

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

### AMS 7: Discussion Section 3

1. Back in the mid-1970s the University of Washington initiated an experimental physical fitness program for its students. To evaluate its effects, all those who registered for the program at the beginning of the first year of operation were tested at registration and retested at the end of the year. Physical fitness improved remarkably. As a result, it was decided to run the program for a second year. It was not considered necessary to test new participants at the beginning of the second year, since the results from the first year were available as a benchmark. At the end of the second year, all the participants were tested. The tests showed a marked deterioration in physical fitness by comparison with the scores at the beginning of the first year. Does this mean that the second year of the program was a failure? If so, what changes would you have recommended to the program leaders to revitalize it? If not, what other explanation do you have for the results, and what would you have done to improve the design of the study?

2. According to a study done at Kaiser Permanente in Walnut Creek, CA, and published in the *American Journal of Epidemiology* in 1977, users of oral contraceptives have a higher rate of cervical cancer than non-users, even after adjusting for age, education, and marital status. The investigators who ran the study concluded that taking the pill causes cervical cancer.

- (a) Is this a controlled experiment, an observational study, or neither? Explain briefly.
- (b) Why did the investigators adjust for age? education? marital status? Explain briefly.
- (c) Women using the pill were likely to differ from non-users on another factor which may well affect the risk of cervical cancer. What factor is that? Explain briefly.
- (d) Were the conclusions of the study justified by the data? Explain briefly.

3. People who get lots of vitamins by eating five or more servings of fresh fruit and vegetables each day (especially *cruciferous* vegetables like broccoli) have much lower death rates from colon cancer and lung cancer, according to many observational studies. These studies were so encouraging that two randomized controlled experiments were performed, with the results published in the *New England Journal of Medicine* in 1994: treatment groups were given large doses of vitamin supplements, while people in the control groups just ate their usual diet. One experiment looked at colon cancer, the other at lung cancer.

The first experiment found no difference (large enough to matter in practical or statistical terms) in the death rate from colon cancer between the treatment and control groups. The second experiment found that beta carotene (as a dietary supplement) actually increased the death rate from lung cancer.

- (a) True or false, and explain:
- (i) The experiments confirmed the results of the observational studies.
  - (ii) The observational studies could easily have reached the wrong conclusions, due to confounding — people who eat lots of fruit and vegetables have lifestyles that are different in many other ways too.
  - (iii) The controlled experiments could easily have reached the wrong conclusions, due to confounding — people who eat lots of fruit and vegetables have lifestyles that are different in many other ways too.
- (b) This problem is from a statistics book by Freedman, Pisani and Purves (1998), who evidently believe that the observational studies were wrong. However, as far as colon cancer is concerned, did the controlled experiment address all relevant aspects of the “treatment” of eating five or more servings of fruits and vegetables each day by focusing only on vitamin supplements, or might there have been another aspect of the fruits-and-vegetables treatment that the controlled experiment missed? Explain briefly, and describe what this suggests for a follow-up controlled experiment.

4. A study of young children found that those with more body fat tended to have more “controlling” mothers; the *San Francisco Chronicle* (9 Nov 1994) concluded that “Parents of fat kids should lighten up.”

- (a) Was this an observational study, or a controlled experiment, or neither? Explain briefly.
- (b) Did the study find an association between the behavior of mothers and the body fat levels of their children? Explain briefly.
- (c) If controlling behavior by the mother causes her child to eat more, would that explain an association between controlling behavior by the mother and her child’s level of body fat? Explain briefly.
- (d) Suppose there is a gene which causes obesity. Would that explain the observed association? Explain briefly.
- (e) Can you think of another way to explain the association? Explain briefly.
- (f) Do the data support the *Chronicle*’s advice on child-rearing? Explain briefly.

5. Two easy questions on probability:

- (a) True, false or meaningless: if a fair coin is tossed twice, the chance of getting exactly one head is 50%.

- (b) In the game of craps, somebody rolls a pair of dice, and that person plus a bunch of people standing around watching bet on the sum of the faces pointing up when the dice come to rest. Each die has six faces, which (when the dice are thrown fairly) come up equally often. Two outcomes of particular interest are rolling a three (that is one way to lose, or “crap out”) and rolling a seven, which at some points in the game is one way to win. What is the chance of getting a three on any single roll of the dice? How about a seven?

6. It is now generally accepted that cigarette smoking causes heart disease, lung cancer, and many other diseases. However, in the 1950s, this idea was controversial. Epidemiological studies in the late 1950s established a strong association between smoking and ill-health, but association is of course not the same thing as causation. The statistician and geneticist R. A. Fisher advanced the “constitutional hypothesis”: there is some genetic factor that disposes you both to smoke and to die.

To test Fisher’s idea, epidemiologists used twin studies. They identified sets of smoking-discordant monozygotic twin pairs (“monozygotic” twins come from one egg and have identical genetic makeup; “smoking-discordant” means that one twin in the pair smokes and the other doesn’t). This sets up a race: which twin dies first — the smoker or the non-smoker? Data from a twin study in Finland were as follows:

Cause of Death	Smokers	Non-smokers
Coronary heart disease	9	0
Lung cancer	2	0

In other words, there were 9 twin pairs where at least one twin died of coronary heart disease, and in all 9 the smoker died first (and similarly for lung cancer, which is a rare disease, even among smokers).

Under Fisher’s constitutional hypothesis  $H$ , each twin in any given pair is equally likely to die first, so that the number of pairs in which the smoker dies first is like the number of heads in IID tossing of a fair (50/50) coin.

- (a) Assuming  $H$ , what is the chance of both smokers dying first of lung cancer? Show your work.
- (b) Repeat for the 9 deaths from coronary heart disease.
- (c) On the basis of these data, can the difference in the rate of first dying for smoking and non-smoking twins be explained by
- (i) chance?
  - (ii) genetics?
  - (iii) the adverse health effects of smoking?

Explain briefly.