

STATS 500 - Homework 6

Due in class on Wednesday, November 8, 2017

1. (a) Using the `sat` data, fit a model with `total` as the response and `takers`, `ratio`, `salary` and `expend` as predictors using the following methods:

1. Ordinary least squares
2. Least absolute deviations
3. Huber's robust regression
4. Least trimmed squares

Compare the results. In each case, comment on the significance of predictors (note that for least trimmed squares you will need to use the bootstrap).

(b) Now use diagnostic methods to detect any outliers or influential points. Remove these points and then use least squares. Compare the results with part (a), and the other robust regression results.

Solutions to problem 1 should not exceed 4 pages.

2. Suppose that the regression model is given as

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i$$

where errors are assumed to be iid with a double exponential distribution, i.e., they have a pdf of

$$f(x) = \frac{1}{2\lambda} e^{-|x|/\lambda} \quad \text{for all } x$$

The log-likelihood function is given as

$$L(\beta_0, \beta_1, \dots, \beta_p, \lambda) = \sum_{i=1}^n \log \left[f \left(y_i - \sum_{j=0}^p \beta_j x_{ij} \right) \right]$$

and the maximizing values of the parameters are referred to as the Maximum-Likelihood Estimates (MLE).

(a) Show that MLE's of $\beta_0, \beta_1, \dots, \beta_p$ are equivalent to the LAD estimates.

(b) Show that $\hat{\beta}_0$ is the median of the values $\{y_i - \sum_{j=1}^p \hat{\beta}_j x_{ij}\}_{i=1}^n$.

(c) What sort of graphical diagnostics might you do after carrying out the LAD regression to determine whether the double exponential model for the errors was approximately valid. Be specific.