

STAT 3008
Homework 1

Due date: 5pm, Sep 30 (Tuesday). Assignment Box at LSB 1/F.

1. Recall that the definitions of separated points, outlier and leverage point:
 - x-separated point: An observation with its x-value far from the rest of data points.
 - y-separated point: An observation with its y-value far from the rest of data points.
 - outlier: An observation that is distinct from what suggested by the model, i.e., that lies far from the regression line.
 - influential point: The regression line obtained from the full dataset is very different from that obtained from the dataset without this observation.

For each of the following, scratch scatterplot(s) to illustrate the event. If the event does not exist, explain why.

- a) x-separated point and outlier.
- b) y-separated point and outlier.
- c) x-separated point and NOT outlier.
- d) y-separated point and NOT outlier.
- e) x-separated and influential point.
- f) y-separated and influential point.
- g) x-separated and NOT influential point.
- h) y-separated and NOT influential point.
- i) Outlier and influential point.
- j) Outlier but NOT influential point.
- k) Influential point but NOT outlier.

2. Given observations (x_i, y_i) for $i = 1, 2, \dots, n$, consider the regression model

$$y = \beta x^3 + e, \quad e \sim N(0, \sigma^2)$$

- i) Propose estimators for the unknown parameters β and σ^2 , i.e. $\hat{\beta}$ and $\hat{\sigma}^2$.
- ii) Suppose that $X = (1, 2, 3, 4, 5)$, $Y = (3.2, 10.8, 28.9, 64.2, 127)$. Find the values of $\hat{\beta}$ and $\hat{\sigma}^2$.
- iii) Draw a scatterplot and a residual plot for ii). Also draw the fitted regression line on the scatterplot.
- iv) Does the fitted regression line pass through (\bar{x}, \bar{y}) ? Is the sum of residuals equals zero?
- (v) Consider the regression model

$$y = \alpha_0 + \alpha_1 x + e, \quad e \sim N(0, \sigma^2)$$

- a) Find a 95% confidence interval for α_1 . Give the p -value of testing α_1 against $\alpha_1 \neq 0$.
- b) Perform a F -test for the effect of x (Give the p -value). Which two regression models are compared by this F -test?