

## How to Experiment Well

### 1. Random Assignment

The remedy for confounding in the SAT Prep course example is to do a **comparative** experiment in which some students are taught in the classroom and other similar students take the online course.

Comparison alone is not enough to produce results that can be trusted. If the treatments are given to groups that differ greatly when the experiment begins, **bias** will result. The solution is **random assignment**.

**Definition:** In an experiment, **random assignment** means that experimental units are assigned to treatments at random, that is, using some chance process.

**Example** - 50 students have agreed to participate in an experiment to compare the online SAT course with traditional classroom instruction. Describe how you would randomly assign the two instructional methods.

LABEL STUDENTS 1-50, USE RANDINT(1,50)

FIRST 25 STUDENTS → ASSIGN TO O/L COURSE  
2ND 25 " → " " C/L "

### 2. Three Principles of Experimental Design

Logic of randomized comparative experiments:

- Ensures that influences *other* than the experimental treatments operate equally on all groups.
  - Controls for the effects of *other variables*.
- Balances out the effects of *lurking variables* that we cannot control or do not think of on treatment groups. *Confounding*
  - Random assignment forms groups of experimental units that should be similar.
- Since groups are roughly equivalent except for treatments, any differences in average response must be due either to the treatments or to the play of chance in the random assignment.



## Principles of Experimental Design

The basic principles for designing experiments are:

1. **Control** for other variables that might affect the response: Use a comparative design and ensure that only the systematic difference between the groups is the treatment administered.
2. **Random Assignment**: Use impersonal chance to assign experimental units to treatments. This helps create roughly equivalent groups of experimental units by balancing the effects of other variables that are not controlled on the treatment groups.
3. **Replication**: Use enough experimental units in each group so that any differences in the effects of the treatments can be distinguished from chance differences between the groups.

**Example** - Many students regularly consume caffeine to help them stay alert. Thus, it seems plausible that caffeine might increase an individual's pulse rate. One way to investigate this is to have volunteers measure their pulse rates, drink some cola with caffeine, measure their pulse rate after 10 minutes and calculate difference. Unfortunately, even if every student's pulse rate went up, we could not attribute it to caffeine.

Explain how to use all three principles of experimental design in the caffeine experiment.

CONTROL: CONTROL GROUP THAT RECEIVES NONCAFFEINE COLA. SAME AM'T OF COLA @ SAME TEMP. LOOK + TASTE EXACTLY SAME. DRINK @ SAME RATE + WAIT SAME AM'T OF TIME.

RANDOMIZATION: RANDOMLY ASSIGN SUBJECTS TO 2 TREATMENTS. THIS SHOULD BALANCE OUT EFFECTS OF CONFOUNDING VARIABLES THAT WE CANNOT CONTROL: BODY SIZE, CAFFEINE TOLERANCE, AM'T OF FOOD EATEN.

REPLICATION: USE AS MANY SUBJECTS AS POSSIBLE (LARGER SAMPLE, BETTER REPRESENTATION OF POP.)

### 3. Completely Randomized Designs

This diagram details the SAT prep experiment: random assignment, the sizes of the groups and which treatment they receive, and the response variable.

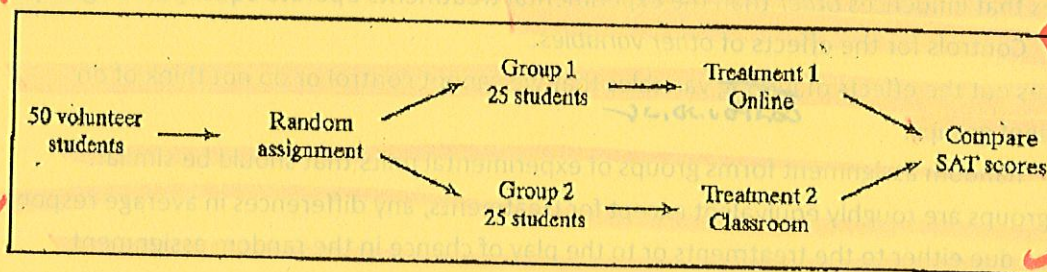


FIGURE 4.4 Outline of a completely randomized design to compare online and classroom instruction.

**Definition:** In a completely randomized design, the treatments are assigned to all the experimental units completely by chance.

- Completely randomized design does *not* require that each treatment have equal number of experimental units.
- Assignment of treatments *must* occur *completely* at random. (The best bet is to choose them using the "hat method.")

DISCUSS  
DESCRIBING  
EXPERIMENTS



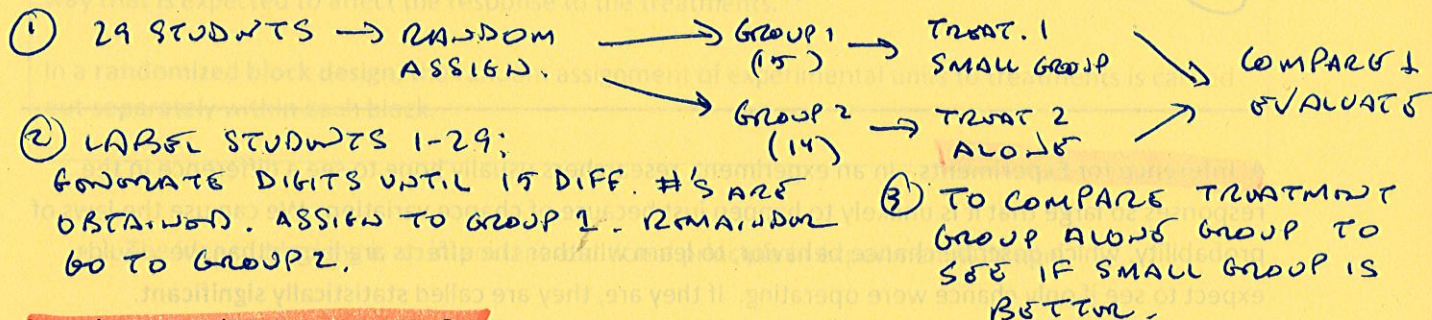
Some experiments include a **control group**. The primary purpose of a control group is to provide a baseline for comparing the treatments of the other treatments. Refer to example on p. 246.

It should be noted that although many experiments include a control group that receives an inactive treatment, a control group can be given an active treatment. For example, if researchers are only concerned with comparing two active treatments and do not care to determine if they are different than no treatment.

### CHECK YOUR UNDERSTANDING

Music students often don't evaluate their own performances accurately. Can small-group discussions help? The subjects were 29 students preparing for the end-of-semester performance that is an important part of their grade. Assign 15 students to the treatment: videotape a practice performance, ask the student to evaluate it, then have the student discuss the tape with a small group of other students. The remaining 14 students form a control group who watch and evaluate their tapes alone. At the end of the semester, the discussion-group students evaluated their final performance more accurately.<sup>22</sup>

1. Outline a completely randomized design for this experiment. Follow the model of Figure 4.4.
2. Describe how you would carry out the random assignment. Provide enough detail that a classmate could implement your procedure.
3. What is the purpose of the control group in this experiment?



### Experiments: What Can Go Wrong?

The logic of a randomized comparative experiment depends on our ability to treat all the subjects the same in every way except for the actual treatments being compared. Good experiments require careful attention to details to ensure that all subjects are really treated identically.

**Placebo Effect** - The placebo effect is the tendency in humans to show a response whenever they think a treatment is in effect. Well-designed experiments use a control group so that the placebo effect operates equally on both the treatment group and the control group, thus allowing us to attribute changes in the response variable to the explanatory variable.

**Double-Blind** - In a double-blind experiment, neither the subjects nor those who interact with them and measure the response variable know which treatment a subject received.

DISCUSS.



**Single-Blind** - Some experiments cannot be carried out in a double-blind manner. Sometimes the subjects know what treatment they are receiving. If those who interact with them do not know how the individuals are treated, the experiment is **single-blind**.

### CHECK YOUR UNDERSTANDING

In an interesting experiment, researchers examined the effect of ultrasound on birth weight. Pregnant women participating in the study were randomly assigned to one of two groups. The first group of women received an ultrasound; the second group did not. When the subjects' babies were born, their birth weights were recorded. The women who received the ultrasounds had heavier babies.<sup>27</sup>

1. Did the experimental design take the placebo effect into account? Why is this important?
2. Was the experiment double-blind? Why is this important?
3. Based on your answers to Questions 1 and 2, describe an improved design for this experiment.

① NO, IT IS POSSIBLE THAT WOMEN WHO THOUGHT THAT THEY WERE GETTING AN U/S WOULD HAVE DIFF. REACTIONS TO PREGNANCY THAN THOSE WHO KNEW THAT THEY HAD NOT.

② NO. WEIGHTS DID NOT KNOW BUT MOTHERS DID. MOTHERS MAY HAVE AFFECTED RESULTS.

③ DOUBLE BLIND: U/S MACHINE ON + OFF.

**4. Inference for Experiments** - In an experiment, researchers usually hope to see a difference in the responses so large that it is unlikely to happen just because of chance variation. We can use the laws of probability, which describe chance behavior, to learn whether the effects are larger than we would expect to see if only chance were operating. If they are, they are called **statistically significant**.

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

SIGNIFICANT.

If we observed statistically differences among groups in a randomized comparative experiment, we have good evidence that the treatments actually caused these differences. **A statistically significant association in data from a well-designed experiment does imply causation.**