STAT22000 Summer 2020 Homework 3

All page, section, and exercise numbers below refer to the course text (*OpenIntro Statistics*, 3rd edition, by Diez, Barr, and Cetinkaya-Rundel.).

Reading: Section 1.7, 1.1, 1.3, 1.4, 1.5 (in this order)
Problems for Self-Study : (Do Not Turn In)

- Exercise 1.9, 1.11, 1.13, 1.15, 1.21, 1.23, 1.25, 1.27, 1.37 on p.58-64
- Answers can be found at the end of the book (p.405-407).
- Self-study problems are as important as Turn-In problems. We don't require submission because we think you can learn from those problems by doing them yourself and checking the answers, without grading feedbacks. If having questions about those problems, you are welcome to ask the instructor or CAs.

Problems to Turn In: due midnight of **Wednesday, July 1, on Gradescope**.

1. The North Carolina births data is a random sample of 1000 birth records in the state of North Carolina in 2004. The description of variables can be viewed at

   ~~https://www.openintro.org/stat/data.php?data=ncbirths~~

   http://www.stat.uchicago.edu/~yibi/s220/labs/lab08.html

   We are interested in whether unmarried mothers were more likely to have low-birth-weight babies (under 2500 grams) than married mothers. The variables are `marital` with two levels `married` and `not married`, and `lowbirthweight` with two levels `low` and `not low`.

   ```
   nc = read.table("https://www.openintro.org/stat/data/csv/ncbirths.csv",sep=",", header=T)
   ```

   There is a birth record in the data that the `marital` status of the mother is missing. Let's exclude that record and examine the remaining 999 records.

   ```
   nc = subset(nc, is.na(marital) == FALSE)
   ```

   (a) Obtain the two-way contingency table between `marital` and `lowbirthweight`. Please ignore the observation with a missing marital status (with a value of `NA`) here and in all the parts below.

   (b) What proportion of babies were born out of marriage? What proportion of babies were born with a low birth weight? Show your calculation.

   (c) Among the babies born to unmarried moms, what proportion had a low birth weight. Show your calculation.

   (d) Make a segmented bar chart that displays the two-way table in part (a). Use the bars to represent the variable `marital` and the segments of the bars to represent the levels of `lowbirthweight`.

   (e) Make a mosaic plot that compares the percentage of babies born with a low birth weight between those born to married women and those born to unmarried women.

   (f) Based on the plot in the previous part, were babies born within a marriage or those born outside of marriage more likely to have a low birth weight?

   (g) Are the two variables `marital` and `lowbirthweight` independent?

2. (North Carolina births data continued) It is known that smoking during pregnancy can cause low-birth-weight. Whether the mother is a smoker may confound the relationship between mothers' marital status and whether the babies have low birth weights.

(a) The variable `habit` in the NC births data indicates whether the mother was a `smoker` or `nonsmoker` during pregnancy. Obtain the two-way contingency table between `marital` and `habit`. What percentage of married mothers smoked during pregnancy? How about unmarried mothers? Were married or unmarried mothers more likely to smoke during pregnancy?

(b) Let's examine the relation of `marital` and `lowbirthweight` after adjusting for the mother's smoking habit.

```
tab = tally(~marital + lowbirthweight + habit, data=nc)
ftable(tab, col.vars = c("habit","marital"))
prop.table(ftable(tab, col.vars = c("habit","marital")), 2)
```

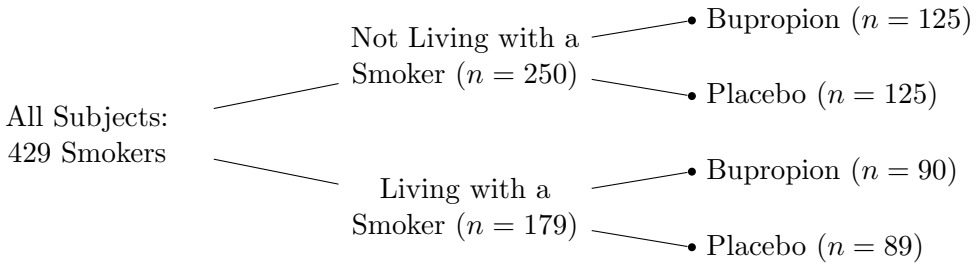
i. Consider babies with NONSMOKING mothers only, were those born to married mother or those born to unmarried mother more likely to have low birth weights?

ii. Consider babies with SMOKING mothers only, were those born to married mother or those born to unmarried mother more likely to have low birth weights?

iii. Is out-of-marriage birth associated with low birth weights after adjusting for mothers' smoking habit?

(c) Can you claim that marriage reduces the chance of births having low birth weights? Why or why not?

3. Smokers may more difficult to quit smoking if they live with another smoker. A study comparing bupropion (an antidepressant and smoking cessation aid) with a placebo tried to take this into account in their design. The researchers first split the subjects based on whether they lived with another smoker. The subjects who live with another smoker were randomly assigned to take bupropion or a placebo, and those who didn't live with smokers were also randomly assigned to take bupropion or a placebo. The figure below shows a flow chart of the design, when 250 of the 429 study subjects lived with nonsmokers and 179 lived with another smoker.



The percentages of subjects who relapsed were then compared among the four groups.

(a) If the two groups of subjects that received bupropion had substantially lower relapse rates than the two placebo groups, can we claim that bupropion is effective as a smoking cessation aid?

(b) Has blocking been used in this study? If so, identify the blocks.

(c) If the placebo group of subjects who lived with another smoker had a higher relapse rate than the placebo group of subjects who did not live with smokers, can we claim that living with smoker(s) makes it harder to quit smoking? Why or why not?