Mayfield School Mathematics Statistics Coursework

Introduction

I have been assigned to complete a statistical investigation around the fictitious data of Mayfield High School, which has data of a real school. I will be completing this investigate for the subject Mathematics: Statistics. By completing the task that I have been set this will help me achieve my aim which is to gain a General Certificate of Secondary Education in this very particular subject. I will be using various techniques that I have recently studied and learnt and captured to produce a successful & efficient coursework. Alternating Statistical Methods will be used throughout this assignment to prove if my hypothesis is either correct or incorrect.

Task Situation

I have decided to investigate majorly between the relationship between the height and weight of the pupils and to tell whether or not there is any correlation between them. I will take many actions as possible in achieving pure and efficiently results to meet the needs and requirement of my assignment. To meet my particular aim I will use many statistical interpretations and methods to help me form sufficient conclusions on what I have gained and obtained from the evidence that I will be collecting for this project.

My Hypothesis

A hypothesis is the outline of the idea/ideas which I will be testing and below are the following hypothesis I have decided to investigate for this particular assignment:

- **' Boys at Mayfield School are Taller and Weigh more on average in comparison to females'**
- **'Key Stage 4 Students who watch more hours of television on average have a lower IQ Level'**
- **' Left Handed Students have higher IQ levels and Key Stage 2 Results in Comparison right handed student at Mayfield High School'**

I will now investigate the correlation between the hypothesis that I have decided to investigate and proceed with a full investigation.

HYPOTHESIS 1: 'Boys at Mayfield School are Taller and Weigh more on average in comparison to females'

Planning

I will need the data of Mayfield High School between the Years of 7 to 11 and this is due to the fact for a wider sampling range and sufficient and unbiased results and uses many sampling methods to make my assignment unique and unbiased. The total number of students in the school is 1183. Here is a Table that I have produced which contains the number of boys and girls in each year.

Two Way Tables

The Table Below is a two way table due to the fact there are two variables shown at the same time and helps view results and data conclusively.

Year Group	Number of Boys	Number of Girls	Total
7	151	131	282
8	145	125	270
9	118	143	261
10	106	94	200
11	84	86	170
TOTAL	**604**	**579**	**1183**

I will use Stratified Sampling to investigate my First Hypothesis. This is because it took into thought all our needs of the sampling of the data; and this methods was easily accessible and can be easily manipulated and carried and only asked for a simple understanding of the subject.

The variables for the sample are gender and age so I had to do separate samples for boys and girls and vary the amount of samples taken from each year to keep the sample unbiased and insufficient. This was done as the different year groups had different numbers of pupils and it would be unfair to take the same number of samples from each year group i.e. 5 samples out of 55 is be more than 5 samples out of 200 so stratified sampling will be helpful this is due to the fact the

number of student in each year and so there is less chance of unequal representation. I will be investigating 20 boys and Girls and Boys from each Year altogether.

I will be calculating my Stratified Sampling using the Table above now and I will need calculation proportion to stratify my data spread/range.

Stratified Sampling Calculations

I want 40 Students from the from the school total 1183.

I want 20 Boys from 604 Boys and my aim is to work out how many from each year and below is my calculation

Year 7 Boys = 151 Divided By 604 multiplied by 20 = **5**

Year 8 Boys = 145 Divided by 604 multiplied by 20 = 4.8 rounded of to **5**

Year 9 Boys = 118 Divided By 604 multiplied by 20 = 3.9 is rounded of to **4**

Year 10 Boys = 106 Divided By 604 multiplied by 20 = 3.4 is rounded of to **3**

Year 11 Boys = 84 Divided By 604 multiplied by 20 = 2.7 rounded of to **3**

Note * I then applied the same methods in calculating the number of Girls for my stratified sampling and here is a table of results that I will be using for my data sampling for this particular hypothesis.

Year Group	Number of Boys	Number of girls
7	5	5
8	5	4
9	4	5
10	3	3
11	3	3

Within this process the decision of choosing which students in particular to choose from within each year was done via Random Sampling.

These was done through the process in which I used a Hat to enter all the Year Group (Boys and then Girls) separately and then I mixed and entered all the names and choose an unbiased person to draw names out of the hat and below are the results of the people chosen for this sample.

I obtained these results. These results were typed up on a excel spreadsheet. Below is the Data I have gained from my Sample. Since I have investigated the Height and Weight I will only show these particular fields in my results:

1) **Male Stratified Sample Results**

Year Group	Surname	Forename	Height (m)	Weight (kg)
7	Austin	Steven	**1.54**	**43**
7	Lloyd	Mark	**1.61**	**56**
7	Mills	Robert	**1.63**	**50**
7	Pearce	Stuart	**1.50**	**34**
7	Thorpe	Billy	**1.53**	**40**
8	Freeman	Ian	**1.82**	**64**
8	Jones	Kevin	**1.62**	**49**
8	McGrail	Craig	**1.38**	**35**
8	Peter	Zakir	**1.63**	**41**
8	Wilson	Christopher	**1.53**	**32**
9	Davidson	Jimmy	**1.61**	**45**
9	Jones	Jimmy	**1.52**	**46**
9	Jones	Brian	**1.32**	**38**
9	Rowley	Geoff	**1.56**	**53**
10	Bates	Markus	**1.80**	**60**
10	Edd	Michael	**1.68**	**59**
10	Javidson	Carlos	**1.70**	**57**
11	Fairfax	Jacob	**1.62**	**51**
11	Little	James	**1.65**	**47**
11	Vincent	Nigel	**1.8**	**62**

2) **Female Stratified Sample Results**

Year Group	Surname	Forename 1	Height (m)	Weight (kg)
7	Butler	Leanne	**1.65**	**40**
7	Earnshaw	Catherine	**1.45**	**41**
7	Meager	Becky	**1.64**	**47**
7	Miles	Amanda	**1.46**	**40**
7	Richards	Abbie	**1.53**	**47**
8	Burton	Prudence	**1.59**	**52**
8	Campbell	Julia	**1.41**	**30**
8	Kudray	Rebecca	**1.54**	**52**
8	Water	Rebecca	**1.55**	**57**
9	Atkins	Patience	**1.57**	**40**
9	Bagnall	veronica	**1.49**	**37**
9	Dixon	Mary	**1.49**	**52**
9	Kelsey	Sannita	**1.62**	**46**
9	Mosler	Samantha	**1.58**	**36**
10	Bhatti	Hannah	**1.72**	**56**
10	Durst	Freda	**1.75**	**60**
10	Hall	Jane	**1.51**	**36**
11	Marta	Sana	**1.52**	**45**
11	Ratty	Louise	**1.65**	**59**
11	Peckeleka	Chantel	**1.56**	**38**

Tally and Frequency Chart

Now that I have my data I will put them into frequency/tally tables to make it easier to read and it is a useful way of representing and helps view trends within my sampling that I have produced:

1) Boys Heights

BOYS		
Height (cm)	**Tally**	**Frequency**
130≤h<140	II	2
140≤h<150		0
150≤h<160	IIIIII	6
160≤h<170	IIIIIIII	8
170≤h<180	I	1
180≤h<190	III	3
190≤h<200		0

A Pattern I have spotted in this particular Frequency/Tally table is that nearly 75 percent of boys are of the height between 150 to 170 from my sample and this shows me that students in my sample are rather tall and a steady size or above for their age group

2) Boys Weights

BOYS		
Weight (kg)	**Tally**	**Frequency**
30≤w<40	IIII	4
40≤w<50	IIIIIII	7
50≤w<60	IIIIII	6
60≤w<70	III	3
70≤w<80	0	0

This Tally and Frequency table shows the boy's weight and the spread of data are large and not as compact as the height whereas the height are rather scattered and vary. This however notifies me that some of the students in my sample and tall for their age but weight are average in relationship to their height.

Girls Heights

Tally and Frequency charts are regularly used to process raw data making easier to spot irregularities and patterns.

GIRLS		
Height (cm)	Tally	Frequency
130≤h<140		0
140≤h<150	IIIII	5
150≤h<160	IIIIIIII	9
160≤h<170	IIII	4
170≤h<180	II	2
180≤h<190		0

This table shows me nearly 50 percent of students are between the height of 150-160 and there are not many student who excel over 180 cm tall which and there are also not many if any student with a height between 130 to 140 cm which shows me that the spread of data is compact which makes it easier to view trends and also it shows that there are not many irregularities in height in the student in my sample that I have taken.

Girls Weight

GIRLS		
Weight (kg)	Tally	Frequency
30≤w<40	IIIII	5
40≤w<50	IIIIIIII	8
50≤w<60	IIIIII	6
60≤w<70	I	1
70≤w<80		0

From this I have found that many girls weigh 40-49 kg and also the girl's weigh in this particular region or slightly above on most occasions and this shows me that some girls may stop putting weight rapidly after a particular age

Comparisons of Tally & Frequency Table

I will now compare each of the tables of girls & boys height and then weight and view the differences in trends and if there are and mistakes or bias.

Height Comparison: From the Sample I have taken I have come to find that boys grow rapidly at a later age whereas girls grows faster from an earlier age and stop at a particular age also. This is shown since I have found that 2 students are of a height between 130- 139 cm, which shows that my data may be misleading or there is a lapse in growth. Both Girls and Boys have average heights and are fairly balanced and of equal size. Although my Sample shows that boys on some occasion grow to above average height such as over 180 cm whereas in Girls this is rare and unique.

Weight Comparison: Most Boys and Girls weigh in the region of 40- 60 from my sample and I have found that not many girls are over the weight of 60 whereas in Males usually are borderline 60 or above when they come to the age of 16.

I have produced a Composite Bar Chart which is attached to the upcoming page

The Pie Chart shows the percentage of each year in the School and these help me form my sample and this form of representing data is efficient accurate and eye catching and help form divisions in my data.

Calculations:

Year 7: 282 Divided by 1183 Multiplied by 360
= 86 Degrees

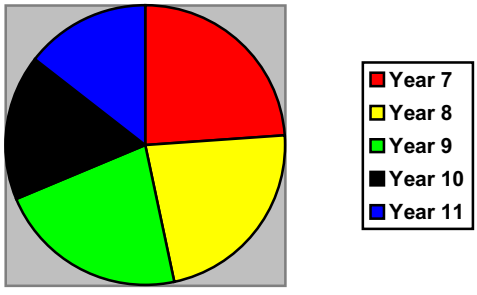
Year 8: 270 Divided by 1183 Multiplied by 360
= 82 Degrees

Year 9: 261 Divided by 1183 Multiplied by 360
= 79 Degrees

Year 10: 200 Divided by 1183 Multiplied by 360
= 61 Degrees

Year 11: 170 Divided by 1183 Multiplied by 360
= 52 Degrees

Total 360 Degrees



Mean and Mode of Frequency Data

I will now find the Mean, Median and Mode of the Frequency that I have found and this will be quick efficient and reliable and will help me gain evidence on whether boys are taller and weigh more in comparison to girls.

Mean of Girls and Boys Weight

BOYS				
Weight (kg)	Tally	Frequency (f)	Mid-point (x)	fx
30≤w<40	IIII	4	35	140
40≤w<50	IIIIIII	7	45	315
50≤w<60	IIIIII	6	55	330
60≤w<70	III	3	65	195
70≤w<80		0	75	0
TOTAL		**20**		**980**

Mean = Total Frequency Times Midpoint Divided by Total Frequency

980 Divided by 20 = **49 kg for BOYS**

GIRLS				
Weight (kg)	Tally	Frequency (f)	Mid-point(x)	fx
30≤w<40	IIIII	5	35	140
40≤w<50	IIIIIIII	8	45	360
50≤w<60	IIIIII	6	55	330
60≤w<70	I	1	65	65
70≤w<80		0	75	0
TOTAL		**20**		**895**

Mean = 895 divided by 20 = 44.75 rounded off to **45 kg for GIRLS**

Mean of Boys and Girls Height

BOYS				
Height (cm)	**Tally**	**Frequency**	**Mid-point**	**Fx**
130≤h<140	II	2	135	270
140≤h<150		0	145	0
150≤h<160	IIIIII	6	155	930
160≤h<170	IIIIIIII	8	165	1320
170≤h<180	I	1	175	175
180≤h<190	III	3	185	555
190≤h<200		0	195	0
TOTAL		**20**		**3250**

Mean= 3250 Divided by 20 = 162.5 = **163 cm = 1.63m**

GIRLS				
Height (cm)	**Tally**	**Frequency**	**Mid-point**	**Fx**
130≤h<140		0	135	0
140≤h<150	IIIII	5	145	725
150≤h<160	IIIIIIIII	9	155	1395
160≤h<170	IIII	4	165	660
170≤h<180	II	2	175	350
180≤h<190		0	185	0
TOTAL		**20**		**3130**

Mean: 3130 Divided by 20 = 156.5 which is 157 cm = **1.57m**

Comparison of Mean of Girls & Boys (Height and Weight)

In comparison of the mean I have found for both boys and girls heights the differences vary. This shows that on average from my sample boys are weigh more then the girls in my data although it may not be by a large amount. Boys from my Stratified Sample also on average are 6 cm taller than Girls.

Mode of Girls and Boys Weight

The arithmetic mean of a group of numbers is found by dividing their sum by the number of members in the group; e.g., the sum of the seven numbers 4, 5, 6, 9, 13, 14, and 19 is 70 so their mean is 70 divided by 7, or 10. Less often used is the geometric mean (for two quantities, the square root of their product; for *n* quantities, the *n*th root of their product).

Modal Weight Stem and Leaf Diagram

Boys weight		
Stem	**Leaf**	**Frequency**
3	4 5 2 8	**4**
4	3 0 9 1 5 6 7	**7**
5	6 0 3 9 7 1	**6**
6	4 0 2	**3**
7		**0**
8		**0**

I have come to find from this stem and leaf diagram shows that the 40 -49kg weights is the most frequent weight from the sample that I have taken for this particular hypothesis.

Key 3 4 = 34 kg

Girls weight		
Stem	**Leaf**	**Frequency**
3	0 7 6 6 8	**5**
4	0 1 7 0 7 0 6 5	**8**
5	2 2 7 2 6 9	**6**
6	0	**1**
7		**0**
8		**0**

From this Stem and Leaf Diagram I have also found that the samples of both boys and girls have this there most frequent weight common which although there irregularities or the uncommon heights are less regular which shows that the data is less biased and compact

To find the modal weight I will now look at which frequency seems to have appeared the most often and for the Boys Modal Weight it is:

Modal Group for Boys: 40 ≤w< 50 **Modal Weight for Girls: 40 kg**

Below is Similar Data but Different Calculation methods to find the Mean, Median and also the Mode as this will help me towards proving my hypothesis also this data will help me find the spread and the average of the data which will be helpful throughout this portfolio:

Boys: Height

Mean

$$\frac{1.38 + 1.61 + 1.63 + 1.50 + 1.53 + 1.82 + 1.62 + 1.38 + 1.63 + 1.53 + 1.61 + 1.52 + 1.32 + 1.56 + 1.80 + 1.68 + 1.70 + 1.62 + 1.65 + 1.8}{20}$$

= 31.28 Divided by 20 = 1.6m

Median: 1.65 m (Calculated with the Use of Microsoft Excel)

Range: 1.8 – 1.32 = 0.48 m

Boys Weight

Mean: 48.1 kg

Median: 48 kg

Range: 64-32 =32 kg

Girls: Weight

Mean

$$\frac{30 + 36 + 36 + 37 + 38 + 40 + 40 + 40 + 41 + 45 + 46 + 47 + 47 + 52 + 52 + 52 + 56 + 57 + 59 + 60}{20}$$

= 45.55 which is rounded to **46 kg**

Median: 46kg

Range: **60 – 30 = 30**

Girls Height

Mean

1.56 m

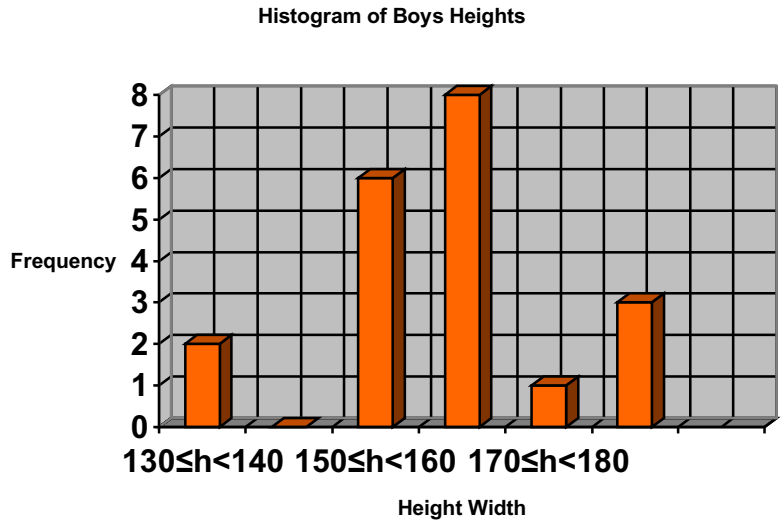
Median 1.56m

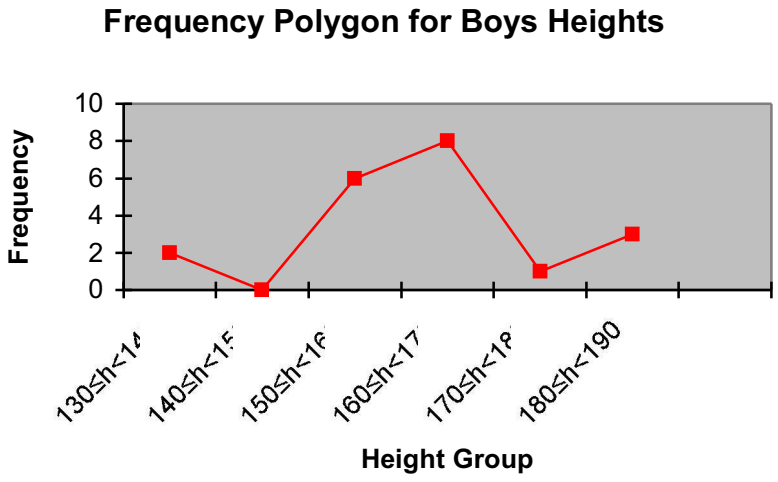
Range **1.75 – 1.41 = 0.34m**

Histogram and Frequency Polygons

From the data I have collected and formed through my frequency tables and mean averages and many more I will now produce a Frequency Polygon and a Histogram that shows the Boys & Girls Height and Weights From my sample that I have taken a for my assignment. The Frequency Polygon will clearly identify the shape of my variations and both these forms of representing data will help me form a sufficient analysis.

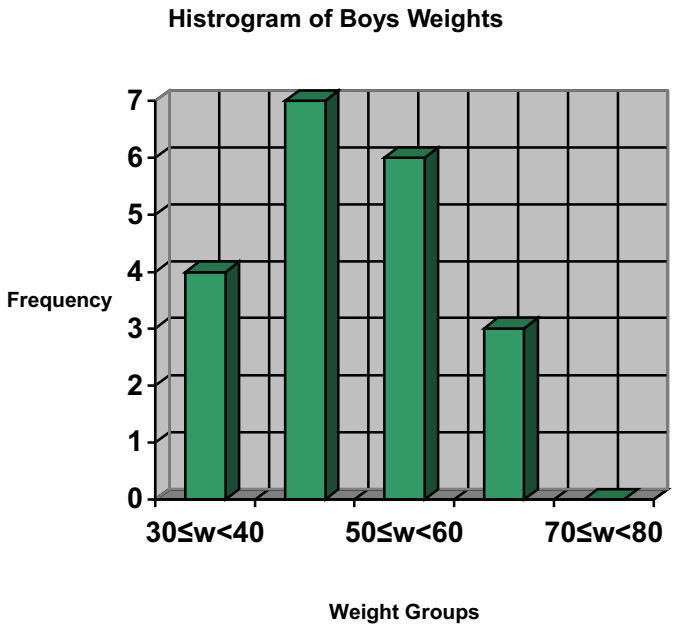
Boys Height

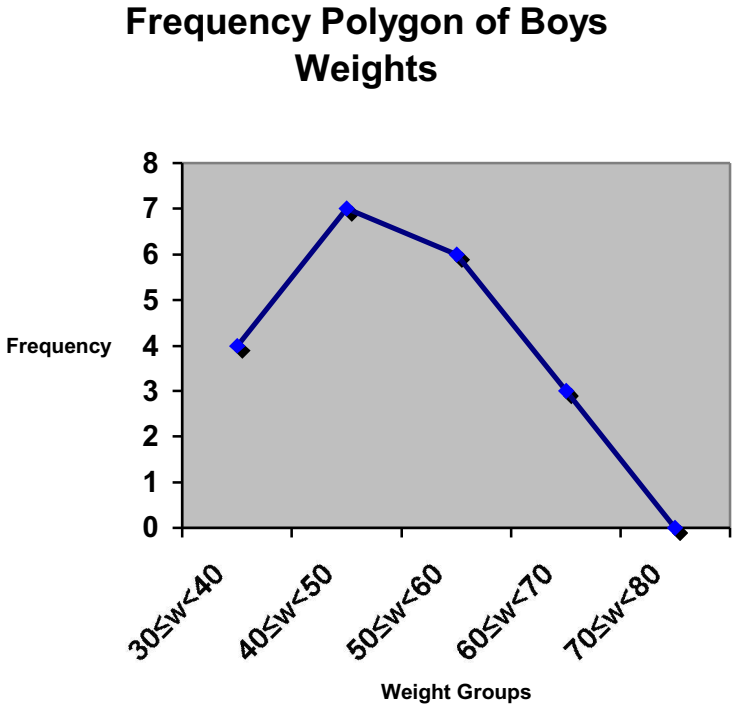




The Following Histogram and Frequency Polygon clearly shows to me that the most common height is between 160 to 169 cm which I believe is average height which in my opinion shows that the hypothesis that I have formulated is correct and is not unbiased. Although both graphs are representing the same data I believe that each graph has its value and help me provide and further in depth analysis and conclude efficiently

Histogram for Boys Weights



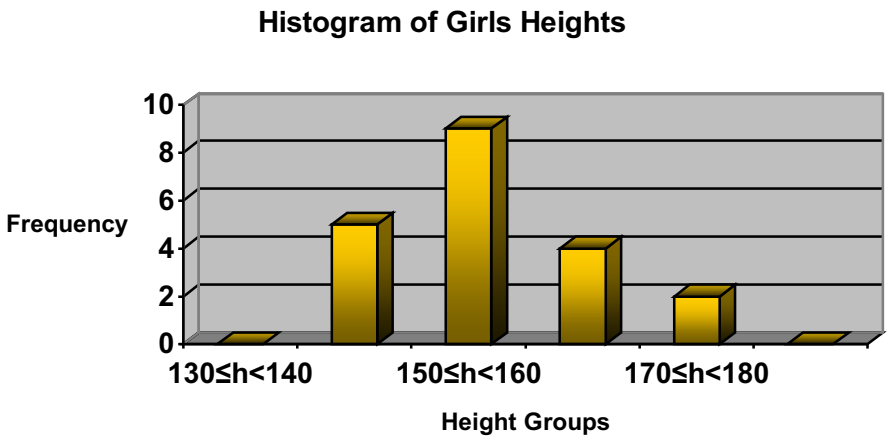


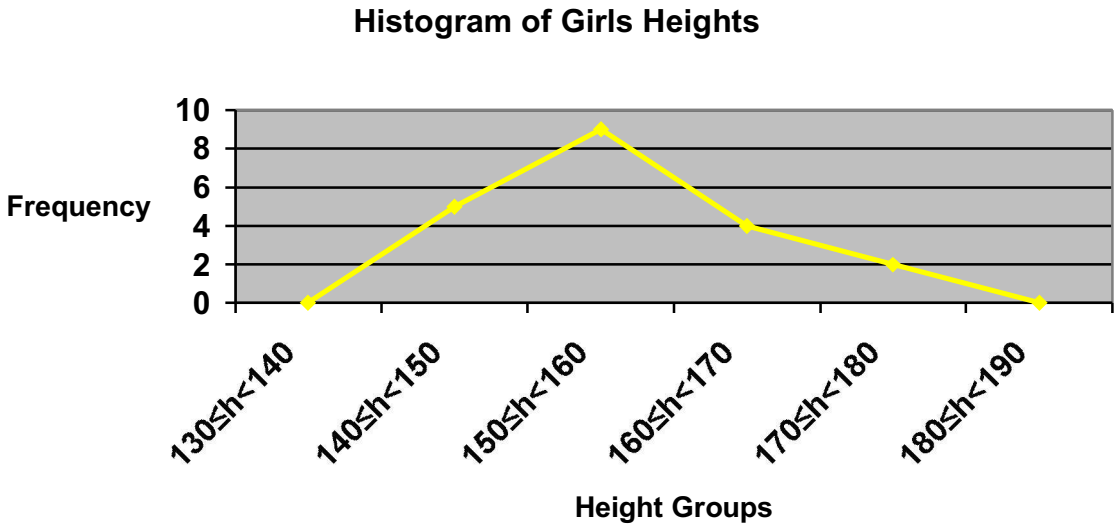
This particular histogram helps me identify the most common weight group which you can clearly view from the histogram an d also the frequency polygon which is 40 – 49 kg for boys weight which shows me that is of an average weight for an student studying in the school year of 10 and I have this opinion due to the fact personal experiences although also I have come to find that no one from the sample that I have taken is above the weight does not excel over 79 kg as this polygon clearly expresses and the trend is also lower in frequency from 40- 49 kg onwards.

Girls Height and Weight Frequency Polygons & Histogram

These histograms now give me a clear picture of the data distribution. For the sample there is an even distribution of data. The middle group has the highest frequency which is expected. For the data to be evenly distributed, the other two sides must be fairly symmetrical. It is clear that the histogram do not show this. This shows that the majority of scores were above the median.

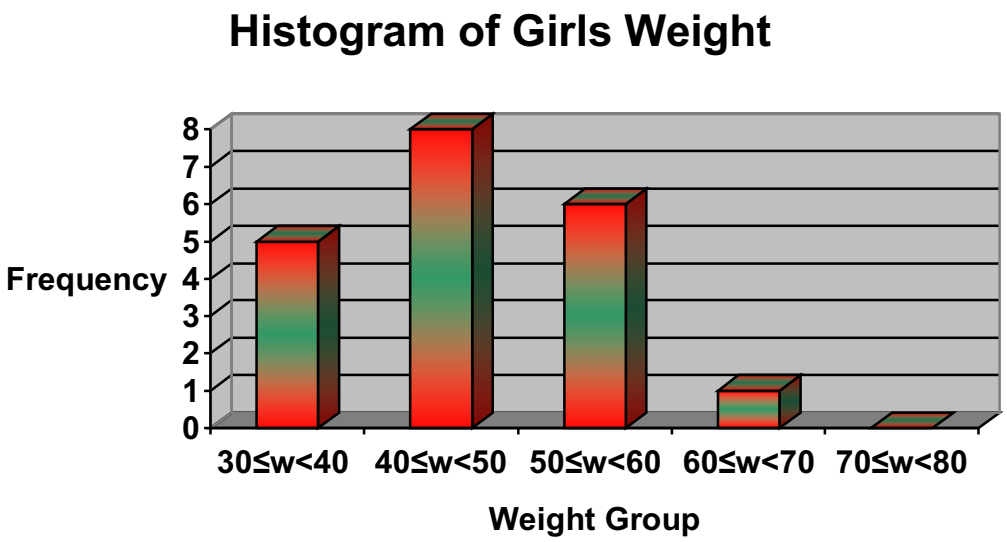
Girls Height

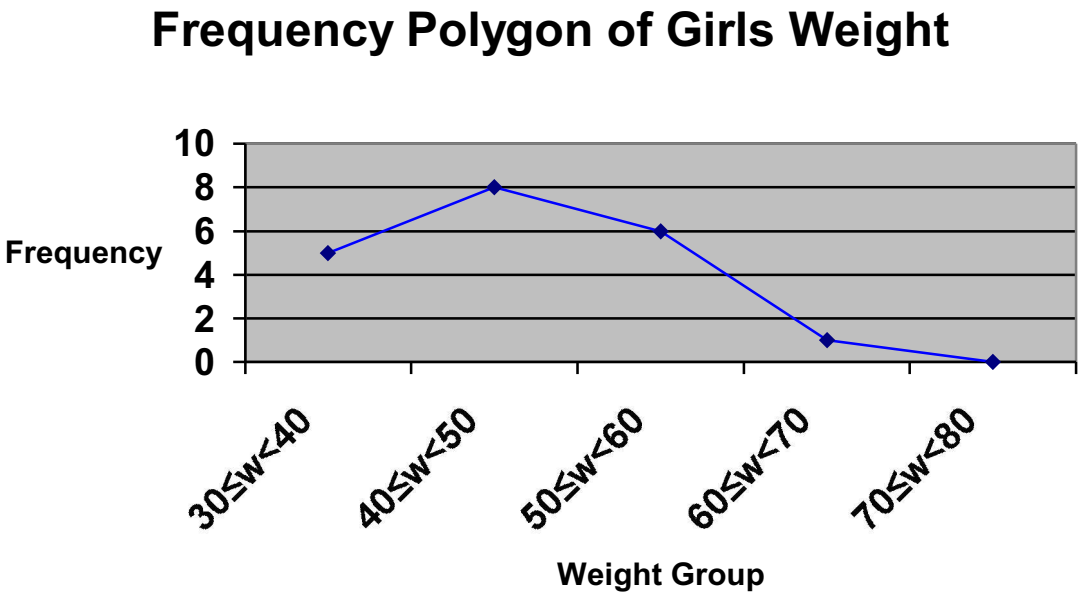




These Representations of Data shows admirably that the average height is 150 to 160 which I believe is slightly above average in my honest opinion for my sample and also I have come to find that the trend is rather varied although the frequency are upward to a certain point and downward from the peak onwards.

Girls Weight





This data shows with great intent that the highest frequency I 40 – 49 kg which shows that both boys and girls from the stratified sample that I have taken for this area of my assignment and this hypothesis in particular that they have a lot In common in terms of frequent data sources. In addition to this I have also come to find that the girls do not have many students above 69 kg whereas for boys there are 3 as times as many students above this height.

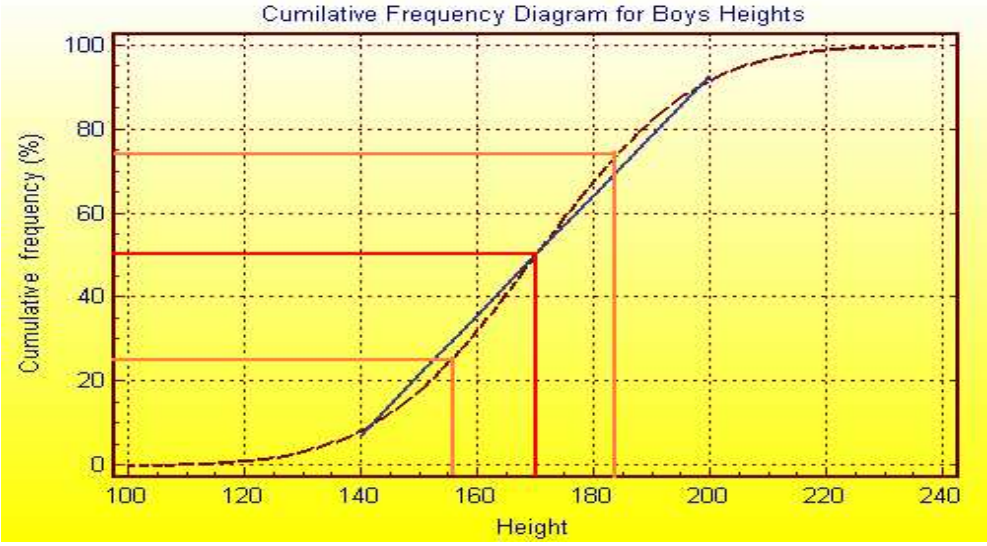
Cumulative Frequency Diagrams & Tables

I will now produce cumulative frequency diagram for both girls & boys height and weight and this will help me gain sufficient evidence towards forming my conclusion and I will also find percentiles and will produce box and whisker plots as this will help me view my data and trend efficiently.

1) BOYS HEIGHTS:

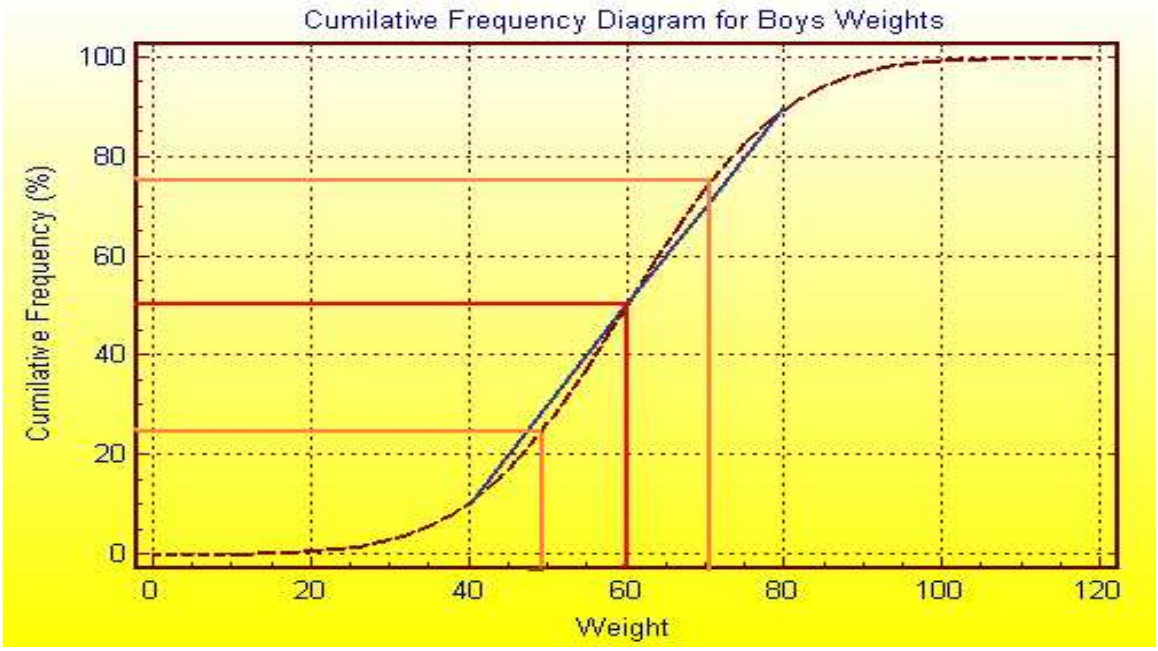
BOYS	
Height (cm)	**Cumulative Frequency**
Less than 140	2
Less than 150	2
Less than 160	8
Less than 170	16
Less than 180	17
Less than 190	20
Less than 200	20

Using a Statistical Program that I have downloaded I entered my data into the appropriate field and below Is a diagram of for the Cumulative Frequency which will help me identify percentiles and trends in my data. The graph will also include a line which helps me identify the trend clearly



2) BOYS WEIGHT

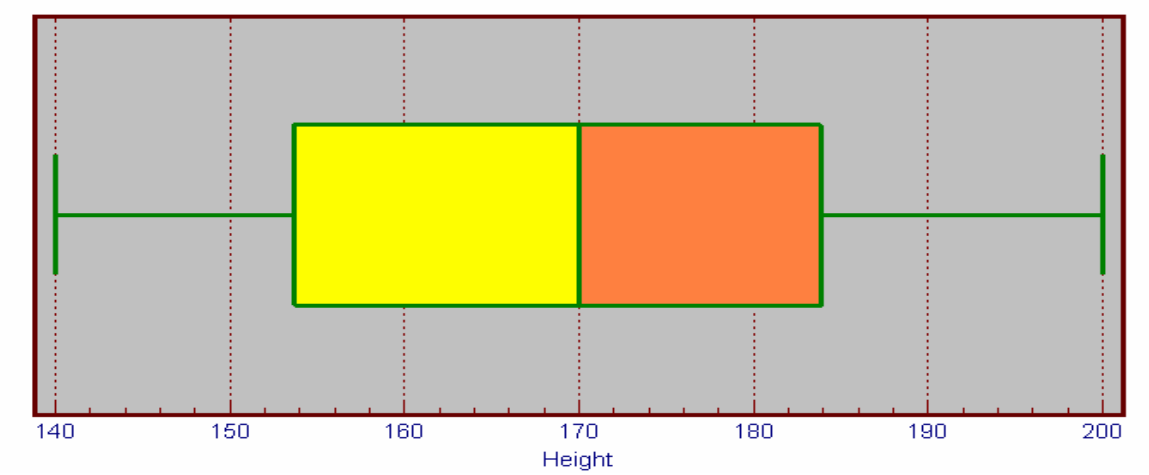
BOYS	
Weight(kg)	**Cumulative Frequency**
Less than 40	4
Less than 50	11
Less than 60	17
Less than 70	20
Less than 80	20



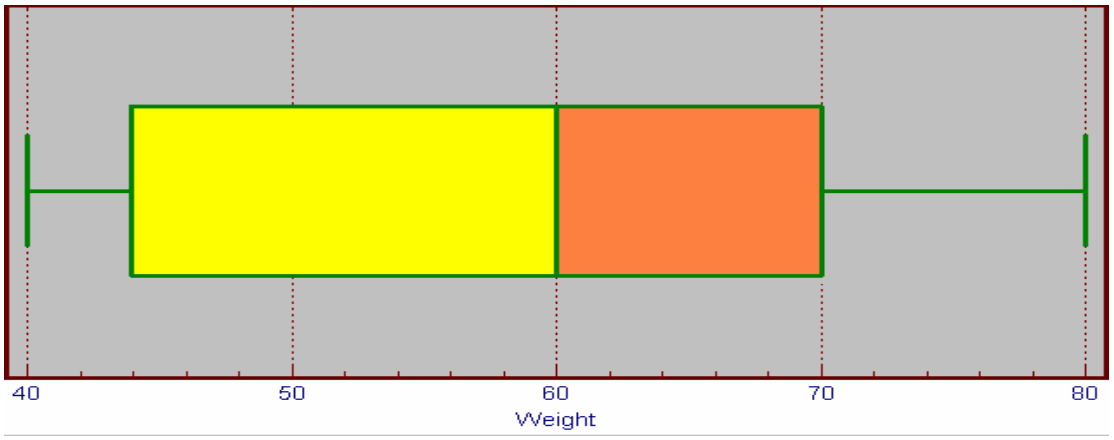
From this particular Cumulative Frequency Diagram I have been able to find that the average weight from the sample that I have taken is 60kg which I my opinion is generally quite high and the Interquartile Range that I will be finding at a later stage will help me view the spread of data and the margin of error. I believe that all the graphs that I will produce will help me complete my objective and conclude efficiently and successfully.

Boys Height Box and Whisker Plot

I then used the same statistical software I created a box and whisker which is a vital representation of data and it will



Boys Weight Box and Whisker Plot



Comparison of Box and Whisker Plots

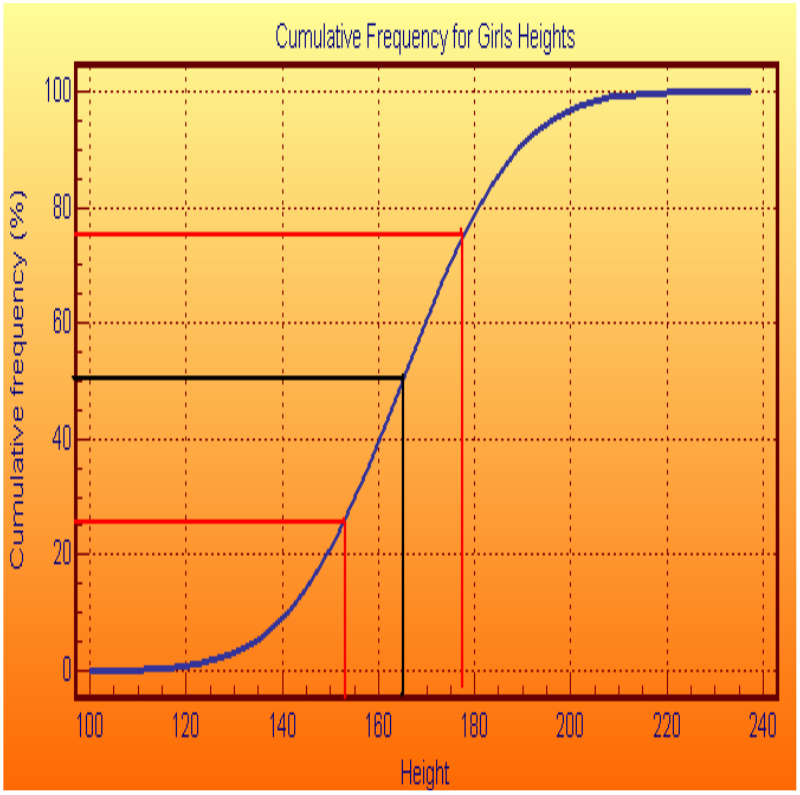
Form this Box and Whisker Plot I will be able to find the median which will show the middle frequency of my data and also will be able to view the maximum and minimum values for both height and the weight and find the percentiles and the quartiles. For the Height the Median is **170cm** and the interquartile range is **30 cm** and the maximum value is 200 cm and minimum value is 140cm. whereas for the weight I have found the Median as **60kg** which is respectively what I had predictable and is suitably accurate although the range of data for the weight is less and the data is negatively skewed in comparison to the height for the boys.

I will no be producing Cumulative Frequency Diagram and Tables for the Girls Height and Weight since this will help me form a sufficient and reliable comparison in height between Girls and Boys and form a successful and accurate conclusion to my hypothesis.

GIRLS HEIGHT

GIRLS	
Height (cm)	**Cumulative Frequency**
Less than 140	0
Less than 150	5
Less than 160	16
Less than 170	17
Less than 180	20
Less than 190	20

Cumulative Frequency Diagram



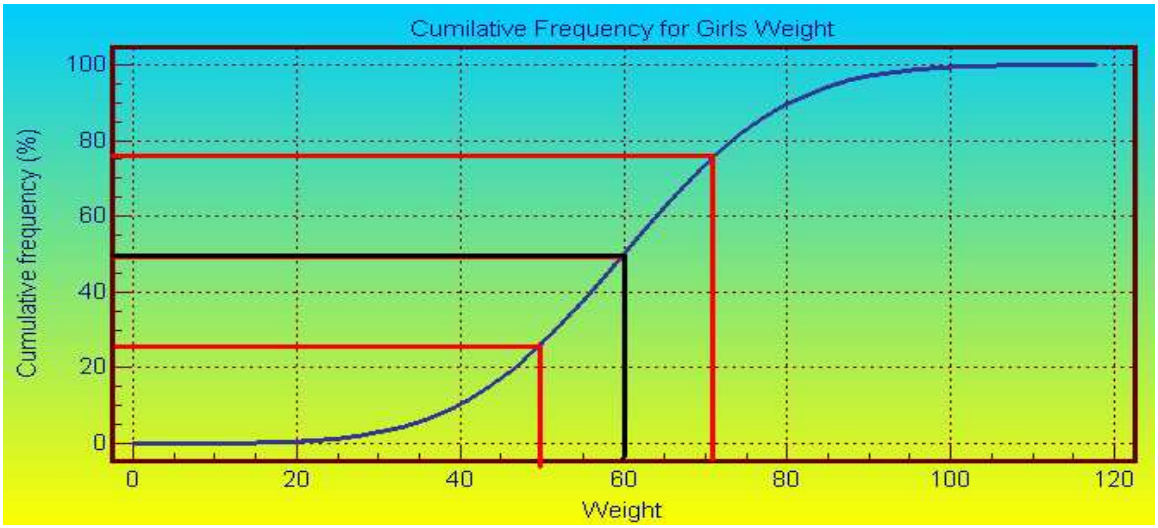
This Cumulative Frequency Diagram I have labeled the Median and Upper and Lower Quartile and through this I will find the spread of data and also I will find the middle number of the data of girls heights that I am investigating and the Median for this graph is **165 cm** and the Interquartile range is 177 cm – 154 cm = **33 cm** and this will help toward my final analysis

GIRLS WEIGHT

GIRLS	
Weight (kg)	Cumulative Frequency
Less than 40	5
Less than 50	13
Less than 60	19
Less than 70	20
Less than 80	20

Cumulative Frequency Diagram

My variable will be the Height



Median 60 kg

Lower Quartile 50 kg

Interquartile Range 22g

Upper Quartile 72 kg

From this I have found the spread on data efficient and on the following page I will compare my results from the cumulative frequency against both boys and girls height and weight and make a suitable conclusion from this representation.

Comparison of Height and Weight of Boys and Girls from C.Frequency and Box Plots

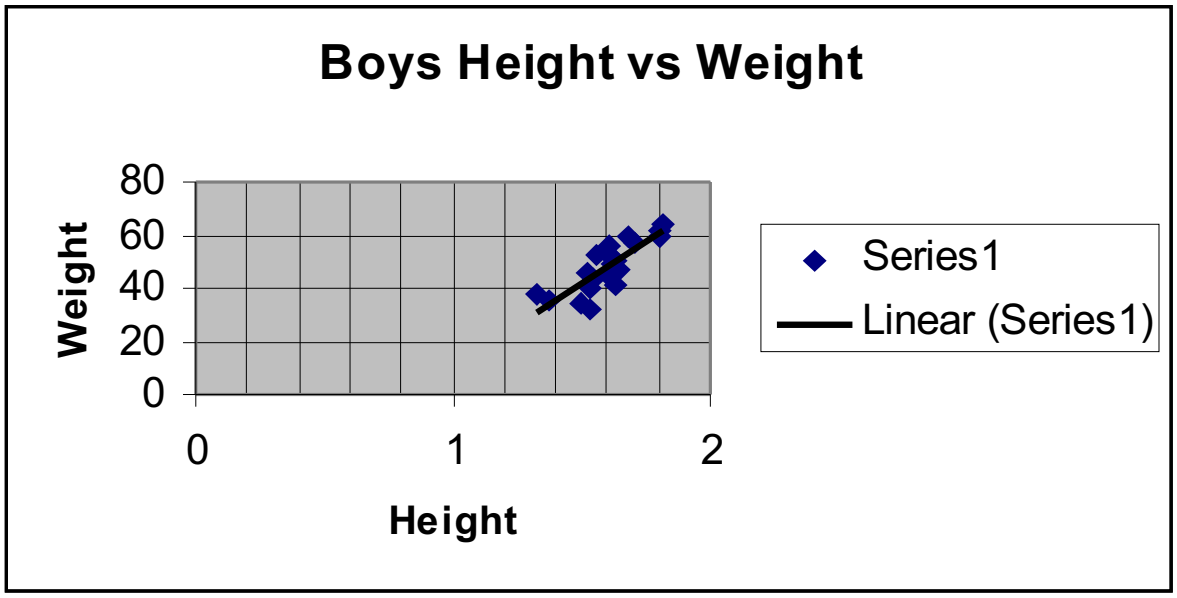
From the Cumulative Frequency Diagram I have come to find that the Median height for boys in the sample that I have taken is larger in comparison to the girl's median height. Boys Median Height is **170 cm** whereas Girls Median Height is **162 cm** and shows that there is an 8cm difference in the middle figure from both sets of data I have collected although this may on some occasion be accurate due to the fact my sample may not be efficient and also the range of data varies where the IQ range for Boys is 21 cm and Girls in 33 cm which shows that the spread of data that I have from my sample for boys is narrower whereas girls have a wider spread and make the results reliable as a whole. As far as the Weights are concerned the range is similar and the range is rather symmetrical also and this may be since my graphs may be irregular.

Scatter Diagram

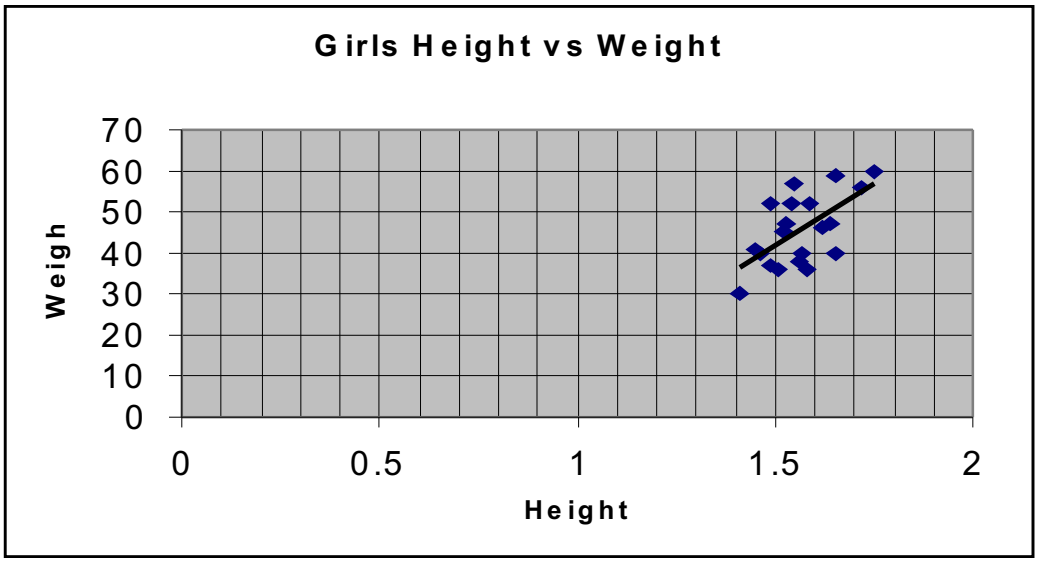
I will now produce a Scatter Graph which shows the Height vs Weight for all the Boys and girls data that I have collected and I will be able to find whether there is a pure correlation and I will then compare my results and I will see whether my two sets of data are related from my sample:

I will conclude after both graphs

Boys Height vs Weight



Girls Height vs. Weight



Conclusion of Scatter Graphs

Boys & Girls Height vs. Weight = From the Scatter Graph that I have produced I have come to find that there is a Correlation or a Trend between both variables which are Height and Weight and there is a Fairly Strong Positive Correlation and this shows me that the taller the person the higher the weight although there are always some irregularities which is unique as a symmetrical trend is rather impossible as every human being has various growth period. In addition to this will be the major form of representation in my and I believe this shows the trend clearly and efficiently.

Body Mass Index

Calculation= WEIGHT / (HEIGHT) ^{2}

I will be Doing this calculation for 5 people randomly from the sample that I have taken:

7	Austin	Steven	**1.54**	**43**
7	Lloyd	Mark	**1.61**	**56**
9	Bagnall	Veronica	**1.49**	**37**
10	Bhatti	Hannah	**1.72**	**56**
10	Durst	Freda	**1.75**	**60**

Calculation= WEIGHT / (HEIGHT) ²

1) 43/(1.54) ² = 43/2.3716 = **18.13**

2) 56/(1.61) ² = 56/2.5921 = **20.64**

3) 37/(1.49) ²= 37/2.2201 = **16.67**

4) 56/(1.72) ²= 56/2.9584 = **18.93**

5) 60/(1.75) ²= 60/3.0625 = **19.6**

From this calculations I can conclude that the Higher the Height in Equivalent to the Weight the Higher the Value of the Body Mass Index which shows me that both variable have a certain trend and also shows that on some occasions the height is parallel to the Weight.

Spearman Rank Correlation Co-Efficient

I will now using the same data sample of five random students to find the spearman rank correlation. The Spearman's Rank Correlation Coefficient is used to discover the strength of a link between two sets of data:

$$r = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$

X Value	Y Value	X Rank	Y Rank	D	D²
1.54	43	4	4	0	0
1.61	57	3	2	1	1
1.49	37	5	5	0	0
1.72	56	2	3	-1	1
1.75	60	1	1	0	0

R = 1 – 6(2) / 5(25-1) = 1- 12 / 24 Times 5= 120 = 1 – 12/120 = 1 -1/10 = **0.9**

Evaluation of Hypothesis 1

In my honest opinion I feel that I successfully completed and analyzed my hypothesis and I have gained a sufficient evidence to back up my theories. I would like to remind you that my main objective for this hypothesis was to find out whether I was correct or incorrect in my thinking that Boys at Mayfield School are taller and weigh more on average than the Girls at the same school. Within this aim I was also aiming to find whether there is a certain trend or relationship between the height and weight of the students that I have chosen to analyse and as I explained earlier due to the large number of students I was not possible to analyse all students so I gained a sufficient sample which I made as unbiased as possible. MY HYPOTHESIS WAS CORRECT

Conclusion of Hypothesis

- The Histograms, frequency polygons proved that the results were more accurate and made more sense than that from the random sampling.
- There is a positive correlation between height and weight. In general tall people will weigh more than smaller people.
- In general boys tend to weigh more and be taller then girls.
- By doing stratified sampling, there were a fewer exceptional values caused by different year groups and therefore ages. I was bound to find irregularities within my data
- The cumulative frequency curves confirm that boys have a more spread out range in weight, with more girls having smaller weights. In height, boys tend to be taller.
- The spearman rank correlation coefficient shows that the correlation between height and weight is strong.
- My Body Mass Index showed that there is a strong trend between height and weight
- In general the taller a person is, the more they will weigh.
- There is a positive correlation between height and weight. In general tall people will weigh more than smaller people.
- There therefore is a positive correlation between height and weight across the school as a whole. This correlation seems to be stronger when separate genders are considered
- If I had taken larger samples my hypothesis may become more accurate.

Hypothesis 2 'Key Stage 4 Students who watch more hours of television on average have a lower IQ Level'

Planning

For this particular hypothesis I will be only be using the data of the School Year of 10 and 11 due to the fact I have specifically chosen to investigate Key Stage 4 and this is the Stage studied by these 2 school years. I have chosen to only base my investigation on this key stage since I believe if I used the complete Mayfield High School data for this hypothesis the range and spread and range of data would be too large to make a sufficient analysis of the results that I will gain. I will be a Sampling Method is which I can sufficiently break down the number of student and meanwhile keep the investigation fair as possible.

Sampling

The sampling method that I have chosen to use for this very particular hypothesis is RANDOM SAMPLING. This is where every item in the population will have an equal chance of being selected. Below is the method that I chose to do the successfully and conveniently. I believe this method will help my results and outcome stay unbiased.

I. I Printed a Copy of the Students and there personal qualities and detail.

II. I then cut out the student First Name and Surnames.

III. As you may remember from my two ways table earlier in this portfolio the total number of student that I have in this particular investigation is 370.

IV. In my sampling the student Gender is not an issue of comparison so I will be using all the names in one sample

V. I then used a Hat which I had and put the all the names into this Hat

VI. I then shook the hat so that I cannot tell or neither can anyone else see the order of the names, which are in the hat.

VII. I then instantly decided to withdraw 30 students from the hat. An unbiased person who was besides me while I was completing this sample adjudged this procedure.

After I had sufficiently completed the sample below are the results that I had obtained. I then typed the results into a Microsoft Excel Spreadsheet. Below is the data that I have gained from my sample. I have decided to only include the Name, Year, Number of Hours of TV, and Favourite TV Show and IQ Level since these is the field, which are important to me whereas other fields are invaluable.

Random Sample Results:

Year Group	Surname	Forename	Favourite TV programme	Average number of hours TV watched per week	IQ
10	Air	Jason	Match Of The Day	2	116
10	Black	Mia	Ali G	14	103
11	Compass	Sharon	Big Brother	40	106
11	Dixon	Graham	The Simpsons	30	102
10	Doens	John	The News	16	101
10	Ewards	Michael	Match Of The Day	27	104
11	Flawn	Elise	Neighbours	10	101
10	Grimshaw	Katie	Blind Date	17	104
11	Jackson	Debi	Eastenders	22	90
11	McCreadie	Jenny	The Simpsons	25	104
10	McDonald	Harold	The Simpsons	21	100
11	McDonald	James	The Simpsons	21	122
10	Edd	Michael		0	81
11	Fillstin	Rowena	Holly Oaks	14	104
10	Jones	Nathan	Bad Girls	8	92
11	Zarrent	Donna	Buffy	36	103
11	Thompson	Kamara	Big Brother	6	89
11	Thomson	Jade	Brookside	18	96
11	Solomons	Ian	The Simpsons	16	100
10	Sosay	Kaiser	Angel	8	124
10	Grimshaw	Jane	Bad Girls	10	95
10	Fox	Serena	Bad Girls	7	108
11	Cripp	Justin	M.T.V. Base	24	100
10	Dolt	Anthony	Big Brother	28	100
11	Donald	Adam	The Simpsons	28	103
10	Quershi	Abduraheim	Charmed	17	103
11	Slim	Andre	Ali G	20	99
10	Slone	Mark	Wrestling	15	100
11	Thompson	Kamara	Big Brother	6	89
11	Acton	Jenny	Coronation Street	22	108

I will now begin to investigate the hypothesis that I have chosen to the fullest extent possible I will first begin by producing frequency tables to firstly find the mean and the average IQ Level of the Data that I have chosen and also the average number of hours of TV watched per student.

Mean IQ of Student in my sample

I will now find the Mean, Mode of the Frequency that I have found and this will be quick efficient and reliable and will help me gain evidence on the average IQ and see whether each student has a IQ high or low in comparison to the mean of the sample.

Mayfield High School (SAMPLE)				
IQ Level	Tally	Frequency (f)	Mid-point (x)	fx
80≤w<90	III	3	85	255
90≤w<100	IIIII	5	95	475
100≤w<110	IIIIIIIIIIIII IIIIII	19	105	1995
110≤w<120	I	1	115	115
120≤w<130	II	2	125	0
TOTAL		**30**		**2840**

Mean can be calculated using this formulae shown below:

Σfx Divided by Σf = 2840 Divided by 30 = 95.7 1d.p

Average IQ Level is **96.**

From this excellent presentation of data I have come to find that 7 Students out of the 20 students are below average that is rather a concerning since this is a large sum of students. Also in addition to this I will be able to now compare these student hours of TV and this will help me towards forming a conclusion on the trend between both the number of hours of TV and the IQ level and prove whether my theory was correct or incorrect. I can also gain from this that the most IQ group or median group is of **100≤w<110** and this group consists of over 60 percent of the student that I have sampled.

Mean TV per week of Student in my sample

As with the first mean frequency table that I had produced this will help me toward making a comparison and to realize a trend that there may be between the number of hours of TV watched per weeks and the particular students IQ this is my aim from producing the Mean Frequency Tables

Mayfield High School (SAMPLE)				
Average TV Time(Hours)	Tally	Frequency (f)	Mid-point (x)	fx
0≤w<10	IIIIIII	7	5	35
10≤w<20	IIIIIIIII	10	15	150
20≤w<30	IIIIIIIII	10	25	250
30≤w<40	III	3	35	105
TOTAL		**30**		**540**

Mean can be calculated using this formula shown below:

Σfx Divided by Σf = 540 Divided by 30 = 18 Hours

Average Hours of TV watched by a student in my sample is **18 Hours**

From this particular presentation of data I have come to find that over 50 percent of Students watch more TV than the average that I have found of my sample after producing a mean frequency table.

Comparison

I can compare that from the tables that I have produced and also by looking straightforwardly at the sample I have created which is that the children who watch a large sum or amount of TV on average per week rather peculiarly have an IQ which is above average in contrast to the average IQ I had found for my sample. Although I have also found using my general knowledge and understanding and interpretation of the data that the students who watch a

reasonable amount of TV have a good IQ level also so this shows me that there is a certain limit to the hours of TV to be watched. Contrary to this TV can help the mind be stimulated and also help to take immediate action in everyday life situation and make you more aware of what is occurring in your surrounding and environment whatever program it may be.

Mode of Girls and Boys Weight

The arithmetic mean of a group of numbers is found by dividing their sum by the number of members in the group; e.g., the sum of the seven numbers 4, 5, 6, 9, 13, 14, and 19 is 70 so their mean is 70 divided by 7, or 10. Less often used is the geometric mean (for two quantities, the square root of their product; for *n* quantities, the *n*th root of their product).

Modal Weight Stem and Leaf Diagram

Student IQ Level		
Stem	**Leaf**	**Frequency**
8	9 9 1	**3**
9	9 5 6 2 0	**5**
10	8 0 3 3 0 0 8 0 3 4 0 4 4 1 4 1 2 6 6 3	**19**
11	6	**1**
12	2 4	**2**

I have come to find from this stem and leaf diagram shows that the 100-109 IQ Level is the most frequent IQ Level from the sample that I have taken for this particular hypothesis. The Modal IQ Level is 100

Key 13 0 = 130 kg

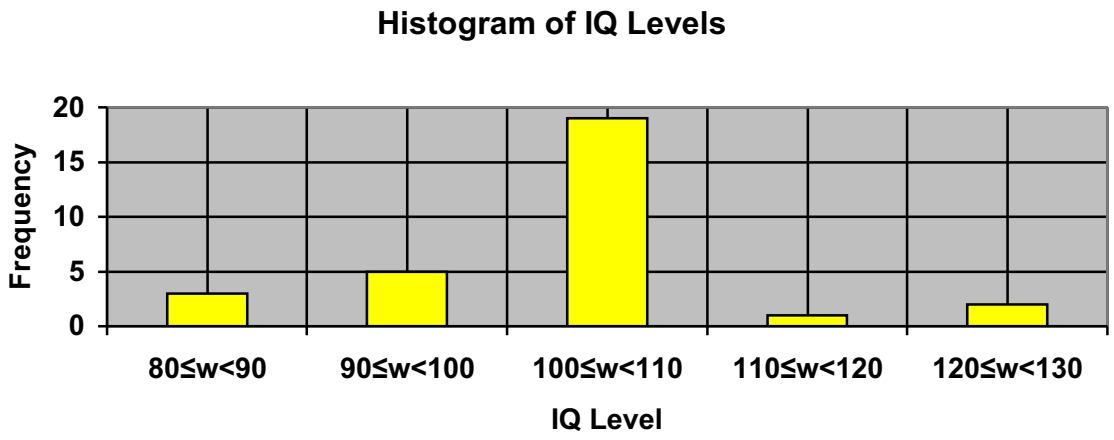
Average TV hours per week		
Stem	**Leaf**	**Frequency**
0	2 0 8 6 8 7 6	**7**
1	4 6 0 7 4 8 6 0 7 5	**10**
2	7 2 5 1 1 4 8 8 1 2	**10**
3	0 6	**2**
4	0	**1**

This particular Stem and Leaf Diagram has shown both the frequency groups of 10-19 and 20-19 have both 10 frequencies and also find that the modal Average Hour is 21

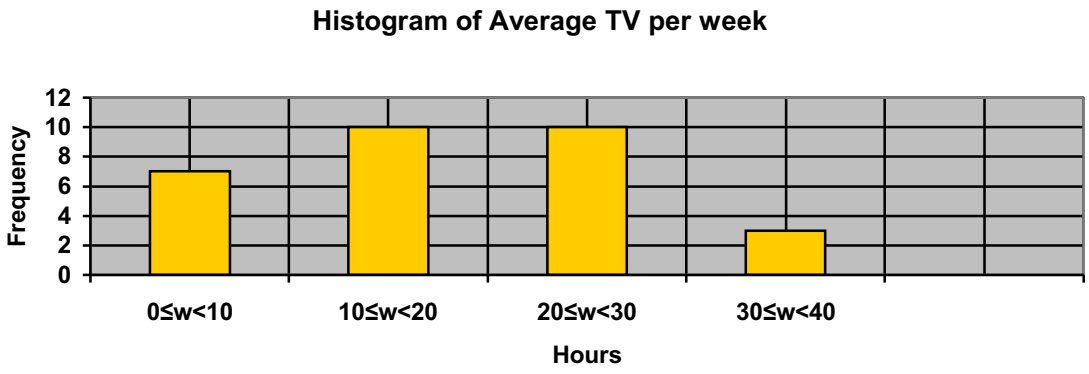
Histogram

This form of data representation will help me find the spread of data and also the trend in data and the results and firstly below is a histogram showing the IQ levels of my sample:

IQ



Average Hours of TV per week

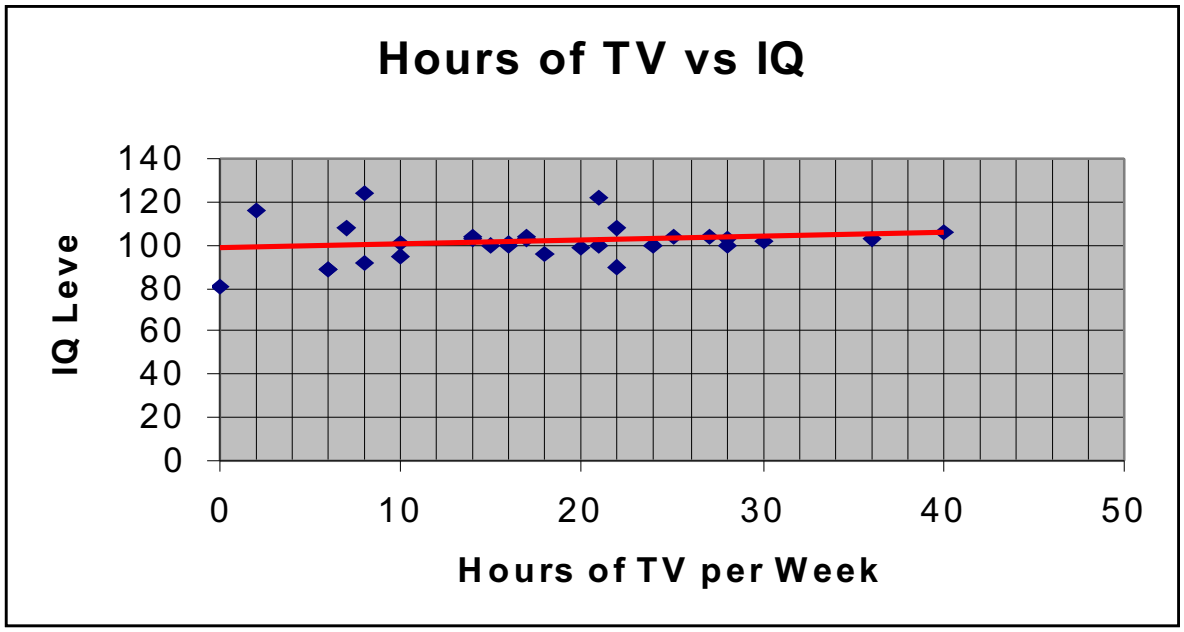


This shows me the Median Group for IQ Level 100-110 and the Number of Average hours varies Median group of 10-30

Scatter Graph

I will now produce a Scatter Graph. I will use this to compare the two sets of data when there are two variables, which are IQ/Hours of TV. Also in addition to this a straight line is used to join the dots and this will help me view any trend between the two variables that there may be. The Data I will input in Microsoft Excel and this will help me produce an efficient and successful graph in which I can then evaluate on the trend.

LINE OF BEST FIT TO SHOW TREND CLEARLY IS ALSO INCLUDED.



This Scatter Graph that I have produced shows that there is a very weak positive correlation between the two variables that I am investigating for this hypothesis and this means that that the increase in Hours of TV the higher the IQ on average which is rather bemusing as I had expected the opposite using my general knowledge and everyday theory although the statistics and my sample have proven me wrong at this current moment. It can also be easily viewed that although I may be incorrect there are some irregularities in the sample or variances in that particular person characteristics and personality. Also I feel this trend may be since the TV helps stimulate a child mind to a certain degree and whichever program it may be it has some effect on the way a child lives there live and in my sample the effect is largely positive as I have duly come to find. The variances also show that my sample is not unbiased and reliable as it is natural and unique for they're to be some minor irregularities. Overall there is a casual relationship between the IQ Level and Hours of TV watched for Children at Mayfield School

Variance and Standard Deviation

Variance is a measure of the spread of the distribution and the square root of the result in the Standard Deviation. This can be calculation using a certain formula, which is shown below. By completing this statistical task it will help me measure the spread of the data and the mean of the distribution.

Standard Deviation Formula:

$$\sigma = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{\Sigma f}}$$

Below I will now produce a table in which I will calculate the standard deviation for firstly the IQ and then the Average Hours of TV:

IQ Levels

IQ Level (Midpoint of Group)	85	95	105	115	125
Frequency	3	5	19	1	2

The Mean Is

$$\frac{(3 \text{ X } 85) + (5 \text{ X } 95) + (19 \text{ X } 105) + (1 \text{ X } 115) + (125 \text{ X } 2)}{30} = 3090/30 = 103$$

And the standard deviation will be

$$\sqrt{\frac{[2 \text{ X } (85\text{-}103)] + [5 \text{ X } (95 - 103)] + [19 \text{ X } (105\text{-}103)] + [1 \text{ X } (115\text{-}103)] + [2 \text{ X } (125\text{-}103)}{30}}$$

$\sqrt{}$ 18/30 = 0.7746 4.dp

I will now complete the same Standard Deviation Calculation as I had produced but in this particular case I will do the Deviation and Spread of Data on Average Hours per Week per student

Average Hours Per week TV

Average Hours Per TV week (Midpoint of Group)	5	15	25	35
Frequency	7	10	10	3

The Mean Is

$$\frac{(5 \text{ X } 7) + (10 \text{ X } 15) + (10 \text{ X } 25) + (3 \text{ X } 35)}{30} = 540/30 = 18$$

And the standard deviation will be

$$\sqrt{\frac{[7 \text{ X } (5\text{-}18)] + [10 \text{ X } (15 - 18)] + [10 \text{ X } (25\text{-}18)] + [3 \text{ X } (35\text{-}18)]}{30}}$$

$\sqrt{}$ 1/30 = 0.0333 4.dp

Conclusion

From the Standard Deviation that I have come to find how spread out the data that I have used in my sample for this particular hypothesis and measure of the dispersion of the data and has helped me see whether the data I have is unbiased and of a sufficient spread and size and it will also help me measure the reliability of the sample that I have taken is that it is reliable which is usual on a scale between –1 to 1 and also that the spread of data in the IQ levels in larger than the Average TV hours watched which shows me that in my final evaluation the results for Average TV hours watched per week will be generally more accurate and conclusive.

Standardised Scores

I will now find the Standardised score for 5 students from my sample at Random. It will help me compare the values.

5 Student that I have chosen for this calculation are:

11	**Flawn**	**Elise**	**Neighbours**	**10**	**101**
10	**Grimshaw**	**Katie**	**Blind Date**	**17**	**104**
11	**Cripp**	**Justin**	**M.T.V. Base**	**24**	**100**

Standardised Scores Calculation/Formula

$$\frac{\textbf{Score} - \textbf{Mean}}{\textbf{Standard Deviation}}$$

	Elise	Katie	Justin	Mean	Standard Deviation
IQ Level	101	104	100	**103**	**0.7746**
Average Hours TV Per Week	10	17	24	**18**	**0.0333**

Elise Standardised Scores

IQ Level = (101-103) / 0.7746 = **- 2.58**
Average TV Hours = (10-18)/0.033 = **-240**

Katie Standardised Scores

IQ Level = (104-103) / 0.7746 = **1.29**
Average TV Hours = (17-18)/0.033 = **-30.30**

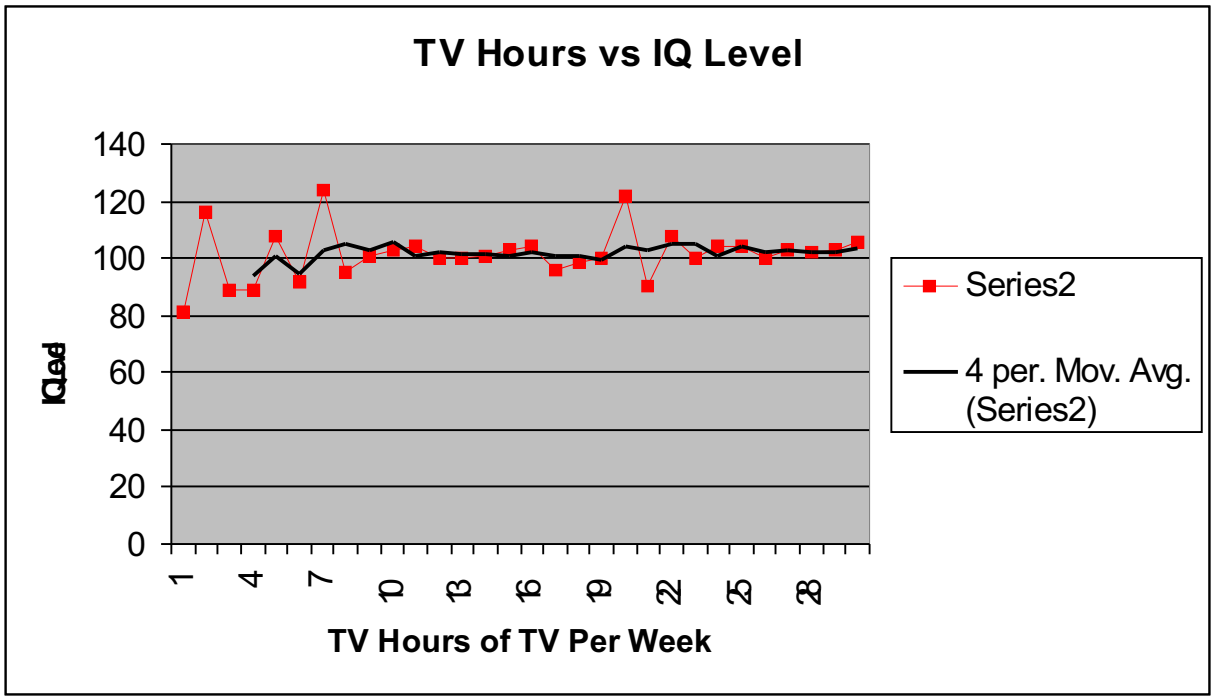
Justin Standardised Scores

IQ Level = (100-103) / 0.7746 = **- 3.87**
Average TV Hours = (24-18)/0.033 = **181**

From this I find that that Katie has the better standardized result since her results is a positive figure whereas as Justin's standardized score was a negative value and Elise has the best Average TV house standardized score.

Time Series Line Graph

A Line graph is used to display data when the two variables are not related by an equation and you are not certain what happens from one point to another. I will need to plot this graph extremely carefully due to the fact I cannot enter all the data in my sample and produce this graph since there will no trend which I will be able to view so I will viewing student at different Average Heights in Order and then Plot their IQ Levels on a Line Graph and from this I will able to make a trend comparison. The Line graph will consist of TWO Variables.



On the Whole it is very compact and difficult to give a definite conclusion according to this graph due to the fact the trend is fluctuation up and down and very slightly and upward trend although it is very difficult to tell and verify and form an definite answer to the hypothesis that I have chosen to investigate. Also in addition to this the verification are rather minor and steady and fluctuate largely. Also I can add to this that the basic IQ level is not efficient and excessively effected by the Number of Hours of TV you but generally your age and your understanding and knowledge overall.

***I WILL BE CONCLUDING AND EVALUATING THIS PARTICULAR HYPOTHESIS AFTER THE COMPLETION OF THE THIRD HYPOTHESIS**

Hypothesis 3

' Left Handed Students have higher IQ levels and Key Stage 2 Results in Comparison right handed student at Mayfield High School'

Planning

For my third and final hypothesis I will need the data of Mayfield High School between the years of 7 to 11 and this is due to the fact I need a wide sampling frame so that my analysis and results are not unbiased and as accurate and efficient as possible. Also in order to make my portfolio unique I will be forming a SAMPLE in which I can accurately and efficiently minimise the range of the data but meanwhile keep the hypothesis as unbiased as possible as I had explained at an earlier stage.

The Sampling Method I will be using to do this is STRATIFIED as with the first hypothesis since I duly believe that this sampling method is the most accurate and reliable and meets the needs and requirements of this particular hypothesis to the fullest extent. I will now explain its benefits and the method in which I calculated and completed the Sample.

I will be calculating my Stratified Sampling using the Table above now and I will need calculation proportion to stratify my data spread/range.

Here is another copy of the two-way table of the table, which helps me form my sampling method.

Year Group	Number of Boys	Number of Girls	Total
7	151	131	282
8	145	125	270
9	118	143	261
10	106	94	200
11	84	86	170
TOTAL	**604**	**579**	**1183**

On the upcoming page you will find the calculation of my Stratified Sample.

Stratified Sampling Calculations

I want 20 Students from the school total 1183.

I have come to find using Microsoft Excel that **251 Students throughout the school are Left Handed** whereas **Right Handed Students are 922** and there are **10 Both Handed** so these student have been left out of my investigation for this hypothesis. As I have found Right Handed Student is common for most student.

Right Handed Students= 922 Divided By 1173 multiplied by 20 = **16**

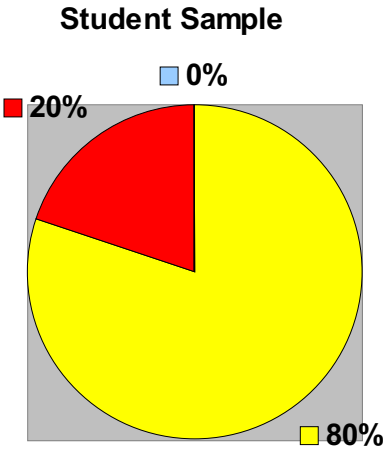
Left Handed Students= 251 Divided by 1173 Multiplied by 20= **4**

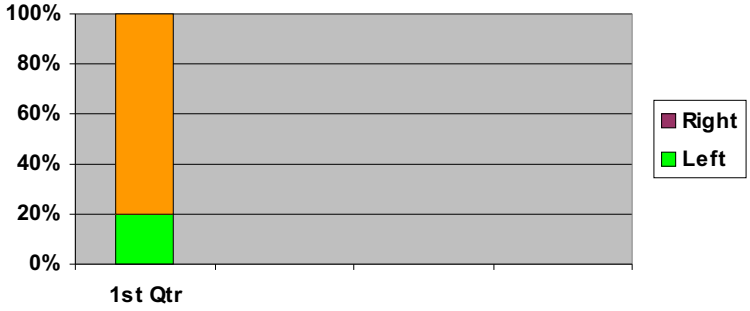
Below are the Results of my Sample

Surname	Forename	Right Handed	IQ Level		SAT's Results	
Robinson	Luke	Left	100	4	4	4
Lister	Kuta	Left	104	4	5	4
Abejurouge	Henry	Right	89	3	3	3
Aberdeen	Richard	Right	103	4	5	4
Acton	Jenny	Right	108	5	5	5
Agha	Shohaib	Right	91	4	3	4
Ali	Aisha	Right	103	5	4	4
Ali	Amera	Right	90	4	4	4
Alsam	Samia	Right	97	2	4	4
Coleman	Jenifer	Left	100	4	4	4
Rooster	Hally	Left	103	5	4	4
Kennedy	Matthew	Right	103	4	5	5
Khan	Jamal	Right	100	4	4	4
Lall	Alex	Right	92	3	4	3
Leonard	Robert	Right	102	4	3	4
Muppeteal	Nubaid	Right	100	4	4	4
Salah	Summy	Right	100	4	4	4
Sammy	Singh	Right	104	4	5	5
Shady	Philip	Right	93	3	4	4

I will now begin my Statistical Investigation on this particular hypothesis and I will represent my data in many various methods and also produce analyse after each graph that I will be producing. I will begin with firstly forming a pie chart at the number of students in my Sample that are Right and Left Handed and also a multiple bar chart showing the number of students investigated.

Pie & Bar Charts



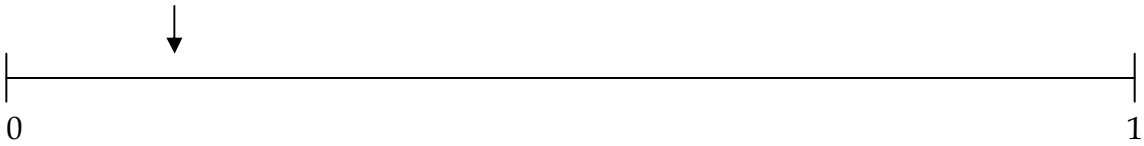


The Graphs that I have composed besides this particular text is that 80% of the students in my sample are Right Handed which are basically a large sum of students from a very small sample and also I have come to find that in every 5 students there is 1 Left Handed student and 4 right handed student. Also in addition to this I have expressed this in two separate forms of data representation and it can be clearly viewed the differences and its appeal rather than text.

Probability

Probability is an area that is used to predict the chance of something happening in the future or the likelihood of something to occur successfully and efficiently.

Here is the likelihood of a student being Left Handed from my sample:



I will now analyse data further in depth the sample that I have collected and form a Two Way Table that shows the boys and girls that are Right & Left Handed:

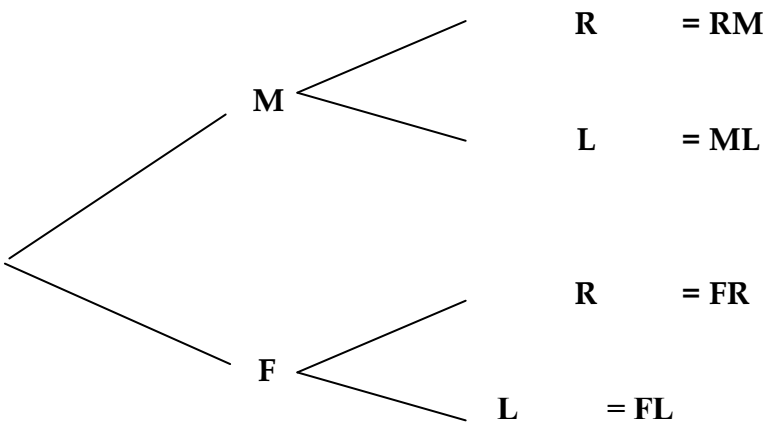
	Male	**Female**	**Total**
Right –Handed	11	6	16
Left -Handed	1	3	4
Total	11	9	**20**

Probability Tree Diagram

My variables are Male and Female and Left and Right Handed.

M= MALE
F=FEMALE

R= RIGHT
L= LEFT



Calculations have been written by Hand Below:

I have decided that if I am to compare my hypothesis efficient and to a certain high standard then I will have to analyse the data by looking at the data and picking put certain data sample and compare the results and form a hypothesis:

- FIRSTLY I WILL FIND TO SIMILAR STUDENT WHO ARE SIMILAR IN CHARACTERISITICS AND PERSONALITY

- BELOW IS THE DATA THAT I WILL BE USING:

Surname	Forename	Gender	Left or Right Handed?	IQ		SATS Results	
Matthew	David	Male	Left	102	4	4	4
Smith	Amrit	Male	Right	100	4	3	4
Muppeteal	Nubaid	Male	Right	101	4	4	4
O'Neill	Krisila	Female	Left	100	4	4	5

- AS YOU CAN SEE FROM THE DATA THE DATA IS REASONABLY SIMILAR AND I WILL BE ABLE TO FORM A SUCCESFUL CONCLUSION FROM THIS.

***Conclusion for both Hypothesis 1 and Hypothesis 2 can be found on the upcoming page**

Conclusion/Evaluation of Hypothesis 2

'Key Stage 4 Students who watch more hours of TV on average have a Lower IQ Level.

I feel that I have successfully completed and investigated my hypothesis to an extent in which I can be sure of my accuracy of my conclusion and I gained many sufficient forms of evidence. As stated above my hypothesis was to find whether Students who spent a large sum of time per week watching TV have a lower IQ Level and I had come to find that my theory was incorrect to a certain level. I had used many statistical representations to prove my theory.

I was also aiming to find whether there was a regular trend between the two variable and aimed to make my hypothesis as unbiased as possible.The Histogram that I had produced showed that my results were unbiased and relatively accurate in comparison the stratified choice of methods and made more sense on the whole.There is a very weak positive correlation between IQ and TV hours although it is not enough to prove whether my theory was correct and there was not a strong positive correlation that I had expected. In general IQ Level does not increase an Hours of TV increase.Some irregularities were found. The Standardised Score and Standard Deviation help me find the spread of the data so the sample is unbiased and insufficient.

If I had taken large sample my hypothesis may become more accurate and able to form a successful conclusion. The IQ also depends on the persons surrounding, ability and knowledge and stimulation and motivation which can all play a factor in the results.Overall I have found that my Hypothesis was incorrect and the statistical evidence that I had gained did not back up my theory. Another reason behind my misfortune is the range of data from 7-11 is too wide and I should have narrowed the frame down but now helps me in the future.

Conclusion/Evaluation of Hypothesis 3

'Left Handed Students have higher IQ levels and Key Stage 2 Results in Comparison Right handed student at Mayfield High School'

In my honest opinion I feel that I have proved my hypothesis in a short span and I feel that the statistical evidence that I have duly gained is more then sufficient to form a hypothesis and if I used any other forms of representing data it would have meant I was generally repeating. Also I feel that I have used the correct statistical evidence to prove my theory and used the right evidence. In addition to that I feel that Left Handed students are generally have higher IQ Levels according to my sampling frame and investigate in this is what I have based my theory upon.