Chapter 2 - Lessons

Data Collection

MDM4U

Unit Outline

Overall Unit Goal: Be able to describe the characteristics of a good sample, some sampling techniques, and principles of primary data collection, and collect and organize data to solve a problem.

Section	Subject	Learning Goals	Curriculum Expectations
2.1	Thesis Development	- be able to create a mind-map of interests - be able to develop a thesis for your culminating projects	C1.1, C1.2, C1.3, E1.1
2.2	Characteristics of Data	- understand differences between quantitative and qualitative, cross-sectional and longitudinal, population and sample	C1.3, C2.2
2.3	Random Sampling	- understand how to use different random sampling techniques	C2.1, C2.3
2.4	Survey Design and Types of Bias	- know how to create a survey using different question types - know the different types of bias present in survey designs	C2.3, C2.4
2.5	Experiment Design	- understand the difference between observational studies and experiments - understand the four principles of experimental design - know how placebos and blocking can be used within an experiment	C2.1

By the end of the unit, you will be able to:

- Describe principles of primary data collection including various sampling techniques.
- Be able to describe and create an effective survey for an observational study.
- Collect primary data and then organize and analyze it.

Assessments	F/A/O	Ministry Code	P/O/C
Note Completion	A		P
Practice Worksheet Completion	F/A		P
Assignment – School Survey	O	C1.1, C1.2, C1.3, E1.1, **C2.5**	P
PreTest Review	F/A		P
Test – Data Collection	O	C1.1, C1.2, C1.3, E1.1, C1.3, C2.1, C2.2, C2.3, C2.4	P

2.1 - Developing a Thesis

MDM4U

Part 1: ISU Intro

This chapter will prepare you to begin your ISU that is worth 10% of your final grade. For the ISU you will be required to choose a topic that interests you and conduct a study that analyses large amounts of data using:

- one-variable statistics tools (chapter 3)
- two variable statistics tools (chapter 1)
- probability (chapter 4/5)

Part 2: Mind-Map

Before you can begin your project, you must create a thesis:

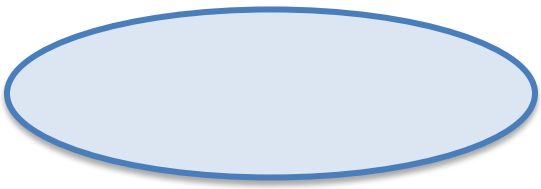
thesis:

To begin creating a thesis, you must first determine what topics interest you and then determine what concepts related to that topic you want to study. A useful brainstorming tool that can illustrate how a topic relates to other concepts is a _____.

mind map:

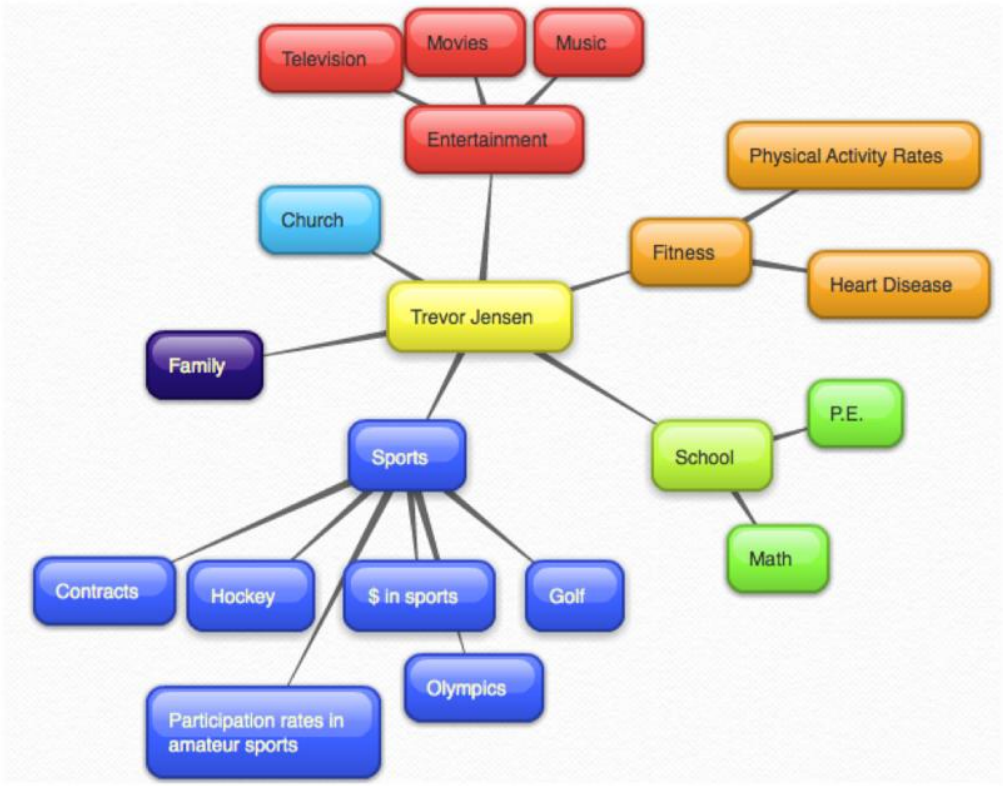
Constructing a Mind Map

1. Start by making a mind map of your interests with you at the centre. Start off as simple as possible and draw arrows to show how topics are connected. Work from the inside out.



Extended Mind Map

2. Pick one of the topics from your mind map and extend it with sub-topics.



Part 3: Thesis Question Development

Once you have narrowed down your topic, you will need to pose a problem that you plan to investigate.

3. Brainstorm and create number of questions that can be explored with the use of statistical information

Thesis Question Analysis

Questions to ask of your Thesis:

4. Once you have chosen your thesis, analyse it using the three questions above to make sure your study will be able to provide an insightful answer.

Thesis:

Analysis:

Project tips:

One way of posing a problem is to generate questions from data. For example, once a topic has been identified, do a preliminary data search. The type and quantity of available data may indicate some possible questions. Data from print sources, the Internet, and E-Stat are some resources that may be used.

2.2 – Characteristics of Data

MDM4U

Part 1: Population vs. Sample

_____ are any collection of numbers, characters, images, or other items that provide information about something.

The entire group of individuals that we want information about is called the _____.

A _____ is an attempt to gather information about every individual member of the population. Problems with census: _____; _____ needed to complete; sometimes testing can _____ items.

A _____ is a part of the population that we actually examine in order to gather information.

Note: It usually isn't practical to collect data from the entire population; instead you should take a representative sample and study it.

Example 1: Determine the population of each of the following questions

a) Whom will you plan to vote for in the next Ontario election

b) What is your favourite brand of hockey stick?

c) Do women prefer to wear ordinary glasses or contact lenses?

Once you have identified the population, you need to decide how you will obtain your data. If the population is _____, it may be possible to survey the entire group (census). For _____ populations, you need to use appropriate sampling technique.

We will discuss different sampling techniques next lesson.

Part 2: Types of Studies

Cross Sectional:

Longitudinal:

Example 2:

For the thesis question:

How do the opinions about the cafeteria change among students from Grade 9 to Grade 12?

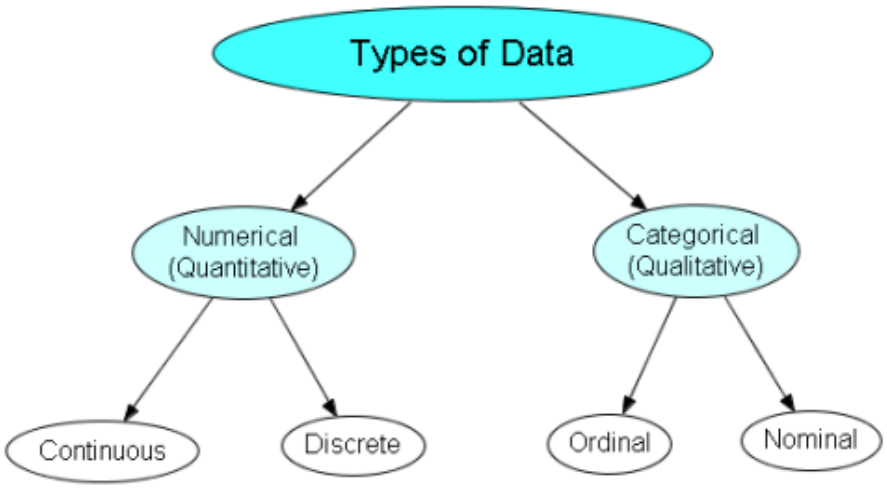
a) How could you conduct a cross-sectional study?

b) How could you conduct a longitudinal study?

c) Which study would be more time efficient?

d) Re-write the thesis question to reflect a cross-sectional study

Part 3: Types of Variables



Quantitative/Numeric Variable: A quantitative variable that takes _____ values for which it makes sense to find an _____. These variables can be either _____ or _____.

Qualitative/Categorical Variable: A variable that places an individual into one of several _____ or _____ (also known as qualitative variables). Categorical variables may have categories that are naturally ordered (_____ variables) or have no natural order (_____ variables).

Example 3: Identify whether each of the following questions measures a qualitative or quantitative variable.

a) How tall are you?

b) What conference are the Leafs in?

c) What colour is your hair?

d) How many students are in this class?

e) What is your favourite school subject?

Part 4: Types of Quantitative Variables

Continuous Variable: A numeric variable that can have an _____ number of values in a given interval. Measurable with all real numbers.

Examples:

Discrete Variable: A numeric variable that can take on only a_____ number of values within a given range. (usually measured with _____ values only)

Examples:

Example 4: Classify each quantitative variable as either continuous or discrete

a) Temperature outside

b) Number of goals scored by Crosby

c) Number of songs on your IPod

d) Speed of Zdeno Chara's slapshot (108.8 mph) https://www.youtube.com/watch?v=vZssDq7lJus

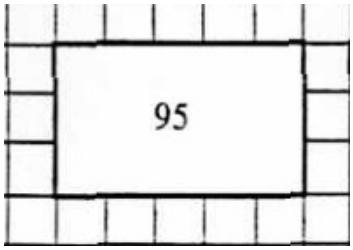
2.3 – Sampling Principles

MDM4U

Part 1: Random Rectangles Activity

1. a. Guess the average area of all rectangles on the page: **(guess)** _____

   b. Choose six rectangles (before you calculate any areas) that you think represent the entire population of rectangles well.



6 rectangles – subjective – "rectangle **expert**":

rectangle number	area
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
average:	_____

2. a. After setting a new seed value on your calculator, use the randint function to choose six random rectangles for you.

6 rectangles – **random**:

rectangle number	b. area
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
average:	_____

3. a. mean of sample averages:

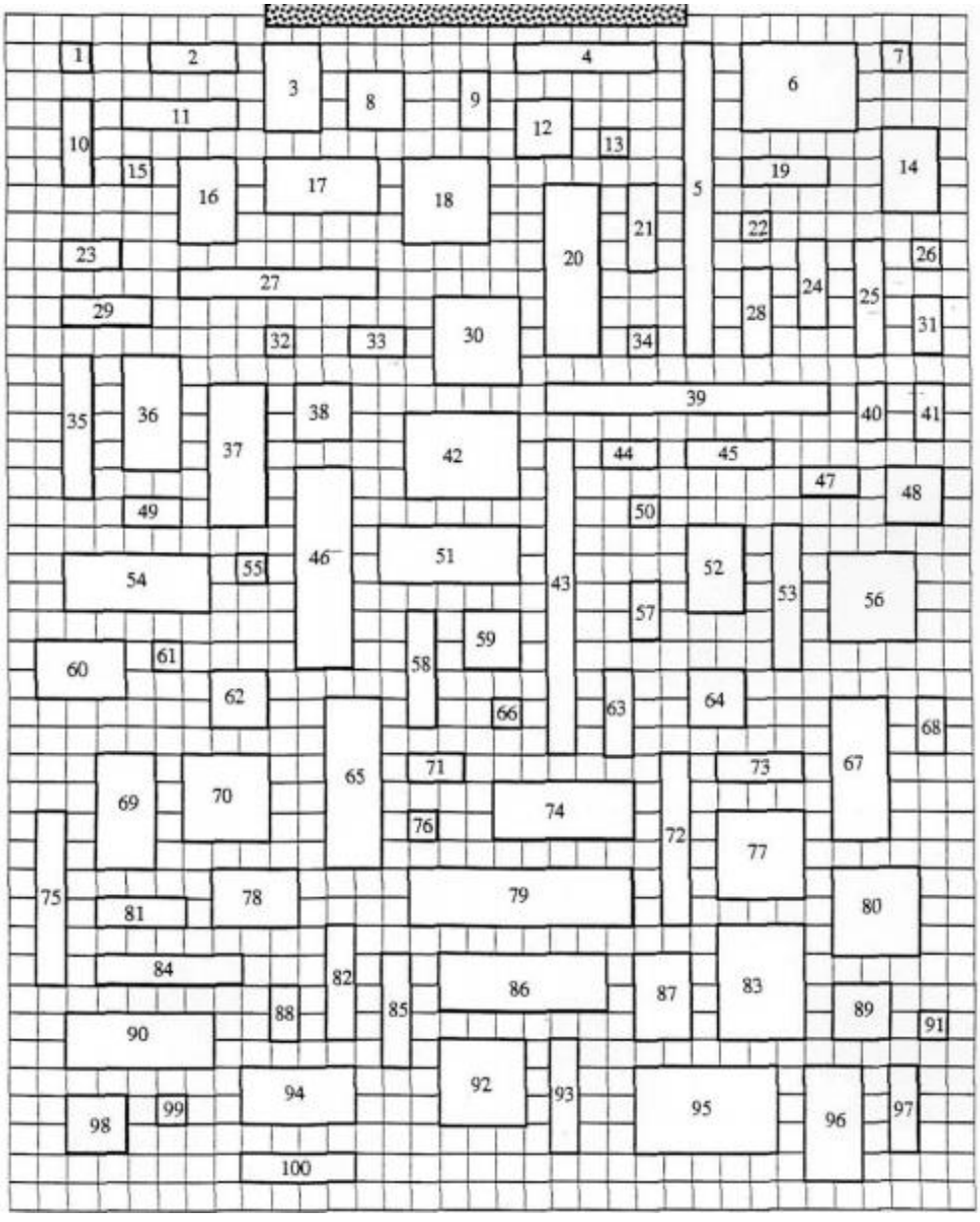
guesses _____

subjective (expert) _____

random _____

c. actual area of 100 rectangles (population): _____

Wrap-up (what have you learned?):



Part 2: Random Sampling Methods

1. Simple Random Sampling

A sample is a _____ if it is selected so that:

- each member of the population is _____ likely to be chosen and the members of the sample are chosen independently of one other;

OR

- every set of n units has an _____ chance to be the sample actually selected.

Example: Put names in hat and draw until have desired sample size; more commonly, number names and use random number generator or other source of random numbers to select sample. Notice that some type of unbiased method must be used; haphazard ≠ random.

2. Systematic Random Sampling

A sample is a systematic random sample if you randomly choose some _____; then select every _____ element in the population, where n is the sampling interval. This guarantees that the sample is taken from throughout the _____ but it requires an ordered list of everyone in the population.

Example: If we wanted to get a systematic random sample of 10% of the students from King's which has approximately 600 students…

- Calculate number of students required for sample: $600 \times 0.10 = 60$
- Calculate the sampling interval: $sampling\ interval = \frac{population\ size}{sample\ size} = \frac{600}{60} = 10$
- Choose a random starting point using a random number generator
- Include every 10^{th} student from the randomly chosen starting point in your sample

3. Stratified Random Sampling

When using a stratified random sample, the population is divided into _____called _____ (e.g. age, geographical areas, grade, etc.)

A _____ of the members of _____ stratum is then taken. The size of the sample for each stratum is _____ to the stratum's size (you must survey the same _____ of people from each stratum).

Example: If we want a stratified random sample of 10% of the 600 King's students, we can divide the population in to four groups based on grade (9, 10, 11, 12) and then take a simple random sample of 10% of the students in each grade.

4. Cluster Random Sampling

When using a cluster random sampling method, divide the population into _____ or _____; randomly select a few of those groups and then sample _____members from the selected groups.

Example: _____ select 5 block C classes—survey _____ students in each class selected.

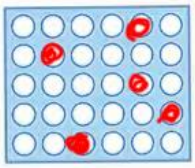
5. Multi-Stage Random Sampling

When using multi-stage random sampling, the population is organized in to groups, a simple random sample of groups is chosen, and then a simple random sample of people within the chosen groups is taken.

Example: _____ select 5 block C classes—survey _____ of the students in each class selected.

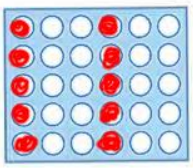
Review of Different Random Sampling Techniques:

Simple Randon



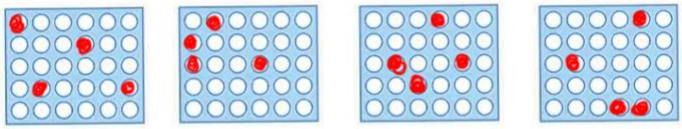
- all selections ar
equally likely

Systematic Random



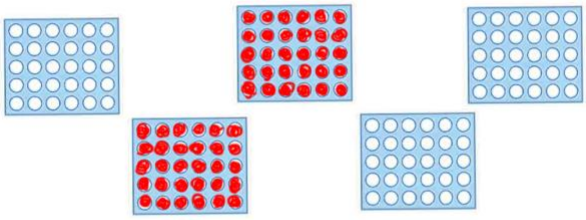
- random starting poi
choose individuals at
interval (every *n*th pe

Stratified Random



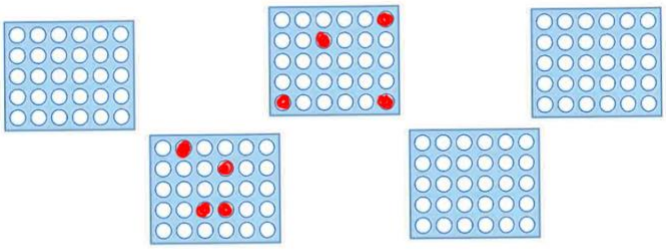
- Divide population into groups then survey an equal percentage of each group.

Cluster Random



- Divide the population into groups. Choose a random sample of groups and then survey **every member** of the groups chosen.

Multi-Stage Random



- Divide the population into groups. Choose a random sample of groups and then choose a random sample of members of the chosen groups.

Part 3: Types of Non-Random Samples

1. Convenience sampling

The easiest way to obtain a sample is to choose it without any random mechanism (also called haphazard sampling). Choosing individuals from the population who are _____ results in a convenience sample. Convenience sampling often produces _____ data.

Example: Suppose we want to know how long students at a large high school spent doing homework last week. We might go to the school library and ask the first 30 students we see about their homework time.

2. Voluntary Response Sampling

A voluntary response sample consists of people who _____by responding to a general _____. Voluntary response samples attract people who feel strongly about an issue, and who often share the same opinion. This leads to _____.

Example: A radio host invites listeners to call in to give opinions on a new band.

2.4 – Bias and Survey Design

MDM4U

If you conduct a survey and collect information firsthand, this is called _____ data. This type of data is easy to work with because you control how it is collected.

Information obtained from similar studies conducted by OTHER researchers is called _____ data.

Part 1: Principles of Survey Design

Basic Principle #1:
A survey is not merely a collection of questions, thrown together without purpose—surveys should be designed around specific needs for information about a _____ topic.

Basic Principle #2:
Both parties to the survey have responsibilities:

- The interviewer's work must be mostly done in advance; identify relevant variables, craft questions, design the flow of the survey.
- The interviewee's task is to—having agreed to answer questions—be _____.

Basic Principle #3:
A prime task of the interviewer at the question design stage is to help the interviewee be honest.

Part 2: Open vs. Closed Questions

1. Open Questions

Examples:

How do you feel about the salaries paid to professional athletes?

What is the most important issue for King's students?

2. Closed Questions

Part 3: Types of Closed Questions

i) _____

Circle the appropriate response:

a) Gender: *M* *F*

b) Age: *under 14* *15 or 16*
17 or 18 *19 and over*

ii) _____

Which of the following sports do you enjoy watching? (check all that apply)

☐ *Basketball* ☐ *UFC*
☐ *Baseball* ☐ *Lacrosse*
☐ *Hockey* ☐ *Soccer*

iii) _____ – asks survey respondents to compare different items using a common scale. It can also be used just to rate one item using a scale.

How satisfies were you with your grade from the first unit test? (check the one that applies)

___ *Very dissatisfied*

___ *Dissatisfied*

___ *Satisfied*

___ *Very Satisfied*

Using a scale of 0 = not at all to 4 = very important, please rate the importance of each of the following aspects of service in a restuarant

	0	*1*	*2*	*3*	*4*
Speed of service	☐	☐	☐	☐	☐
Friendliness of staff	☐	☐	☐	☐	☐
Helpfulness of staff	☐	☐	☐	☐	☐
Value for money	☐	☐	☐	☐	☐
Taste of food	☐	☐	☐	☐	☐

iv) _____ – asks survey respondents to compare a list of objects to one another by ORDERING them

When choosing a restaurant to eat at, please rank the following in order of importance from 1 to 4 where 1 is the most important to you and 4 is the least important to you

___ *Speed of Service* ___ *Ease of parking* ___ *Cleanliness* ___ *Friendliness of staff*

Part 4: Good vs. Bad Questions

Good Questions are:

Good Questions avoid:

Example 1: What's wrong with each of the following questions?

1. Given the increasing problem of obesity amongst teenagers in North America, do you agree that King's should make physical education a mandatory class for every grade?

2. Do you think the NHLPA should have agreed to the last CBA?

3. Which player would you not select first in a fantasy hockey draft?

☐ Ovechkin
☐ Crosby
☐ Malkin
☐ Stamkos

Part 3: Types of Bias

The results of a survey can be accurate only if the sample is _____ of the population and the measurements are objective. The methods used for choosing the sample and collecting the data must be free from _____. Statistical bias is any factor that favours certain outcomes or responses and hence systematically _____ the survey results.

Sampling Bias:

Household Bias:

Non-response Bias:

Response/Measurement Bias:

Example 2: Identifying Bias

You are the campaign manager for your best friend, Rebecca, who is running for student council Prime Minister. You have been asked to determine the overall level of support for Rebecca among the 1500 students at your school. Design a sampling method that will provide the least **sampling bias.**

Potential Solution - Plan A

To save time, you have decided that a sample of about 50 students will provide a good picture of the school's political landscape. Students have lunch periods 2, 3, or 4. By random draw from a hat, you have decided to conduct the survey in the cafeteria during period 4. The first 50 students who enter the cafeteria are given the questionnaire, and you instruct them to fill it out and return it to you before the end of lunch.

What is wrong with this scenario?

Plan B

To fix the problems with Plan A, you have decided to provide a questionnaire to one person from each homeroom (your sample size is now 73). You can wait until the respondent finishes with the questionnaire to collect it. This will eliminate the non-response bias.

What is wrong with this scenario?

Create a Plan C that is free from as much bias as possible:

Example 3: Identifying Sources of Response Bias

Consider the questionnaire below developed by Rebecca's friends. Identify examples of response bias.

Election Survey
(brought to you by the friends of Rebecca committee)

Circle the appropriate response

Gender: Male Female

Grade: 9 10 11 12

On Election Day, I intend to vote for:
Rebecca Mable Jacob

Circle what you would like:

more dances
more theme-dress days
more holidays
more fun

2.5 – Experiment Design

MDM4U

https://www.youtube.com/watch?v=c6FS3D4a_kA

Part 1: Experiment Design Video

http://www.learner.org/courses/againstallodds/unitpages/unit15.html

While watching the video, answer the following questions

1. Why is the study of the effect of humans on the coral reefs not an experiment?

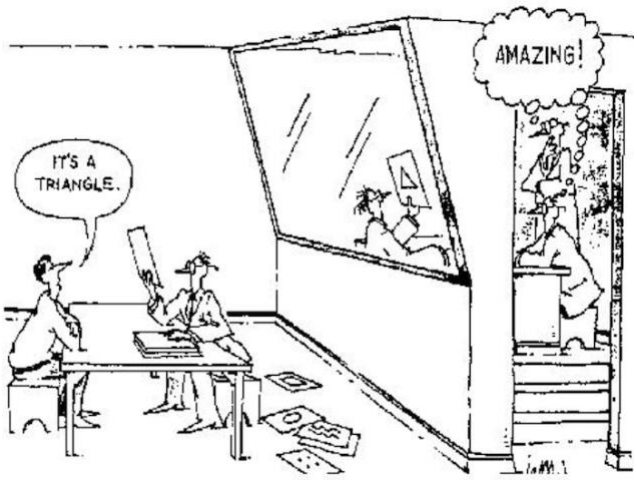
2. Who were the subjects in the Glucosamine/Chondroitin study? What did researchers want to find out?

3. Why were subjects randomly assigned to the treatments?

4. Dr. Confound conducted a very badly designed experiment on mood-altering medication. List some of the problems with his experiment.

Part 2: Observational Studies vs. Experiments

A _____ aims to gather information about a population without disturbing the population in the process. Sample surveys are one kind of _____study. Other observational studies watch the behavior of animals in the wild or the interactions between teacher and students in the classroom. This section is about statistical designs for _____, a very different way to produce data.



In contrast to observational studies, experiments don't just observe individuals or ask them questions. They actively impose some _____ to measure the response. The purpose of an experiment is to determine whether the treatment _____ a change in the response.

When our goal is to understand _____, randomized experiments are the only source of fully convincing data.

An experimenter must identify at least one _____ variable to manipulate (this is the treatment) and at least one _____ variable (response) to measure. The experimenter deliberately manipulates the treatments and must assign subjects to treatments at random.

_____(subjects) are the collection of individuals to which treatments are applied.

Example 1: Observation vs. Experiment

Should women take hormones such as estrogen after menopause, when natural production of these hormones ends? Several major medical organizations thought yes because women who took hormones seemed to reduce their risk of a heart attack 35 to 50%. The evidence in favour of hormone replacement came from a number of observational studies that compared women who were taking hormones with other who were not. But the women who chose to take hormones were richer and better educated and saw doctors more often than women who didn't take hormones. It isn't surprising that they had fewer heart attacks. In this scenario, wealth, education level, and number of doctor visits are _____ (we don't know if it was the hormone or any of these variables that caused a reduce in heart attacks)

To get convincing data on the link between hormone replacement and heart attacks, we should do an experiment. Experiments don't let women decide what to do. They assign women to either hormone replacement pills or to placebo pills that look and taste the same as hormone pills. The assignment is done by a coin toss, so that all kinds of women are equally likely to get either treatment.

By 2002, several experiments with women of different ages agreed that hormone replacement does not reduce the risk of heart attacks. In fact, some studies concluded that hormone replacement with estrogen carried increase risk of stroke.

Example 2: In 2007, deaths of a large number of pet dogs and cats were ultimately traced to contamination of some brands of pet food. The manufacturer now claims that the food is safe, but before it can be released, it must be tested. In an experiment to test whether the food is now safe for dogs to eat, what would be the treatments and what would be the response variable measured?

Part 3: Experimental Design

4 Principles of Experimental Design

1. _____ – use a design that compares two or more treatments

2. _____ – Use chance to assign experimental units to different treatments.

3. _____ – Keep other variables (besides the ones you are testing) that might affect the response of the subject the same for all groups.

4. _____ – use enough experimental units in each group so that any differences in the effects of the treatments can be distinguished from chance differences between groups

Example 3: We're planning an experiment to see if the new dog food is safe to eat. We have established that we will feed some dogs the new food and some dogs food that is known to be safe (principle of comparison). In this experiment, how could you implement the principles of control, random assignment, and replication?

Strategies to Improve Experiments

1. Use a control group – researchers vary the independent variable (treatment) for the _____ group but not for the _____ group. Any differences in the dependent variable (response) for the two groups can be attributed to the changes in the independent variable.

Example: A medial researcher wants to test a new drug believed to help smokers quit. 50 people volunteer for the study. The researcher randomly divides the smokers in to two groups. One group is given nicotine patches with the new drug, while the second group uses ordinary nicotine patches. The researcher then measures how many in each group quit smoking.

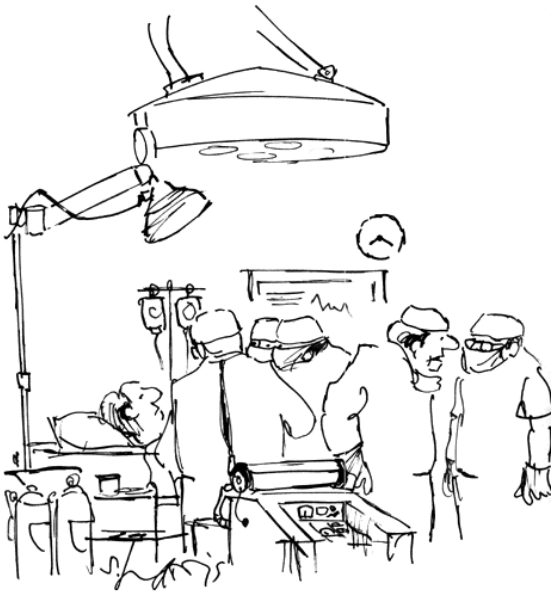
2. Blinding – keep anyone who could affect the outcome of the response from knowing which _____ have been assigned to which _____. A _____ experiment is when both the subject and experimenter don't know which treatment the subject has been given.

Example: in the earlier pet food example, the vet should not be told which dogs ate which food.

3. Use a placebo – often, simply applying _____ treatment can induce an improvement. A _____ treatment that looks just like the treatments being tested is called a placebo. Placebos are the best way to blind subjects from knowing whether they are receiving the treatment or not.

4. Blocking – group _____ experimental units together. Then random assignment of subjects to treatments is carried out separately within each block.

Example: in the previous dog food example, different breeds of dogs may respond differently to the foods. Blocking by breeds can remove that variation.



"We'll just mill around till he's asleep, and and then send him back up. This operation is actually for a placebo effect."

Example 4: Tire Blocking

A firm wishes to test the durability of four tire types that we'll call A, B, C, and D for convenience. Here are four possible studies they might perform. In all cases, the cars are to be driven on a track under controlled conditions until its tires are deemed "worn out". The response variable for each experimental unit (a car) is the number of miles the car drove with the tires. Each of the first three designs contains at least one serious weakness. Comment briefly on them. The fourth design is called a blocked design. State what the blocks are and explain what the advantage is of this design over design number 3.

1. Four Cadillacs of the same type are purchased new from four dealers. One gets tire A (i.e., gets outiftted with four type A tires), one gets B, one gets C, and one gets D.

2. Twelve Cadillacs of the same type are purchased new from four dealers. Three get tire A, three get B, three get C, and three get D.

3. Twelve vehicles of different types are randomly selected from a list of many vehicle types and then are randomly allocated into four groups of three. One group gets tire A, one group gets tire B, one group gets tire C, and one group gets tire D.

4. Four Cadillacs, four Fords, and four Volkswagens are purchased. One of each type of car gets tire A, one gets tire B, one gets tire C, and one gets tire D.