Project Maths Statistics

There are a number of areas that must be studied:

1. **Statistics Key Terms**
2. **Central Tendency and Averages: Mean, Mode and Median**
3. **Spread of Data (variability/dispersion): Range, Standard Deviation and Interquartile Range**
4. **Stem and Leaf Plots**
5. **Histograms**
6. **Shape of Distribution**
7. **Scatter Plots**
8. **Correlation Coefficient**
9. **The Empirical Rule and Normal Distribution**

1. Statistics Key Terms:

- Univariate Data: This type of data deals with one variable at a time, for example the height of a group of people. (Uni means one)

- Bivariate Data: This type of data deals with two variables at a time, for example the height **and** ages of a group of people. (Bi means two)

- Categorical Data: Data that does not refer to a number. This type of data comes from survey questions that do not have numerical answers. For example the colour of your hair or the grade you got in the Junior Cert.

- Ordinal Data: Is categorical data that can be ordered, for example the grades in the Junior Cert can be ordered from A to NG.

- Nominal Data: Is data that cannot be ordered, for example the colour of a group's hair.

- Numerical Data: Is data that refers to numbers. The survey question answers have a numerical value. For example your height, the amount of rainfall in a day or your shoe size.

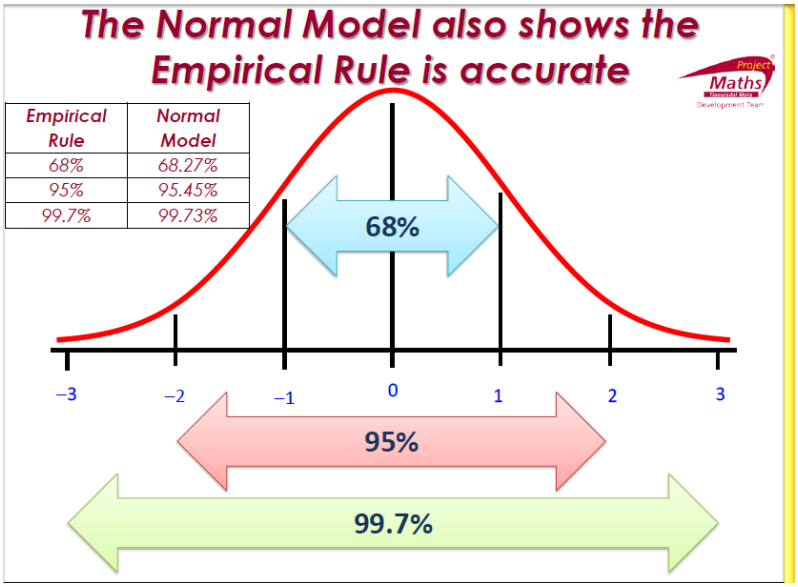
- Discrete Data: (Pg 2 textbook) Data that can only have a fixed numerical value, for example the shoe size you wear can only be a fixed size: 5, 5.5, 6, etc.

- Continuous Data: (Pg 2 textbook) Data that can have any value inside a fixed range. For example the rainfall on a particular day could be 14.567689mm.
  (Discrete Data & Continuous Data have been asked on both SEC Sample and past exam papers.)

- Population: Is the entire group that is being surveyed. For example the height of people in Ireland or the colour hair of students in the school.

- Sample: Is a group selected from the population. This can lead to Bias if the group is not selected randomly. (A random sample should be selecting without bias.)

- Outlier: In a selection of data, an outlier is a value that lies outside the main body of information.
- Shape of Distribution: What is the general shape of a graph? Is it symmetrical or skewed (unbalanced on one side)?
- Central Tendency: Is the way in which data cluster around some value. Often the central tendency is called the average. You can measure the central tendency with the mean, mode or median value.
- Mean: Is only one type of average (mode and median are two others), however when we say the average we often talk about the mean.
- Mode: That value which occurs most often/frequently.
- Median: If a list of values is written in order, the median is the middle value.
- Spread of Data: This is sometimes called the **variability or dispersion of data**. It refers to how the data is spread out. It can be measured by the range, standard deviation or interquartile range.
- Range: The range of a set of data is the difference between the maximum value and the minimum value in a set. The range can be unreliable as a measure of spread as outliers can affect the spread.
- Interquartile Range: This is a more reliable measure of the spread as it is not affected by outliers. **Formula: Upper Quartile Value - Lower Quartile Value = Interquartile Range**.
- The Lower Quartile ($\mathbf{Q_1}$) of an ordered set of data is a value such that one quarter of the values are less than or equal to.
- The Upper Quartile ($\mathbf{Q_3}$) is a value such that three quarters of the values are less than or equal to.
- Standard Deviation: This measure the average spread or dispersion from the mean of all values. It is a reliable measure of spread as it takes account of all values. **The higher the standard deviation the greater the spread of data; a high standard deviation means that the data is very spread out/dispersed and a low standard deviation means that the data tends to be close to the mean.** However if there are extreme values (outliers) it is best to use the interquartile range.
- Empirical Rule: This is sometimes called the 68-95-99.7 Rule. This states that for a normal distribution nearly all values lie within 3 standard deviations of the mean.



2. Central Tendency and Averages: Mean, Mode and Median

- **Mean**: You need to be able to find the mean of an array (list) of numbers and the mean of frequency distribution.
- Symbols for mean: μ or $\bar{x}$ (x-Bar).
- The mean formulas are in the Log tables on Pg 33.
- Formula for mean from a list of *n* numbers: $\frac{\sum x}{n}$
- The 2010 Project Maths Paper 2 Q4 asks students to calculate mean using a calculator (or otherwise).

Example:

2002

7 (a) Calculate the mean of the following numbers:

1, 0, 1, 5, 2, 3, 9.

SOLUTION

The MEAN or AVERAGE of a set of numbers is calculated by adding the numbers together and dividing by the number of numbers.

$$\text{Mean} = \frac{\text{Sum of the numbers}}{\text{Number of numbers}}$$

The mean is denoted by $\bar{x}$.

$$\bar{x} = \frac{1+0+1+5+2+3+9}{7}$$

$$\Rightarrow \bar{x} = \frac{21}{7} = 3$$

$$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots\ldots\ldots + x_N}{N} = \frac{\text{Sum of the Numbers}}{\text{Number of Numbers}} = \frac{\sum x}{N} \quad \ldots\ldots (1)$$

Formula for mean from a frequency distribution: $\frac{\sum fx}{\sum f}$

$$\frac{\sum fx}{\sum f} = \frac{\textit{The sum of each bottom line } (\textit{frequency}) \textit{ multiplied by the top line } (x)}{\textit{The sum of the bottom line } (\textit{frequency})}$$

If it is a grouped frequency distribution, as in the example below, you must find the mid interval value for each group.

Example:

7 (c) The age of each person living in one street was recorded during a census. The information is summarised in the following table:

Age (in years)	0 – 20	20 – 30	30 – 50	50 – 80
Number of people	16	12	32	12

(i) How many people were living in the street?

(ii) Using mid-interval values, calculate the mean age.

(iii) What is the greatest number of people who could have been aged under 40 years?

SOLUTION

Draw up a frequency table using the mid-interval values. To get a mid-interval value add the two numbers together and divide by 2.

Ex. Class interval: 50 – 80

Mid-interval value: $\frac{50+80}{2} = 65$

x	f	fx
10	16	160
25	12	300
40	32	1280
65	12	780
	72	2520

7 (c) (i)

Number of people (frequency) = 16 + 12 + 32 + 12 = 72

7 (c) (ii)

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots\ldots + f_Nx_N}{f_1 + f_2 + \ldots\ldots\ldots + f_N} = \frac{\sum fx}{\sum f} \quad \ldots\ldots (2)$$

Mean age: $\bar{x} = \frac{\sum fx}{\sum f} = \frac{2520}{72} = 35$

Central Tendency continued:

- The **Mode** is the value or the group in a grouped frequency that occurs the most often.

- The **Median** is the middle value in a list of ordered data. You can be asked to find the median of a frequency distribution table, a histogram or a stem leaf plot.
- The median satisfies the following conditions:
  - The median separates points of data set into two groups: an upper half and a lower half. Each half contains the same number of points.
  - Points in the upper half are larger or equal to the median and points in the lower half are less than or equal to the median.
  - **The median may or may not be a point in the given data set** and is not in either the upper or lower half.
- **If there are an even number of data points, the median is the mean (average) of the two points in the middle.**
- **If there are an odd number of data points, the median is the point in the middle.**
- **Sample question:**Finding the median of an odd number of values,2011 Project Maths Paper 2

Question 4 **(25 marks)**

Below is a stem-and-leaf plot showing the number of sweets in each of nineteen packets of sweets.

2	5
2	6 6 7
2	8 8 8 9 9 9 9 9
3	0 0 0 0 1 1
3	2

Key: 2 | 5 means 25 sweets.

(a) What is the *median* number of sweets?

Solution: In this stem and leaf plot there are an odd number of packets of sweets (19). The median value is the middle value of 1 to 19, which is 10. This can also be found by dividing 19 by 2 and rounding up if there are an uneven number of values.**Always round up when finding the median of an uneven number of values.**

The 10^{th} value is 29. **Median:** 29 sweets.

- To find the Median of an even number of pieces of data divide the number of values by 2 and usethis value and the next value up to get the median values. For example to find the median of 30 pieces of data divide 30 by 2. This gives us 15. The median value will be found at the 15^{th} and 16^{th} piece of data. We must now get the mean (average) of the 15^{th} and 16^{th} piece of data.

- **Sample question:** Finding the median from an even number of data values.

The table below shows the marks out of 100 that **50 students** received in an exam. Calculate the median value.

Marks	0-39	40 - 54	55-69	70 -84	85 - 100
Frequency	6	10	15	15	4

Solution: As there are an **even** number of students (50) the median will be the mean of two values:

$\frac{50}{2} = 25$ ⟶ The Median will be from the 25^{th} and 26^{th} students. ⟶ Now we must find what marks did the 25^{th} and 26^{th} student receive.

Using the information from the frequency distribution table we can order the students from the lowest grade to the highest.
6 students received marks between 0 – 39, these students are students **1 to 6**.
In the next group, 40 – 54 marks, there are 10 students, these are the **7 to 16** and so on.

Marks	0-39	40 - 54	55-69	70 -84	85 - 100
Frequency	6	10	15	15	4
Students:	**1 to 6**	**7 to 16**	**17 to 31**	**32 to 46**	**47 to 50**

→ 25^{th} student mark = 55-69
26^{th} student mark = 55-69 → In this question the **Median**mark is 55-69

3. Spread of Data (variability/dispersion):
Range, Standard Deviation and Interquartile Range

- **Range:**As described on page 2 of this handout, the **range** of a set of data is the difference between the maximum value and the minimum value in a set.

- **Standard Deviation:** This measures the average spread or dispersion (**variability)**from the mean of all values. It is a reliable measure of spread as it takes account of all values. **The higher the standard deviation the greater the spread of data; a high standard deviation means that the data is very spread out/dispersed and a low standard deviation means that the data tends to be close to the mean.**
  You can be asked to get the standard deviation of a list of numbers or the standard deviation of a frequency distribution table.
  The formulae are given in the Log Tables on Pg 33.

From a list of numbers: $\sqrt{\frac{\sum(x-\mu)^2}{n}}$

From a frequency table: $\sqrt{\frac{\sum f(x-\mu)^2}{\sum f}}$

In the syllabus it states that you should be able to do this on your calculator:

Syllabus:	− recognise standard deviation as a measure of variability − use a calculator to calculate standard deviation

You should have notes on how to use your calculator to find standard deviation. You should be able to use your calculator to find the standard deviation of aboth a list of numbers and of a frequency distribution.
You can also find them on http://mrmoynihan.pbworks.com

NB**The higher the standard deviation the greater the spread of data; a high standard deviation means that the data is very spread out/dispersed and a low standard deviation means that the data tends to be close to the mean.**

Example (i) The standard deviation of a list of numbers without using a calculator.

7 (a) The mean of the five numbers 2, 4, 7, 8, 9 is 6.
Calculate the standard deviation of the five numbers, correct to one decimal place.

SOLUTION

> STEPS
> 1. Find the mean.
> 2. Draw up a table of x, d and d^2.
> 3. Apply the standard deviation formula.

1. This is done for you.
$\overline{x} = 6$

2.

> The deviation, d, is given by the formula:
> $d = (x - \overline{x})$ = (Number − Mean).
> To work out d, get the difference between each number, x, and the mean, $\overline{x}$.

x	d	d^2
2	−4	16
4	−2	4
7	1	1
8	2	4
9	3	9
		34

3. $\sigma = \sqrt{\frac{34}{5}} = 2.6$

$$\sigma = \sqrt{\frac{\text{Sum of (Deviations)}^2}{\text{Number of numbers}}} = \sqrt{\frac{\sum d^2}{N}} \quad \ldots\ldots \text{(4)}$$

Example (ii) The standard deviation of a frequency distribution without using a calculator.

7 (b) There are fourteen questions in an examination.
The table below shows the performance of the candidates.

Correct responses	0 – 2	3 – 5	6 – 8	9 – 11	12 – 14
Number of candidates	1	2	6	8	3

(i) Using mid-interval values, calculate the mean number of correct responses.

(ii) Calculate the standard deviation, correct to one decimal place.

SOLUTION

If you are asked to find the mean and standard deviation of a frequency distribution, set it out in a table as shown.

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + \ldots\ldots + f_Nx_N}{f_1 + f_2 + \ldots\ldots + f_N} = \frac{\sum fx}{\sum f} \quad \ldots\ldots \text{(2)}$$

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f}} \quad \ldots\ldots \text{(5)}$$

x	f	fx	d	d^2	fd^2
⋮	⋮	⋮	⋮	⋮	⋮
	$\sum f$	$\sum fx$			$\sum fd^2$

Work out the mean first. Then work out d using $d = (x - \overline{x})$. Finally, calculate the standard deviation.

CONT....

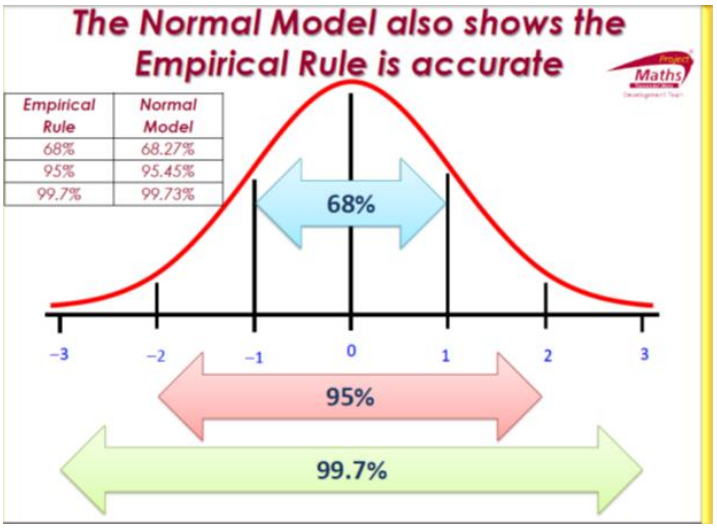
x	f	fx	d	d^2	fd^2
1	1	1	− 7·5	56·25	56·25
4	2	8	− 4·5	20·25	40·5
7	6	42	− 1·5	2·25	13·5
10	8	80	1·5	2·25	18·0
13	3	39	4·5	20·25	60·75
	20	170			189

Draw up a table in the way as shown on the left. The mid-interval values, x, are obtained by adding the class interval values together and dividing by two.

(i) $\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{170}{20} = 8 \cdot 5$

(ii) $\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f}} = \sqrt{\frac{189}{20}} = 3 \cdot 1$

- We will look at the Empirical Rule in more detail later. It is sometimes called the 68-95-99.7 Rule and it states that for a normal distribution nearly all values lie within 3 standard deviations of the mean.



68% of values lie within 1 standard deviation of the mean.

95% of values lie within 2 standard deviations of the mean.

99.7% of values lie within 3 standard deviations of the mean.

Past exam question:**2010 SEC Paper 2 Q4**

(a) Using a calculator, or otherwise, find the mean and standard deviation of the data in the following frequency table.

x	20	30	40	50
f	16	38	26	20

Mean = _____ Standard deviation = _____

- **Interquartile Range:** This is a more reliable measure of the spread than the range as it is not affected by outliers. **Formula: Upper Quartile Value (Q_1) - Lower Quartile Value (Q_3) = Interquartile Range**.

The lower quartile (Q_1) is the median of the lower half of values.

The upper quartile (Q_3) is the median of the upper half of values.

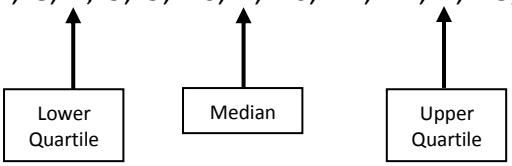
{ 6, 7, 8, •, 9, 9, 10, •, 10, 12, 14, •, 15, 15, 22}

Lower Quartile | Median | Upper Quartile

Finding the Interquartile Range by analysing a set of values:

{ 6, 7, 8, 9, 9, 10, 10, 12, 14, 15, 15, 22}

{ 6, 7, 8, •, 9, 9, 10, •, 10, 12, 14, •, 15, 15, 22}



Step 1: Locate the Median.
Step 2: Locate the lower quartile by finding the middle of the lower half of numbers.
Step 3: Locate the upper quartile by finding the middle of the upper half of numbers.

There are an even amount of numbers in the list above, so the median is between two values.
The Lower quartile and upper quartile are also between two values in this example.

Step 4: To get the quartiles we have to get the mean of the two values at each point (as you would also do for the median)

Lower quartile (Q_1): $\frac{8+9}{2} = 8.5$ Upper quartile (Q_3): $\frac{14+15}{2} = 14.5$

Step 5: Interquartile Range = Upper Quartile Value – Lower Quartile Value → 14.5 – 8.5
Interquartile Range = 6

To mathematically calculate the Interquartile Range from a set of numbers take the following steps:

Step 1. Lower quartile/Q_1: To find the lower quartile of an even set of numbers find ¼ of the number of data values given. **This will result in one of two outcomes:**

1. If you get a decimal number (not a whole number) round up (always round up). The whole number you round up to gives you the exact place in your set of numbers to locate the lower quartile.
2. If you get a whole number, the lower quartile lies between this value and the next value up on the given set of numbers.Take the two matching values from the set of numbers and find the mean to get the lower quartile.

Step 2. Upper quartile/ Q_3: To find the upper quartile of an even set of numbers find ¾ of the number of data values given. **As in Step 1, this will also result in one of two outcomes:**

1. If you got a decimal number in **Step 1** (**Q_1**), you will also get a deciamal for **Q_3**. Again round the decimal up (always round up). The whole number you round up to gives you the exact place in your set of numbers to locate the upper quartile.
2. If you got a whole number in **Step 1** (**Q_3**), you will alos get a whole number for **Q3.** The upper quartile lies between this value and the next value up on the given set of numbers. Take the two matching values from the set of numbers and find the mean to get the upper quartile.

Step 3. Interquartile Range=Upper quartile – lower quartile.

On the next page are two examples which explain each of the outcomes in Step 1 & 2.

Example 1: When finding ¼ **and** ¾of the set of numbers gives a **decimal** number.

2	5
2	6 6 7
2	8 8 8 9 9 9 9 9
3	0 0 0 0 1 1
3	2

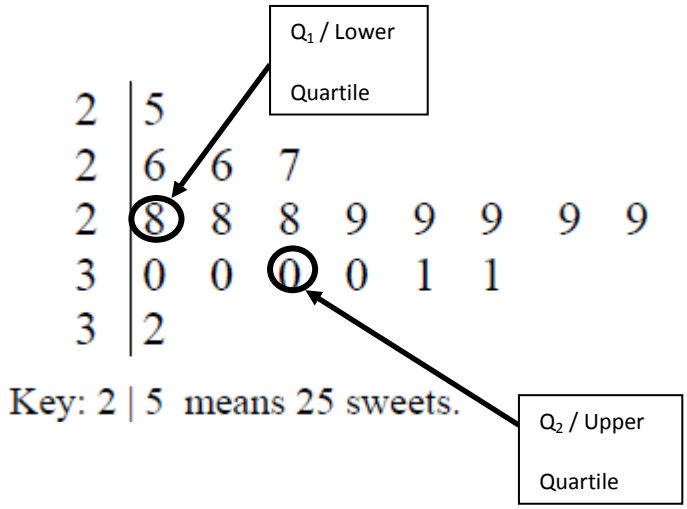
Key: 2 | 5 means 25 sweets.

In the stem and leaf plot above there are nineteen pieces of data.

Step 1: Lower quartile/Q_1
¼ of 19 = 4.75
As 4.75 is a decimal round up to 5. (Always round up)
Find the fifth value in the plot, which is 28.

Step 2: Upper quartile/Q_3
¾ of 19 = 14.25
As 14.25 is a decimal round up to 15.
Find the fifteenth value in the plot, which is 30.

Step 3: Interquartile range
$Q_3 - Q_1 = 30 - 28 = 2$
Interquartile range = 2



Example 2: When finding ¼ **and** ¾of the set of numbers gives a **whole** number.

Using the data from before: {6, 7, 8, 9, 9, 10, 10, 12, 14, 15, 15, 22}(**Note: the 12 pieces of data are ordered**)

Step 1: Lower quartile/Q_1:
¼ of 12 = 3
As 3 is a whole number, the lower quartile will lie midway between two values: in this example the third and fourth values.

{6, 7, 8 9 9, 10, 10, 12, 14, 15, 15, 22} Third value = 8, Fourth value = 9,

Lower quartile: $\frac{8+9}{2} = 8.5$

Step 2: Upper quartile: ¾ of 12 = 9
As 9 is a whole number, the upper quartile will lie midway between two values: in this example the ninth and tenth values.

{6, 7, 8, 9, 9, 10, 10, 12 14 15 15, 22} Ninth value = 14, Tenth value = 15,

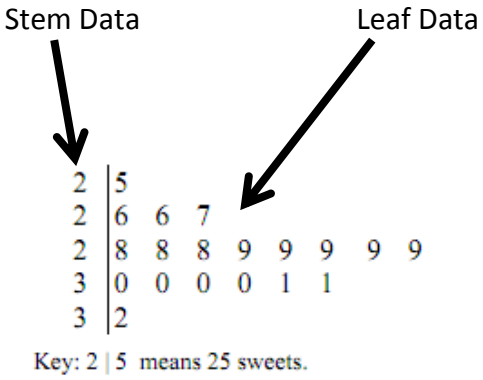
Upper quartile: $\frac{14+15}{2} = 14.5$

Step 3: Interquartile Range: 14.5 – 8.5 = 6 Interquartile range = 6

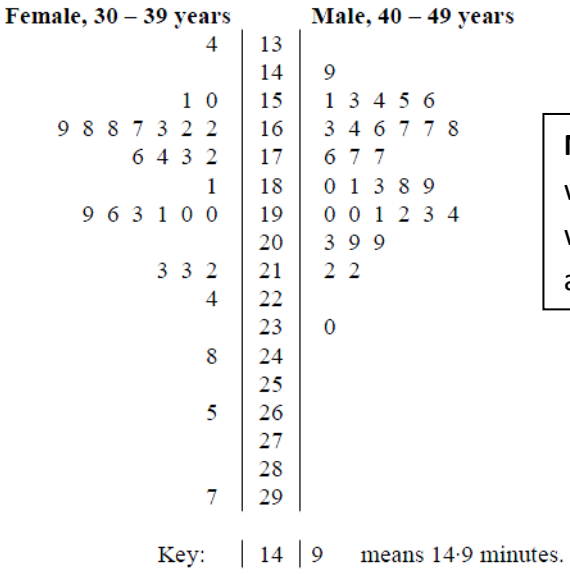
4. Stem and Leaf Plots

A stem and leaf plot is a type of graph that presents information. It represents data by separating each value into two parts: the stem (the vertical line of figures) and the leaf (horizontal lines of figures).

NB You should arrange the data in ascending order when drawing a stem and leaf diagram.



- You can be also be asked to draw a regular stem and leaf plot or a back to back stem and leaf plot:



NBYou should always write the **Key** used when drawing a stem and leaf plot.

- You can also be asked a number of questions about stem and leaf plots:
    - Is the data univariate or bivariate
    - Is the data discrete or continuous
    - Find the median
    - Find the mode
    - Find the range
    - Find the interquartile range
    - Describe differences and similarities between back to back stem and leaf plots
    - Describe the shape of the distribution (Using terms such as symmetrical or skewed)
    - Describe the Location of Data/Central Tendency/Average (Using terms such as mean, mode, median)
    - Describe the spread of Data/Dispersion/Variability (Using terms such as Range, Interquartile Range, Standard Deviation)

Example 1/Sample question: 2011 Project Maths Paper 2

2	5
2	6 6 7
2	8 8 8 9 9 9 9 9
3	0 0 0 0 1 1
3	2

Key: 2 | 5 means 25 sweets.

Q4

(i) What is the median?

(ii) What is the range?

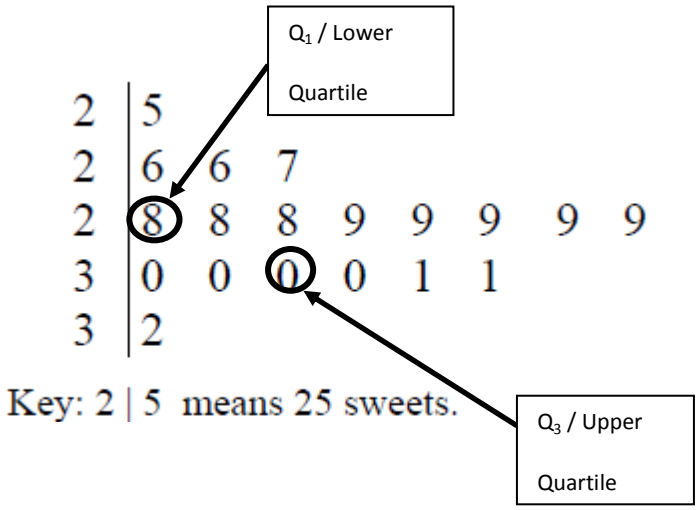
(iii) Find the Interquartile range.

Solution:

Step 1: Lower quartile/Q_1
¼ of 19 = 4.75
As 4.75 is a decimal round up to 5. (Always round up)
Find the fifth value in the plot, which is 28.

Step 2: Upper quartile/Q_3
¾ of 19 = 14.25
As 14.25 is a decimal round up to 15.
Find the fifteenth value in the plot, which is 30.

Step 3: Interquartile range
$Q_3 - Q_1 = 30 - 28 = 2$
Interquartile range = 2



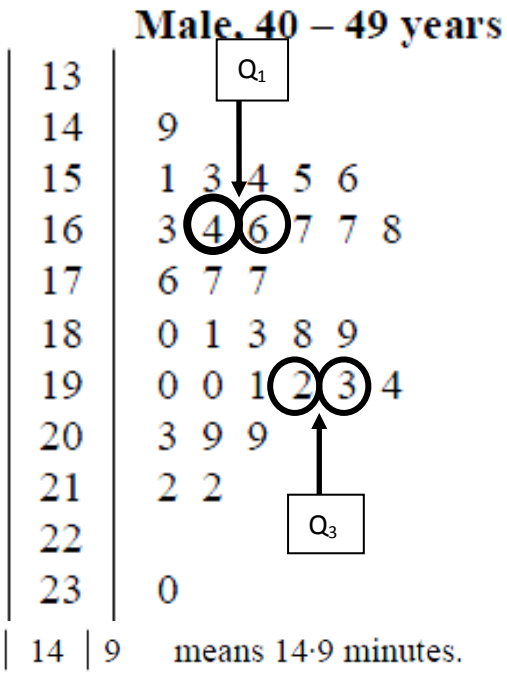
Example 2: There are 32 data items.
Step 1: Lower quartile/Q_1:
¼ of 32 = 8
As 8 is a whole number, the lower quartile will lie midway between this value and the next: in this example the eight and ninth values.
Eighth value = 16.4, Ninth value = 16.6

Lower quartile: $\frac{16.4 + 16.6}{2}$ = 16.5

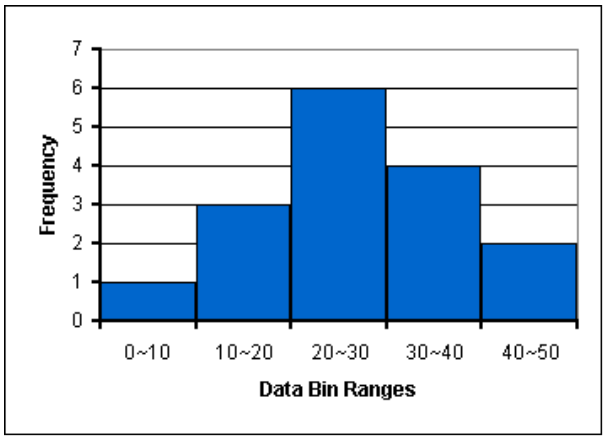
Step 2: Upper quartile/Q_3:
¾ of 32 = 24
As 24 is a whole number, the upper quartile will also lie midway between two values: in this example the 24th and 25th values.
24th value = 19.2, 25th value = 19.3,

Upper quartile: $\frac{19.2 + 19.3}{2}$ = 19.25

Step 3: Interquartile Range:
$Q_3 - Q_1 = 19.25 - 16.5$ **= 2.75**
Interquartile Range = 2.75



5. Histograms



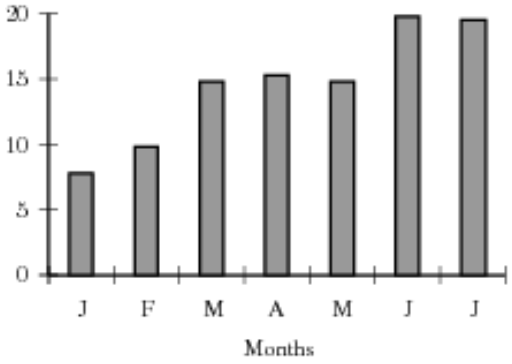
The area of each bar in a histogram represents the frequency.
In ordinary level bars will have equal widths for all histograms.
Histograms can represent discrete or continuous data (for continuous data bars represent grouped values, eg 1-10)
Drawing a histogram:

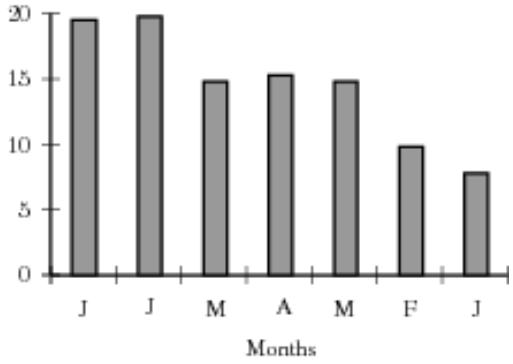
1. Each bar should have equal width.
2. There are no gaps between each bar; each bar touches its neighbour.
3. Label each axis and each rectangle, use units to label if given.
4. Always give a title.

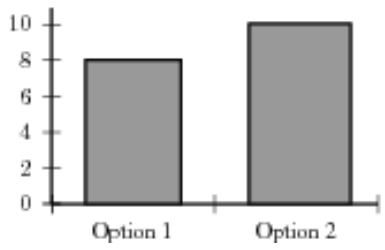
Sample and Past Project Maths papers have asked the following questions:

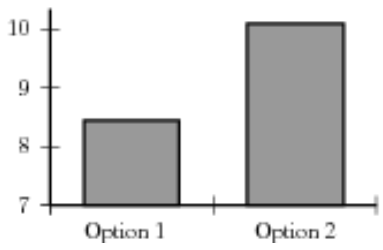
1. Describe the differences between two histograms.
2. Find the median.
3. Describe the shape/distribution of a histogram.
4. Using standard deviation with a histogram.

Misuse of Statistics:





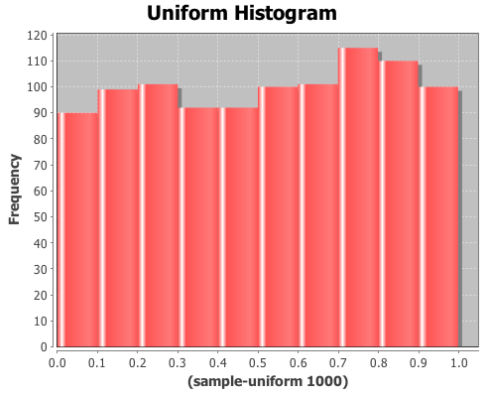




6. Shape of Distribution

The following terms are used to describe the shape of a distribution:

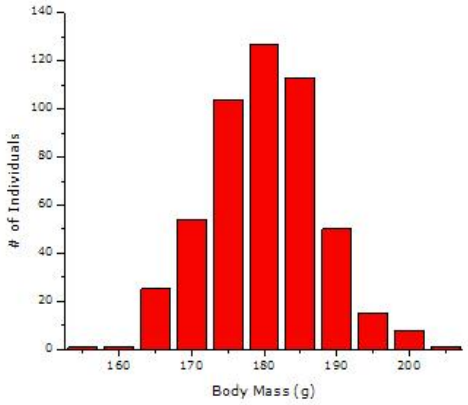
1. Uniform Distribution



2	1	2	3	4
3	3	4	5	
4	1	1	3	4
5	6	7	7	8
6	1	2	2	
7	1	5	6	9

Key **2|1** = 2.1

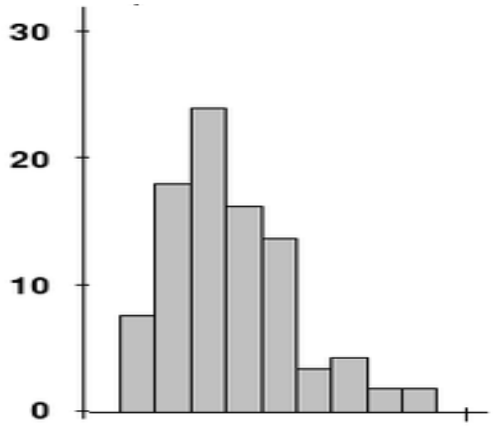
2. **Symmetrical Distribution**: Describe a graph as **roughly symmetrical** if it is not exactly symmetrical.



2	1					
3	3	4				
4	1	1	3	4	4	
5	6	6	6	7	8	8
6	1	2	2	4		
7	1	5				

Key **3|3** = 33

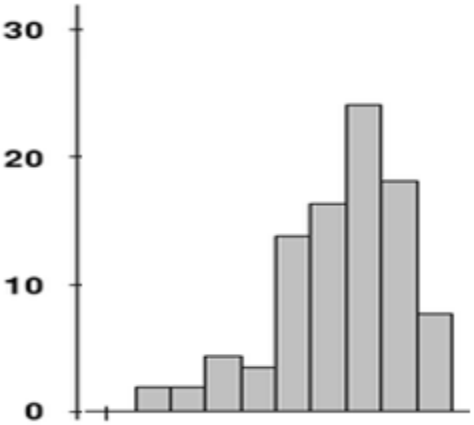
3. Positive Skew: (the information tends towards the lower values)



2	1					
3	6	6	6	7	8	8
4	1	1	3	4	4	
5	3	4	4	5		
6	1	2	2			
7	1	5				

Key **3|3** = 33

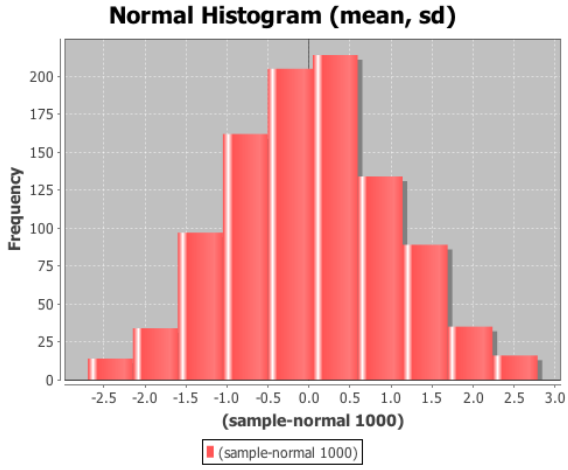
4. Negative Skew



2	1					
3	2	3				
4	3	4	4	5		
5	1	1	3	4	4	
6	6	6	6	7	8	8
7	1	5				

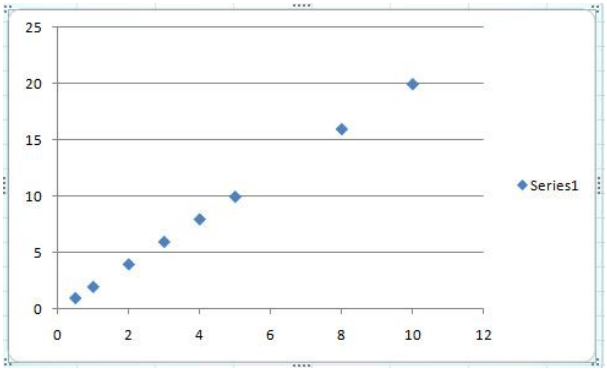
Key **3|3** = 33

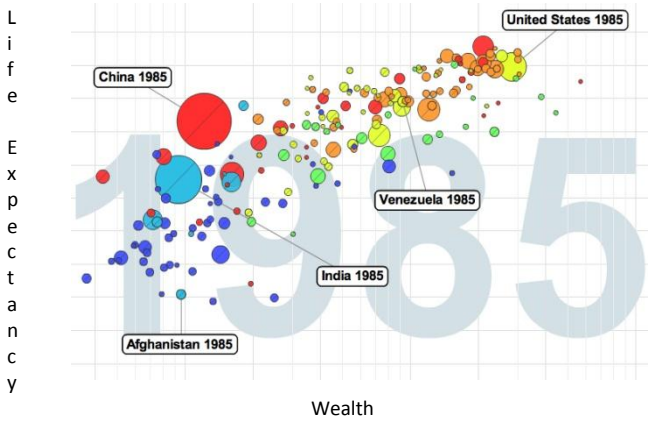
5. **Normal Distribution**: In a normal distribution the mean = mode =median.



7. Scatter Plots

The Scatter Plots will we study have only two variables: they are **bivariate.**



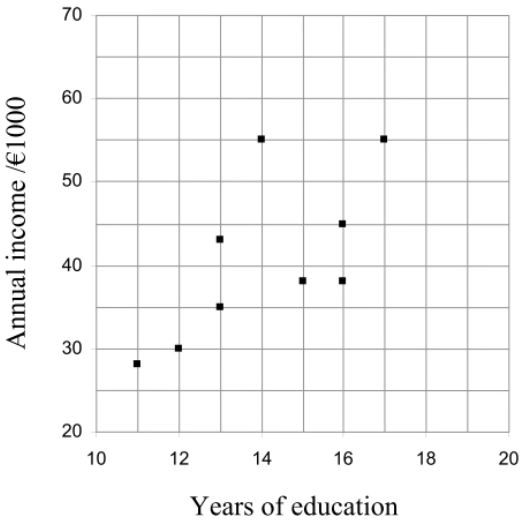


To draw a scatter plot you plot dots on an x and y axis.

2012 Sample Paper:

(f) An economics student wants to find out whether the length of time people spend in education affects how much they earn. The student carries out a small study. She asks twelve adults to state their annual income and the number of years they spent in full-time education. The data are given in the table below, and a partially completed scatter plot is given.

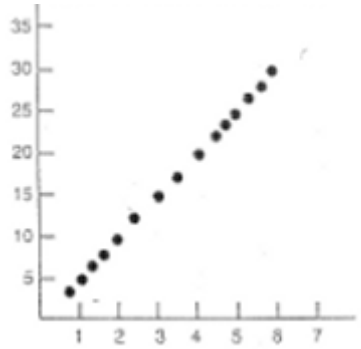
Years of education	Income /€1,000
11	28
12	30
13	35
13	43
14	55
15	38
16	45
16	38
17	55
17	60
17	30
19	58



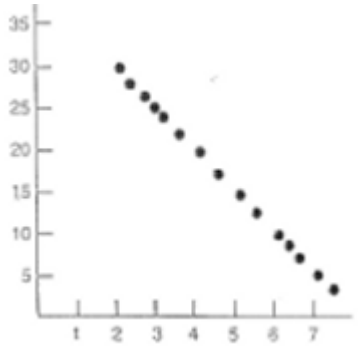
(i) The last three rows of data have not been included on the scatter plot. Insert them now.

8. Correlation Coefficient

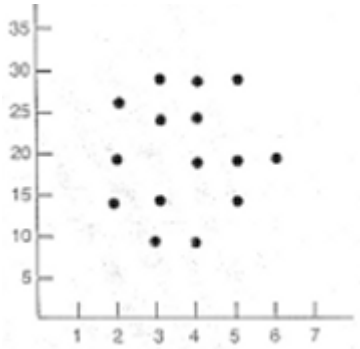
The correlation coefficient is a value between – 1 and + 1. You can be asked to match correlation coefficient values to scatter plots.



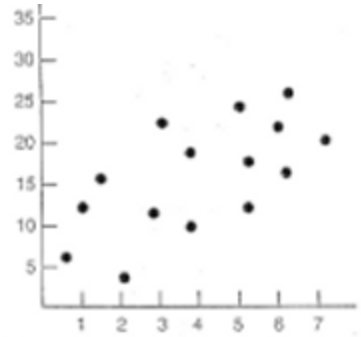
Correlation Coefficient =+1



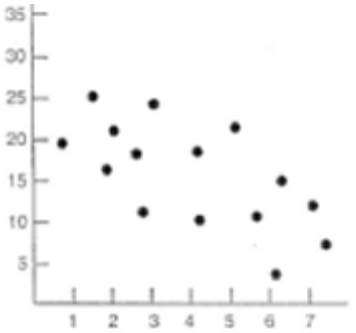
Correlation Coefficient = – 1



Correlation Coefficient = 0



Correlation Coefficient = + 0. 5

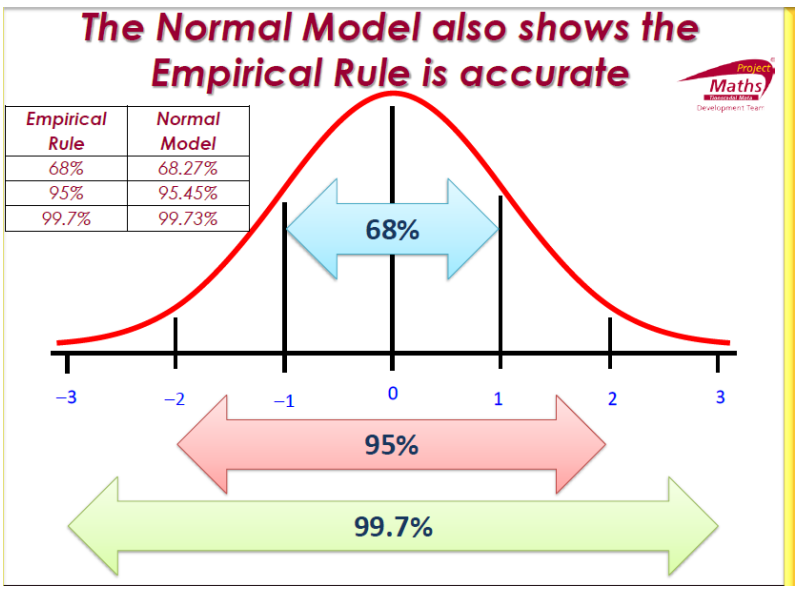


Correlation Coefficient = – 0.5

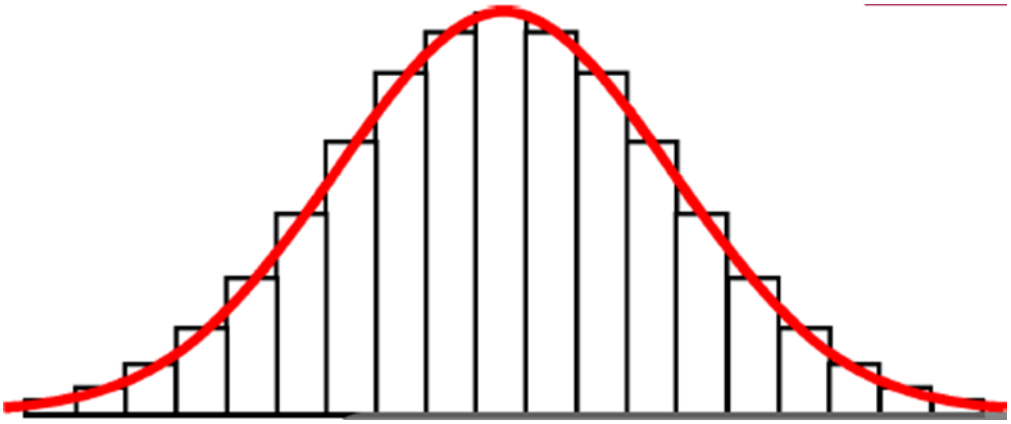
Causality: In a scatter plot one value **may** affect the other value (**cause** it to increase/decrease).
For example if you had a scatter plot which showed height and age of a population from the ages of 1 to 16, as age increases so too does the height increase. In this case age can affect height.
However, just because the values on a scatter plot have a correlation coefficient of +1 or – 1 it does not mean that there is a relationship where one value that affects the other.

8. Empirical Rule

Empirical Rule: This is sometimes called the 68-95-99.7 Rule. This states that for a normal distribution nearly all values lie within 3 standard deviations of the mean.



Normal Distribution:



A symmetric, unimodal, bell-shaped distribution of continuous random data is called a **normal distribution**.
It is the most common model in statistics, because so often, with a big enough sample, we see the data mostly collect around the mean.

Short Questions on **The Empirical Rule**:

A. (i) When data is normally distributed describe the chances of a piece of data being near the middle?
(ii) When data is normally distributed what are the chances of being within 1 standard deviation of the mean?
(iii) If we assume data is normally distributed what are the chances of being within 2 standard deviations of the mean?

B. The mean height is 166.6cm and the standard deviation is 13.3cm. Based on an assumption that the distribution of heights is approximately normal, use the empirical rule for the following questions:
(i) 68% of this school's students have heights between ___ and ___ cm.
(ii) What percentage of students have heights between 140cm and 193.2cm?
(iii) A school tour is being organised. All students can apply to go on it. There is a rollercoaster at one location on the tour. You have to be over 140cm to be allowed on the rollercoaster. What percentage of students wouldn't be tall enough?

C. To enter a particular college course, candidates must complete an aptitude test. In 2010 the mean score was 490 with a standard deviation of 100. The distribution of the scores on the aptitude test is a normal distribution.
(i) What percentage of candidates scored between 390 and 590 on this aptitude test?
(ii) One student scored 795 on this test. How does this student's score compare to the rest of the scores?
(iii) The college admits only students who were among the highest 16% of the scores on this test. What score would a student need on this test to be qualified for admission to this college? Explain your answer.

2012 SEC Sample Paper

(c) The heights in 2011 of Irish males born in 1992 are normally distributed with mean 178·8 cm and standard deviation 7·9 cm.

(i) Use the *empirical rule* to complete the following sentence:

"95% of nineteen-year-old Irish men are between _____ and _____ in height."

(ii) Use the empirical rule to make one other statement about the heights of nineteen-year-old Irish men.

Calculator:

You can use your calculator to get a random sample of a study population: