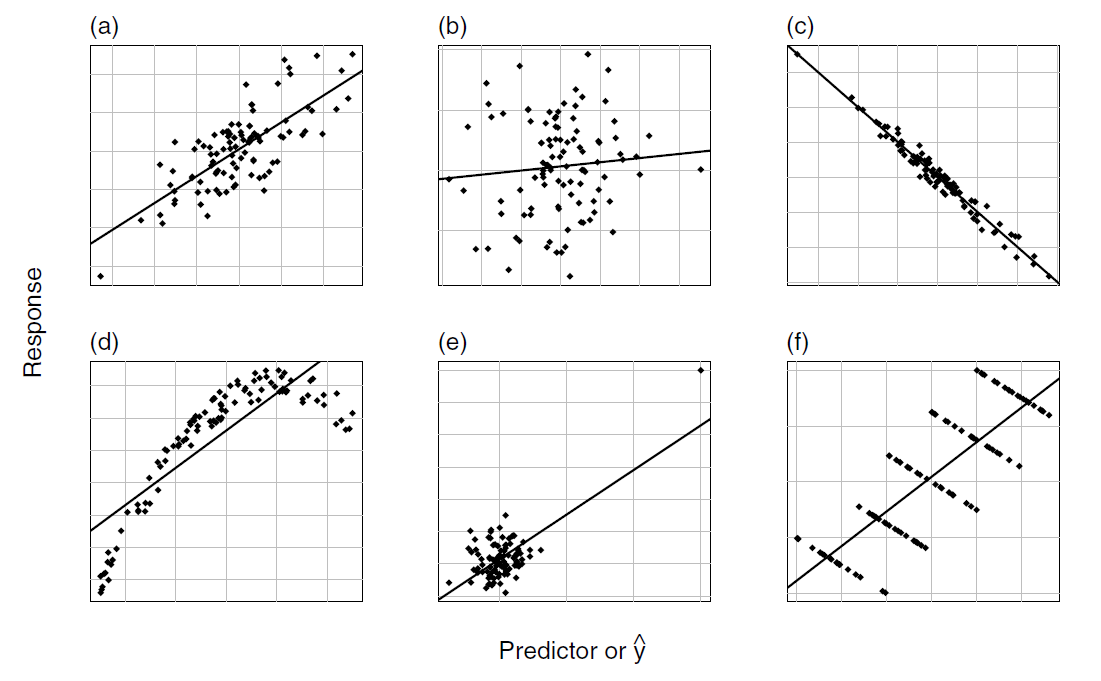
Question Bank

Linear Regression

1. What is linear regression and what are its assumptions?
2. Explain the difference between simple linear regression and multiple linear regression.
3. What is the purpose of the coefficient of determination (R-squared) in linear regression?
4. How do you interpret the slope and intercept coefficients in linear regression?
5. What is the difference between correlation and regression?
6. What is the difference between ordinary least squares (OLS) and gradient descent in linear regression?
7. How do you detect outliers in a linear regression model?
8. What is multicollinearity and how does it affect linear regression?
9. Explain the difference between the training set and the test set in linear regression.
10. What is regularization and how does it help to prevent overfitting in linear regression?

Logistic Regression

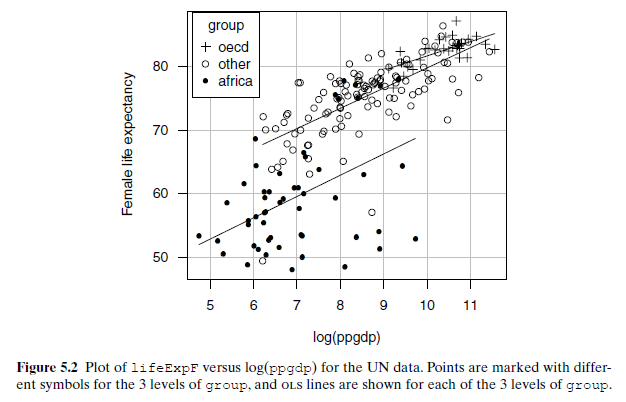
1. What is logistic regression and how is it different from linear regression?
2. Explain the sigmoid function and how it is used in logistic regression.
3. What is the purpose of the odds ratio in logistic regression?
4. How do you interpret the coefficients in a logistic regression model?
5. What are the assumptions of logistic regression and how do you check them?
6. What is the difference between binary logistic regression and multinomial logistic regression?
7. What is overfitting in logistic regression and how can it be prevented?
8. How do you evaluate the performance of a logistic regression model?
9. What is regularization and how does it help to prevent overfitting in logistic regression?
10. In linear regression problems, we can always determine the appropriateness of R-squared as a summary by examining the summary graph of the response versus the regressor. The figure below shows the six summary graphs.



a.) Which graphs are showing R-squared is an appropriate measure?

b.) Which graphs are showing R-squared is not an appropriate measure and why? What would you recommend to do after seeing these graphs?

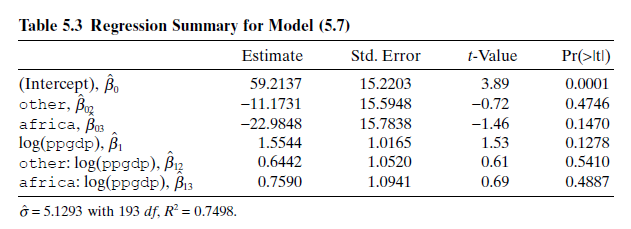
1. a.) Based on the figure 5.2 below, is there a sign of interaction effect between log(ppgdp) and group on female life expectancy and why?



b.) Suppose you are building a linear model for modeling female life expectancy and obtain a model below.

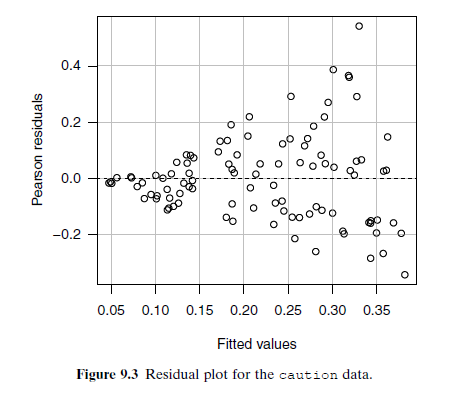


Assume that all assumptions of linear model are satisfied and you obtained a summary table for a model below:



i.) Give an interpretation of , and .

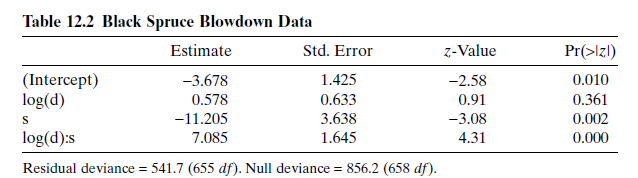
ii.) What conclusion(s) can you make from the table.

1. Based on the plot below. What assumption of linear regression is violated?   
   
2. **Blowdown**: Data from the Boundary Waters Canoe Area Wilderness Blowdown. The data frame Blowdown includes nine species of trees, but this file only includes black spruce, grouped by diameter.

* y: status of a tree after blowdown. 1: death and 0: survive
* d: tree diameter, in cm
* s: local severity of the storm (as the fraction of the total basal area of trees)

You perform a logistic model to model the probability of death by the blowdown for a tree using logit link and obtain the summary table below:

Model: y ~ log(d) + s + log(d):s



a.) Give an interpretation of coefficient estimate of log(d), .

b.) What conclusion(s) can you make from the table.