# Homework 6

Jeffrey Liang

11/27/2020

Problem 1(15p)

A hospital administrator wishes to test the relationship between 'patient's satisfactionscore' (Y) and three potential predictors:'age', 'severity of illness', and 'anxiety level' (see dataset'PatSatisfaction.csv'). The administrator randomly selected 46 patients, collected the data, and asked for your help with the analysis.

1. Create a correlation matrixfor all variablesand interpret your findings. Focus onthe correlationvalues between each predictor and the outcomeof interest. (2p)
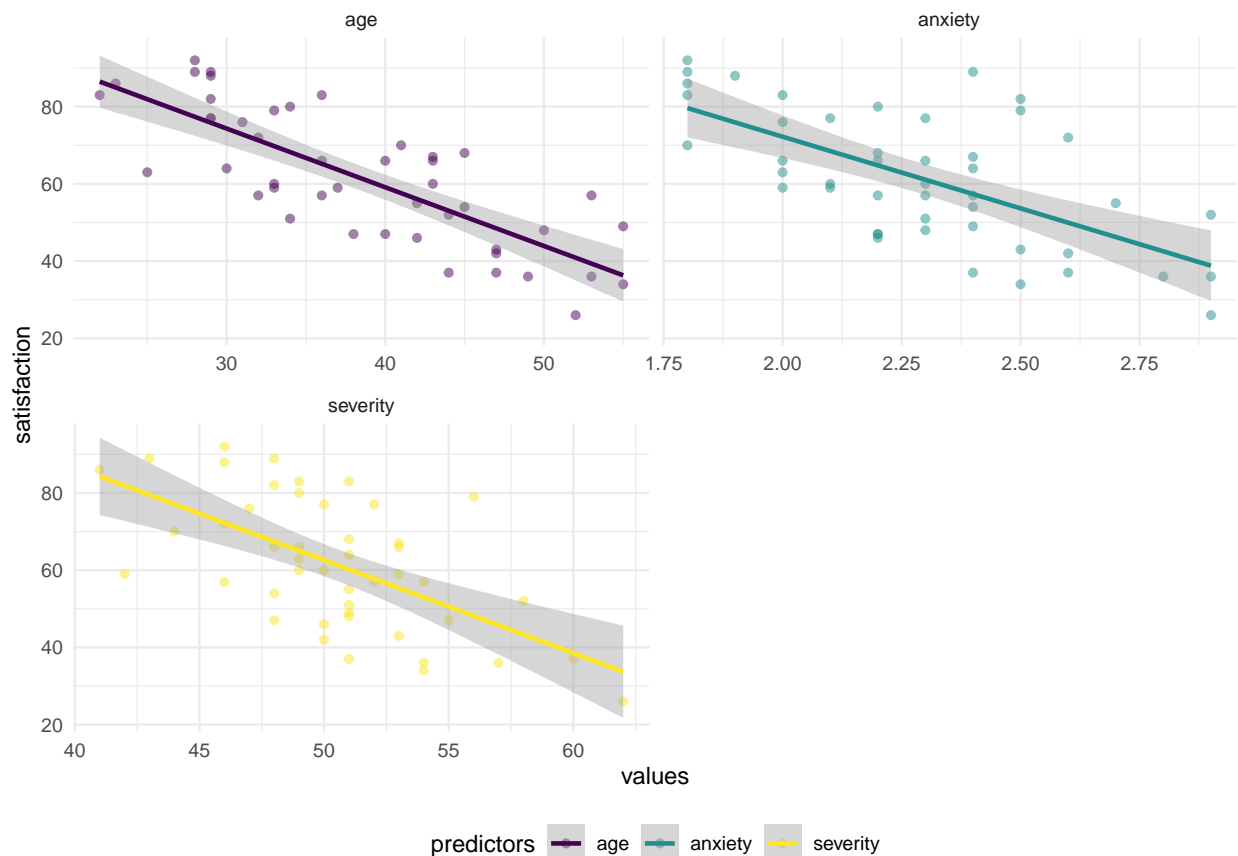
2. Fit a multiple regression model including all three predictorsand test whether at least one of these variables is significant. State the hypotheses, test-statistic, decision rule and conclusion. (3p)

To test if at least one of these model is significant, we propose hypothesis:

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

H_1 : at least one of the coefficient not equal to 0

The model we fit is:

$$statisfaction = 158.49 - 1.14age - 0.44severity - 13.47anxiety$$

We campared with the model with only intercept, and with ANOVA, we have

$$F^* = \frac{MSR(0|X1X2X3)}{MSE(X1X2X3)} \sim F_{df_L-df_S, df_L}$$

```
## Analysis of Variance Table
##
## Model 1: satisfaction ~ 1
## Model 2: satisfaction ~ age + severity + anxiety
##   Res.Df   RSS Df Sum of Sq  F  Pr(>F)
## 1     45 13369
## 2     42  4249  3      9120 30 1.5e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With F statistics of 30, at 95% confidence level, we have critical value of r, we reject the null hypothesis and conclude that at least one predictor have coefficent not equal to zero.

3. Show the regression results for all estimated slope coefficients with 95% CIs. Interpret the coefficient and 95% CI associated with 'severity of illness'. (5p)

term	conf.low	conf.high
(Intercept)	121.91	195.071
age	-1.57	-0.708
severity	-1.44	0.551
anxiety	-27.80	0.858

4. Obtain an interval estimate for a new patient's satisfaction with the following characteristics:Age=35, Severity=42, Anxiety=2.1. Interpret the interval. (2p)

5. a) Test whether 'anxiety level' can be dropped from the regression model, given the other two covariates are retained. State the hypotheses, test-statistic, decision rule and conclusion. (1.5p)

   b) How are R2/R2-adjusted impacted by the action that youtook in part 5-a)? (1.5p)

Problem 2(15p)

Obesity is very common in American society and is a risk factor for breast cancer in postmenopausal women.One mechanism explaining why obesity is a risk factor is that it may raise estrogen levels in women. Inparticular, one biomarker of estrogen, serum estradiol, is a strong risk factor for breast cancer. To better assessthese relationships, researchers studied a group of 210 premenopausalwomenand recorded the followinginformation('Estradl.csv'):

- Estradiol hormonalserumlevels(Estradl);
- BMI = weight (kg)/height2(m2); measure of overall adiposity(used to indicate obesity, e.g., BMI>30; note that for this analysis we will use thecontinuous measurements);
- Ethnicity (Ethnic= 1 if African American, = 0 if Caucasian)- Age (Entage);
- Number of children (Numchild);

- Age at menarche (Agemenar= age when menstrual periods began).

1. Is there a crude association between BMI and serum estradiol? a) Generate a scatter plot with the overlaid regression line. Comment.(2.5p) b) Provide the summary regression output and comment on the nature of the relationship (i.e., sign, magnitude, significance).(2.5p)

2. How does the relationship between BMI and serum estradiol change after controlling for all the other risk factors listed above? Provide the summary regression output and comment on the relationships observed for each of the predictors.(5p)

3. Now focus only the relationship between BMI and serum estradiol by ethnicity. Is there any evidence that these relationships vary for African American and Caucasian women?

   a) Use graphical displays and numerical summaries to sustain your conclusion.(2.5p)
   b) Based on your response in part 3-a), take additional steps to quantify the relationship between BMI and serum estradiol by ethnicity. Comment on your findings.(2.5p)