

1. A study done at Kaiser Permanente in Walnut Creek, CA showed that users of oral contraceptives have a higher rate of cervical cancer than nonusers, even after adjusting for age, education and marital status. Researchers concluded that oral contraceptives causes cervical cancer.

(a) (3 pts) Is this a controlled experiment or an observational study? Explain briefly.

***Observational study.** The women decide for themselves if they want to take oral contraceptives (they aren't assigned to the treatment group by the investigators).*

(b) (2 pts) Why did the investigators 'adjust for age, education and marital status'?

To control for confounding factors.

(c) (3 pts) Were the researchers conclusions justified, or did they miss a confounding variable? Explain.

*The conclusions were **not** justified, because they missed one (or two) significant confounding variables that directly affect the incidence of cervical cancer. As we know (from reading Chapter 2 for example), cervical cancer is caused mostly by a sexually transmitted virus – HPV. Women in the study who chose to take oral contraceptives were likely to be more sexually active than women who did not. Moreover, women who were sexually active and **not** taking oral contraceptives were probably more likely to be using condoms during sex, which would reduce the chance of infection with HPV.*

2. The FBI reports state-level and national data on crimes.

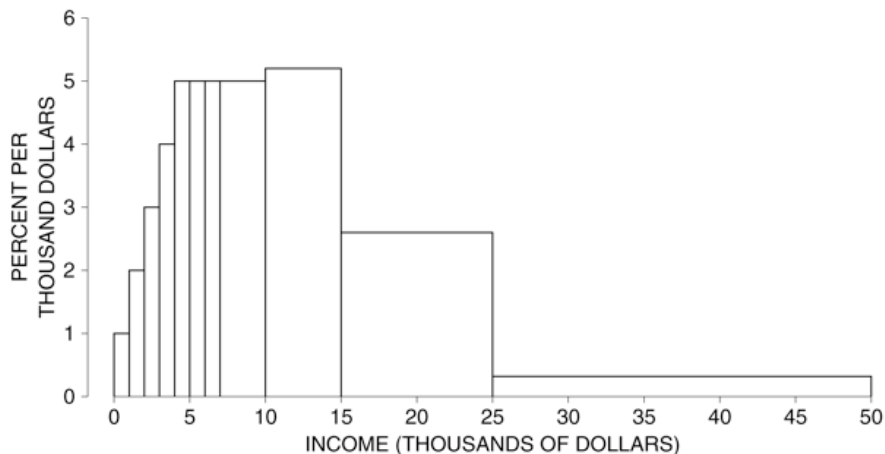
(a) (3 pts) An investigator find that in 2001 there were 3584 crimes in Minnesota, compared to 4082 in Michigan. He concludes that the people of Minnesota are more law-abiding. Do you agree, or is something missing? Explain briefly.

The investigator did not take the relative sizes of the populations of Minnesota and Michigan were not taken into account. If the populations were roughly the same size or if Minnesota had a larger population than Michigan, then the conclusion would be supported by the data. But Michigan is a much more populous state than Minnesota, so the conclusion is not warranted.

(b) (3 pts) Another investigator notes that there were 28,000 crimes committed in the U.S. in 1991, compared to 22,000 crimes committed in 2001. She concludes that the U.S. has become more law-abiding over that decade. Do you agree? Explain.

*At first glance, it may appear that this conclusion is just as unwarranted as the previous one, except that in this case it is reasonable (and correct) to assume that the population of the U.S. was larger in 2001 than in 1991, so in this case there were in fact fewer crimes **and** a bigger population in 2001, so the investigator was right in this instance.*

3. The histogram below gives the distribution household income of 50,000 US households in 1973. Use the histogram to answer the following questions. *Explain your answers and show your work.*



- (a) (3 pts) Is the percentage of households with income below \$7,000 closer to 25%, 35% or 45%?

The percentage of households with income below \$7,000 is equal to the sum of the areas of the bars to the left of 7 in the histogram above, which is (from left to right):

$$1\% + 2\% + 3\% + 4\% + 5\% + 5\% + 5\% = 25\%.$$

- (b) (3 pts) What is the approximate *median* income in this data?

The median income is the point on the horizontal axis below which exactly 50% of the households lie. We already know that 25% of the households have incomes below \$7000, and continuing in the same way we find that the

$$1\% + 2\% + 3\% + 4\% + 5\% + 5\% + 5\% + 15\% = 40\%$$

of the households have incomes below \$10000. Next, approximately

$$(5 \cdot 5.2)\% = 26\%$$

of the households have incomes between \$10000 and \$15000, so the median income is somewhere between \$10000 and \$15000. Moreover, assuming that the incomes are evenly spread between \$10000 and \$15000, it stands to reason that the median income is closer to \$10000 than to \$15000, since we only need 10% more to get from 40% to 50%, so the median income is likely to be approximately \$12000.