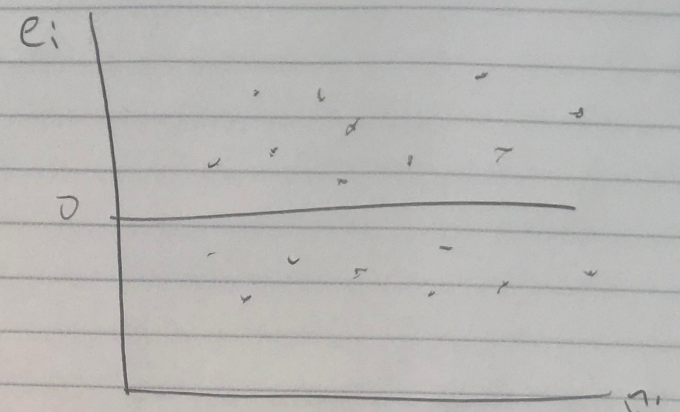
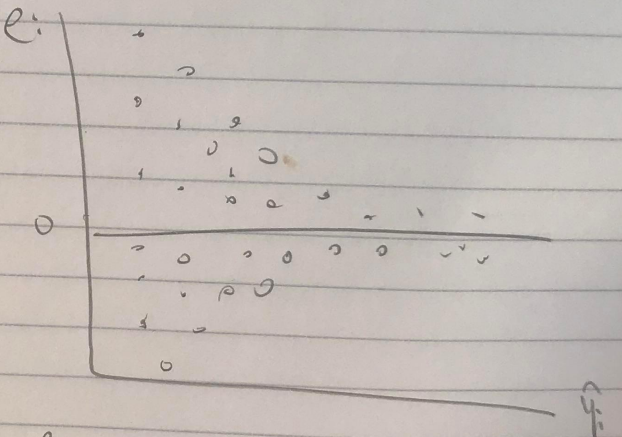
Q1

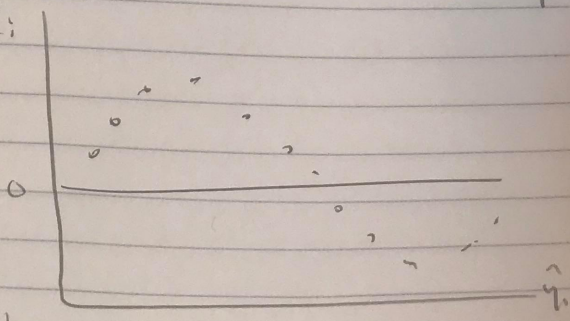
1 Constant variance & linearity



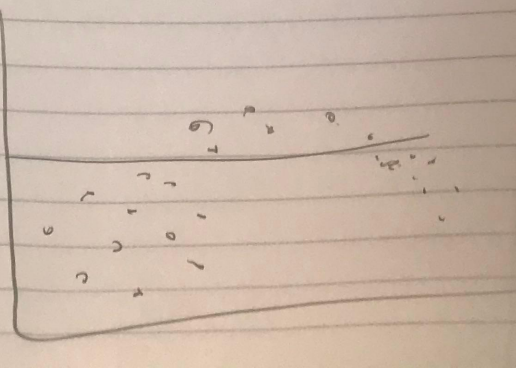
2 non-constant variance & linearity



3 Constant variance & nonlinearity



4 Non constant variance & nonlinearity



Q2

1 X & Y are uncorrelated then $\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = 0$

& $Cov[X,Y] = 0$

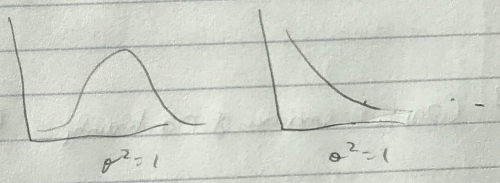
2 if X & Y are independent then $P(X,Y) = P(X)P(Y)$

if X & Y are independent they are uncorrelated

* if X & Y are uncorrelated they are not necessarily independent

3 X & Y have same variance $Var(XY) = E[X^2Y^2] - E[XY]^2$

→ don't necessarily have same distribution if they have same variance



4 X & Y have same distribution

Stat 021 Homework 5

Christina Holmgren

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 horizantal 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 defendent's race (white or non-white), the victim's race (white or non-white), and a binary variable indicating whether or not the defendent 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?

(from table 4) The overall conviction rates for minority defendants is 29/89 which is 32.58% The overall conviction rates for white defendants is 45/131 which is 34.35%

(from table 2) For cases with minority victims conviction rate of minorities is 19/64 or 29.69% conviction rate of whites is 5/24 or 20.83%

(from table 3) For cases with whitre victims conviction rate of minorities is 10/25 or 40% conviction rate of whites is 40/107 or 37.38%

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.

This shows Simpson's Paradox minorities have a higher conviction rate than whites when the victims are minorities (29.69% > 20.83%) and when the victims are white (40% > 37.38%), but when you look at totals (not conditioned by race of victems) the conviction rate is lower (32.58% < 34.35%). This paradox is can be explained by seeing how the number of obersvational units varies when conditioned by race (107 obs for white victim and 24 obersvations for minority victems). This shows us that there is racial bias that can be seen in the convictions conditioned by race, that can't be seen in overall conviction rates. This also means that depending you can use these same obersvational points to argue that there is or isn't a racial bias in the criminal justice system depending on which side you want to prove.

```
stand_your_ground_data <- read_csv(file = "stand_your_ground.csv")
```

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

```
convicted <- table(stand_your_ground_data$Convicted,stand_your_ground_data$Accused)
convicted
```

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

```
minority_data <- filter(stand_your_ground_data, MinVictim == 1)
minority <-table(minority_data$Convicted,minority_data$Accused)
minority
```

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

```
white_data <- filter(stand_your_ground_data, WhiteVictim == 1)
white <-table(white_data$Convicted,white_data$Accused)
white
```

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

```
merge(minority, white)
```

```
## [1] Var1 Var2 Freq
## <0 rows> (or 0-length row.names)
```

```
last_table <- white + minority
last_table
```

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