Homework - Week 3 (solution)

Questions marked with "(OS3: X.X)" are from the textbook with "X.X" as the exercise number. The answers to the odd questions (odd by book numbering that is) will be in the back of the book.

1. (OS3: 2.17) A 2010 Pew Research poll asked 1,306 Americans "From what you've read and heard, is there solid evidence that the average temperature on earth has been getting warmer over the past few decades, or not?". The table below shows the distribution of responses by party and ideology, where the counts have been replaced with relative frequencies.

		Response			
		Earth is warming	Not warming	Don't Know Refuse	Total
Party and Ideology	Conservative Republican	0.11	0.20	0.02	0.33
	Mod/Lib Republican	0.06	0.06	0.01	0.13
	Mod/Cons Democrat	0.25	0.07	0.02	0.34
	Liberal Democrat	0.18	0.01	0.01	0.20
	Total	0.60	0.34	0.06	1.00

   a. Are believing that the earth is warming and being a liberal Democrat mutually exclusive?

      No.

   b. What is the probability that a randomly chosen respondent believes the earth is warming or is a liberal Democrat?

      Let A be the event a randomly chosen respondent believes the earth is warming and B be the event that a randomly chosen respondent is a liberal Democrat.

      $$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.60 + 0.20 - 0.18 = 0.62$$

   c. What is the probability that a randomly chosen respondent believes the earth is warming given that he is a liberal Democrat?

      $$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.18}{0.20} = 0.9$$

   d. What is the probability that a randomly chosen respondent believes the earth is warming given that he is a conservative Republican?

      Let C be the event a randomly chosen respondent is a conservative Republlican.

      $$P(A \mid C) = \frac{P(A \text{ and } C)}{P(C)} = \frac{0.11}{0.33} = 0.33$$

   e. Does it appear that whether or not a respondent believes the earth is warming is independent of their party and ideology? Explain your reasoning.

      No. Probability of believing the earth is warming is different for members of different party and ideology.

   f. What is the probability that a randomly chosen respondent is a moderate/liberal Republican given that he does not believe that the earth is warming?

      Let D be the event a randomly chosen respondent believes the earth is not warming and E be the event that a randomly chosen respondent is a moderate/liberal Republlican.

      $$P(E \mid D) = \frac{P(E \text{ and } D)}{P(D)} = \frac{0.06}{0.34} = 0.18$$

2. (OS3: 2.18) The Behavioral Risk Factor Surveillance System (BRFSS) is an annual telephone survey designed to identify risk factors in the adult population and report emerging health trends. The following table displays the distribution of health status of respondents to this survey (excellent, very good, good, fair, poor) and whether or not they have health insurance.

		Health Status					
		Excellent	Very good	Good	Fair	Poor	Total
Health Coverage	No	0.0230	0.0364	0.0427	0.0192	0.0050	0.1262
	Yes	0.2099	0.3123	0.2410	0.0817	0.0289	0.8738
	Total	0.2329	0.3486	0.2838	0.1009	0.0338	1.0000

a. Are being in excellent health and having health coverage mutually exclusive?

No.

b. What is the probability that a randomly chosen individual has excellent health?

Let A be the event a randomly chosen individual has excellent health.

$$P(A) = 0.2329$$

c. What is the probability that a randomly chosen individual has excellent health given that he has health coverage?

Let B be the event a randomly chosen individual has health coverage.

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.2099}{0.8738} = 0.2402$$

d. What is the probability that a randomly chosen individual has excellent health given that he doesn't have health coverage?

$$P(A \mid \bar{B}) = \frac{P(A \text{ and } \bar{B})}{P(\bar{B})} = \frac{0.0230}{0.1262} = 0.1823$$

e. Do having excellent health and having health coverage appear to be independent?

No. Probability of excellent health is different depending on whether someone has health coverage.

3. (OS3: 2.23) Swaziland has the highest HIV prevalence in the world: 25.9% of this country's population is infected with HIV. The ELISA test is one of the first and most accurate tests for HIV. For those who carry HIV, the ELISA test is 99.7% accurate. For those who do not carry HIV, the test is 92.6% accurate.

a. If an individual from Swaziland has tested positive, what is the probability that he carries HIV?

Let A be the event an individual from Swaziland has HIV and let B be the event that an individual from Swaziland tested positive.

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \bar{A})P(\bar{A})} = \frac{(0.997)(0.259)}{(0.997)(0.259) + (1 - 0.926)(1 - 0.259)} = 0.825$$

b. (Additional questions, not in book) What is the sensitivity of the ELISA test? What is the specificity?

$$\text{Sensitivity} = P(\text{positive test} \mid \text{has HIV}) = 0.997$$

$$\text{Specificity} = P(\text{negative test} \mid \text{does not have HIV}) = 0.926$$

4. (OS3: 2.26) About 30% of human twins are identical, and the rest are fraternal. Identical twins are necessarily the same sex – half are males and the other half are females. One-quarter of fraternal twins are both male, one-quarter both female, and one-half are mixes: one male, one female. You have

just become a parent of twins and are told they are both girls. Given this information, what is the probability that they are identical?

Let A be the event that twins are identical and let B be the event that twins are both girls.

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \bar{A})P(\bar{A})} = \frac{(0.5)(0.3)}{(0.5)(0.3) + (0.25)(0.7)} = 0.462$$

5. (OS3: 1.17) A large college class has 160 students. All 160 students attend the lectures together, but the students are divided into 4 groups, each of 40 students, for lab sections administered by different teaching assistants. The professor wants to conduct a survey about how satisfied the students are with the course, and he believes that the lab section a student is in might affect the student's overall satisfaction with the course.

   a. What type of study is this?

      Observational, cross-sectional

   b. Suggest a sampling strategy for carrying out this study.

      A stratified sample that selects random students from each lab section would help provide a belanced sample that could help answer the question.

6. (OS3: 1.20) A study that surveyed a random sample of otherwise healthy high school students found that they are more likely to get muscle cramps when they are stressed. The study also noted that students drink more coffee and sleep less when they are stressed.

   a. What type of study is this?

      Observational, cross-sectional

   b. Can this study be used to conclude a causal relationship between increased stress and muscle cramps?

      No. This kind of study can only show associations.

   c. State possible confounding variables that might explain the observed relationship between increased stress and muscle cramps.

7. (OS3: 1.25) Identify the flaw(s) in reasoning in the following scenarios. Explain what the individuals in the study should have done differently if they wanted to make such strong conclusions.

   a. Students at an elementary school are given a questionnaire that they are asked to return after their parents have completed it. One of the questions asked is, "Do you find that your work schedule makes it difficult for you to spend time with your kids after school?" Of the parents who replied, 85% said "no". Based on these results, the school officials conclude that a great majority of the parents have no difficulty spending time with their kids after school.

   b. A survey is conducted on a simple random sample of 1,000 women who recently gave birth, asking them about whether or not they smoked during pregnancy. A follow-up survey asking if the children have respiratory problems is conducted 3 years later, however, only 567 of these women are reached at the same address. The researcher reports that these 567 women are representative of all mothers.

   c. An orthopedist administers a questionnaire to 30 of his patients who do not have any joint problems and finds that 20 of them regularly go running. He concludes that running decreases the risk of joint problems.

8. (OS3: 1.26) A city council has requested a household survey be conducted in a suburban area of their city. The area is broken into many distinct and unique neighborhoods, some including large homes, some with only apartments, and others a diverse mixture of housing structures. Identify the sampling

methods described below, and comment on whether or not you think they would be effective in this setting.

a. Randomly sample 50 households from the city.

Simple random sample

This would be a good sample for this study.

b. Divide the city into neighborhoods, and sample 20 households from each neighborhood.

Stratified sample

While this method would likely represent the diversity of the city, it would notnecessarily be a representative sample, i.e. a sample which match the demogra[hic make up of the city. For example, the neighborhoods with many one-family homes likely have far fewer households than neighborhoods that are predominately apartments. Thus, 20 households represents a larger portion of the former than the latter.

c. Divide the city into neighborhoods, randomly sample 10 neighborhoods, and sample all households from those neighborhoods.

Cluster sample

This sample might be problematic. While each neighborhood has an equal chance of being selected, in the final sample their might be important demographic groups that are completely unrepresented.

d. Divide the city into neighborhoods, randomly sample 10 neighborhoods, and then randomly sample 20 households from those neighborhoods.

Mixed (cluster, then random)

Same criticisms as above.

e. Sample the 200 households closest to the city council offices.

Convenience sample

This is the worst sample of these scenarios. It is likely that the 200 households closest to the city council offices have many traits in common that are not shared with many other households in the city. This sample would lead to very biased results.