Statistics Coursework

For my initial investigation I am going to compare results on the following hypothesis, to see if people's intelligence affects their height and their ability to memorise a certain amount of objects in a pre-prepared memory test. I hope I ill be able to prove the following:

- The higher maths set in school lessons a person is, the taller they are.
- The higher maths set in school lessons a person is, the better their memory is.
- The taller someone is, the better their memory is.

For each hypothesis, I will encounter different problems, and I have to approach each hypothesis with a different technique to ensure I will get a fair and conclusive results. For each one I am going to use primary data, using pupils form Hurworth School.

The first hypothesis is: The higher maths set in school lessons a person is, the taller they are. I think this will be true because the most intelligent people are usually older and more mature. I will design a section of my survey sheet as drawn below:

Height	Set

For my next hypothesis, I am going to investigate whether the set a person is in affects their memory. I will prepare a memory test of 10 random objects and get their teacher to hold this up in front of the class for 30 seconds. The pupils then have to write down as many objects as they can remember on the survey sheet I will be sending round the tutor groups. My memory test will consist of non-related objects, such as envelope, a moon, diamond, etc.

My third and final hypothesis is that the taller someone is, the better their memory is. I think this will be true because it will relate to my previous hypothesis of the taller you are the more intelligent you are the taller you are. I will simply compare results from the last tests I did on the survey sheet, the height records and the score from the memory test.

There are a few problems that could possibly arise during the investigation, I will list them below and state how I intend to deal with each possible problem.
- People could possibly cheat by conferring during the memory test
- I have to get the teachers to co-operate throughout the test, by making sure no cheating occurs
- People may feel uncomfortable about disclosing their height, or may not know their actual height

I intend to solve these problems by attaching a personal note to each teacher asking for them to check there is no cheating or conferring

happening in the room, and make sure nobody talks or confers throughout the memory test. I will send out a pre-test first, to see how I can manage the problem if people do not disclose their height.

Having designed my survey, I found out that it was going to take a very long time to get every year 7 to answer each question, and considering there was 31 people in our class, to get everyone's survey answered and returned was impractical and unnecessary. To solve this problem, we combined everyone's hypotheses into one large questionnaire that included all tests, and sent it out after doing a pre-test. It is classed as primary data, because we collected the data ourselves.

Pre-test

We pre-tested our first design of the survey on ourselves as a class, and did the memory test, measuring lines and angles and answering simple questions on ourselves e.g. height and weight. We then handed in our surveys and our scores were recorded. However, simple problems occurred which need to be prevented before the actual survey is sent out. Some of the problems, on the other hand, will automatically be solved because they only happened because everybody in our class knew each other well therefore found it quite simple to confer throughout the survey. Whilst doing the measuring lines section on the survey, quite a few people were hiding rulers to give them a more accurate measurement of lines. However this ruins the whole aspect of doing the investigation, because the test then relies on who used a ruler and who didn't, rather than what affects your estimating of lines.

Next

Firstly I needed to make a table of the amount of females and males in each set. From this I need to use the data provided to create a stratified sample of however many pupils I need out of my total of 40 males and 40 females. I used stratified sampling because I needed a representation of the whole year group, with correct proportions for each set, males and females etc. I have shown the table I created below:

Codes for Sets and Gender	No. of People in Set	(No of people in each set divided by total males or females multiplied by 40)	Stratified Sample of 40 males or 40 females
Females set 1	22	18.33333333	18
Females set 2	12	10	10
Females set 3	6	5	5
Females set 4	6	5	5
Females set 5	2	1.666666667	2
Males set 1	9	7.826086957	8
Males set 2	12	10.43478261	10
Males set 3	14	12.17391304	12
Males set 4	7	6.086956522	6
Males set 5	4	3.47826087	4

The following table is an example of how I worked the first stratified sample out, of Females Set 1:

| Codes for Sets and Gender	No. of People in Set	(No of people in each set/total males or females*40)	Stratified Sample of 40 males and 40 females
Females set 1	22	18.33333333	18

The formula I used to work out the actual amount of people I will be using in my investigation is: 22/48*40. 22 is the amount of people in Set 1, 48 is the total amount of females in all sets, and 40 is the amount of people I want at the end, to use their results.

Once I have found the amount of people I want to use, I will then use a random sampling technique to select the actual pupils. The formula I will use is:

$$=RAND()*(22-1)+1$$

22 is the amount of females in the actual set, (for example in this case there is 22 people in Set 3). This formula then brings up a random number, for example 3.4567, or 1.25698. I rounded the number up to the nearest whole integer to get my random sample.

Graphs

Histogram to show heights of people in lower sets   Histogram
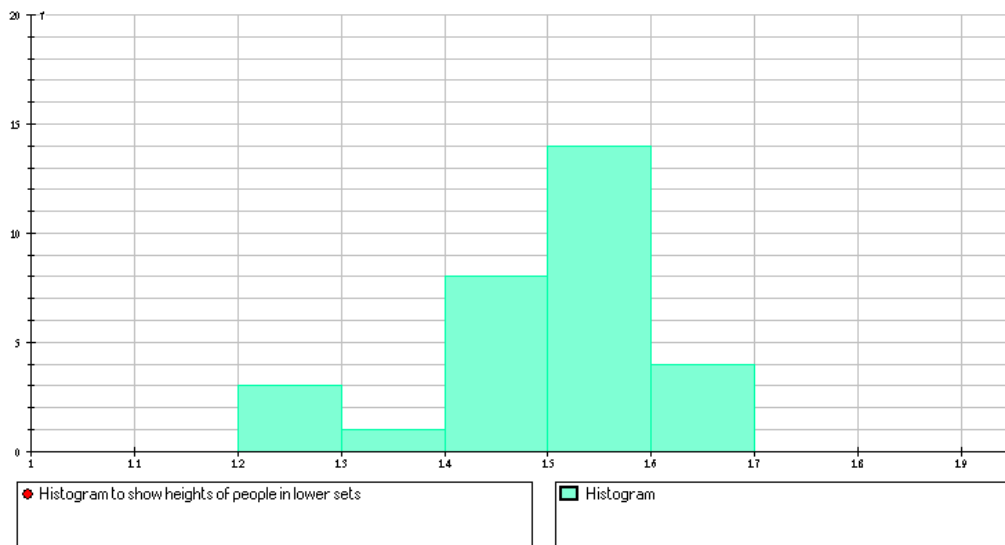
Histogram to show the heights of people in Lower Sets. I have used histograms for most of my 'height' graphs, as I am dealing with mainly continuous data, which is data that can always get more accurate. Later on, however, for discrete data, such as the memory test results, where they can only be one definite answer, I have used graphs such as box plots or scatter diagrams.

The statistic box for the graph (above) shows that:
- The mean height of people in lower sets is 1.57m.
- The standard deviation of heights is 0.47, which means that on average, each height is 0.47m away from the mean.
- The median height is 1.524m

I will now show the graph for the heights of people in higher sets, and then I will be able to compare my graphs and see whether my first hypothesis (That people in higher sets are taller than people in lower sets) is true.

● Histogram to show heights of people in high sets   □ Histogram
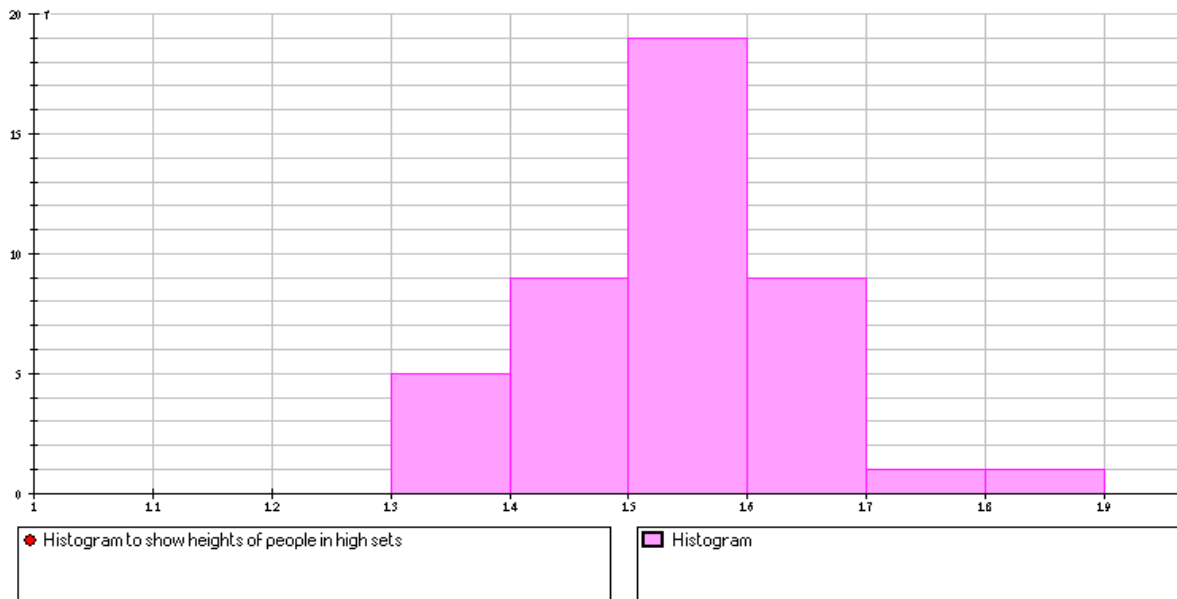
This is my graph to show the heights of people on higher sets. Some results, as seen in the 'statistics box' are as follows:
- The mean height of the students is 1.54m
- The standard deviation of heights is 0.09m, which means that on average, each height is 0.09m away from the mean.
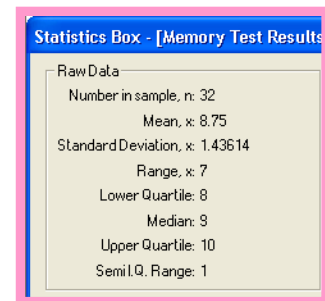- The median height is 1.549m

I think this is probably because in hindsight, my first hypothesis was not the most realistic one, and I could have maybe thought it through a bit more, with more statistical explanations to back up my theory.

My second hypothesis is: The higher maths set in school lessons a person is, the better their memory is. I will now present my graphs and explain what each one shows.
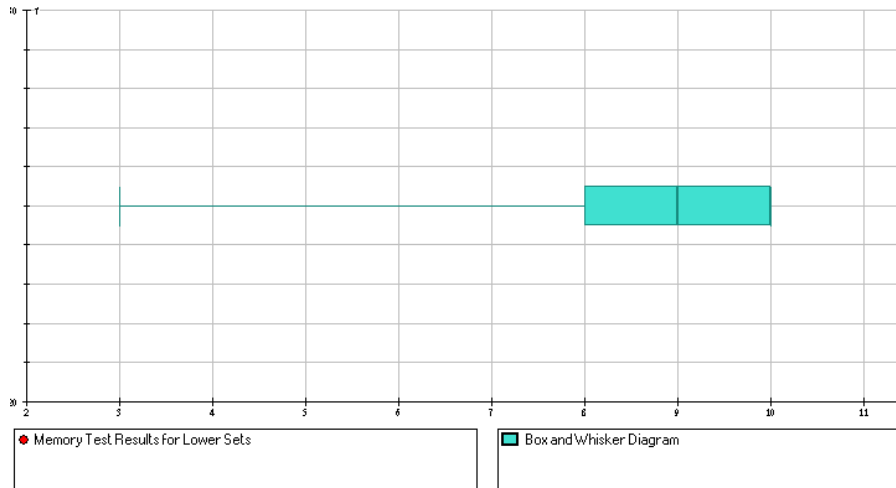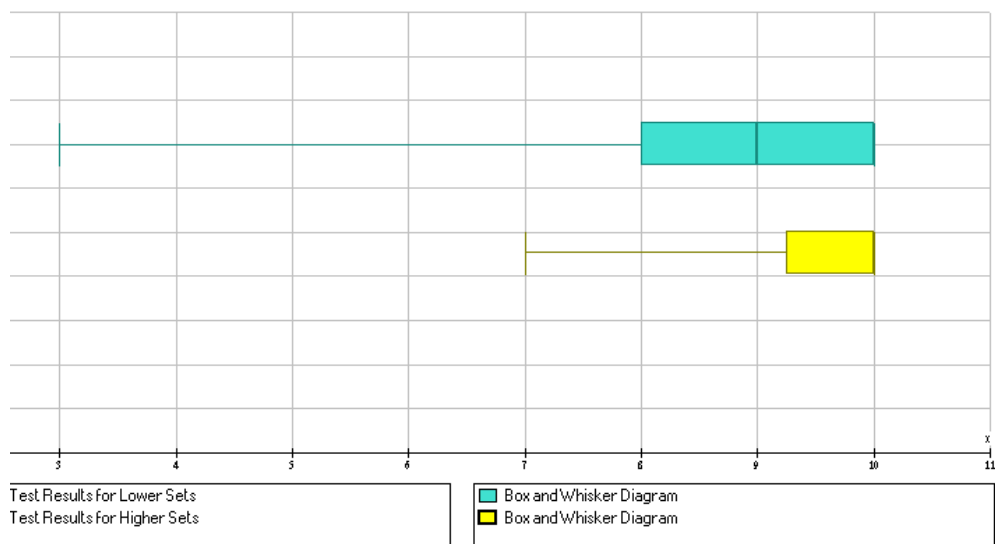
The box plot above shows the Memory Test Results for people in Lower sets. Because the data is not continuous data, it is discrete; I have created a box plot rather than a Histogram. By looking at the box-plot's statistics, I can see that:

Statistics Box - [Memory Test Results]
Raw Data
Number in sample, n: 32
Mean, x: 8.75
Standard Deviation, x: 1.43614
Range, x: 7
Lower Quartile: 8
Median: 9
Upper Quartile: 10
Semi I.Q. Range: 1

- The mean memory test result for people in lower sets is 8.75
- The standard deviation is 1.44, which means that on average, each person was roughly 1.44 away from the mean score
- The median result was 9
- The Inter-quartile range is 2, which means the difference between the upper and lower quartiles is 2. The smaller the spread of the data the better the results are.
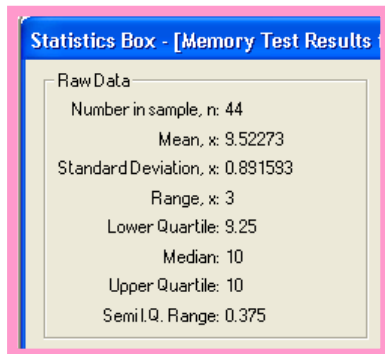
I will now put the memory test results for higher set's results onto the same graph as the lower test results, to make it easier to compare results between the sets.

Test Results for Lower Sets
Test Results for Higher Sets

Box and Whisker Diagram
Box and Whisker Diagram

The graphs above shows the memory test results for both lower and higher sets. I will firstly explain some statistics for the Higher Sets, and then compare both sets' results.

The yellow graph is Memory Test Results for Higher Sets. It shows:

Statistics Box - [Memory Test Results
Raw Data
Number in sample, n: 44
Mean, x: 9.52273
Standard Deviation, x: 0.891593
Range, x: 3
Lower Quartile: 9.25
Median: 10
Upper Quartile: 10
Semi I.Q. Range: 0.375

- The mean memory test result for higher sets is 9.5.
- The standard deviation for higher set's results is 0.89, which means that on average, each person's result deviates 0.89 from the mean.
- The inter-quartile range is 0.375.

Comparing the Box Plots

I can now see that for higher sets, the mean test result is 9.5, whereas the mean result for lower sets is 8.75. This proves my hypothesis is. The standard deviation for lower sets was 1.44, whereas for higher sets it was only 0.89. This means that as well as having a higher average score; the standard deviation is reduced by 0.55 in higher sets, making the overall memories of the people better and more accurate. Memory test results for the Higher Sets also have a lower inter-quartile range, meaning the data is less spread out, and the smaller the spread of the data, the better and more accurate the results are.

I will now attempt to prove my third hypothesis true, the taller someone is, the better their memory is. I can already make an accurate prediction as to whether my theory will be true or not. Because I have already looked at the hypothesis that relates the height of people to their maths set, I can say that my third hypothesis will probably be UNTRUE. This is because the shorter people in the year were in the lower sets, but the higher sets had the better memory, therefore the taller someone is, according to my previous investigations, the worse their memory will be. I will now compare graphs and see if my hypothesis is true or untrue.

Scatter graph to compare people in higher sets memory test result and height is below:

Graph to Show Height and Memory Test Results of People in Higher Sets
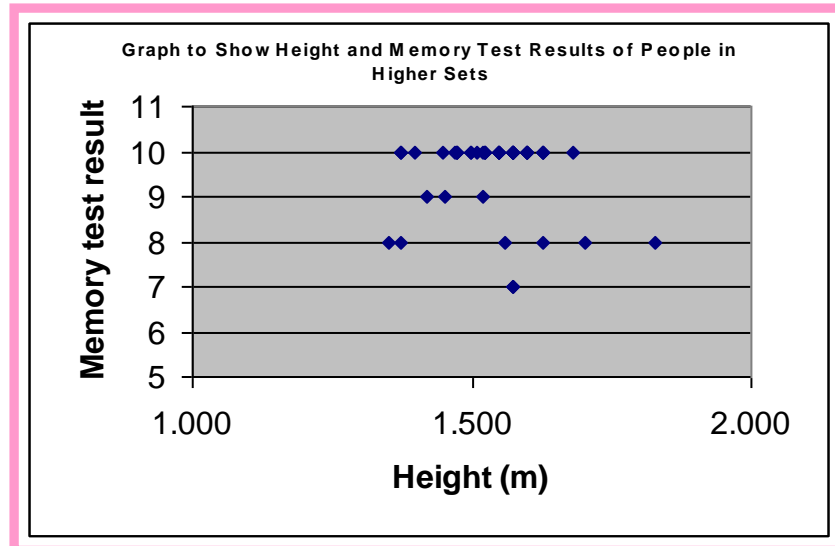
This shows no correlation between the height of people and their memory test results. I will now do the same graph for lower set people:

Graph to Show Height and Memory Test Results of People in Lower Sets
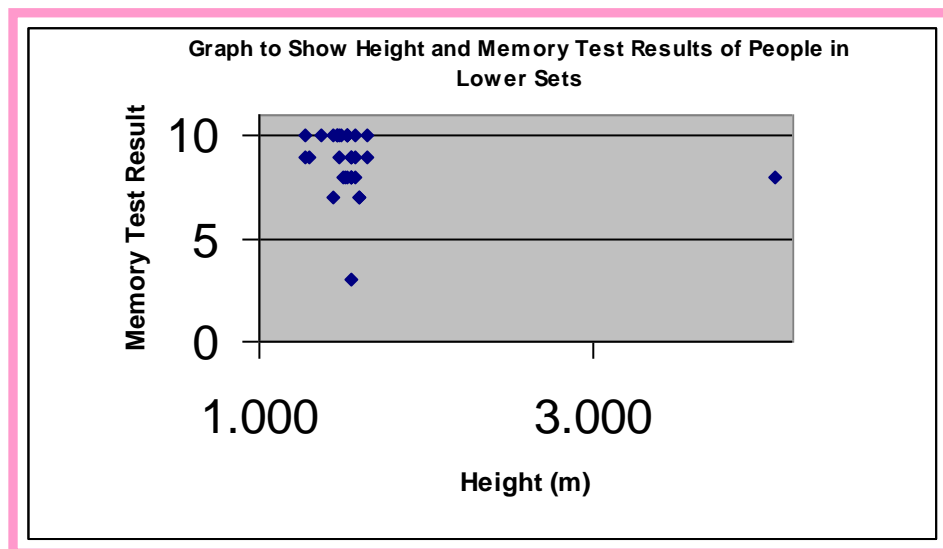
There are, however, two anomalous results on this graph, which I am going to discard otherwise it will affect my results and therefore possibly my conclusion. The new graph is below:

Graph to Show Heights and Memory Test Results for People in Lower Sets
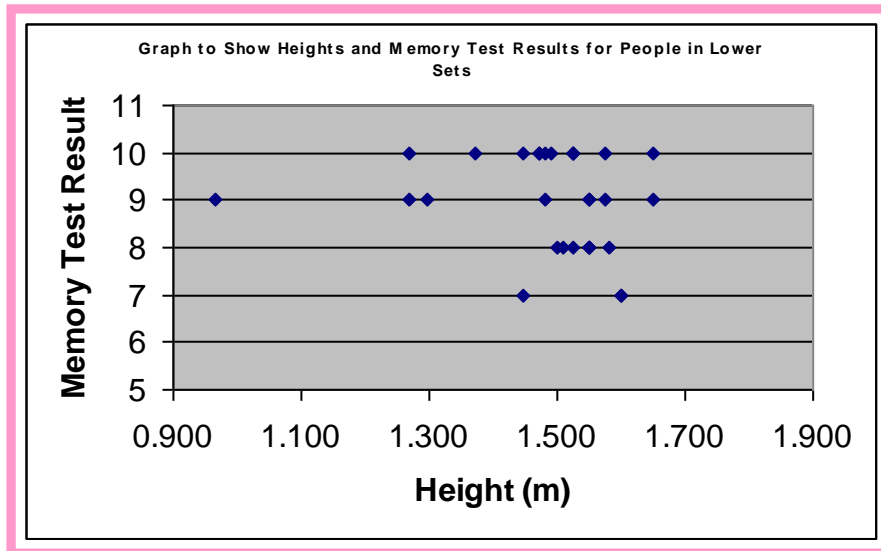
You can now see that there is also no correlation between the heights of people and their memory in Lower Sets, so I will now use the statistic information to compare my results and prove my hypothesis. By looking at my previous data I have concluded that: The Higher set in maths a person is, the better their memory is. Also, I have concluded that the LOWER set in maths a person is, the taller they are. Therefore, my third hypothesis is. I think this is because again, my hypothesis was not the most realistic one, and I only chose it because it compared my first two hypotheses and I thought a conclusion could have possibly been made. However, because my first hypothesis proved untrue, the direction of my other two hypotheses was uncertain, resulting in the third hypothesis being also untrue.

Conclusion and Evaluation

After carrying out the surveys and completing my statistics coursework, I have looked at some of the problems that arose and how I could improve these if I was to do the same survey again. One of the main problems was my first and third hypothesis, being that they were both proved to be untrue. This, in itself, isn't a problem; however I think in hindsight I could have maybe thought them through slightly more to ensure more realistic prediction therefore possibly true hypotheses. However, this did make my conclusions slightly more interesting and different, even though they didn't prove my hypothesis to be right. In doing the actual surveys, I discovered the main problems were people who didn't interpret the questions correctly, or were not sure of the average height they would be, therefore guessing at heights of up to 4.5 metres!

 I did manage to just eliminate these results, because these looked likely to affect my final results. Because my hypotheses involved using the heights of people, I couldn't afford to have any mistakes or anomalies on my graphs because when it came to draw my conclusion, it was almost certain my hypothesis would be proved different to what it eventually was. So I discarded these results and used the ones that didn't deviate too much from

the average height of people. The things I would improve are probably – asking for the heights in feet and inches because people tend to know their heights in this form rather than in metres and centimetres. However, this would probably take up too much time, as I would then have to convert each height into metres and centimetres, as this is much easier to work with. Something else I could probably improve on is the memory test, as shapes are probably the simplest form of pictures and maybe a slightly more complex test would make the scores more accurate because you would be able to notice the difference between lower and higher sets more. Overall, however, although two of my hypothesis proved untrue, I was pleased with the survey and my results, as my graphs and tables all turned out correct and I was able to show statistical data quite well.

An alternative hypothesis that could possibly have been proved true, could have involved estimating the length of lines or size of angles. For example I could compare memory test results to results of line measuring. 'The better people are at measuring lines, the better memory they will have' would be a better hypothesis to compare the two results. This would possibly be more realistic and provide more sensible results.