PROJECT 2 SOLUTIONS

Statistics 401: Fall 2016

1. Regarding *Sharp Racial Divisions in Reactions to Brown, Garner Decisions* reported by the LA Times, the Washington Post, and CNN in November 2014:

   (a) (1 pt) The articles report on an observational study since no treatment was applied to the people who were interviewed by phone. Although the interviewees were asked about events that occurred in the past (i.e., retrospectively), research questions were formed (regarding the American public's view of the grand jury decisions) and data were collected prospectively to address the research questions. So this was a **prospective observational study** of the public's view of the grand jury decisions regarding Garner and Brown.

   (b) (1 pt) The population of interest is the US population in November of 2014.

   (c) (2 pts) The "landline and cell phone random digit dial samples" used for this study are potentially fraught with selection and non-response bias. **Selection bias** is possible since US citizens with neither landline nor cell phones were not considered for the sample. **Non-response bias** is possible since respondents may have not answered their phone or hung up when contacted by the interviewers.

   (d) (2 pts) Table 1 shows the explanatory variables and Table 2 shows the response variables considered in the study. Although Table 1 reports age as a discrete numerical variable, the article summarizes ages into 4 categories: 18-29, 30-39, 40-64, 65+.

Table 1: Explanatory variables in the Pew Research Study of the American public's opinions regarding the grand jury decisions in Garner and Brown.

Variable	Type	Properties
Race	Categorical	3 levels: Hispanic, black, white
Political party	Categorical	3 levels: Democrat, Independent, Republican
Age	Numerical	discrete: 18, 19, 20, ...

Table 2: Response variables in the Pew Research Study of the American public's opinions regarding the grand jury decisions in Garner and Brown. DK = "don't know"

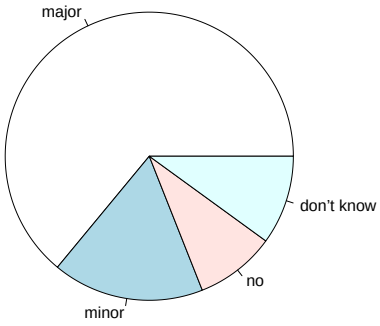
Variable	Type	Properties
Jury decision re: Michael Brown case	Categorical	3 levels: Right, wrong, DK
Jury decision re: Eric Garner case	Categorical	3 levels: Right, wrong, DK
Race was a factor in Michael Brown case	Categorical	4 levels: major, minor, none, DK
Race was a factor in Eric Garner case	Categorical	4 levels: major, minor, none, DK

   (e) (2 pts) Because the sample was assured to include individuals from all 50 states and DC, the sample design was likely a stratified random sample. In this case, there would be 51 strata (one for each state and one for DC) and then a SRS would have been taken from each stratum.

   (f) (1 pt) The following R code was used to generate the pie chart in Figure 1 that describes African American opinions re: the grand jury decision in the case involving Michael Brown.

```
pie(c(64,17,9,10), labels=c("major","minor","no","don't know"))
```

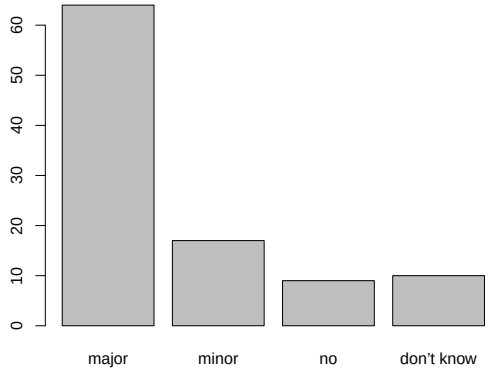
Figure 1: Pie Chart depicting the percent of interviewees who say race was (a) _____ factor in the grand jury decision in the case of Michael Brown



(g) (1 pt) The following R code was used to generate the barplot in Figure 2. Like Figure 1, Figure 2 is just a different way describe African American opinions re: the grand jury decision in the case involving Michael Brown.

```
barplot(c(64,17,9,10), names=c("major","minor","no","don't know"))
```

Figure 2: Bar plot depicting the percent of interviewees who say race was (a) _____ factor in the grand jury decision in the case of Michael Brown



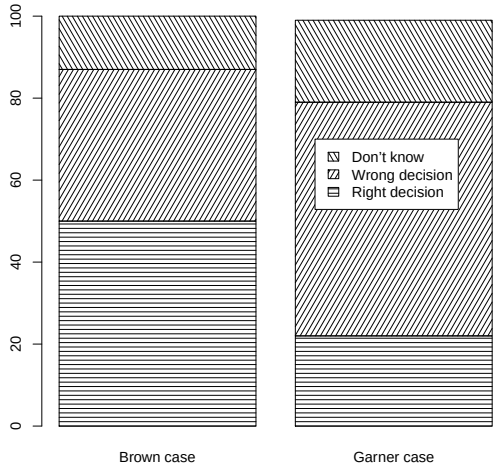
(h) (1 pt) The following R code was used to generate the side-by-side segmented barplot in Figure 3.

```
> percent = matrix(c(50,37,13,22,57,20),ncol = 3,byrow=T)
> rownames(percent) = c("Brown case","Garner case")
> colnames(percent) = c("Right decision","Wrong decision","Don't know")
> percent=t(percent)
> percent
              Brown case Garner case
Right decision         50          22
Wrong decision         37          57
Don't know             13          20

> ang = c(0,60,120)
> index=c(3,2,1)    # change the order of the labels in the legend pane
```

```
> barplot(percent,angle=ang,density=20,col="black")
> legend(1.5,70,legend=rownames(percent)[index],angle=ang[index],
        density=20,bg="white")
```

Figure 3: Side-by-side segmented bar plot comparing the percent of <u>all</u> interviewees who agreed or disagreed with the grand jury decision in the cases of Michael Brown and Eric Garner



2. Regarding the article *In praise of gratitude* reported by Harvard Medical School in November of 2011:

   (a) (1 pt) An experiment was conducted because Emmons and McCullough applied treatments to 3 separate groups of individuals.

   (b) (2 pts) A little research on the internet shows that Emmons and McCullough applied the experiment to university students enrolled in class. This is an example of a **convenience sample**. This sample could be fraught with selection and response bias. **Selection bias** is possible because the sample does not include humans outide of those who took the university class. **Response bias** is possible because if the students knew what their teachers were studying, they may have adapted their answers on happiness to not disappoint their teachers.

   (c) (1 pt) There were 3 treatment groups: one group wrote journal entries that focused on gratitude; a second group wrote wrote journal entries that focused on daily irritations or things that had displeased them; and the last group wrote journal entries with no explicit recommendation on the focus of the entries.

   (d) (2 pts) The third group was the control group. Their writing assignment could be presumed to be a placebo because simply writing is not expected to have any inherent benefit, and because simply writing resembled as much as possible the "real" treatments applied to the other two groups.

   (e) (1 pt) It was important to include the control group to assess the effect of writing each week (regardless of the topic), and also to assess the effect of simply being in the study.

   (f) (1 pt) Random assignment of the students to each of the three treatment groups helps make sure that the values of extraneous factors do not occur more often in one group. Thus, we can make cause-and-effect conclusions regarding the population

from which the sample was drawn. In other words, we may be able to make cause-and-effect conclusions regarding university students who take classes from Emmons and McCullough.

3. Designing a gratitude study on Bozeman's K-12 children:

   (a) (2 pts) Because the article identified that the emotional maturity of a child may interfere with the positive effects of gratitude, we might conclude that children of different ages will respond differently to the 3 treatments of Emmon and McCullough. Thus, we might first stratify Bozeman students by grade level; then cluster students within each grade by classes. Then take a SRS of 1 or 2 classes in each grade level. All of the students in the selected classes will be included in the experiment.

   (b) (2 pts) The article identified the emotional maturity of a child as an extraneous factor that might confound with the treatment. To control the effect of age, we can block by grade level and randomly assign students in each block to each of the 3 treatments.

   (c) (2 pts) Extraneous variables that can be directly controlled (i.e., held constant for all treatment groups) in this experiment are: (i) the time of day and (ii) the day of the week when the treatment (i.e., the writing assignment) is administered; (iii) the location where the treatment is administered (e.g., the writing assignment must be performed at school); and (iv) the timing of curricular events, like exams, that might affect the students' happiness.

   (d) (1 pt) Replication typically refers to the number of individuals (i.e., in this case students) who participate in a study. Other "levels of replication" in this study are the number of classes who participate, and also the number of separate experiments that were conducted.

   (e) (2 pts) The kids in this K-12 study will be told exactly which writing assignment, and hence exactly which treatment, they are receiving. Hence, the students can not be blinded. If the teacher were to be blinded, then someone else would have to come into the class to administer the writing assignments. If the teacher is not blinded, perhaps the teacher would coach students to write a certain way, or would unconsciously interact with students differently depending on which treatment group the students belonged to.

4. (1/2 pt) If one is interested in making an inference from a sample to a population, then random sampling is most important.

5. (1/2 pt) If one wishes to make a causal inference then random assignment is most important.

6. (1 pt) A cluster random sample is more efficient than a simple random sample when the population is "naturally" divided into roughly equally sized groups, each of which is representative of the total population.

7. (1 pt) A systematic random sample is more efficient than a simple random sample if the individuals in the population can be listed and the list is not related to the response variable.