

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 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.

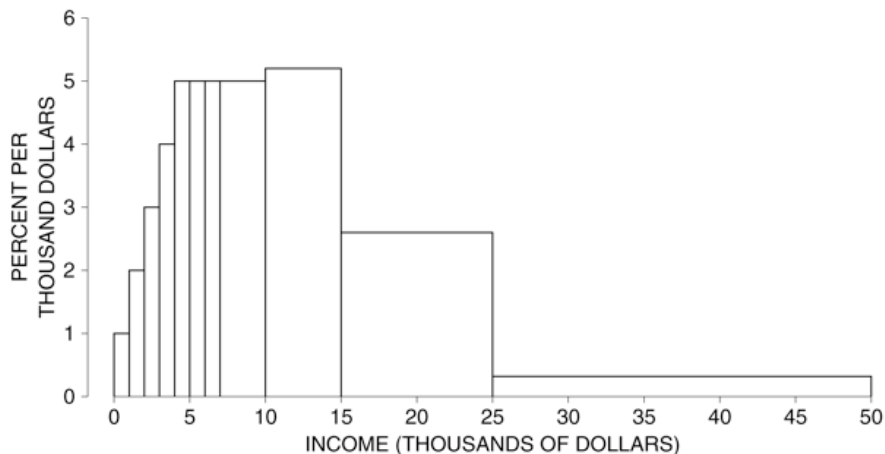
*The raw number of crimes is not the best measure of 'law-abiding'. What we want is the number of crimes relative to the population size (i.e., crimes per capita). On the other hand, in this case it is reasonable (and correct) to assume that the population of the U.S. was larger in 2001 than in 1991, so there were fewer crimes **and** a bigger population in 2001, which means that the number of 'crimes per capita' was lower in 2001, so the investigator is right in this instance.*

(b) (3 pts) Another investigator finds 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.

*Once again, the number we want here is 'crimes per capita', and in this case we can't just assume that one state has a bigger or smaller population than the other – we need the actual numbers. I.e., we can't agree with the investigator based on the given numbers alone.*

*(In fact, Michigan is (and was in 2001) a much more populous state than Minnesota. In 2001, the population of Michigan was about 10 million, while the population of Minnesota was about 5 million, so the per capita crime rate in Minnesota was considerably higher (about 7 crimes per 10000 residents) than in Michigan (about 4 crimes per 10000 residents) in 2001.)*

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 above \$7,000 closer to 35%, 50% or 75%?

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

$$(3 \cdot 5)\% + (5 \cdot 5.2)\% + (10 \cdot 2.5)\% + (25 \cdot (0.4))\% = 76\%.$$

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

*The median income is the point on the horizontal axis above which exactly 50% of the households lie. The last two bars on the right (incomes between \$15000 and \$50000) have a combined area of about 35%, and the bar for incomes between \$10000 and \$15000 has an area of about 26%. Since  $35\% + 26\% = 61\%$ , we know that the median income is somewhere in the range \$10000 to \$15000. But we can be more precise.*

*We only need 15% more to get from 35% to 50%, and 15 is about  $3/5$  of 26, so we need to go about  $3/5$  of the way down from \$15000 to \$10000, which is \$12000 — this is the approximate median.*