STAT GU5205 Homework 1 [100 pts]
Due 8:40am Monday, September 23rd

Problem 1 (**1.22 KNN**)

Sixteen batches of the plastic were made, and from each batch one test item was molded. Each test item was randomly assigned to one of the four predetermined time levels, and the hardness was measured after the assigned elapsed time. The results are shown below; X is the elapsed time in hours, and Y is hardness in Brinell units. Assume the first-order regression model (1.1) is appropriate (**model (2.1) in the notes**).

Data not displayed

Perform the following tasks:

i. Use `R` to obtain the estimated regression function.

ii. Use `R` to create a scatter plot with the line of best fit. Make the line of best fit red.

iii. Use `R` to calculate the best point estimate of σ^2.

iv. Use `R` to calculate the sample correlation coefficient and coefficient of determination.

Problem 2

Recall the sample residual is defined by $e_i = y_i - \hat{y}_i$, where y_i is the *ith* response value and $\hat{y}_i$ is its corresponding fitted value computed by least squares estimates $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Prove the following properties:

i.
$$\sum_{i=1}^{n} x_i e_i = 0$$

ii.
$$\sum_{i=1}^{n} \hat{y}_i e_i = 0$$

Problem 3

Recall that the *ith* fitted value $\hat{Y}_i$ can be expressed as a linear combination of the response values, i.e.,

$$\hat{Y}_i = \sum_{j=1} h_{ij} Y_j,$$

where

$$h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}},$$

and

$$S_{xx} = \sum_{i=1} (x_i - \bar{x})^2.$$

Prove the following properties of the hat-values h_{ij}.

i.
$$\sum_{j=1} h_{ij}^2 = h_{ii}$$

ii.
$$\sum_{j=1} h_{ij} x_j = x_i$$

Problem 4

Consider the *regression through the origin model* given by

$$Y_i = \beta x_i + \epsilon_i \qquad i = 1, 2, \ldots, n \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2). \tag{1}$$

The estimated model at observed point (x, y) is

$$\hat{y} = \hat{\beta} x,$$

where

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}. \tag{2}$$

Complete the following tasks

i. Show that
$$\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$$
is an unbiased estimator of β.

ii. Compute the standard error of estimator $\hat{\beta}$.

iii. Identify the probability distribution of estimator $\hat{\beta}$.