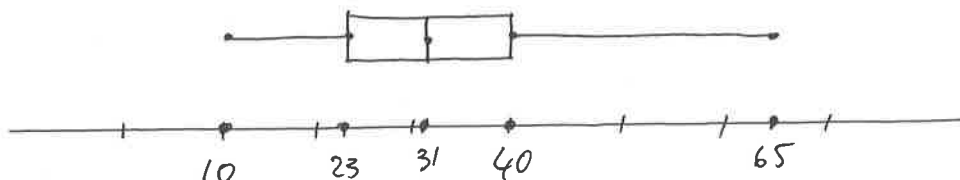
Since all data values lie inside the interval $(14.5, 48.5)$,

There are no outliers.

h. What is the five-number-summary for this data set?

10, 23, ~~30~~, 40, 65.

i. Make a box plot for this data set (include the outliers as separate points).



3. Consider the following data set: 1, 4, 5, 7, 3.

a. Find the mean.

$$\frac{1+4+5+7+3}{5} = \frac{20}{5} = 4.$$

b. Find the average deviations.

$$\frac{3+0+1+3+1}{5} = \frac{8}{5} = 1.6$$

c. Find the variance.

$$\frac{9+0+1+9+1}{5} = \frac{20}{5} = 4$$

d. Find the standard deviation.

$$\sqrt{4} = 2.$$

e. What does the standard deviation measure?

Standard deviation measures how spread out the data is. Larger the standard deviation, more spread out the data will be.

Data values	distances from the mean	squares of distances
1	3	9
4	0	0
5	1	1
7	3	9
3	1	1

4. Generate a data set with at least 5 values having a mean that is less than the first quartile score.

0, 10, 10, 10, 10.

first quartile score = 10

$$\text{mean} = \frac{0+10+10+10+10}{5} = \frac{40}{5} = 8 \quad 8 < 10. \checkmark$$

5. Linda went to a craft sale and bought five items. She spent an average of \$10 per item.

- a. Total of how much money did Linda spend?

$$5 \times 10 = 50.$$

- b. Generate a set of five prices where exactly one item cost \$10.

~~5 5 10 15 15~~
5 5 10 15 15

- c. Generate a set of five prices where all items cost different amounts.

2 5 10 15 18

- d. Generate a set of five prices where all items cost the same amount.

10 10 10 10 10

- e. Generate a set of five prices where all but one item cost more than \$10.

~~2 12 12 12 12~~

2 12 12 12 12

only extra condition is that the prices should add up to \$50.

6. Explain the following concepts and support your explanations with examples.

a. Statistical variable.

A statistical variable is a property on which we wish to collect data.

Example: height or weight of a person, prices for a home, ...

b. Categorical data.

Data that come from assigning objects or individuals to categories are called categorical data.

Example:

Data gathered in response to a question "What is your favorite color?" is going to be categorical data.

c. Measurement data.

Measurement data is data that can be measured, ordered and operated on.

Example: Data gathered in response to a question "What is your height?" is going to be measurement data.