

DESCRIPTIVE STATISTICS

3.1 Classification of data

- There are two types of data:
 - 1) Qualitative data – this data is descriptive or categorical. E.g. red or blue, tall or short, overweight or underweight etc.
- Quantitative data – this data is countable or measurable. This will largely be our focus for this topic.

Quantitative data

- Quantitative data is either **discrete** (countable for example litter size, number of children etc. These are basically whole numbers.) or **continuous** (measurable for example height, weight, volume etc. These can have in-between values).
- After you collect your data, it has to be summarized in a format that makes it easier to understand.

- **3.2 Summarizing data**
- Data can be summarized in the following ways:
 - a) Graphical Summary
 - Numerical Summary

3.2.1 Graphical summary of discrete data

- You can summarize discrete data using the following:
- Frequency tables
- Cumulative frequency tables
- Bar diagrams
- Dot diagrams/ dot plots

Example 1.

Twenty five litters are observed and the number of male piglets in a litter of 5 piglets are given as follows:

2, 3, 3, 2, 0, 3, 2, 3, 2, 1, 2, 3, 0, 5, 1, 4, 2, 4, 1, 3, 2, 2, 3, 4, 4.

- Sampling frame: litter with 5 piglets.
- Summarize the above discrete data graphically.

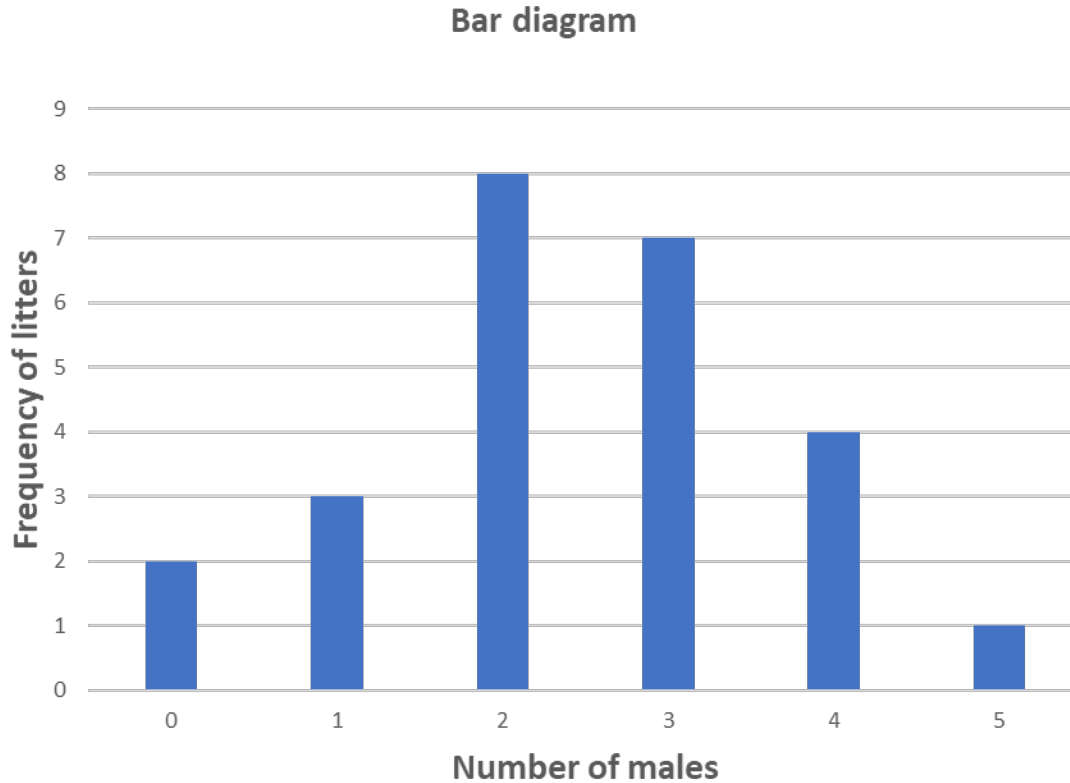
Frequency and Cumulative frequency table:

# male piglets in a litter of 5	0	1	2	3	4	5
Frequency	2	3	8	7	4	1
Cumulative frequency	2	5 (3+2)	13 (5+8)	20 (13+7)	24 (20+4)	25 (24+1)

- *Example summary statement for frequency table above:* 2 and 3 males in a litter of 5 piglets is most common while 0 and 5 males in a litter of 5 piglets is least common.
- *Example summary statement for cumulative frequency table above:* The cumulative frequency for any value (x) is the total frequency having values less than or equal to (x). An example from above is; 13 is the total frequency of litters having 2 or less male piglets in a litter of 5 piglets.

- **Bar diagram**

Plot frequency against number of males

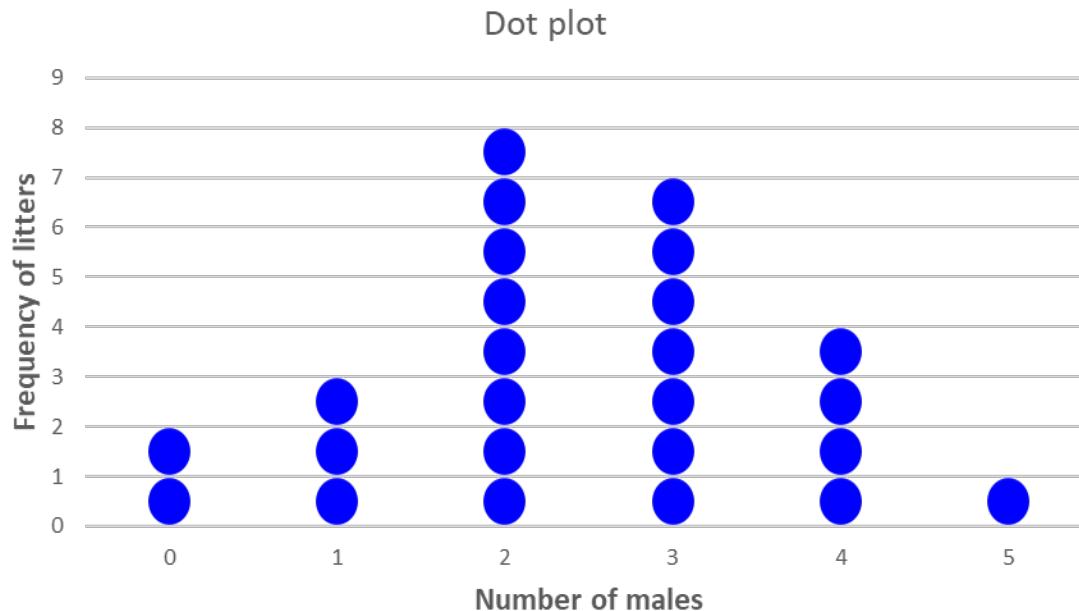


- Summary: most frequent is 2 males / litter and least frequent is 5 males / litter.

- **Dot plot**

A dot plot distributes numerical values where each dot represents a value. If a value occurs more than once, the dots are placed one above the other so that the height of the column of dots represents the frequency of that value.

Plot frequency of litters against males



- Summary: most frequent is 2 males / litter and least frequent is 5 males / litter.
- Summarizing data in the above different ways is done for easy interpretation or extraction of data.

- **3.2.2 Graphical summary of continuous data**
- You can summarize continuous data using the following:
 1. Stem and leaf
 2. Histogram
 3. Frequency tables
 4. Frequency polygons
- Frequency curve

Example 1. The time (seconds) taken by 15 students, to read a given passage is as follows:

- 47, 61, 53, 43, 46, 46, 68, 48, 72, 57, 48, 54, 41, 63, 49
- Summarize the above continuous data graphically.

- ***Stem and Leaf***
- Refer to you tube link and fill in blank below ([https://youtu.be/ 7m0Q_m2ppg](https://youtu.be/7m0Q_m2ppg))

Summary: Most students took between 40 and 50 seconds to read the passage. Only one student took more than 70 seconds.

- **Histogram**
- To come up with a histogram, one needs to determine the required interval using the formular:

$$\text{Interval} = \text{range} / \sqrt{n}$$

Where:

- Range = the difference between the highest and lowest values and n = the number of observations.

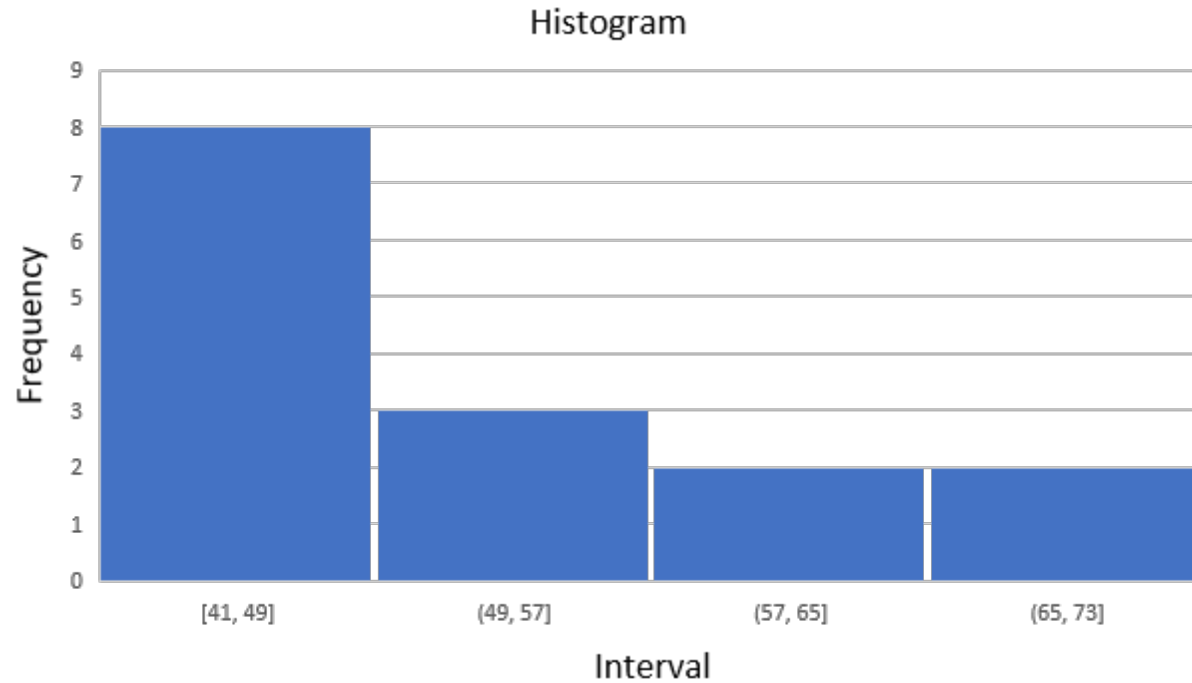
In the example; range is $72 - 41 = 31$ and n is 15

- Interval is therefore $31 / \sqrt{15} = 8$

The value “8” will then be used to partition the intervals starting from the smallest value of observations

Interval	41 - 49	50 - 58	59 - 67	68 - 76
Frequency	8	3	2	2

- You then plot interval (on x-axis) against frequency (on y-axis)



Summary: Most students took between 40 and 50 seconds to read the passage.
Only one student took more than 70 seconds.

- This is not the only way to calculate intervals. There are several other ways. However, the smaller the interval, the more the pillars you have and vice versa.

- ***Cumulative frequency table***

You use the frequency table used to create your histogram.

Interval	41 - 49	50 - 58	59 - 67	68 - 76
Frequency	8	3	2	2
Cumulative frequency	8	(8 + 3) 11	(11+2) 13	(13+2) 15

Example summary: There are 13 observations that fall in the interval up to 67 seconds.

- ***Frequency polygon***

These plot interval mid-points against frequencies. You then connect the plotted points with straight lines. The heights of the points represent the frequencies.

In the below example, plot the corresponding frequency polygon:

Interval	26-30	31-35	36-40	41-45	46-50	51-55	56-60
Frequency	4	5	23	58	61	30	3
Mid-point of interval	28	33	38	43	48	53	58

- To get your frequency polygon. Plot mid-point of interval (x-axis) against frequency below: (plot it and give summary)

- ***Frequency curve***

This curve is obtained by drawing a smooth curve connecting the frequency polygon mid-points. Using the above frequency polygon example, plot a frequency curve below and give summary: