

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.

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)

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.

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?

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.