

## Statistics Coursework- Plan

I am going to be basing my course work on swimming time, I am using swimming times because all of the data is continuous so that it is easy to work with and allows simple and easy collection of primary data.

I am going to find out the following things:

- Are men faster at swimming than women?
- Does the amount of people from a certain club in an event affect the speed of the people in that club?
- How drastically does the age of a swimmer effect the time of the swimmer?
- What is the relationship between the entry time and finishing time of an event?

For the question, are men faster at swimming than women? I will need information and data on the following. The times of both men and women in the same event. For example, a list of 30 times taken at random from the men's 100m backstroke and a list of 30 times taken at random from the women's 100m backstroke.

I will use line graphs to illustrate the data that I have collected because they will be easy to compare with each other. I intend to use average and standard deviation as calculations to represent my data.

The hypothesis for the question, are men faster at swimming than women? Will be, "men are faster than women at 100m freestyle than women.". I am proposing the hypothesis in this way simply because men have a larger physical strength than women.

The question, does the amount of people from a certain club in an event effect the speed of the people in that club? , will require data from two different clubs, one of the clubs with less swimmers in it than the other, the question is asking of the popularity of a club effect how good the swimmers in it are, so I will need to collect the same amount of data from two clubs as said before and process the data to answer the question.

I will need to use line graphs and possibly scatter graphs to illustrate my data. For the calculations I will used the average.

The hypothesis for this question is going to be: "The more popular a club is the faster the times of its swimmers are." I am proposing the hypothesis in this way because for a club to become popular it needs to produce good swimmers.

For the question, how drastically does the age of a swimmer effect the time of the swimmer, I will use the following to illustrate my data; scatter graphs and line graphs. Calculations necessary to use to process my data are average/mean and spearman's coefficient.

The Hypothesis for this question will be: "the age of a swimmer drastically effects the time of the swimmer in the way of them being faster the older they get." For this question I will need to get data from two of the same events but use two different age categories.

For the question 'What is the relationship between the entry time and finishing time of an event?' I will use scatter graphs and the average to illustrate the data that I collect and I will also use standard deviation to analyse my results. To answer the question I will need the entry time and the finishing time of one event and make comparisons between them. The hypothesis will be: "the relationship between the entry and finishing times of swimmers is that the finishing time is always faster".

### **Hypothesis 1**

For this hypothesis I am going to use data I have acquired from the Ponds Forge North East Swimming Gala. I am using the results sheets from the events 110m Women's Freestyle and 100m Men's freestyle. With this data I will collect a random sample of 30 people and use that data to put into my calculations and into my graphs. Etc. I think that a sample size of 30 is large enough to represent my population without affecting the results of my calculations.

I will use random sampling to collect my sample by generating a random number on my calculator and using that number to collect every multiple of that numbers item of data. Eg. If the number generated is 0.5621 then I will use every 5<sup>th</sup> item of data until I have 30 items.

The number my calculator generated was 0.512 so I will be using every fifth item of data. The data that I have collected is continuous data.

### **Hypothesis 2**

For this hypothesis I am going to use data from two different clubs and compare the times of people in the same event in each club. For example, I could use the event men 100m backstroke and choose one club with lots of people in the event and then use one club with not many people in an event. I would then collect 30

items of data from each club recording their club name and their time it took them to complete the event. I will then use the data and input it into the calculations and graphs that I have decided to use earlier and compare the results I get and then compare the results to find out which team is the best team and see if it does or doesn't prove my hypothesis.

To collect my data I will use a random sampling method. I will, again use my calculator to generate a random number and then use every item of data along corresponding to the number as before. But before I do this I will only use the data from 2 individual clubs.

### **Hypothesis 3**

For this hypothesis I am going to collect data from two different events of two different sexes and use two of the same age groups to collect the data from both events. For example men's 50m backstroke and women's 50m backstroke and then collect data from two different age groups e.g. 14-16 and 17-19. By doing this I get a wider range of data to compare, I will have men's and women's time and the times from two different age groups. Once I have collected the data I will process it into the necessary calculations and graphs.

I will collect my data by random sampling as before and use the random sampling method by generating a random number on my calculator etc. as before. Once I have collected the data I will put it into appropriate tables so that it is easier to process.