Adam Wright

# <u>Maths Statistics Coursework</u>

<u>Body Statistics</u>

As we grow up we change - or do we? Boys are different from girls - or are they?

Use samples from different age groups for both sexes. Choose two measurable statistics that can be sampled easily and sensitively. You need one statistic you think will be different and one you do not think will be different and use a sample for each.

Make a hypothesis.

<u>Collecting the data</u>

 We collected data from those who were both older and younger than us. I was directly involved in data capture of body statistics for two classes of year sevens who had maths simultaneously as us. We chose to survey the whole of the accessible population because of the relatively small size of it. We chose to sample to classes because the amount of data we would collect, about 60 sets, would be about correct. It is large enough to collect in a reasonable amount of time. It is also large enough to be able to discount any obvious discrepancies without the data figure falling below 50.

We presented the children with a form which they complete themselves with basic information such as shoe sizes. As a class they collectively took there own pulse rate. Although this may not seem over accurate with children of this age most of the data we received falls within human medical boundaries.

<u>Sampling the data</u>

- I will take a random sample of 30 students from each year group we have information.
- The sample will consist of 16 boys and 14 girls. This is to most accurately represent the 53%: 47% split in the population births. I found this information out by e-mailing the national statistics office. The sample will therefore be a random quota sample.
- To generate my sample I will assign within each year group each sex set of data a two digit number and then by using the random number generator on my calculator with the first two numbers it creates will relate to a piece of data. If the first two digits either fall outside the boundaries or are a repeated I will redo the process.
- The process is to be repeated until the right number, either 14 or 16, pieces of data have been collected.

**Hypothesis A**    *As age increases therefore height increases. Boys throughout are taller than there female counterparts*
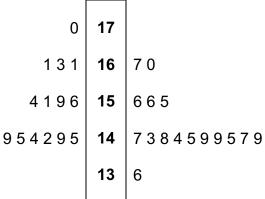
Initially I will be collecting my data in a stem and leaf diagrammatic form. This is because
- It is a simple process to initially sort data
- Straight away it can give us a rough idea of the mean with a small amount of calculation. Also we can find out a mean and an exact figure of median with little calculation.
- We can see the rough shape a graph would take.

The first stem and leaf diagram will be unordered. I will then do a second and order my data.

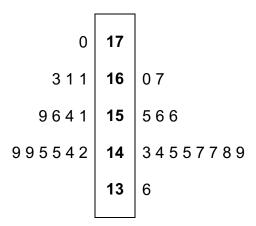Year 7 Data

Males                                                                                                    Females
47% population                                                                                   53% population

| Males | Stem | Females |
|------:|:----:|:--------|
| 0 | **17** | |
| 1 3 1 | **16** | 7 0 |
| 4 1 9 6 | **15** | 6 6 5 |
| 9 5 4 2 9 5 | **14** | 7 3 8 4 5 9 9 5 7 9 |
| | **13** | 6 |

Straight away it is possible to see that on average males are slightly taller than females at this age group. This is a simple observation.

I have now sorted my data.

| Males | Stem | Females |
|---:|:---:|:---|
| 0 | **17** | |
| 3 1 1 | **16** | 0 7 |
| 9 6 4 1 | **15** | 5 6 6 |
| 9 9 5 5 4 2 | **14** | 3 4 5 5 7 7 8 9 |
| | **13** | 6 |

Year 11 Data

| Males | | Females |
|---|---|---|
| 47% population | | 53% population |

| Males | Stem | Females |
|---:|:---:|:---|
| 5 5 0 0 3 | **18** | 2 |
| 5 1 2 8 0 2 | **17** | 0 |
| 8 5 7 | **16** | 4 2 4 2 8 0 0 5 4 |
| | **15** | 8 5 8 8 8 |

Males are clearly taller than females in year 10

| Males | Stem | Females |
|---:|:---:|:---|
| 5 5 3 0 0 | **18** | 2 |
| 8 5 2 2 1 0 | **17** | 0 |
| 8 7 5 | **16** | 0 0 2 2 4 4 4 5 8 |
| | **15** | 5 8 8 8 8 |

Year 12 Data

Males                                                                      Females
47% population                                                      53% population

```
  5 5 0 0 3 | 18 | 0
  5 1 2 8 0 2 | 17 | 2
      7 5 8 | 16 | 8 6 4 1 6 7 0 0 5 4
            | 15 | 5 8 5 8
```

Sorted data

```
  5 5 3 0 0 | 18 | 0
  8 5 2 2 1 0 | 17 | 2
      8 7 5 | 16 | 0 0 1 4 4 5 6 6 7 8
            | 15 | 5 5 8 8
```

Drawing Conclusions from the data

| | | Year 7 | | Year 11 | | Year 12 | |
|---|---|---|---|---|---|---|---|
| | | Male | Female | M | F | M | F |
| ↑ | Highest Value | 170 | 157 | 185 | 182 | 185 | 180 |
| | Upper quartile | 160 | 155.5 | 180 | 164.5 | 181.5 | 166.5 |
| | Median | 152.5 | 148.5 | 173.5 | 162 | 173.5 | 164 |
| ↓ | Lower Quartile | 147 | 145 | 170.5 | 158 | 170 | 159 |
| | Lowest value | 142 | 136 | 165 | 155 | 165 | 155 |

As we can see from the box and whisker graph that accompanies this work that the data is visually pleasing. We can draw several conclusions from the box and whisker diagrams when they are laid out side by side to compare. The most simple observation we can make is that males are on average taller than females in the years we have surveyed. We can see that on average boys experience height changes in puberty earlier and have finished growing on average by year 11. The male means of year 11 and 12 are almost identical whereas the means of the year 11 and 12 girls are still increasing.  With more data across more age groups we could successfully produce a mean age of puberty. With the amount of data available to us we can say that women and females grow later in there teenage years.

**Hypothesis B**    *Hand span and Height are directly related*

To see if these two are directly related I will use *Spearman's Rank Coefficient*. I will take data from one year group and see if it related through the coefficient and whether they are ranked closely together.

| Hand Span | Height | Rank A | Rank B | A-B | A-B squared |
|-----------|--------|--------|--------|-----|-------------|
| 21 | 170 | 30 | 29 | 1 | 1 |
| 21 | 161 | 25.5 | 29 | -3.5 | 12.25 |
| 21 | 160 | 23.5 | 29 | -5.5 | 30.25 |
| 20.6 | 157 | 19 | 27 | -8 | 64 |
| 20 | 163 | 27.5 | 24 | 3.5 | 12.25 |
| 20 | 160 | 23.5 | 24 | -0.5 | 0.25 |
| 20 | 158 | 22 | 24 | -2 | 4 |
| 20 | 154 | 15.5 | 24 | -8.5 | 72.25 |
| **20** | **154** | 15.5 | 24 | -8.5 | 72.25 |
| 19 | 161 | 25.5 | 20 | 5.5 | 30.25 |
| 19 | 155 | 17 | 20 | -3 | 9 |
| 19 | 151 | 13.5 | 20 | -6.5 | 42.25 |
| 18.5 | 156 | 18 | 18 | 0 | 0 |
| 18.3 | 149 | 10.5 | 17 | -6.5 | 42.25 |
| 18 | 157 | 19 | 14.5 | 4.5 | 20.25 |
| 18 | 151 | 13.5 | 14.5 | -1 | 1 |
| 18 | 150 | 12 | 14.5 | -2.5 | 6.25 |
| 18 | 145 | 5.5 | 14.5 | -9 | 81 |
| 17.6 | 157 | 19 | 12 | 7 | 49 |
| 17 | 145 | 5.5 | 10.5 | -5 | 25 |
| 17 | 144 | 3 | 10.5 | -7.5 | 56.25 |
| 16.5 | 144 | 3 | 9 | -6 | 36 |
| 16 | 167 | 29 | 6.5 | 22.5 | 506.25 |
| 16 | 163 | 27.5 | 6.5 | 21 | 441 |
| 16 | 149 | 10.5 | 6.5 | 4 | 16 |

| | | | | | |
|---|---|---|---|---|---|
| 16 | 144 | 3 | 6.5 | -3.5 | 12.25 |
| 15.9 | 143 | 1 | 4 | -3 | 9 |
| 15 | 147 | 8 | 3 | 5 | 25 |
| 14 | 147 | 8 | 1.5 | 6.5 | 42.25 |
| 14 | 147 | 8 | 1.5 | 6.5 | 42.25 |
| | | | | Total | 1761 |

I will use the following formulae to find out if there is any correlation between the two sets of ranking.

$$P = 1 - \frac{6\sum d^2}{N(n^2-1)}$$

Where N is the number of pairs of data.

The result of the calculations is 0.61 that is an encouraging result when the population size is taken into account. This figure is well above the minimum point at which rank coefficient is likely. This is a good result to prove the theory above. The minimum figure below 1 of which rank coefficient is probable is 0.46. Therefore in this case there is a clear link between the rankings of the two sets of data. This proves the hypotheses B by proving strongly that there is a clear link between hand span and height. Therefore we can say quite confidently that the all bodies are roughly within set proportions.

**Hypothesis C**    *Heart rate and age are unrelated.*

To find out if this hypothesis is true or untrue I will work out the mean for each year. The reason I think heart rate and age are unrelated is because I believe that heart rate is directly related to the physical fitness of the subject.

I will work out the mean and standard deviation for each year group I am working with. I will then be able to appreciate whether we have a normal spread across the data.

Year 7

| X | X | X-X | (X-X)² |
|---|---|---|---|
| 71 | 67.5 | 3.85 | 14.8225 |
| 82 | | 14.85 | 220.5225 |
| 69 | | 1.85 | 3.4225 |
| 11 | | -56.15 | 3152.823 |
| 113 | | 45.85 | 2102.223 |
| 42 | | -25.15 | 632.5225 |
| 38 | | -29.15 | 849.7225 |

| | | -8.15 | 66.4225 |
|---|---|---|---|
| 59 | | -8.15 | 66.4225 |
| 63 | | -4.15 | 17.2225 |
| 51 | | -16.15 | 260.8225 |
| 45 | | -22.15 | 490.6225 |
| 88 | | 20.85 | 434.7225 |
| 81 | | 13.85 | 191.8225 |
| 57 | | -10.15 | 103.0225 |
| 56 | | -11.15 | 124.3225 |
| 114 | | 46.85 | 2194.923 |
| 66 | | -1.15 | 1.3225 |
| 63 | | -4.15 | 17.2225 |
| 60 | | -7.15 | 51.1225 |
| 114 | | 46.85 | 2194.923 |
| | | **Total** | **13124.55** |

The total is 143.43. I need to divide this by the number of sets of data, which is twenty.

Then I need to square root the answer, 13124.55.

This gives us the standard deviation figure of 25.5.

  Standard Deviation  25.5
            Mean  67.5

Year 11

At this stage I also realised I could save myself a lot of time by creating a spreadsheet in Microsoft Excel that would provide me with an answer for Standard Deviation by just entering the original data.

Launch Excel

| | X | X-xbar | X-xbar squared |
|---|---|---|---|
| | 89 | 19.45 | 378.3025 |
| | 59 | -10.55 | 111.3025 |
| | 75 | 5.45 | 29.7025 |
| | 81 | 11.45 | 131.1025 |
| | 67 | -2.55 | 6.5025 |

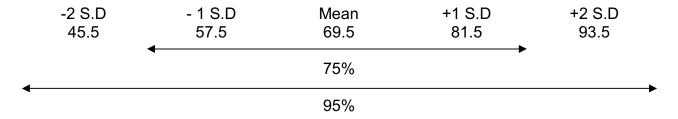|  | 53 | -16.55 | 273.9025 |
|---|---|---|---|
|  | 50 | -19.55 | 382.2025 |
|  | 60 | -9.55 | 91.2025 |
|  | 80 | 10.45 | 109.2025 |
|  | 79 | 9.45 | 89.3025 |
|  | 60 | -9.55 | 91.2025 |
|  | 66 | -3.55 | 12.6025 |
|  | 67 | -2.55 | 6.5025 |
|  | 63 | -6.55 | 42.9025 |
|  | 62 | -7.55 | 57.0025 |
|  | 96 | 26.45 | 699.6025 |
|  | 70 | 0.45 | 0.2025 |
|  | 84 | 14.45 | 208.8025 |
|  | 60 | -9.55 | 91.2025 |
|  | 70 | 0.45 | 0.2025 |
| total | 1391 |  | 2812.95 |
| **Mean** | **69.55** |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  | **Standard Deviation** | **11.85949** |  |

## Year 12

|  | **X** | **X-xbar** | **X-xbar squared** |
|---|---|---|---|
|  | 84 | 3.35 | 11.2225 |
|  | 72 | -8.65 | 74.8225 |
|  | 90 | 9.35 | 87.4225 |
|  | 96 | 15.35 | 235.6225 |
|  | 84 | 3.35 | 11.2225 |
|  | 93 | 12.35 | 152.5225 |
|  | 60 | -20.65 | 426.4225 |
|  | 72 | -8.65 | 74.8225 |
|  | 74 | -6.65 | 44.2225 |
|  | 88 | 7.35 | 54.0225 |
|  | 80 | -0.65 | 0.4225 |
|  | 96 | 15.35 | 235.6225 |
|  | 84 | 3.35 | 11.2225 |
|  | 73 | -7.65 | 58.5225 |
|  | 81 | 0.35 | 0.1225 |
|  | 73 | -7.65 | 58.5225 |
|  | 114 | 33.35 | 1112.223 |
|  | 65 | -15.65 | 244.9225 |

| | | | |
|---|---|---|---|
| | 72 | -8.65 | 74.8225 |
| | 62 | -18.65 | 347.8225 |
| total | 1613 | | 3316.55 |
| **Mean** | **80.65** | | |
| | | | |
| | | | |
| | **Standard Deviation** | | **12.8774** |

Do my results fit into a standard distribution?
We know that their a certain facts surrounding a normal distribution and standard deviation. This diagram represents a standard spread. I have decided to use my year 11 results as my example to fit into a normal distribution and to compare it. Below is how my data fits into a standard distribution.

| -2 S.D | - 1 S.D | Mean | +1 S.D | +2 S.D |
|---|---|---|---|---|
| 45.5 | 57.5 | 69.5 | 81.5 | 93.5 |

75%

95%

My data fits well into a standard distribution, there is only one result out of the twenty, 5%, that falls outside plus or minus 2 points of Standard Deviation. 75%, a little over the average 68% fall inside plus or minus one point of standard deviation which is also pleasing.

Drawing Conclusion
The high rate of standard Deviation for the year sevens may indicate either that there ability to carry out a seemingly simple task is hampered by there level of maturity or that there is a wide range in the answers and the class included both athletes and the less athletic couch potato character. The closeness of the two standard deviation results for years eleven and twelve and is interesting because of the distance the two respective means are apart. This mean would suggest that over the course of a year the health of the year 11 to 12 deteriorates a quite a rate. This may suggest that the stress of so many exams is taking its toll on the student population. If this is not the case then it highlights a problem with my statistical methods. For example; the selection/population is not big enough for the results I am trying to correlate.

In response to the to my original hypotheses the age and pulses are related the results would suggest so. But I think we need to take into account the spread of the ages we are looking at. Although between all three there is an increase of sorts the difference between the first two age groups is large the difference in the results is relatively minor. The age gap between the year 11 and 12 data is just a single year and we see a vast

increase. When we see this on a graph with the means plotted the absurdity of the results is all to apparent. Therefore in future if repeating this I would have to use a bigger population as to find better results. As the population size increases it will better represent the full population. Therefore I would safely dismiss any relationship between age and pulse rate over the age we have covered.