

Problem Set 3

Elana Nelson

April 8, 2019

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.

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.

term	estimate	p.value	conf.low	conf.high
(Intercept)	0.9848485	0.9303778	0.6985457	1.3880778
race	0.4749380	0.0089549	0.2694239	0.8250258

By this model, we see that a unit increase in race (meaning from white to black) results in a -52.5062035% decrease in odds of receiving a prostatectomy.

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.

term	estimate	p.value	conf.low	conf.high
(Intercept)	1.4523810	0.0627011	0.9838382	2.1646297
hospital	0.2642013	0.0000034	0.1492365	0.4598822

By this model, we see that a unit increase in hospital (meaning from University to VA) results in a -73.5798704% decrease in odds of receiving a prostatectomy.

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.

term	estimate	p.value	conf.low	conf.high
(Intercept)	1.4526892	0.0681971	0.9758192	2.1830747
race	0.9981802	0.9959192	0.5006556	2.0381436
hospital	0.2644648	0.0001241	0.1313651	0.5145046

We now see that not only is the odds ratio very close to 1 (suggesting no change by race), but it is also insignificant (again suggesting no change by race). This implies that race is not very important in explaining the odds of receiving a prostatectomy. In fact, it is the hospital that is chosen that has a statistically significant effect on the outcome. The confidence interval for hospital has also not increased by very much, and the estimate has remained the same, suggesting this is a good estimate for the effects of hospital choice.

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

When viewed individually, each subgroup (race and hospital) was associated with a significant odds ratio, both showing a significant % decrease in odds of receiving a prostatectomy with one unit increase in category. However, when viewed together, race was left insignificant, and hospital became the most important variable on the outcome. This occurrence is similar to Simpson's Paradox, in which, when viewed all together, one subgroup dominates the other subgroups.

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

There is such a change because when viewed separately, both race and hospital choice have a significant effect on the outcome, but when viewed together, hospital choice has a more overwhelming effect on the outcome.

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

term	estimate	p.value	conf.low	conf.high
(Intercept)	1.4594595	0.0764751	0.9643615	2.2330756
race	0.9592593	0.9467884	0.2844572	3.4566718
hospital	0.2598978	0.0011176	0.1117342	0.5711112
race:hospital	1.0607600	0.9380930	0.2313888	4.7009517

For understanding the different components in the interaction model: (Intercept) -> White men in University Hospital race -> Black men in University Hospital hospital -> White men in VA Hospital race:hospital -> Black men in VA Hospital

Because the interaction term is insignificant, we should just use the smaller trend model. However we can test this with a Likelihood Ratio Test.

Resid..Df	Resid..Dev	df	Deviance	p.value
219	279.7774	NA	NA	NA
218	279.7714	1	0.006027	0.9381194

The p-value of this test suggests that we should favor the null hypothesis, meaning that the larger model with the interaction term is no better than the smaller model. Thus, we should go with the smaller adjusted

model.

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

Hosmer and Lemeshow goodness of fit (GOF) test

data: mod3\$y, fitted(mod3) X-squared = 5.8668e-23, df = 1, p-value = 1

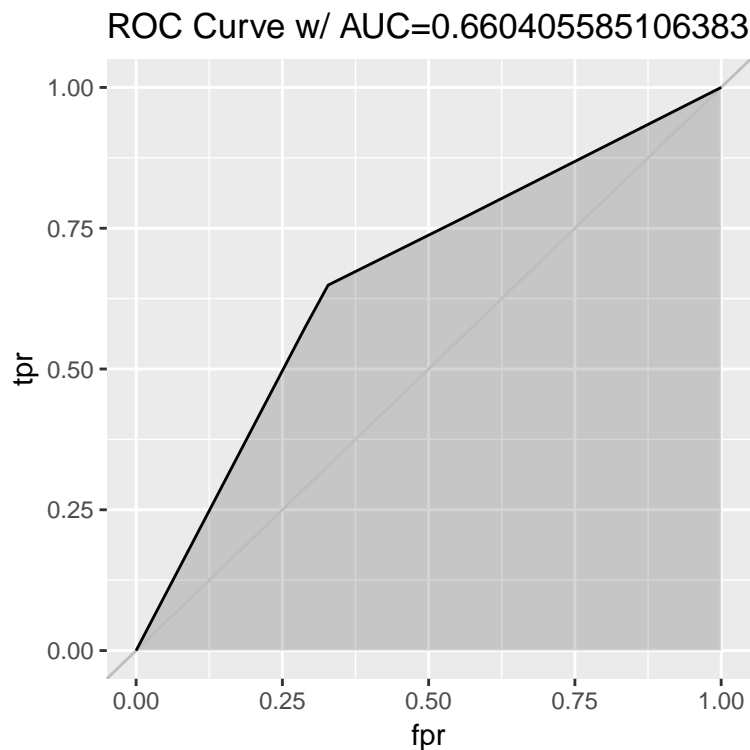
Note: Some suggest that “using a small value of g ought to give less opportunity to detect mis-specification. However, if we choose g to large, the numbers in each group may be so small that it will be difficult to determine whether differences between observed and expected are due to chance or indicative of model mis-specification”. As a result, it is suggested that $g > p+1$, where p is our number of predictors.

The Hosmer-Lemeshow test is a test of calibration. It determines whether or not our predicted probabilities fit the distribution of the data. Because our p-value is very high (in fact, it is 1 which is odd), we can reject the alternative hypothesis in favor of the null hypothesis, meaning that our model is correct.

8. Test for discrimination in the Logistic model.

To test for discrimination, we consider the area under the Receiver Operating Characteristic (ROC) curve, or the AUC.

AUC	Model Fit
0.5	Random
0.6	Mediocre
0.7	Decent
0.8	Good
0.9	Excellent



We see that the AUC is about 0.66, which implies a mediocre fit, but not awful. This means that this is not

as discriminatory as would be ideal (AUC of 1).

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

The significant components in the model suggest that the disparity lies within the choice of hospital. When viewed individually, both race and hospital choice played a significant role in the outcome of receiving a prostatectomy. However, when viewed together, as is done in this model, we see that hospital choice plays the most significant role. Furthermore, since both the calibration and discrimination tests of this model suggest a decent fit, all evidence suggests that the biggest disparity is in hospital choice. However, due to the significant change in model statistics, this suggests that there is high collinearity between race and hospital choice, implying a relationship between the two variables. In other words, hospital choice may be influenced by race.

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

One challenge is the very few number of predictors included in the study. In my opinion, identifying health disparities means analyzing a process with many different layers and relationships and interactions between a wealth of identifying variables including hospital, race, age, health, etc.