***Curriculum Topic II – DATA COLLECTION***

***Chapter 5***

**Data must be collected according to a well-developed plan if valid information on a conjecture is to be obtained.**

The plan must identify important variables related to the conjecture and specify how they are to be measured. From the data collection plan, a model can be formulated from which inferences can be drawn. There are three primary ways to gather data: sampling, observational studies, and designed experiments.

**Sampling**

* Students should be able to design and implement a procedure for gathering a simple random sample from a given population, either using a table of random digits or a random number generator.
* They should recognize at least three sources of bias in any sampling procedure: selection bias, response bias, and non-response bias.
* Students should understand other methods of sampling, including cluster and stratified sampling. They should understand the ideas of sampling with replacement or without replacement.

**Observational studies**

In an observational study, an investigator is often using a sample from one or more groups to draw conclusions about these groups or the differences between these groups. It is important to recognize that while associations between two groups or variables may be obtained through an observational study, no conclusions can be reached about the reasons for these differences, due to the possible presence of confounding variables. A confounding variable is related to both group membership and the response variable being measured. A variable is confounded if its effects on the response variable cannot be separated from the effects of the explanatory variable.

**Designed experiments**

 In a designed experiment, an investigator manipulates one or more factors and measures the effect on a response variable. With random allocation to the experimental groups, the effects of any confounding groups should be spread equally among the groups. This allows a conclusion of a cause and effect relationship between the manipulated factor and the response variable.

 Students should understand the role of randomization in this equalizing of the effects of potentially confounding variables. Randomization helps prevent effects from unanticipated variables.

 Blocking is often used in an experiment if there is some factor that can be anticipated as having an effect on the response. The subjects of the experiment are first separated into blocks, which are as alike as possible in the anticipated factor (homogenous). The variation in outcome due to different randomizations can be made smaller, on average. This increases the likelihood that true differences between the two groups can be accurately detected. Blocking reduces variation by distributing anticipated factors that might affect the response equally to the treatment groups.

 For experiments that are conducted on human subjects, the use of a placebo and blinding can control for some psychological effects that might otherwise bias the results.

 At this point in the curriculum, students should understand the broad idea of statistical significance as a result that is not likely to be explained by random selection or random assignment.

**Surveys, observational studies, and experiments**

Know what is required for a sample to be a **simple random sample** (SRS). If each individual in the population has an equal probability of being chosen for a sample, it doesn't follow that the sample is an SRS. Consider a class of six boys and six girls. I want to randomly pick a committee of two students from this group. I decide to flip a coin. If "heads," I will choose two girls by a random process. If "tails," I will choose two boys by a random process. Now, each student has an equal probability (1/6) of being chosen for the committee. However, the two students are not an SRS of size two picked from members of the class. Why not? Because this selection process does not allow for a committee consisting of one boy and one girl. To have an SRS of size two from the class, each group of two students would have to have an equal probability of being chosen as the committee.

SRS refers to how you obtain your sample; random allocation is what you use in an experiment to assign subjects to treatment groups. They are not synonyms.

Well-designed experiments satisfy the principles of **control, randomization, and replication**.

* Control for the effects of lurking variables by comparing several treatments in the same environment. **Note:** Control is not synonymous with "control group."
* Randomization refers to the random allocation of subjects to treatment groups, and not to the selection of subjects for the experiment. Randomization is an attempt to "even out" the effects of lurking variables across the treatment groups. Randomization helps avoid bias.
* Replication means using a large enough number of subjects to reduce chance variation in a study.
**Note:** In science, replication often means, "do the experiment again."

Distinguish the language of surveys from the language of experiments.      Stratifying:sampling::Blocking:experiment

**It is not enough to memorize the terminology related to surveys, observational studies, and experiments**. You must be able to **apply** the terminology in context. For example:

**Blocking** refers to a deliberate grouping of subjects in an experiment based on a characteristic (such as gender, cholesterol level, race, or age) that you suspect will affect responses to treatments in a systematic way. After blocking, you should randomly assign subjects to treatments within the blocks. Blocking reduces unwanted variability.

An experiment is **double blind** if neither the subjects nor the experimenters know who is receiving what treatment. A third party can keep track of this information.

Suppose that subjects in an observational study who eat an apple a day get significantly fewer cavities than subjects who eat less than one apple each week. A possible **confounding variable** is overall diet. Members of the apple-a-day group may tend to eat fewer sweets, while those in the non-apple-eating group may turn to sweets as a regular alternative. Since different diets could contribute to the disparity in cavities between the two groups, we cannot say that eating an apple each day causes a reduction in cavities.