Chapter 9. Producing Data: Experiments

Experiments

Definition. The **individuals** studied in an experiment are often called **subjects**, particularly when they are people. The explanatory variables in an experiment are often called **factors**. A **treatment** is any specific experimental condition applied to the subjects. If an experiment has several factors, a treatment is a combination of specific values of each factor.

Example. Example 9.2 page 214. Notice the individuals (subjects), factors, treatments, and response variables.

Example S.9.1. Women Hate the Stooges!
It is commonly held that males are much fonder of the Three Stooges than females. Explain an experiment that could be performed to test this idea. What are the individuals, factors, treatments, and response variables of your experiment?

How to Experiment Badly

Note. A common way to perform an experiment (see Example 9.3) is outlined as:

$$\text{Subjects} \longrightarrow \text{Treatment} \longrightarrow \text{Measure response}$$

In a controlled laboratory setting, this may be suitable, but "in the field" it is quite possible that such a simple experiment can be misleading and confounded by lurking variables. The next subsection describes a better approach.

Randomized Comparative Experiments

Definition. An experiment that uses both comparison of two or more treatments and chance assignment of subjects to treatments is a **randomized comparative experiment**.

Definition. In a **completely randomized** experimental design, all the subjects are allocated at random among all the treatments.

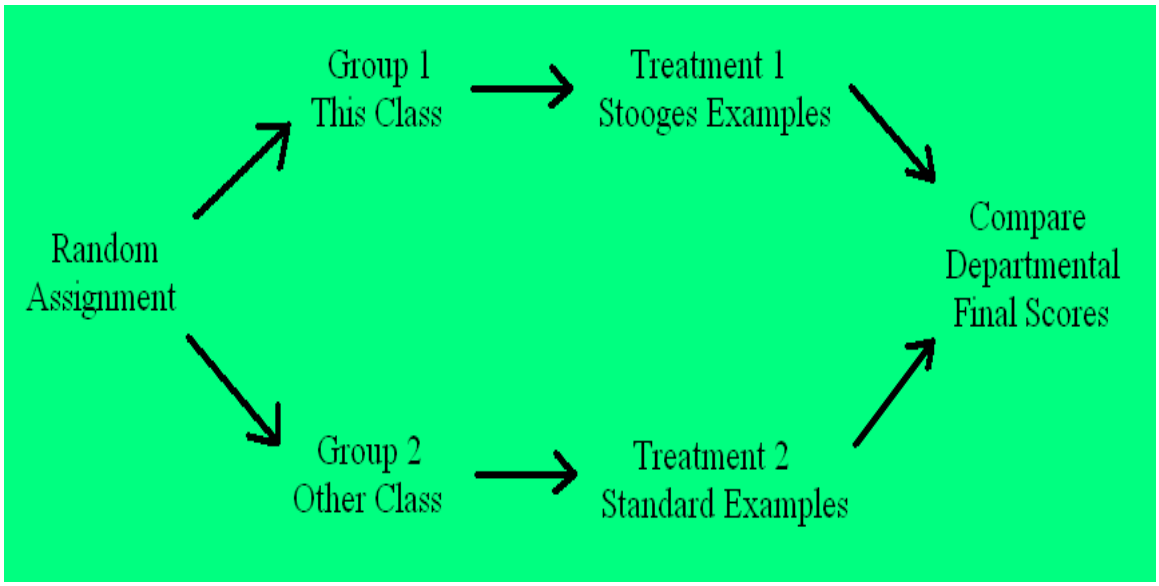
Note. It is common in many experiments, especially those testing drug treatments, to have a group which is not subject to any treatment. This group is called a **control group**.

Note. Random assignment is essential any any statistical study. Personal choice of the person designing the experiment or the subject can strongly influence results of an experiment. The fact that online surveys which many of the news channels love these days are based solely on voluntary response means that these "surveys" **are totally worthless** in a statistical sense.

Example S.9.2. Learnin' from the Stooges.
There are some who contend that by incorporating a topic from popular culture, in this case the films of the Three Stooges, into the college classroom experience, students are inspired to be more attentive, more involved, and ultimately more successful. Your knuckleheaded instructors make this claim. Design an experiment for us to test this idea.

Solution. Let's do this in the "real world" of this class and this semester. There are two particular sections of Probability and Statistics being taught at the same time—our section and another section (this is because we go back and forth between the Stat Cave and the Mansion). Let's assume that 40 students are registered for each class. We number the students 1 through 80 (maybe based on alphabetic order) and select a random sample of size 40 from this set. Let's step aside

and use the *Simple Random Sample* Applet provided by the publishers of our book to generate the numbers—notice that it simulates a lottery with numbered balls (the Applet will generate no more than 40 numbers in a sample)... Next, let's assign the 40 subjects from the sample (we call them "Group 1") to this class and the remaining 40 subjects ("Group 2") to the other section of Probability and Statistics which is taught at the same time. The treatment for Group 1 is exposure to examples and data gathering experience based on the films of the Three Stooges. The treatment for Group 2 is a standard non-Stooge presentation of the same material (the control group). To measure responses, we can use student score on the common departmental final. We then have the following outline of this completely randomized comparative experiment:



The subjects are the students, the treatments are the two different types of classes (Stooges/Standard), and the response variable is the score on the departmental final. Notice that by performing the experiment on classes that are taught at the same time, we are avoiding the lurking variable of "time of day of the class." However, there is still one very obvious lurking variable: the instructor(s). Since different instructors can have dramatically different teaching approaches, this certainly could have an effect much greater than the treatment! An alternative experiment might adjust for this by moving students from one section of our Probability and Statistics class to another section we are teaching at a different time. In practice, this would be a logistical nightmare, since it would potentially change students schedules and it would also bring back concerns over the lurking variable of "time of day." By the way, any such actual manipulation of students of this sort would require approval by ETSU's Institutional Review Board (IRB).

The Logic of Randomized Comparative Experiments

Note. Randomized comparative experiments are designed to give good evidence that differences in the treatments actually *cause* the differences we see in the response.

Definition. The basic principles of statistical design of experiments are:

1. **Control** the effects of lurking variables on the response, most simply by comparing two or more treatments.
2. **Randomize**—use impersonal chance to assign subjects to treatments.
3. **Use enough subjects** in each group to reduce chance variation in the results.

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

Note. A statistically significant association in data from a well-designed experiment *does* imply causation. Exactly how to determine if our results yield statistical signigicance will be dealt with later, starting in Chapter 14.

Example. Exercise 9.9 page 222.

Cautions about Experimentation

Note. A **placebo** is a dummy treatment. Many medical patients respond favorably to any treatment, even a placebo, perhaps because they trust doctors. The response to a dummy treatment is called the **placebo effect**.

Definition. In a **double-blind** experiment, neither the subjects nor the people who interact with them know which treatment each subject is receiving.

Note. The text comments: "The most serious potential weakness of experiments is lack of realism: the subjects or treatments or setting of an experiment may not realistically duplicate the conditions we really want to study."

Matched Pairs and Other Block Designs

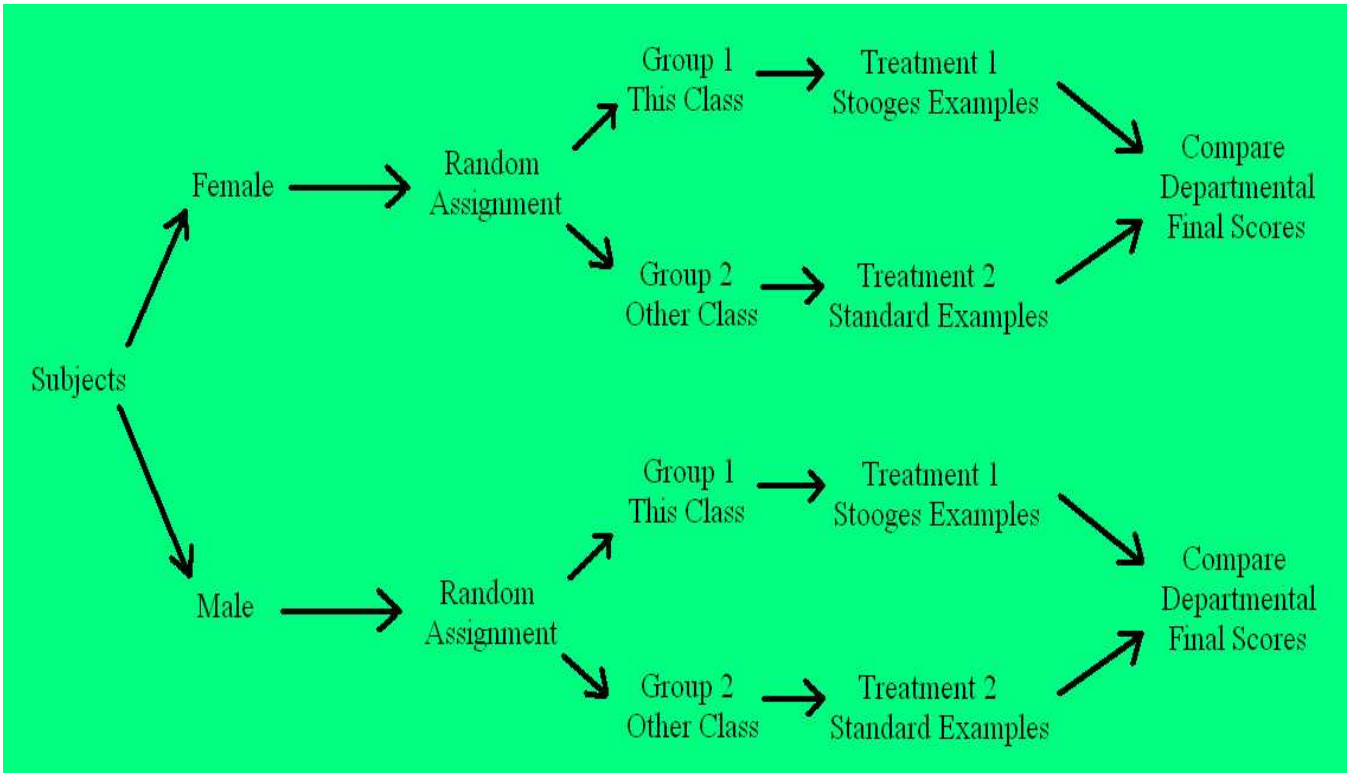
Note. The text states (pages 224 and 225): "A match pairs design compares just two treatments. Choose pairs of subjects that are as closely matched as possible. Use chance to decide which subject in a pair gets the first treatment.... That is, the random assignment of subjects to treatments is done within each matched pair, not for all subjects at once. ... Matched

pairs are one kind of *block design*, with each pair forming a *block*."

Definition. A **block** is a group of individuals that are known before the experiment to be similar in some way that is expected to affect the response to the treatments. In a **block design**, the random assignment of individuals to treatments is carried out separately within each block.

Example S.9.3. Women Hate the Stooges 2.
Let's reconsider the question raised in Example S.9.1: Do men like the Three Stooges more than women? If we find this credible, then we might think that using the Three Stooges films in Probability and Statistics would affect male and female students differently. Create a block design where the blocks are based on gender, but the subjects are then assigned to the two possible treatments (Stooge/Standard) at random.

Solution. An outline of the block design is:



Notice that the blocks are "Female" and "Male" and that assignment to the blocks is not done at random. However, assignment to the treatment is done randomly. To answer the question concerning the effect of a "Stooges intense" class on the two genders, we would need to make an additional comparison not in the outline. We would need to compare the results from the female population to the results for the male population.

Example. Exercise 9.25 page 229.

Example. Exercise 9.42 page 233. Repeat, but in the context of a Three Stooges related question.

rbg-3-7-2009