UCSC AMS 5 **Quiz 1 — Solutions** **June 23, 2016**

1. (5 pts) Various studies have found an association between smoking and liver cancer. These studies note that alcohol consumption is a *confounding variable.* This means...

(a) Drinking alcohol causes liver cancer.

(b) People who smoke also drink alcohol, and drinking alcohol causes liver cancer.

Choose one and explain briefly.

The correct choice is (b).

The fact that alcohol consumption causes liver cancer is not enough by itself to make it a confounding variable. The relation between drinking and liver cancer, ***together*** *with the fact that smoking and drinking are associated means that the higher rate of liver cancer among smokers might be explained by their alcohol consumption. This is the confounding effect.*

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) (2 pts) Is this study 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.*

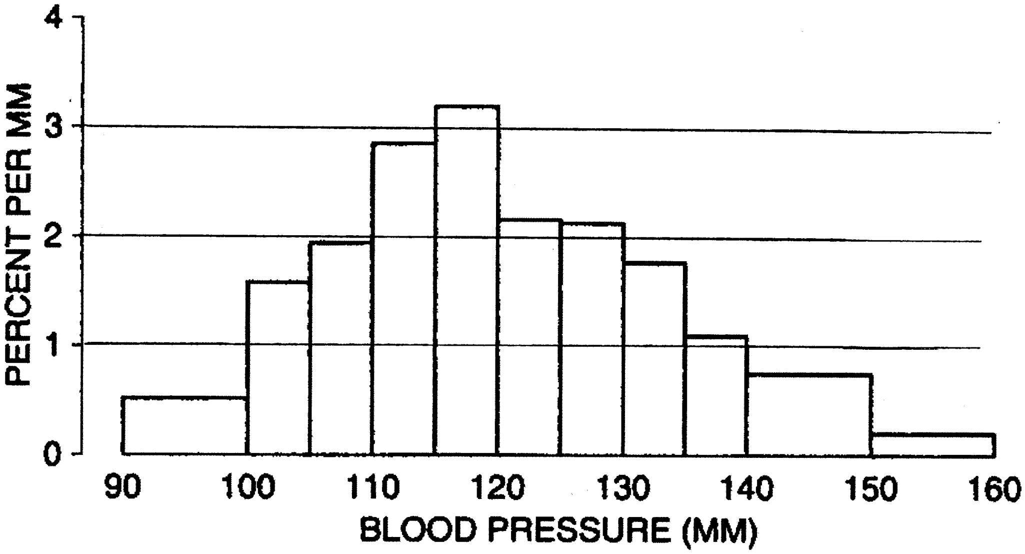
(b) (2 pts) What is the treatment group according to the spokesman's statement? What is the control group?

The treatment group in this study is the group of inmates who volunteered for the boot camp ***and completed the program****. The control group includes the inmates who did not volunteer for the boot camp and the inmates who did not complete the boot camp.*

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

No — 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 120mm closer to $\boxed{47\%}$, 57% or 67%?

Explanation: the percentage of women with blood pressure above 120mm is equal to the area of the histogram between 120 and 160. The heights of the bars from 120 to 140 are (approximately) $2.2, 2.1, 1.8,$ *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 (approximtely)*

$$5 \times (2.2 + 2.1 + 1.8 + 1.1) + 10 \times (0.8 + 0.2) = 46,$$

which is closest to 47 *.*

(b) (3 pts) Is the percentage of women with blood pressure between 105mm and 120mm closer to 36%, $\boxed{41\%}$ or 46%.

Explanation: the percentage of women with blood pressure between 105mm and 120mm is equal to the area of the histogram between 105 and 120. The heights of the bars from 105 to 120 are (approximately) $1.95, 2.9$ *and* 3.2*, all with width* 5*. So the total area in question is (approximately)*

$$5 \times (1.9 + 2.9 + 3.2) = 40$$

which is closest to 41*.*

(c) (2 pts) In what bin will we find the ***median*** blood pressure for women in this study?

Answer: *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 40% of the data lies between* 105 *and* 120 *(from (b)) and about* $10 \times 0.5 + 5 \times 1.6 = 13\%$ *of the data lies between* 90 *and* 105*, so about 53% of the data lies below* 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* $37\% = 53\% - 16\%$ *of the data lies below* 115*. This means that the median is somewhere between* 115 *and* 120*.*