Unit 4 – Experimental Design

Lesson 2
Observational studies of the effect of one variable on another often fail because of **confounding** between the explanatory variable and one or more **lurking variables**.

**Definition:**

A **lurking variable** is a variable that is not among the explanatory or response variables in a study but that may influence the response variable.

**Confounding** occurs when two variables are associated in such a way that their effects on a response variable cannot be distinguished from each other.
**Lurking Variables:**
Influences change in the two variables of interest creating the impression of an association between them. Often with observational studies. For example, in a study where the number of tv's per household was compared to the life expectancy, it was found that there was a positive relationship - i.e. the more tv's, the longer the life expectancy. What are some possible variables that would explain high values for both of these variables?

**Confounding Variables:**
Often with experiments - you can't tell which variable had an affect on the response. For example: suppose you are interested in whether or not fertilizer helps your lawn grow better. You spread fertilizer on half of your lawn which just so happens to be the sunny half. You can't determine if it was the sun or the fertilizer that caused a change in how the lawn looks compared to the non-fertilizer/non-sun side.
There is a positive association between ice cream sales and drowning. Can you think of a lurking variable in this situation?
During the summer, there was a week of rainy days in a row. A grocery store was also running a special on movie rentals. During that week there was an increase in video rentals. The following week rentals went down but sunny weather returned. What is this an example of? What does this mean for the marketing folks at the grocery store?
A car battery company wanted to test their car battery against their competitors for how quickly a car engine started with their battery as opposed to the competition. They took a car and placed their battery in a car in a garage and put their competitor's battery in a similar car parked on the street. What is wrong with this design?
A study showed that people who drive more expensive cars are better golfers. Can you think of a lurking variable for this situation?
How to Experiment Well:
The Randomized Comparative Experiment

The remedy for confounding is to perform a *comparative experiment* in which some units receive one treatment and similar units receive another. Most well designed experiments compare two or more treatments.

Comparison alone isn’t enough, if the treatments are given to groups that differ greatly, *bias* will result. The solution to the problem of bias is **random assignment**.

**Definition:**
In an experiment, **random assignment** means that experimental units are assigned to treatments at random, that is, using some sort of chance process.
**Four Principles of Experimental Design**

1. **Control**
   - sources of variation - experimental factors and other sources

2. **Randomize**
   - helps to equalize the effects of unknown or uncontrollable sources among all experimental units
   - if you do not randomize you CANNOT use statistical techniques to draw conclusions

3. **Replicate**
   - two kinds
   - use multiple subjects in order to determine the variation among the responses
   - experimental units need to represent the population so that the experiment could be replicated with other representative samples

4. **Block (not required)**
   - use when attributes of the experimental units that are NOT being studied might affect the outcomes
   - similar to strata when sampling
Eye cataracts are responsible for over 40% of blindness around the world. Can drinking tea regularly slow the growth of cataracts? We can't experiment on people so we use rats as subjects. Researchers injected 18 young rats with a substance that causes cataracts. One group of the rats also received black tea extract; a second group received green tea extract; and a third got a placebo, a substance with no effect on the body. The response variable was the growth of cataracts over the next six weeks. They found that both tea extracts did slow cataract growth. Outline the design of this completely randomized experiment. Use the table of random digits, staring at line 124, to assign rats to treatments.
The law allows marketers of herbs and other natural substances to make health claims that are not supported by evidence. Brands of ginkgo extract claim to "improve memory and concentration." A randomized comparative experiment found no evidence for such effects. The subjects were 230 healthy people over 60 years old. They were randomly assigned to ginkgo or a placebo pill. All the subjects took a battery of tests for learning and memory before treatment started and again after six weeks.

(a) Outline a completely randomized design of this experiment.

(b) Assign the first 5 members of the ginkgo group using the line of random digits given below:

45149 32992 75730 66280 03819 56202 02938 70915 61041 77684 94322
24709 73698
**Randomized Block Designs** - subjects are placed into blocks based on some similarity that is known prior to the experiment and is expected to affect the response to the treatments. Within each block the subjects are randomly assigned to treatments.

**FOR EXAMPLE:**
Women as a group develop heart disease much later than men. You have recruited 300 adults aged 45 to 65 who are willing to follow your orders about alcohol consumption over the next five years. You want to compare the effects on heart disease of moderate drinking of just wine, just beer, or just spirits. Your 300 subjects include 120 women and 180 men. Outline a block design for comparing wine, beer and spirits. Be sure to say how many subjects you will put in each group in your design.
**Matched Pairs design** -
- A completely randomized design that compares two treatments
- Subjects are grouped in pairs that are as closely matched as possible and receive different treatments which are assigned randomly within the pair.
- Sometimes these pairs are actually a single subject who receives both treatments and therefore serve as their own control.
- The effect of the treatments is then compared within each pair.

**FOR EXAMPLE:** the standard Coke vs. Pepsi taste test

*Can you think of another example?*
Which of the following statements are true?

I. A completely randomized design offers no control for lurking variables.

II. A randomized block design controls for the placebo effect.

III. In a matched pairs design, participants within each pair receive the same treatment.

(A) I only
(B) II only
(C) III only
(D) All of the above.
(E) None of the above.
Consider the following hypothetical experiment. Acme Medicine is conducting an experiment to test a new vaccine developed to immunize people against the common cold. To test the vaccine, Acme has 1000 volunteers - 500 men and 500 women. The participants range in age from 21 to 70.

Describe three experimental designs - a completely randomized design, a randomized block design, and a matched pairs design - and show how each design might be applied by Acme Medicine to understand the effect of the vaccine, while ruling out confounding effects of other factors.
The idea of a randomized comparative experiment is to give good evidence that differences in the treatments actually cause the differences we see in the response.

An observed effect that is SO large that it would rarely occur by chance is called statistically significant.

**Statistically significant association (from a well-designed experiment) DOES imply causation**