STAT22000 Summer 2020 Homework 4 Solutions

Problems to Turn In: due midnight of **Friday, July 3, on Canvas**.

1. The table shows results of whether the death penalty was imposed in murder trials in Florida between 1976 and 1987. For instance, the death penalty was given in 53 out of 467 cases in which a white defendant had a white victim[1].

Defendant's Race	Victim's Race	Death Penalty		
		Yes	No	Total
White	White	53	414	467
	Black	0	16	16
Black	White	11	37	48
	Black	4	139	143

   (a) First, consider only the cases with white victims. What proportion of the white defendants received the death penalty? What proportion of the black defendants received the death penalty? Was a white or a black defendant more likely to receive death penalty when the victim was white?

   (b) Repeat part (a) for cases with black victims. Was a white or a black defendant more likely to receive death penalty when the victim was black?

   (c) Now add up the two tables together to get a summary contingency table that describes the association between the death penalty verdict and defendant's race, ignoring the information about the victim's race. What proportion of white defendants received the death penalty? What proportion of black defendants received the death penalty? Which proportion is higher?

   (d) How can you explain the association in part (c), whereby white defendants were more like to receive death penalty? How can this association be so different from the ones you found in part (a) and (b)? *Hint: Your explanation must address the following.*

      i. When the defendant was white, was the victim more likely to be black or white?

      ii. When the defendant was black, was the victim more likely to be black or white?

      iii. Was the death penalty more likely to be given when the victim was white or black, or no apparent difference?

      Use the three observations above to explain how the association in (c) is so different from in the one in part (a) and (b).

Answer:

(a) [*1pt*] When the victim was white, black defendants were more likely to receive death penalty than white defendants ($11/48 \approx 22.9\%$ v.s.$53/467 \approx 11.3\%$).

(b) [*1pt*] When the victim was black, ditto ($4/143 \approx 2.8\%$ for black defendants and $0/16 = 0\%$ for white defendants).

(c) [*2pts*]

	Death Penalty			
	Yes	No	Total	% Death Penalty
White defendants	$53 + 0 = 53$	$414 + 16 = 430$	483	$53/483 \approx 11.0\%$
Black defendants	$11 + 4 = 15$	$37 + 139 = 176$	191	$15/191 \approx 7.9\%$

[1]Source: Originally published in *Florida Law Review*. Michael Radelet and Glenn L. Pierce, Choosing Those Who Will Die: Race and the Death Penalty in Florida, vol. 43, *Florida Law Review* 1 (1991).

When the information about the victim's race is ignored, it looks like white defendants were more likely (11.0%) to get a death penalty than black defendants (7.9%).

(d) [4pts] This is an example of Simpson's paradox. From the data, we observe that:

- Florida juries were much more likely to give the death penalty in cases involving white victims (11.3% for white defedants and 22.9% for black defedants) than black victims (0% for white defedants and 2.8% for black defedants).
- Both races are much more likely to be involved in murders in which the victim is the same race as the defendant.

As whites were more likely to kill whites, and the victims being white making the white defendants more likely to get death penalty, and hence death penalty rate appears higher for white defendants. Thus, the overall racial bias against whites could be due to the confounding factor of the victims's race.

[Give 2pts only if the student merely mentioned "Simpson's paradox" without further explanation. To receive full marks (4pts), students must point out the association between victim's race and whether juries gave the death penalty, as well as the association between victims' race and the defendant's race, and use them to explain whey the association in (c) is different from the association observed in (a) and (b).]

---

2. In each of the following situations, identify the sampling method as one of the following: simple random sampling, stratified sampling, multistage sampling, or voluntary response sampling.

   (a) There are seven sections of an introductory statistics course. A random sample of three sections is chosen and then random samples of 8 students from each of these sections are chosen.

   (b) An online poll asks people who visit this site to choose their favorite television show.

   (c) Separate random samples of male and female first-year college students in an introductory psychology are selected to receive a one-week alternate instructional method.

Answer:

(a) [2pts] Multistage sampling. We first sample 3 of the 7 course sections, then 8 from each chosen section.

(b) [2pts] Voluntary response sampling. People who visit the site can choose whether to participate.

(c) [2pts] Stratified random sampling: Males and females (the strata) are sampled separately.

---

3. A survey is carried out by the finance department to determine the distribution of household size in a certain city. They draw a simple random sample of 1,000 households. After several visits, the interviewers find people at home in only 653 of the sample households. Rather than face such a low response rate, the department draws a second batch of households, and uses the first 347 completed interviews in the second batch to bring the sample up to its planned strength of 1,000 households. The department counts 3,087 people in these 1,000 households, and estimates the average household size in the city to be about 3.1 persons. Is this estimate likely to be too low, too high, or about right? Why?

Answer: [3 points] This estimate is likely to be too high. The smaller the size of the household, the less likely the interviewer is to find a person at home. It is likely that the households with fewer residents were under-counted in the survey. This results in an estimate that is too high.

4. Suppose you are on the staff of a member of Congress who is considering a bill that would provide government-sponsored insurance for nursing-home care. You report that 1128 letters have been received on the issue, of which 871 oppose the legislation. "I'm surprised that most of my constituents oppose the bill. I thought it would be quite popular," says the congresswoman. Are you convinced that a majority of the voters oppose the bill? How would you explain the statistical issue to the congresswoman?

Answer: *[3 points]* The sample constituted by 1128 letters is a **voluntary response sample** since people chose by themselves whether to send the letters. Such a sample is not a proper representation of all the constituents, because those who oppose the bill would have much stronger motivation to send letters to the the congresswoman than those who support or are neutral about the bill. The actual percentage of the constituents that oppose the bill is probably much lower than $871/1128 \approx 77\%$.

---

5. Carol Dweck is a noted psychologist from Stanford who believes that we should praise students' effort and not their intelligence. In her seminal study (2006), she gave students a test and then randomly divided the students into two groups. She praised the effort of one group and praised the intelligence of the other group. For their next exercise, the groups were given a choice of a challenging task or an easy task. Of those who were praised for their effort, 90% chose the challenging task, whereas fewer than half of the students who were praised for their intelligence chose the challenging task, fearful of losing their smart status.

   (a) What are the treatments in this study? What is the response variable?

   (b) Can we conclude from the study that praising the effort rather than the intelligence prompted the students chose the challenging task?

   (c) Were the subjects of study randomly sampled from some population at large? Can the results of the study be generalized to the population at large?

Answer:

(a) *[2pts = 1pt + 1pt]*

(1pt) The treatments are whether the instructor praised the effort or praised the intelligence.

(1pt) The response is whether the student chose the challenging task.

(b) *[1pt]* Yes, we can make s cause-and-effect conclusion that praising the effort rather than the intelligence prompted the students chose the challenging task since the students were randomly assigned to the praising-the-effort group or the praising-the-intelligence group.

(c) *[2pts = 1pt + 1pt]*

(1pt) The subjects were all students in Stanford. They are not randomly sampled from some population at large (students in all colleges or general people)

(1pt) As the subjects were not randomly sampled from some population at large, we were not sure if similar results could be observed if the subjects were students in other universities or general people.
