Unit 3-1 Introduction to Experimental Design

* Introduction to Experimental Design
  + Two common sources of data are observational studies and experimental studies.
    - **Observational studies** measure the characteristics of a population by studying individuals in a sample without manipulating or influencing the variables of interest. A common type of observational study is a survey. Surveys sample a particular group and ask it questions.
    - **Experimental studies** try to determine what effect a particular treatment has on an outcome by applying a treatment to some individuals (experimental units or subjects). A common type of experimental study is a drug trial, in which a study applies a drug treatment to one group of individuals (also known as experimental units or subjects) and compares the health outcome of interest (for example, blood pressure) to a group of individuals who are given a placebo treatment (the control treatment).
  + **Experimental design** refers to the plan for assigning units to treatment conditions in experimental studies. Good experimental design serves three purposes:
    - **Causation**: We can make causal inferences about the relationship between the independent variable and the dependent variable.
    - **Control**: We can rule out alternative explanations due to the surprising effects of unmeasured variables.
    - **Variability**: When we reduce the variability within the treatment conditions, it's easier to detect differences in treatment outcomes.
* Frequentist Hypothesis Testing: Drug Efficacy Example
  + Let’s say we're interested in testing the efficacy of a new drug. This drug is supposed to lower blood pressure, but we'll test its efficacy using frequentist statistics and the concepts of experimental design.
  + We randomly select 50 people to be in the placebo control group and 50 people to receive the drug treatment.
  + We know our sample is selected from the broader, unknown population pool.
  + We can imagine that, in a hypothetical, parallel world, we could have ended up with a different random sample of subjects from the population pool.
  + We're interested in the average difference in blood pressure levels between the treatment and control groups, so we'll measure each subject's systolic blood pressure at the end of the trial. In order to determine if our drug if effective, we'll see if the difference between the groups' average systolic blood pressure measurements is statistically significant.
* Defining a Testable Hypothesis
  + To perform our experiment, we first must design a testable hypothesis. A testable hypothesis is a statement that can be either supported or falsified based on data.
  + A counterexample must be possible, meaning that the hypothesis must be falsifiable. It must be possible to observe whether the hypothesis is true or false. You should be able to design an experiment to test the hypothesis.
  + A non-testable hypothesis is more of a value judgment or an opinion. It may also be a statement in which the researcher has left out a basis for comparison.
* The Null Hypothesis and Its Alternative
  + The null hypothesis is a general statement or a default position that there is no relationship between two measured phenomena.
  + In frequentist statistics, we assume the null hypothesis is true until evidence indicates otherwise.
  + The null hypothesis is denoted as H0.
  + In our drug efficacy experiment example, our null hypothesis is that there is no difference between the treatment and control groups.
    - H0: The mean difference between treatment and control groups' blood pressures is zero.
  + The alternative hypothesis is a statement that there \_is\_ a relationship between two measured phenomena. It's usually the outcome of the experiment that we hope to show.
    - In our example, the alternative hypothesis is that there is a mean difference in blood pressure between the treatment and control groups.
    - H1: The parameter of interest — the mean difference between treatment and control blood pressure — is different than zero.
  + Let's review the null and alternative hypotheses for our drug efficacy experiment.
    - H0: The mean difference between the treatment and control groups' blood pressures is zero.
    - H1: The parameter of interest — the mean difference between treatment and control blood pressure — is different than zero.
* Experimental Design
  + An experiment is performed to decide whether or not the observed differences among the treatments are due only to chance.
  + There are four important principles of experimental design:
    - Random sampling.
    - Replication
    - Randomization among treatments.
    - Local control.
* Random Sampling
  + Using a randomized sample is the best way to obtain a sample distribution that's representative of the entire population. Giving all experimental units an equal chance of being chosen for the study will eliminate systematic bias. Additionally, the mathematical theorems that justify most frequentist statistical procedures apply only to random samples.
* Replication
  + Replication allows the experimenter to provide an estimate of experimental error. The number of replications (sample size) is the number of experimental units that receive each treatment. These samples should be independent, meaning that one measurement is not connected to another.
  + This sample size should be small enough that negligible treatment differences aren't declared statistically significant but large enough that meaningful treatment differences are.
  + For our drug efficacy experiment, we place 50 people in the placebo control condition and 50 people in the drug treatment. These are 50 independent samples, because one person's blood pressure measurement does not affect another's.
* Randomization Among Treatments
  + Randomization is introduced into the experiment to ensure the estimate is statistically valid. Treatments should be assigned to experimental units randomly to prevent the introduction of systematic bias.
  + For our drug efficacy experiment, we randomly place 50 people in the placebo control condition and 50 people in the drug treatment condition. That way we don't bias our results by accidentally placing people with higher blood pressure in the control group.
* Local Control
  + Local control reduces experimental error. This means that treatments should be applied uniformly under standardized conditions.
  + For our drug efficacy experiment, we'll standardize all of the other conditions to ensure we don't introduce other differences between the treatment and control groups that may affect the experiment. Other factors that may also influence blood pressure, such as diet or exercise, should be standardized to avoid the introduction of hidden factors.
* Control Group and Experimental Group
  + In the context of experiments, we often talk about the control group and the experimental or treatment group. In our drug efficacy example, the control group is given the placebo and the treatment group is given the actual drug. We're interested in the difference in mean blood pressure levels between the treatment and control groups.
  + We randomly select 50 people to be in the placebo control condition and 50 people to receive the drug treatment. We keep all of the other conditions standardized for the people involved in the experiment.
* Why Are Randomized Control Trials Important?
  + Randomized control trials randomly assign participants into either an experimental group or a control group. As the study is conducted, the only difference between the groups is the outcome variable being studied.
  + Randomization minimizes selection bias. With all of the other variables kept constant, the researchers will be able to more accurately determine the effects of the treatment in comparison to the control group.
  + In our drug efficacy trial, the people participating in the trial will be randomly placed in either the group receiving the drug treatment or the group receiving the placebo treatment (the control). All conditions except for those being tested will be kept constant. With randomized control trial experimental design, we minimize selection bias and the potential for another unmeasured variable to cause the measured difference between the study groups.