Stat 021 Homework 5

Suzanne Thornton

Due: Friday, Nov. 1, 12:00pm

Instructions: A **pdf** version of your homework must be submitted to Gradescope by **noon** on the due date. The course passcode is **MPKJ4Z**. If you are having trouble getting your *.Rmd* file to compile, you need to get help with this **before** the due date.

You are allowed to hand in **only one** late homework assignment throughout the semester. If you need to hand in this particular assignment late, you must let me know via email by noon on the due date.

You are encouraged to study with your peers to help complete the homework assignments but no copying is allowed. If I see that two or more homework assignments are copied, all students involved will receive a grade of 0 on that assignment and will forfeit (perhaps retroactively) the opportunity to hand in a late homework.

Q1) Sketch (by hand) residual plots (with $\hat{y}_i$, predicted response values, on the horizontal axis) that show each of the following: (5 points) 1. constant variance and linearity; 1. non-constant variance and linearity; 1. constant variance and non-linearity; 1. non-constant variance and non-linearity.

Solution: All residual plots must be labeled and the horizontal axis must be $\hat{y}$ and the vertical axis, e. To indicate non-linearity, there must be some non-rectangular spread of the residuals about $\hat{y} = 0$. To indicate constant variance, the range of values the residuals can take on should remain the same regardless of where we are with respect to the horizontal axis.

Q2) Suppose we have two random variables X and Y. What are the differences among the following assumptions regarding X and Y:

- X and Y are uncorrelated,
- X and Y are independent,
- X and Y have the same variance, and
- X and Y have the same distribution? (5 points)

Solution:

- X and Y are uncorrelated if they are not linearly related to one another.
- X and Y are independent of the value of one variables affects the value of the other.
- X and Y have the same variance if they have the same spread, i.e. if they can take on the same range of possible values.
- X and Y have the same distribution when their distributions follow the same shape.

Q3) Read the Wikipedia page for Simpson's Paradox: https://en.wikipedia.org/wiki/Simpson%27s_paradox. Then, import the "Stand your ground" data set uploaded on Moodle. This data (from 2015) is related to the Stand Your Ground law in Florida. Each observational unit consists of a case where the Stand Your Ground law was a part of the defense strategy, the defendant's race (white or non-white), the victim's race (white or non-white), and a binary variable indicating whether or not the defendant was convicted. With this categorical data we are not going to fit a regression model but we are going to examine this data and look out for Simpson's paradox. (10 points)

a) Create and print the following tables to summarize the data:
    1. Defendant's race vs convicted for all observational units;

2. Defendant's race vs convicted for cases with minority victims only;
3. Defendant's race vs convicted for cases with white victims only;
4. The table created by adding Tables 2 and 3 together.

```r
setwd("~/Google Drive Swat/Swat docs/Stat 21/Homework")
FL_data <- read_csv("stand_your_ground.csv")
```

```
## Parsed with column specification:
## cols(
##   Convicted = col_character(),
##   Accused = col_character(),
##   WhiteVictim = col_double(),
##   MinVictim = col_double()
## )
```

```r
(table1 <- table(FL_data$Accused, FL_data$Convicted))
```

```
##
##             No Yes
##   Minority 60  29
##   White    86  45
```

```r
## Note: putting something in between ()'s is a shortcut for the "print()" function
(table2 <- FL_data %>% filter(MinVictim == 1) %>% select(Accused, Convicted) %>% table())
```

```
##           Convicted
## Accused    No Yes
##   Minority 45  19
##   White    19   5
```

```r
(table3 <- FL_data %>% filter(WhiteVictim == 1) %>% select(Accused, Convicted) %>% table())
```

```
##           Convicted
## Accused    No Yes
##   Minority 15  10
##   White    67  40
```

```r
table2 + table3
```

```
##           Convicted
## Accused    No Yes
##   Minority 60  29
##   White    86  45
```

b) What are the overall conviction rates for minority and white defendants, respectively? What are the conviction rates for minority and white defendants among the cases with minority victims? What are the conviction rates for minority and white defendants among the cases with white victims?

```r
prop.table(table1, 1)
```

```
##
##                   No       Yes
##   Minority 0.6741573 0.3258427
##   White    0.6564885 0.3435115
```

```r
# Note: the 1 indicates row percentages, 2 indicates column percentages
prop.table(table2, 1)
```

```
##           Convicted
```

```
## Accused           No       Yes
##   Minority 0.7031250 0.2968750
##   White    0.7916667 0.2083333
```

```r
prop.table(table3, 1)
```

```
##           Convicted
## Accused           No       Yes
##   Minority 0.6000000 0.4000000
##   White    0.6261682 0.3738318
```

Solution:

Let p_M = conviction rate for a minority defendant and p_W = conviction rate for a white defendant. Overall: $p_M = 0.326$ and $p_W = 0.344$. For minority victims: $p_M = 0.297$ and $p_W = 0.208$. For white victims: $p_M = 0.4000$ and $p_W = 0.374$.

c) Explain what is going on here in terms of Simpson's paradox and interpret what this means with respect to racial bias in the criminal justice system.

Solution:

This is an example of Simpson's paradox because we see the relationship between the conviction rates of defendants based on the defendant's race only is the opposite of the relationship we see when we also consider the race of the victim. Based on our domain-specific knowledge, we understand that raceism is not only harmful to one group, but it also is beneficial to another. That's exactly what what we see going on here when we account for the equally important variable, race of the victim.