Homework 3

Your Name

April 8, 2019 at 11:59pm

Homework Guidelines:

Please read Homework Guidelines. You must follow these guidelines.

Turning the Homework in:

Please turn the homework in through canvas. You may use a pdf, html or word doc file to turn the assignment in.

PHP 1511 Assignment Link

PHP 2511 Assignment Link

For the R Markdown Version of this assignment: HW3.Rmd

Part 1: Both PHP 1511 and 2511

Health disparities are very real and exist across individuals and populations. Before developing methods of remedying these disparities we need to be able to identify where there are disparities.In this homework we will consider a study by (Asch & Armstrong, 2007). This paper considers 222 patients with localized prostate cancer. The table below partitions patients by race, hospital and whether or not the patient received a prostatectomy.

Undergoing Prostatectomy

Race

Yes

No

University Hospital

White

54

37

Black

7

5

VA Hospital

White

11

29

Black

22

57

You can load this data into R with the code below:

```
phil_disp <- read.table("https://drive.google.com/uc?export=download&id=0B8CsRLdwqzbzOXlIRl9VcjNJRFU", 
```

This dataset contains the following variables

Variable	Description
hospital	0 - University Hospital
	1 - VA Hospital
race	0 - White
	1 - Black
surgery	0 - No prostatectomy
	1 - Had Prostatectomy

1. First, use logistic regression to obtain a crude estimate (i.e. collapsed over hospital) of the relationship of Black vs White men in Philadelphia with risk of receiving a prostatectomy. You can use the odds ratio for this purpose. Report the odds ratio, a 95% confidence interval and provide a brief interpretation.

2. Second, use logistic regression to obtain a crude estimate (i.e. collapsed over race) of the relationship of VA hospital vs University Hospital in Philadelphia with risk of receiving a prostatectomy. You can use the odds ratio for this purpose. Report the odds ratio, a 95% confidence interval and provide a brief interpretation.

3. Thirdly, use logistic regression to obtain an estimate of the relative odds of prostatectomy by race adjusted for hospital. Report the relative odds ratio, a 95% confidence interval and provide a brief interpretation for race.

4. How did the odds ratio change between questions 1 and 3? (**Hint:** *Simpson's Paradox*)

5. Why is there such a change in the odds ratio between 1 and 3?

6. Consider adding an interaction between race and hospital into the model. Perform an appropriate model comparison test and choose the appropriate model.

7. Perform a Hosmer-Lemeshow Test for this this. Describe what this is testing as well as stating the results.

8. Test for discrimination in the Logistic model.

9. What does this model tell you about disparities within this study?

10. What challenges does this study provide in trying to identify health disparities?

Part 2: PHP 2511 Only

11. What is the purpose of the link function in a GLM?

12. What is the basic reasoning behind using the logit link for logistic regression?

13. What are the assumptions of a GLM?