# Teacher Tune-up

## Quick Content Refresher for Busy Professionals

### *What are some of the common ways people ineffectively use statistics?*

When the practices of statistics are important to the analysis of experimental data, people make two very common mistakes: *non-representative sampling* and *experimental bias*.

**Non-representative sampling**

You cannot draw valid conclusions about a population from a sample if the sample is not representative of the population, *no matter what its size*. Sample size matters, however, in the sense that the larger the (representative) sample, the more precisely the sample will mirror the population.

One great way to make a sample representative of a population is to choose the sample randomly. In more advanced courses, such as in high-school AP Statistics, getting a random sample is a big deal. Doing so in practice can be amazingly difficult, however.

The key thing to understand is that if you choose your sample based on what data are easy to get—a “convenience sample”—you will almost certainly get *biased* results, and your conclusions will not be correct for the population.

> **A more extended example**:
>
> Suppose you’re trying to figure out what music seventh-graders like. You develop a simple survey. If you give it only to your friends, it’s obvious that you’ll get a biased result. After all, you and your friends are more likely have similar tastes. So suppose you decide to give the survey to the first 30 people entering the cafeteria at lunch. Now you’re biasing the results towards students who get to lunch early. Maybe they come from a class that gets out earlier, or they walk faster, or something like that. In any case, they have something in common, and you don’t know whether that might be associated with some music preference. Your results are still biased. So you come up with a way to choose students randomly in the cafeteria. But now your results are biased towards students who eat in the cafeteria—if anyone brings their lunch and eats somewhere else, they are not represented, resulting in biased data again.
>
> ***Solution***:
>
> *The only way to get a representative sample is to get a list of all seventh-graders, randomly select names from that list, and use only those in your survey.*

As to sample size, *mirroring* the population means this: when you analyze the results from your sample, you will get some *numbers*. The numbers might be *proportions* (what percent of people in your sample prefer soccer to basketball) or they might be other statistics such as averages (the average length of their feet in centimeters). The bigger the sample, on average, the closer that number will be to the proportions or statistics from the population at large.

If the sample is small but representative, you can draw perfectly valid conclusions, but you have to accept that the conclusions (though valid and unbiased) might be imprecise.

**Bias**

In everyday speech, we talk about the bias of individuals, but in statistics terminology, “bias” means that the study *design* promotes *overall results* that might not match the population. The randomized, double-blind study is the “gold standard” for experiments involving human subjects.

If you’re a subject in a medical study and you know you’re in the treatment group (it’s not double-blind), you (the individual subject) might be “biased” (as we use “bias” in everyday speech) towards thinking you’re getting better, mind-over-body, etc. That’s *not* what we mean by statistical bias, even though the experimental design is in fact biased in this case.

To overcome the problem of bias, researchers invented the placebo, so that the subjects do not know (are *blinded*) whether they are getting the treatment or not. That is, the placebo is a mechanism for doing the study correctly, not a consequence of doing it wrong. The “placebo effect” is the strange mind-over-matter result in which placebo subjects generally see improved outcomes *even though they are not getting the treatment*. (That is, they are [individually] “biased” towards reporting improvement as a result of being in the study, even though the study design is not biased. See the distinction?)

While the randomized, double-blind study is the gold standard for experimentation, it applies mostly to experiments with human subjects. Plenty of valid studies are not experiments, and plenty do not involve human subjects.

Two examples of perfectly valid trials:

- You randomly assign tomato plants to the fertilizer A and fertilizer B groups. Every day you measure them and record their heights. The experiment is not double-blind because you know which group each plant is in. But it’s not biased because you’re making a quantitative measurement that you cannot influence by your demeanor. (If you’re worried about a researcher’s thumb being on the scale, you could blind this design, but for this example, it’s not necessary.)

- You study the cancer rates among people who do and do not smoke. This is not an experiment because you did not assign people to smoke or not smoke. But it’s still a valid study showing the association between the two. What you cannot do (immediately) is assign *cause*. After gathering sufficient evidence, however, you can make a pretty good case that smoking causes cancer, as opposed to, for example, drawing the conclusion that people who are cancer-prone have a natural affinity for smoking.