Introduction

Statistics is a useful tool in understanding natural phenomena, both from nature and from human activities.

Many data sources are available and reported in the news. Statistical thinking will help us to understand these reports, what makes sense and what does or might not make sense. Many data types have a natural variability

- life times for people, electronic equipment

- strengths of manufactured goods (steel, concrete used in bridges, buildings, cars, airplanes

- responses to disease, virus or bacteria

- medical response to treatment

- growth of money or economy

Under identical conditions different trials of the same *game* or *experiment* can be different. The outcome is not exactly predicable, but still one can quantify the relative likelihood of outcomes, predicting ranges or intervals for outcomes. For example two medical treatments A and B for the same disease may have 70% and 40% probability of success respectively. In a statistical sense treatment A is superior than B, but using treatment A does not guarantee a better result than B for an individual. Choosing treatment A still increases the odds of a successful treatment.

Polio and Food

Polio is a crippling disease, maiming many thousands of children each year in Canada and hundreds of thousands per year in the US until it was brought under control. Prior to the polio vaccinations success there were many public policy attempts to try to prevent the spread of this disease.

Polio was particularly prevalent in the hot weather months. Was this caused by diet or food or some other mechanism?

When ice cream consumption increased the polio disease increased. The Public Health Office in North Carolina (and other states) issued warnings against eating ice cream.

Was this justified?

The polio virus became more active during hot weather. The same hot weather also drove the consumption of ice cream, but ice cream did not cause polio nor did it help the spread of polio. The claim of causation was not justified and harmed many companies and the lives of people. The study is one of many examples of spurious correlations in observational studies.

Controlled experiments, the scientific method and randomization is used when possible to overcome this in the study of various phenomena.

In 2008 Canada had an outbreak of Listeriosis which killed some 20 people before the cause was discovered. Public Health Officials are responsible for public health warnings and it took some time to understand the cause in this simpler (than polio) serious health problem.

Some more details on the polio and ice cream problem can be found at

http://www.notadoc.org/Articles/NATURALLY

Does ice cream consumption cause murders?

These and the other cartoons here are obtained from http://www.causeweb.org/ and are greatfully acknowledged.
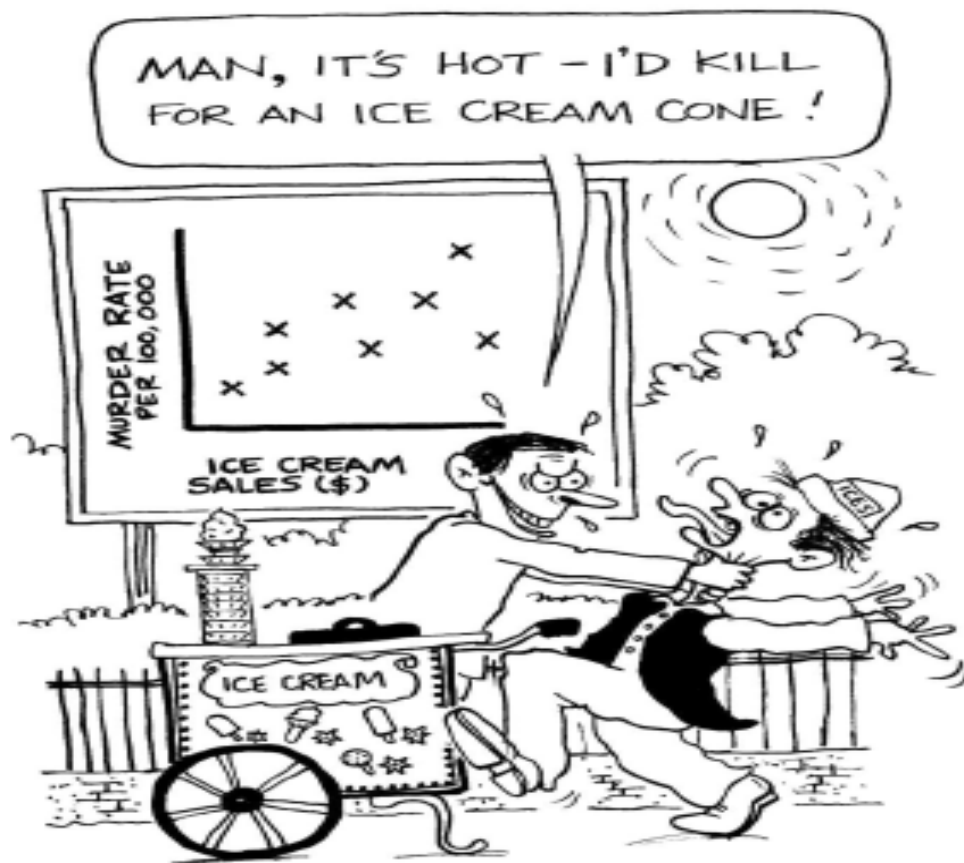
Figure 1: Plot of ice cream sales and murder rates in USA

Population growth and density is an underlying system that drives both the murder rate and ice cream sales.

Fitting a curve to *model* the above murder rate versus ice cream is an example of *regression*. It is easy to misuse and misinterpret such curves. Sometimes it is easy to spot this and sometimes not so easy.

Data may be categorical, integer, or real numbers.

1. **Categorical** : which month are you born in? rate a product as excellent, good, average, poor; is there rain today

2. **Integer** : flip a coin $n = 10$ times and count up the number of heads; observe bird nests and count the number of eggs or eggs that hatch; what is your age in years

3. **Real numbers** : what is the temperature a 12 Noon; what is the weight of ball bearings; what is the rate of water flow in a river; what is the rate of water flow in a city water system; what is your weight in kg (or in pounds)

Figure 2: Do your homework and reading between classes

Between classes read the appropriate material and work on some of the suggested problems.
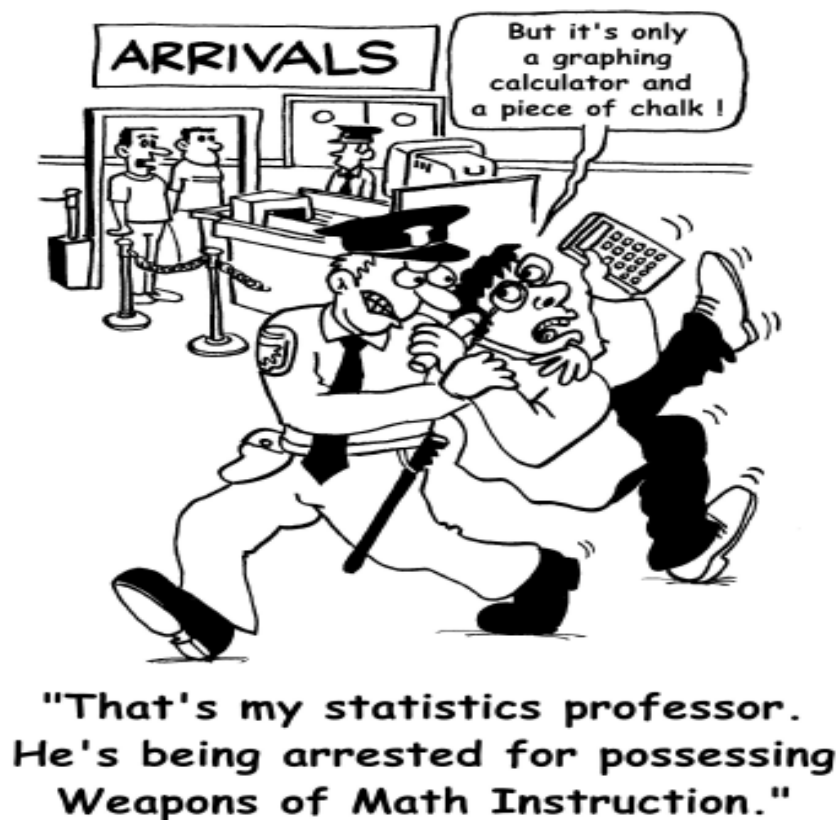
Figure 3: Math Instruction

In this class and for tests you will use a non-programmable calculator. In actual statistical practice one uses other tools, computers and graphics, and mathematics to develop appropriate models.

In this evening section we will have approximately 1 hour of lecture, a 5 minute break, another hour of lecture time, a 5 minute break, and then open questions for the remaining time.