

Prof. David Draper
Department of
Applied Mathematics and Statistics
University of California, Santa Cruz

AMS 7: Discussion Section 2

1. Here is that data set from Zar (1999) again on butterfly wing lengths (in cm), sorted from smallest to largest:

3.3 3.5 3.6 3.6 3.7 3.8 3.8 3.8 3.9 3.9 3.9 4.0 4.0 4.0 4.0 4.1 4.1 4.1 4.2 4.2 4.3
4.3 4.4 4.5

One possible histogram for these data would have 13 bars, centered at 3.3, 3.4, \dots , 4.5; another would have only 7 bars, centered at 3.35, 3.55, 3.75, \dots , 4.55. Draw both of these histograms and comment on which of them seems better to you at displaying the basic features of the distribution (or maybe you'll feel that they're about equally good); explain briefly.

2. In problem 3(c) in Discussion Section 1, I measure the temperatures in $^{\circ}C$ and you decide to work with the same data but in $^{\circ}F$. How will the basic shape of my histogram relate to the shape of yours? Explain briefly.

3. I measure the enzyme activity in the muscles of various animals, including frogs; in my sample of $n = 159$ frogs the mean comes out 1.53, with an SD of 0.16 (the units are $\mu\text{mole}/\text{min}/\text{g}$ of tissue), and the distribution is close to normal.

- (a) If you chose one of these frogs at random, what (approximately) is the chance that its enzyme activity would be greater than 1.77?
- (b) About how many of the frogs in this sample had enzyme activity less than 1.14?
- (c) About what percentage of these frogs had levels of enzyme activity between 1.45 and 1.61?

4. In each of the following examples, thinking about the situation from the point of view of experimental design, identify the treatment and response variables; specify whether this was a controlled experiment, an observational study, or neither; visualize the raw data (by which I mean, as usual: sketch a data table with rows for subjects and columns for variables, fill in the blank in the sentence "1 row for each _____," and (if you know something about the variable(s) in question) fill in some typical data values); summarize the strength of evidence for the effectiveness of the treatment under study; and describe a stronger experimental design for assessing the clinical efficacy of that treatment.

- (a) In the late 1970s in San Francisco, a woman named Dr. Josephine Lo and her associates treated 31 patients suffering from severe headaches resulting from spinal punctures. Conventional treatments did not help. She found that 30 of the 31 patients experienced “complete and permanent relief” after one to five acupuncture treatments.
- (b) One of the leading causes of death in the United States is coronary artery disease, in which the main arteries to the heart break down. The conventional treatment for this disease as of the late 1960s involved drugs and special diets to reduce the patient’s blood pressure and eliminate fatty deposits in the arteries. Several good studies, involving a few hundred patients, had reported by 1970 that about 68% of patients getting this treatment survived for three years or more.

In 1972, a medical researcher named Daniel Ulyot and his associates introduced a radical new treatment, in which the diseased arteries were replaced by veins transplanted from the patient’s own legs (this procedure is called a *coronary artery bypass graft* (CABG); and hundreds of thousands of CABGs are now done in the U.S. every year). On December 9, 1975, the *San Francisco Chronicle* reported preliminary results of Dr. Ulyot’s research: about 100 patients, specially chosen by him, had been treated in the new way by then, and 98% survived three years or more. The *Chronicle* described these results as “spectacular.”

5. Overheard at a coffee shop:

Person 1: I’m convinced that eating cottage cheese makes people fat.

Person 2: What makes you say that?

Person 1: Have you looked at the people who eat it?

Discuss this argument from the point of view of valid conclusion of cause and effect.

6. The requirement that human subjects must give their informed consent to participate in an experiment can greatly reduce the number of available subjects. In randomized controlled trials of treatments for cancer, for example, patients must agree to be randomly assigned to the standard therapy or to an experimental therapy. The patients who do not wish their treatment to be decided by randomization are dropped from the experiment and given standard therapy. Why is it not correct to keep these patients in the experiment as part of the control group, since the control group also receives the standard therapy? Explain briefly.

7. Zar (1999) raises the question of whether the mean left hindleg length of adult deer is larger than the mean left foreleg length, something which we could try to address by taking a sample of deer at the UCSC campus. Here are two possible experimental designs for answering this question (suppose that I’ve worked out that I can afford to collect data on $n = 80$ deer and that this is enough data to answer the question to the accuracy I need):

- *Design 1:* I sample 80 adult deer and measure the left foreleg on 40 of them at random and the left hindleg on the other 40.

- *Design 2*: I sample 80 adult deer and measure the left foreleg and left hindleg on all 80.

Briefly discuss which, if any, of these designs is valid for arriving at correct conclusions about the comparative lengths of fore and hind legs in deer, and (if any of them are valid) which is likely to be more accurate at estimating the mean difference.

8. Every 10 years or so the U.S. Public Health Service (PHS) studies the effects of smoking on health by taking a large representative sample of American households. They study this relationship separately for men and women in each of a number of age groups (18–24, 25–34, ..., 75–84, 85+). For men and for women in each age group, they consistently find that those who have never smoked are on average somewhat healthier than the current smokers, but the current smokers are on average much healthier than those who recently stopped smoking.

- (a) Is this a controlled experiment, an observational study, or neither? Explain briefly.
- (b) Briefly explain why gender and age are potential confounding factors (PCFs) in this study.
- (c) The lesson seems to be that you shouldn't start smoking, but once you've started, don't stop. Briefly explain why this conclusion is wrong and offer an alternative explanation for the PHS findings.