

1. Observational studies have shown that people who ingest a lot of beta carotene, because their daily diet includes five or more servings of cruciferous vegetables, have lower rates of colon cancer than other people. To test the hypothesis that beta carotene lowers the incidence of colon cancer, a large randomized-controlled experiment was run to test the effect of beta carotene on colon cancer. The treatment group received large doses of vitamins to supplement their usual diet and the control group just ate their usual diet.

The experiment found that there was no difference in the death rate from colon cancer between the treatment and control groups. *True or False*, and explain (briefly):

- (a) (3 pts) The observational studies could *easily* have reached the wrong conclusion due to confounding: people who eat lots of fruits and vegetables have lifestyles that are different in other ways too.

True – Observational studies are susceptible to the effects of confounding variables. In this case, it is quite possible that people who eat a lot of fruits and vegetables take better care of themselves in other ways too, and these other habits explain the lower rates of colon cancer.

- (b) (3 pts) The experiment could *easily* have reached the wrong conclusion due to confounding: people who eat lots of fruits and vegetables have lifestyles that are different in other ways too.

False – The controlled experiment was studying the effects of vitamin supplements, not the consumption of fruits and vegetables. Because of the use of randomized controls, it is very likely that the treatment and control groups were very similar in all respects **except** for the vitamin supplements given to the treatment group. E.g., both groups were very likely to have the same distribution of good vs. bad eating and other habits.

2. California is evaluating a new rehabilitation program for prisoners before their release. The goal of the program is to reduce the *recidivism rate* — the percentage of released prisoners who return to prison within two years of their release. The program is *voluntary* and involves several months of “boot camp” (military style basic training with strict discipline). According to a prison spokesman, “Those who complete boot camp are less likely to return to prison than other inmates.”

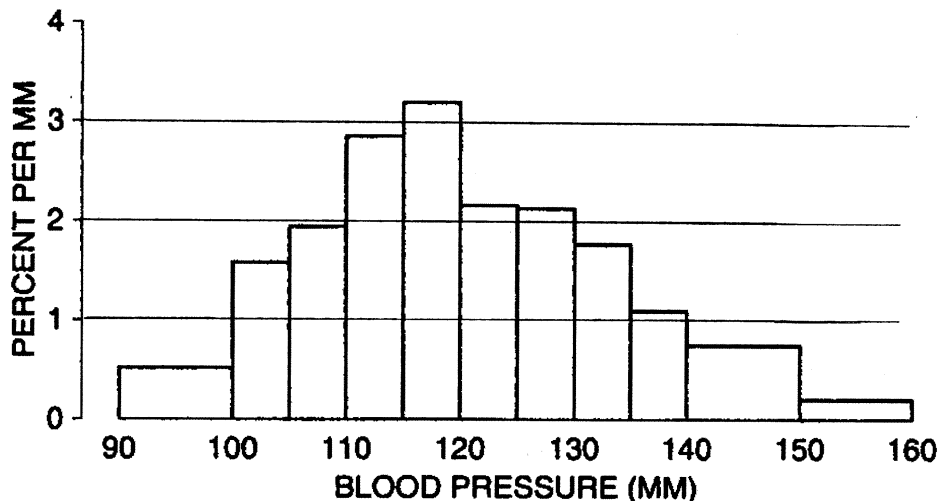
- (a) (3 pts) Is this an observational study or a controlled experiment? Justify your answer.

This is an **observational study** — the program is **voluntary** which means that the subjects in the study choose which group they are in, control or treatment, not the researchers.

- (b) (3 pts) Does the study show that the program is working, or is there another possible explanation for the results? Justify your answer briefly.

This study does **not** conclusively support the claim — there is a possible confounding variable, namely the inmates’ **motivation** not to return to prison. Inmates who volunteer for the boot camp in the first place are likely to be motivated to stay out of prison in the future, and those that complete the program are likely to be the most motivated of all. It is possible that these prisoners would avoid returning to prison on their own, with or without the boot camp.

3. The histogram below shows the distribution of blood pressure for 14,148 women participating in a certain drug study. Use the histogram to answer the following questions. *Explain your answers and show your work.*



- (a) (3 pts) Is the percentage of women with blood pressure above 125mm closer to 36%, 46% or 56%?

Explanation: the percentage of women with blood pressure above 125mm is equal to the area of the histogram between 125 and 160. The heights of the bars from 125 to 140 are (approximately) 2.1, 1.8, and 1.1, all with width 5, and the heights of the last two bars are (approximately) 0.8 and 0.2, both with width 10. So the total area is (approximately)

$$5 \times (2.1 + 1.8 + 1.1) + 10 \times (0.8 + 0.2) = 35,$$

which is closest to 36 .

- (b) (3 pts) In what bin will we find the **median** blood pressure for women in this study?

Answer: The median lies in the bin 115 – 120.

Explanation: The median is the number m such that 50% of the data lies below it (and the other 50% lies above it). About 36% of the data lies above 125 (from (a)) and about $5 \times 2.1 = 10.5\%$ of the data lies between 120 and 125, so about $46.5\% = 36\% + 10.5\%$ of the data lies above 120. This means that the median is less than 120.

Next, $5 \times 3.2 = 16\%$ of the data lies between 115 and 120, so $63\% = 47\% + 16\%$ of the data lies above 115. This means that the median is somewhere between 115 and 120.

- (c) (2 pts) Will the **average** blood pressure for women in this study be greater or smaller than the **median** blood pressure? Justify your answer briefly.

Answer: The mean of this data will be (a little) **greater** than the median because the data are skewed (a bit) to the **right**.