

# STAT22000 Summer 2020 Homework 13

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

**Reading:** Section 6.1, 6.2

**Problems for Self-Study** (Do Not Turn In): Exercise 6.1, 6.5, 6.9, 6.15 6.23, 6.25, 6.29 on p.312-319. The answers can be found at the end of the book.

**Problems to Turn In:** due **5 pm on Thursday, July 23, on Gradescope.**

1. Refer to Exercise 6.8 on p.313 in the textbook. Instead of doing the two parts therein, determine if the following statements are true or false, and explain your reasoning.
  - (a) We are 95% confident that 63% to 69% of American adults in this sample think licensed drivers should be required to retake their road test once they turn 65.
  - (b) We are 95% confident that 63% to 69% of American adults think licensed drivers should be required to retake their road test once they turn 65.
  - (c) If we take many random samples of 1018 American adults, and for each sample, calculated the percentage who think licensed drivers should be required to retake their road test once they turn 65. 95% of those sample percentages will be between 63% and 69%.
  - (d) The margin of error at a 99% confidence level would be higher than 3%.
2. This problem is about the data set in Lab 8:

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

which is a random sample of 1000 birth records in the state of North Carolina. This time we are interested in the percentage of babies that had low birth weight (below 2500 gram or 5 pounds 8 ounces).

- (a) Find the number of babies in the 1000 birth records that had low birth weights using the R commands below.

```
nc = read.csv("https://www.openintro.org/stat/data/csv/ncbirths.csv")
library(mosaic)
tally(~lowbirthweight, data=nc)
```

Estimate the percentage of babies in North Carolina that had low birth weight and calculate a 95% confidence interval for it.

- (b) If we want to reduce the margin of error for the 95% confidence level to 1.5%, how large the size of a sample would you recommend?
  - (c) According to CDC (<https://www.cdc.gov/nchs/fastats/birthweight.htm>), 8.17% of the babies born in 2016 in the U.S. had a low birth weight. Test whether the state of North Carolina had a *higher* percentage of babies having low birth weight than the nationwide percentage 8.17%. Please specify the null and alternative hypotheses, report the test statistic and the  $P$ -value, and make a conclusion using 0.01 significance level.
3. This problem is a continuation of Problem 2. This time we are interested in testing whether babies born to smoking mothers had a percentage of having low birth weight (below 2500 gram or 5 pounds 8 ounces) than those born to nonsmoking mothers
    - (a) Make a two-way table cross classify the 1000 babies by whether they had low birth weights and whether the mother smoked using the R commands below.

```
nc = read.csv("https://www.openintro.org/stat/data/csv/ncbirths.csv")
library(mosaic)
tally(~habit+lowbirthweight, data=nc)
```

- (b) Test if babies born to smoking mothers had a higher percentage of having low birth weights than those born to non-smoking mothers. State the hypotheses. Report the test statistic and the  $P$ -value. What is your conclusion at significance level  $\alpha = 0.05$ ? (Please ignore the observation that the mother's smoking habit is missing.)
4. (Revision of Exercise 6.26) A 2010 Pew Research foundation poll indicates that among 1099 college graduates, 33% watch The Daily Show. Meanwhile, 22% of the 1110 people with a high school degree but no college degree in the poll watch The Daily Show. A 99% confidence interval for  $p_{\text{college grad}} - p_{\text{HS}}$ , where  $p$  is the proportion of those who watch The Daily Show, is (0.06, 0.16).
- (a) Verify that the 99% CI for  $p_{\text{college grad}} - p_{\text{HS}}$  is about (0.06, 0.16). Show your calculation.
- (b) Determine if the following statements are true or false, and explain your reasoning if you identify the statement as false.
- There was strong evidence that  $p_{\text{college grad}} > p_{\text{HS}}$ .
  - 99% of random samples of 1,099 college graduates and 1,110 people with a high school degree but no college degree will yield differences in sample proportions between 6% and 16%.
  - A 95% confidence interval for  $p_{\text{college grad}} - p_{\text{HS}}$  would be wider.
  - A 99% confidence interval for  $p_{\text{HS}} - p_{\text{college grad}}$  is  $(-0.16, -0.06)$ .