Repetition week 8

Significant impact of a single variable

$$H_0 : \beta_j = 0 \quad H_1 : \beta_j \neq 0 \qquad T = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{c_{jj}}} \sim t_{n-k-1}$$

$$H_0 : \beta_j = \beta_{j0} \quad H_1 : \beta_j \underset{<}{\overset{>}{\neq}} \beta_{j0} \qquad T = \frac{\hat{\beta}_j - \beta_{j0}}{\hat{\sigma}\sqrt{c_{jj}}} \sim t_{n-k-1}$$

Criteria for Choice of fitted model

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

$$R_{adj}^2 = 1 - \frac{\frac{SS_E}{n-k-1}}{\frac{SS_T}{n-1}} = 1 - \frac{(n-1)s^2}{SS_T}$$

Let $\hat{\sigma}^2$ be the ML estimator

$$AIC = n\ln\left(\hat{\sigma}^2\right) + 2(k+2)$$

$$BIC = n\ln\left(\hat{\sigma}^2\right) + \ln(n)(k+2)$$

Mallows' $C_p = \frac{SS_E}{s^2} - \left(n - 2(k+1)\right)$

Variable selection methods

Let $\boldsymbol{\beta}$ be the vector of coefficients in the model

Forward selection

Candidate to enter in partial F-test: $\max_j \left\{ SS_R\left(\boldsymbol{\beta}, \beta_j\right) - SS_R\left(\boldsymbol{\beta}\right) \right\}$

Backward elimination

Candidate to leave in partial F-test: $\min_j \left\{ SS_R\left(\boldsymbol{\beta}\right) - SS_R(\boldsymbol{\beta} \setminus \beta_j) \right\}$

Stepwise regression

Combines forward selection and backward elimination

1. Start as with forward selection. Assume x_n and x_m are chosen as the two first variables.
2. Let $\boldsymbol{\beta} = \{\beta_n, \beta_m\}$. Find $\min_{j=n,m} \left\{ SS_R\left(\boldsymbol{\beta}\right) - SS_R(\boldsymbol{\beta} \setminus \beta_j) \right\}$ and check if one of the variables can be taken out
3. Proceed as with forward selection but check in each step if one of the variables can be taken out.

Best subset selection

Find the best model according to given criteria for a model with k parameters, $k = 1, 2, \ldots, n$.

Use $R^2,\ R^2_{adj},\ AIC,\ BIC,\ Mallows' C_p$