Producing Data - Important Terms

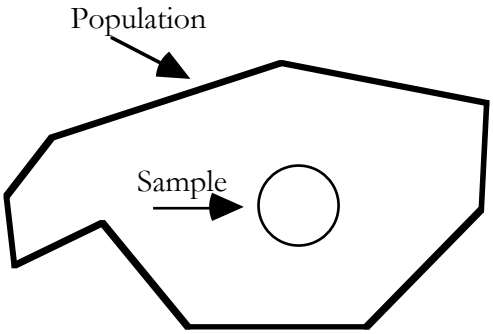
I am not a teacher who believes that memorization is a particularly good way of learning information. But in the case of A.P. stat, I have to make an exception. The terms and concepts below are vital to your success in this course. While you will probably not be asked to define a term in the AP exam, you will be asked, for example, to generate a multistage sample design experiment. If you don't know what that term means how can you do what is required? So as we go through this chapter on experimental design, and discuss each term, you need to refer to this sheet as all the terms and concepts are below giving you a simple sheet off which you can study.

Designing Samples:

- **Population** - the entire group of individuals we want information about. A population can be huge like "all women." It can be small like "all statistics students at Wissahickon."

- **Sample** - a part of the population that we actually examine in order to gather information.

- **Statistical Inference** - coming to a conclusion about the population based on the sample.



Terms associated with Sampling

- **Sample design** - the method used to choose the sample from the population. Poor sample designs can produce misleading conclusions.

- **Voluntary response sampling** - people who choose to be in a sample. This technique usually over represents people with strong opinions, usually negative ones.

- **Confounding** - two variables are confounded when their effects on a response variable cannot be distinguished from each other.

- **Bias** - the design of a study is biased if it systematically favors certain outcomes. Bias does not have to be intentional. But it must be avoided.

- **Simple random sample (SRS)** - choosing n individuals from the population in such a way that every set of n individuals has an equal chance to be the sample actually chosen. Note that it does not give just every *individual* an equal chance to be chosen in the sample - it gives every *group* an equal chance of being chosen.

- **Table of random digits** - a long string of numbers 0 through 9 where each entry is equally likely to be to be any of these 10 digits and the entries are independent of each other.

• **Probability sample** - this gives each member of the population a known chance (greater than zero) to be selected. All SRS's are probability samples with an equal chance to be selected.

• **Stratified Random Sample** - first divides the population into groups of similar individuals called **strata**. Then choose a separate SRS in each stratum and combine these SRS's to form the full sample.

For instance, if we wanted to choose a random sample of students from WHS, we may choose to do it using the stratified random sample approach. First divide the school population into strata, which could be the student's grade. Choose a SRS from each grade and combine these SRS's to form the full sample.

• **Multistage sample design** - choosing the sample in stages. For instance, if we wanted to do a multi-stage sample design in choosing students at WHS, we might do the following.

Stage 1. During period 3, there are 50 classes in session. Take a sample of these classes
Stage 2: Take a sample of students within each class.

If we were to take a sample of boys and girls in each class, then we would have a combined multistage sample design using stratified random sampling.

• **Undercoverage** - when some groups are left out of the process of choosing the sample. This may be intentional or unintentional. For instance, if I were to do some type of survey via email, there would be undercoverage because I would not be including people without computers.

• **Nonresponse** - when the individual chosen for the sample cannot be contacted or refused to cooperate. Think of how many surveys you have decided *not* to fill out and you can see why nonresponse is a problem.

• **Response bias** - respondents lying, usually about illegal or unpopular behavior.

• **Wording of questions** - confusing or leading questions can introduce bias.

• **Sampling frame** - a list of individuals from which a sample is chosen. This should be the entire population but in practice this is difficult.

Terms associated with experiments

• **Observational study** - observes individuals and measures variables but does not attempt to influence the responses. This is a poor way to find the effect of one variable on another.

• **Experiment** - imposes some treatment on individuals in order to observe their responses. For instance, if I wanted to show that intensive scheduling was a more effective way of teaching, I would measure how students do in a traditional length period, then change to a longer period, and compare the two results.

• **Experimental Units** - the individuals on which the experiment is done. If human, they are called subjects.

• **Treatment** - the specific experimental condition applied to the unit. In the example above, the treatment would be the longer class period. Do not think that treatment has to be like getting a medicine. A treatment might be as simple as tasting a new brand of ice cream. Experiments are usually of two types:

Treatment → Observation or Observation 1 → Treatment → Observation 2

• **Factor** - the explanatory variables in an experiment. In the example above, the factor would be the length of class period. There can be several factors in an experiment.

• **Level** - the specific value of each factor. For instance, in the example above, I may have a class period for some subjects to be an hour, others, an hour and a half.

• **Placebo effect** - a placebo is a dummy treatment. Most subjects respond favorably to any treatment, even a placebo. This response to a dummy treatment is called the placebo effect.

• **Control group** - the group of subjects who receive a sham treatment.

• **Matching** - trying to ensure that the experimental group and control group are matched - have the same type of subjects.

• **Completely randomized design** - all the experimental units are allocated at random among all the treatments.

• **Statistical Significance** - an observed effect too large to attribute plausibly to chance.

• **Replication** - repeating each treatment on a large enough number of experimental units or subjects.

• **Principals of Experimental Design**

1. Control of the effects of lurking variables.
2. Randomization, the use of impersonal chance to assign subjects to treatments.
3. Replication, to reduce chance variation in the results

• **Hidden bias** - taking care that all experimental units or subjects are treated in exactly the same way or hidden bias will occur.

• **Blind experiment** - the subjects do not know what treatment they receive.

• **Double blind experiment** - neither the subjects nor the people who have contact with them know which treatment a subject receives.

• **Block Design** - a block is a group of subjects that are similar in ways that are expected to affect the response to treatments (ex: men and women). In block design, the random assignments of units to treatments are carried out separately within each block.

• **Matched Pairs Design** - these compare two treatments. Each block consists of two units or subjects, closely matched (ex: male 27 yrs old, unmarried, female 27 yrs old, unmarried). Subjects are assigned randomly to the treatment. It is possible that each block consists of just one subject who gets both treatments. Each subject serves as his own control. For instance, if I were testing a new brand of salad dressing, the subject might receive both the new brand and an established brand in random order. The established brand is the control.