Data Science 1 IE 594 HMW1

Due Tuesday, Oct 1st, 5:00 pm

1. A hospital wishes to study the relation between patient satisfaction and 3 predictors: patient's age (in years), severity of illness (an index), and anxiety level (an index).
    a. Obtain the scatterplot matrix and correlation matrix.
    b. What do the scatter plot matrix and correlation matrix show about pairwise linear associations among the predictor variables?
    c. Fit linear regression model for 3 predictors without an intercept.
    d. Fit linear regression model for 3 predictors with an intercept.
    e. Compare your findings from c and d and indicate if there is any relationship.
    f. Plot residuals versus the fitted values, each predictor, and each two factor interaction. Prepare a normal probability plot.
    g. Find the coefficient estimate $\hat{\beta}$, the standard error of this coefficient estimate, and the t-statistic and p-value associated with the null hypothesis $H0 : \beta = 0$. Interpret your results.
    h. Is there a relationship between the predictor and the response?
    i. How strong is the relationship between the predictor and the response?
    j. Is the relationship between the predictors and the response positive or negative?
    k. What are the confidence intervals for β0 and β1 based on the original data set, the noisier data set, and the less noisy data set? Comment on your results.
    l. Try a few different transformations of the variables, such as $\log(X), \sqrt{X}, X^2$ Comment on your findings.
    m. Add polynomial features and interactions up to degree= 3 and fit linear regression models. Do any interactions appear to be statistically significant? Is there evidence of non-linear association between any of the predictors and the response?

OBS	SATISFAC	AGE	SEVERITY	ANXIETY
1	48	50	51	2.3
2	57	36	46	2.3
3	66	40	48	2.2
…	…	…	…	…