Chapter 13: Experiments and Observational Studies

The Difference (again)

Remember that in an experiment, you are going to change something and observe the effects—as opposed to an observational study, where you only observe. Don’t let the title of this chapter fool you—it is only about experiments.

The Guiding Principle

I think that experimental design can be reduced to one essential principle: variation. It’s all about variation.

The big idea in an experiment is to make a change to one variable (introduce variation) and observe any changes in the other variable (measure variation). In an ideal situation, that’s all the variation there is…but in reality, there are other sources of variation. Identify them. Control them when you can. Account for them when you can’t.

It’s all about variation.

Vocabulary

Variables

What are you measuring? Weight? Time required to fall asleep? Dosage (mg) of a certain drug? These are your variables. You could have any number of variables, but they fall into three types: explanatory, response, and extraneous. The explanatory variable (or factor) is the one that we will change, in order to observe the resulting effect on the response variable. Extraneous variables are just that: extraneous. We don’t want/care about them, and must attempt to filter out their effects.

Many people sub-classify extraneous variables…they talk about lurking variables and confounding variables. I’m not going to say much about that, since this is just an introduction (if you want/need to know more, take more stats classes at the University!).

Levels

The explanatory variable(s) is/are set by the experimenter: how much of this drug will be used? How much weight will be placed on the girder? How long will we boil the solution? When we decide this, we are setting the levels of the explanatory variable/s. For example, if we are studying the effects of a certain drug, then we might try several dosages—say, 50mg, 100mg and 150mg. This is one variable (Dosage) with three levels (50, 100, 150).

Note that we will NOT decide on levels! That is far beyond our ability. The levels should either be given, or obvious—the most obvious levels being presence/absence…

Treatments

A treatment consists of one level from each explanatory variable. For example, let’s say that you are conducting an experiment on the effects of stirring rate (rpm) and duration (seconds) on the incorporation of an ingredient in a solution/mixture. Furthermore, let’s say that you’ve
decided to test three stirring rates (60, 120 and 240rpm) and two durations (1 and 2 minutes). There would be six treatments—60rpm and 1 minute; 60rpm and 2 minutes; 120rpm and 1 minute; 120rpm and 2 minutes; 240rpm and 1 minute; 240rpm and 2 minutes.

**Experimental Units / Subjects**

The things on which we experiment are called the experimental units—perhaps they are batches of paint, or lots of bearings. When the experimental units are people, we call them subjects.

**Examples**

[1.] A flower farmer is investigating the best ways to produce high quality flowers in a greenhouse. He’s decided to compare three kinds of humidity and four different types of fertilizer. At the conclusion of the experiment, a local florist will evaluate the quality of all of the flowers.

Identify the explanatory and response variables, number of levels of all factors, and the total number of treatments.

The explanatory variables are humidity and fertilizer. There are three levels of humidity and four levels of fertilizer. There will be a total of \(4 \times 3 = 12\) treatments.

The response variable is flower quality.

[2.] A study concerning pain caused by injections measured several possible causes: location (stomach or thigh), volume injected (400 µL, 800 µL, 1200 µL, 1600 µL) and injection speed (150 µL/s, 300 µL/s, 450 µL/s). Pain level was measured using the standard Visual Analog Scale (pain level is measured between 0 and 10).

Identify the explanatory and response variables, number of levels of all factors, and the total number of treatments.

There are three explanatory variables: location, volume, and speed. Location has two levels, volume has four levels, and speed has three levels. There must be a total of \(2 \times 4 \times 3 = 24\) treatments.

The response variable is pain level.

**Comparative Experiments**

It is typically the case that an experiment is designed to determine if changes in A cause (yes, cause) changes in B. The only way that we can be sure that this is the case is to control the effects of any extraneous variables. Sometimes this is possible—often, it is not. Thus, we need to conduct comparative experiments, where we compare results of several treatments that are given under similar conditions.

**Confounding**

One reason why we compare is to minimize/eliminate any confounding. Confounding occurs when we can’t be sure that those changes in B were caused by our changes in A, or by some other reason C. For example…suppose I offer a test camp designed to help students improve their SAT scores. After the camp, every student’s score goes up. Was this change due to
attending the camp, or was it because of some other reason (like the fact that the students are now older, or they’ve completed more material in classes)?

The Placebo Effect

Many experiments compare a treatment against a placebo. A placebo is anything which has a known (or no) effect on the response variable. The word *placebo* refers to a sugar pill, but a placebo does not need to be a pill! For example—does advertising with subliminal messages cause people to have a higher opinion of certain products? When conducting this experiment, we might show people advertising with subliminal messages, and then show some other people advertising without the subliminal messages. The ads with subliminals are an actual treatment; the ads without them are a placebo.

Placebos help to reduce variation due to a subject’s knowledge of the experiment (which is a known condition!).

Historical note: it was once common to include placebos in a doctor’s bag. When patients came and described symptoms which didn’t point to an obvious cause, a doctor would give the patient a placebo. If the patient did not return (if the condition improved), then all was well—either the condition improved on its own, or it never really existed (psychosomatic). If the patient did not get better, *then* the case warranted further attention…

The Control Group

Those individuals that actually receive a treatment are in the *experimental group*. Those that receive a placebo are in the *control group*. It is important that similar individuals are in each group. The best way to do that is…

Randomization

…to randomly assign individuals to these groups! This will “level out” the effects of any unknown extraneous variables in the experiment. In other words, randomization will spread out any extraneous variation amongst the groups—thus, reducing between-group variation.

The Principles of Experimental Design

Our textbook lists four principles. I don’t think you need more than three!

1. **Control**—the effects of extraneous variables must be taken into account (i.e., controlled). Unknown variables are controlled through randomization. Known variables are handled through direct manipulation (e.g., I can directly control the temperature in the room), or blocking (see notes below).

2. **Randomization**—individuals must be assigned to treatment groups at random!

3. **Replication**—larger samples give better results; more individuals in the experiment will provide more reliable results! Replication reduces variation that naturally arises due to chance (from random assignment to groups, for example?).

Really, these can be reduced to a single phrase: **reduce variation**!

OK, I’ll say that it a longer way. Use randomization to reduce variation from unknown sources. Use controls to reduce variation from known sources. Use randomization to reduce chance variation.

*It’s all about variation.*
Additional Precautions

Blinding

Placebos are used as the basis of comparison (and one form of control). Unfortunately, knowing that you might have been given a placebo could introduce bias into the experiment. Thus, it is important that the subjects do not know what kind of treatment has been received. When this is the case, the experiment is single blind.

It is also important for the person taking the measurements not to know what kind of treatment the subjects received—otherwise, there may be some hidden bias. When neither the subject, nor those who take the measurements know which treatment was applied, then the experiment is double-blind.

Realism

In an experiment, we change something. Unfortunately, we can’t always do that in an environment similar to that in which the change might naturally occur…for example: does noise level affect performance on a test? Noise level is explanatory; that’s what we’ll change. It is not easy to make noise, though, that is similar to that which would normally be experienced during a test (Personally, I participated in exactly this type of experiment in college. The noise was simply a TV broadcast, with the volume turned up very loud.). In this case, there is a lack of realism that will have some effect on the results.

Matched Pairs

One way to deal with unknown extraneous variables is randomization, but that isn’t always the best answer. Often, a better choice is to use a matched pairs design. This involves matching one experimental unit in the experimental group with one experimental unit in the control group—these two units are either selected because they are identical or very similar, or they are made to be very similar.

Block Designs

Extraneous variables that are known need to be controlled. Sometimes, this is easy—I can adjust the temperature in the room, or I can cover up the label on the can. There are other extraneous variables that I can’t easily control or affect—I can’t change your gender, nor can I change whether or not you’ve ever taken the SAT. When we encounter extraneous variables that can’t be changed, then we must block. This involves splitting the experimental units into homogenous groups (blocks), and then performing the experiment within each block. This reduces within-group variation.

The Main Types of Experiments

There are three experimental designs that are mentioned by name in the AP Statistics program. You need to know what each of these means so that you can include the proper features when describing your design!

Completely Randomized Design

This is a design where all treatments are assigned at random.
Randomized Block Design

In this design, subjects are first blocked by some variable, then assigned treatments (randomly) within each block. In essence, the experiment is split up into lots of little experiments (one per block).

Matched Pairs Design

This could be thought of as a very special kind of block design, where each block consists of exactly two individuals that have been paired by one or more variables—in other words, they are similar or identical in some manner. Within each pair, one individual is (randomly) assigned to the treatment, and the other to the control.

Examples

[3.] Returning to the flower farmer from Example [1]…he wants to compare the effects of those variables on five different species of flower. What type of design must he choose?

Since the farmer can’t change the species of a flower, he’ll have to use a block design. I hope he has enough greenhouses to do this!

[4.] A consumer group is testing a company’s claim that their newest vehicle gets the best fuel economy (miles per gallon) when using premium gasoline. The consumer group has purchased nine of these vehicles for testing. They have all three of the typical US gasoline versions (regular, mid-grade and premium) available, a professional test driver, and access to a test track in California (where the weather is generally very stable).

Describe a completely randomized design that will help determine if gasoline type affects fuel economy for this model of car.

There is only one factor (gasoline type) with three levels. I’ll write down the VIN for each car on an index card, then put all of the cards into a box and mix thoroughly. I’ll draw out three cards—those three cars will be driven with regular gasoline. I’ll draw three more cards, and assign those cars to be driven with mid-grade gasoline. The final three cars will be driven with premium gasoline. I’ll write each gasoline type on the index card with the VIN, then put all of the cards back in the box.

Now I’ll start drawing out the cards one at a time. The order in which the cards are drawn will be the order in which the cars are driven with their assigned gasoline type. I need to make sure that the driver is unaware of the gasoline type used in the car for each test run. Also, I’ll want to make sure that the same amount of gasoline is put into a car for each trial.

At the conclusion, I’ll calculate the mean fuel economy for each gasoline type and compare the results.

[5.] Describe a matched pairs experiment that would determine if verbal encouragement increases a person’s grip. Assume that there are 50 volunteers available for this study, a force meter is available to measure grip, and one really enthusiastic cheerleader for encouragement.

Let each person be their own matched pair! In other words, each person will test their grip both with and without encouragement. Put the names of all 50 volunteers on index cards and put
them in a box. Mix thoroughly, then draw out 25 cards. These 25 people will test their grip with encouragement first, then their grip without encouragement. The other 25 people will do the opposite (no encouragement, then encouragement).

Put all 50 cards back into the box and mix again. Now draw out the cards one at a time to determine the order in which the volunteers will be tested. Be sure that the volunteer always uses the same (dominant) hand for both tests.