To Do List

**Read Chapter 1
Do Problems 1-20**

**Assignment 1 is due Friday
  September 23
Tutorial Test 1 is on Wednesday
  September 28**

Laptop, iPads, etc.

Latest research shows your non-class laptop use during lecture hinders the learning of your nearby peers.

Laptop multitasking hinders classroom learning for both users and nearby peers

F. Sana, T. Weston, N.J. Cepeda, *Computers and Education*, 62 (2013), 24-31

Please don't use your laptop for Facebook, YouTube, etc. during class time.

Today's Class

Learn some statistical jargon:

Empirical study

Population, process

Variates and types of variates

Attributes

Types of empirical studies

Empirical Studies

An *empirical study* is one in which we learn by observation or experimentation.

We collect data to increase our knowledge about the world around us and/or to make decisions.

Empirical Studies

A key feature of an empirical study is that it involves *uncertainty*.

If we run the experiment more than once we don't get identical results each time.

We will use the probability models you learned about in STAT 230 to try and model this uncertainty.

To talk about empirical studies we need to define some terms.

Statistical Jargon: Population

A *population* is a collection of units.

Example 1:

Example 2:

Statistical Jargon: Population

A *population* is a collection of units.

Example 1: all students taking STAT 231 this term.

Example 2: all persons aged 90+ living in Ontario.

Statistical Jargon: Process

A *process* is a system by which units are produced.

Example 1: The hits on a particular website can be considered as units in a system or process. Of course this system or process would be quite complex and difficult to describe.

Example 2: The claims made by car insurance policy holders could be considered as units in a process.

Both populations and processes are collections of units. A key feature of processes is that they usually occur over time whereas populations are static (defined at one moment in time.)

Example: KW Humane Society

The KW Humane Society collects data about the dogs in their care.

Data collected include:

Breed of dog

Weight of dog (kg)

General health of dog
   (good, satisfactory, poor)

Sex of dog

Number of people who have owned the dog

KW Humane Society Example

What are the units?

What is a suitable population or process for this example?

KW Humane Society Example

The units are the dogs.

Since data are being collected on the dogs over time it is reasonable to think of this example as a process.

The process consists of all dogs in the care of the KW Humane Society now and into the future.

Variates

Variates are characteristics of the units which are usually represented by letters such as *x, y, z.*

Types of variates:
- continuous
- discrete
- categorical
- ordinal
- complex

Note: Determining the type of variate will be useful in identifying a suitable probability model for the data.
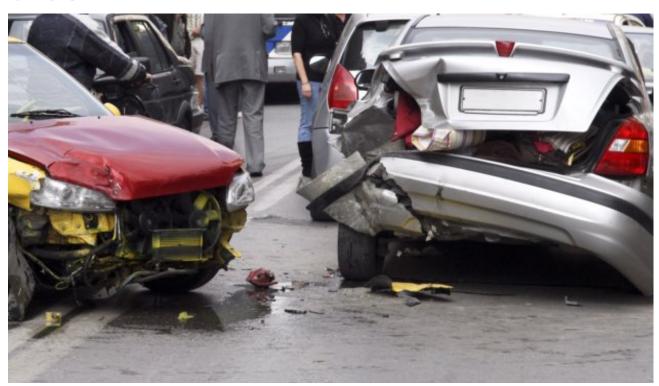
Continuous Variates

Variates such as the height and weight of a person, the lifetime of an electrical component, and the time to reoccurrence of a disease after medical treatment are all examples of continuous variates.

Discrete Variates

Variates such as the number of defective smartphones sold by a particular company or the number of deaths in a year on a certain stretch highway are examples of discrete variates.

Categorical Variates

Variates such as hair colour, university program, or marital status are examples of **categorical** variates.
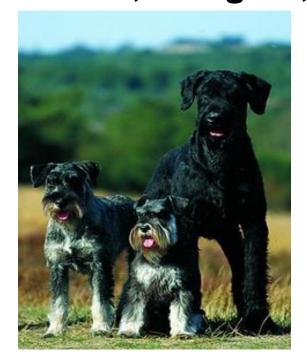
Ordinal Variates

Categorical variates for which an ordering is implied are also called ordinal variates.

Examples:

1) small, medium, large

2) strongly agree, agree, neutral, disagree, strongly disagree

Complex Variates

An image or an open-ended response to a survey question are examples of **complex** variates.

KW Humane Society Example

The KW Humane Society collects data about the dogs in their care.

Data collected include:

Breed of dog

Weight of dog (kg)

General health of dog
   (good, satisfactory, poor)

Sex of dog

Number of people who have owned the dog

Attributes

An *attribute* of a population or process is a function of a variate which is defined for all units in the population or process.

Example 1:

If the population of interest were all persons aged 18-25 living in Ontario an attribute of interest might be the proportion of the population who own a smartphone. Another attribute of interest might be the mean annual income for this population.

Attributes

Example 2:

If the process of interest is the process by which claims are made by car insurance policy holders, the proportion of claims exceeding on million dollars, and the mean claim value might be attributes of interest.

KW Humane Society Example

The KW Humane Society collects data about the dogs in their care.

Data collected include:

Breed of dog

Weight of dog (kg)

General health of dog
   (good, satisfactory, poor)

Sex of dog

Number of people who have owned the dog

Types of Empirical Studies

Sample Survey:

Information about a finite population is obtained by selecting a "representative" sample of units from the population and determining the variates of interest for each unit in the sample.

Types of Empirical Studies

Observational Study:

Information about a population or process is collected without any attempt to change one or more variates for the sampled units.

Example of an Observational Study

In a study on the drinking habits of university students, blood alcohol levels are measured for students at the beginning of an 8:30 Mond lecture using a breathalyzer.

Types of Empirical Studies

Experimental Study:

An experimental study is one in which the experimenter intervenes and changes or sets the values of one or more variates for the units in the study.

Example of an Experimental Study

In a study on the drinking habits of students, blood alcohol levels are measured for students in two different second year classes at the beginning of an 8:30 Monday lecture using a breathalyzer. One class is warned in advance that blood alcohol levels will be measured to test for alcohol while the other class is not warned.

Types of Empirical Studies

These three types of empirical studies (sample surveys, observational studies, experimental studies) are not mutually exclusive.

Opinion polls in which a random sample of people are selected and asked their opinion on a given topic are a good example of a sample survey.

Types of Empirical Studies

The Labour Force Survey (LFS) conducted by Statistics Canada is an important example of a (complicated) survey sample.

(http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3701)

Data collected in the LFS are used to produce Canadian unemployment rates.

Example

Suppose participants in an empirical study on health and diet are asked to complete a questionnaire on their eating habits.

Suppose also that a blood sample is drawn to determine the participant's cholesterol and glucose levels.

What type of study is this?

Example

Suppose participants in an empirical study on health and diet are asked to complete a questionnaire on their eating habits.

Suppose also that a blood sample is drawn to determine the participant's cholesterol and glucose levels.

We would call this an observational study.

Example Continued

Suppose the participants are then randomly assigned to either an increased fruit and vegetable diet or a vegetarian-only diet, and their cholesterol and glucose levels are re-measured after 6 months.

What type of study is this?

Example Continued

Suppose the participants are then randomly assigned to either an increased fruit and vegetable diet or a vegetarian-only diet, and their cholesterol and glucose levels are re-measured after 6 months.

We would now call this an experimental study.