Backward ~~selection~~ elimination

Define $\underline{\beta} \setminus \beta_j = (\beta_0, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k)$

1. Start with all the variables in the model

2. Find $\min_j R(\beta_j \mid \underline{\beta} \setminus \beta_j) = \min_j \{ SS_R(\beta_0, \ldots, \beta_k) - SS_R(\beta_0, \ldots, \beta_{j-1}, \beta_{j+1} \ldots \beta_k) \}$

3. If $\min_j \dfrac{R(\beta_j \mid \underline{\beta} \setminus \beta_j)}{\frac{SS_E}{n-k-1}} \geq f_{\alpha, 1, n-k-1}$ stop, no variable is removed.

4. If $\min_j \dfrac{R(\beta_j \mid \underline{\beta} \setminus \beta_j)}{\frac{SS_E}{n-k-1}} = \dfrac{R(\beta_m \mid \underline{\beta} \setminus \beta_m)}{\frac{SS_E}{n-k-1}} < f_{\alpha, 1, n-k-1}$

remove x_m

Let $\underline{\beta} = \{\beta_1, \beta_2, \ldots, \beta_k\} \setminus \beta_m$, $k = k-1$ and continue until

all the variables are significant.

Stepwise regression

1. Start like forward selection

Assume x_1 and x_2 are chosen to enter the model in

step 1 and step 2. Let $\underline{\beta} = \{\beta_1, \beta_2\}$

2. Find $\min_{j=1,2} R(\beta_j \mid \underline{\beta} \setminus \beta_j) = R(\beta_m \mid \underline{\beta} \setminus \beta_m)$

and investigate if x_m should be removed as for

backward elimination.

3. continue as for forward selection, but test

in each step if any of the variable chosen to enter can

be removed.

12.10. Studying Residuals

$$\hat{\underline{\varepsilon}} = \underline{y} - \hat{\underline{y}} = \underline{X}\underline{\beta} + \underline{\varepsilon} - \underline{X}\hat{\underline{\beta}} = \underline{X}\underline{\beta} + \underline{\varepsilon} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'(\underline{X}\underline{\beta} + \underline{\varepsilon})$$

$$= \underline{X}\underline{\beta} + \underline{\varepsilon} - \underline{X}\underline{\beta} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{\varepsilon} = \underline{\varepsilon} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{\varepsilon}$$

$$= (\underline{I} - \underline{H})\underline{\varepsilon} \implies E[\hat{\underline{\varepsilon}}] = 0$$

$$Cov(\hat{\underline{\varepsilon}}) = E[(\underline{I} - \underline{H})\underline{\varepsilon}\,\underline{\varepsilon}'(\underline{I} - \underline{H})'] = \sigma^2(\underline{I} - \underline{H})(\underline{I} - \underline{H}) = \sigma^2(\underline{I} - \underline{H})$$

Det synes at $Var(\hat{\varepsilon}_i) = \sigma^2(1 - \underbrace{h_{ii}}_{\text{i-te diagonal element i } \underline{H}})$

$$h_{ii} = \underline{x}_i'(\underline{X}'\underline{X})^{-1}\underline{x}_i \qquad \text{where} \quad \underline{x}_i' = [1, x_{1i}, x_{2i}, \dots, x_{ki}]$$

Since $Var\left(\dfrac{\hat{\varepsilon}_i}{\sqrt{\sigma^2(1-h_{ii})}}\right) = 1$, we often use

$r_i = \dfrac{e_i}{s\sqrt{1-h_{ii}}}$ for studying the residuals (studentized residuals)

in MINITAB they are called standardized.

One other alternative is R-studentized residuals, in MINITAB (studentized)

$t_i = \dfrac{e_i}{s_{-i}\sqrt{1-h_{ii}}} \sim$ t-distributed with $n-k-2$ degrees of freedom. Can be used to test if $E[\hat{\varepsilon}_i] = 0$

s_{-i} is estimated standard deviation when the i-th observation ~~set~~ set $(y_i, x_{1i}, \dots, x_{ki})$ is taken out. Useful for diagnosing outliers.

Check of residuals give information about

1. Outliers
2. Heterogenity in variance
3. If the model is misspecified (and if other explanatory variables should be in the model;
4. Normal distribution

12.11. Other methods for choosing the „best model".

Cross-validation. Predictive sum of squares (PRESS)

Let the PRESS-residuals be defined as:

the model fit when $(y_i, x_{1i}, \ldots, x_{ki})$ is taken out.

$$\delta_i = y_i - \hat{y}_{i,-i}, \quad i = 1, 2, \ldots, n$$

It can be shown that $\delta_i = \dfrac{e_i}{1-h_{ii}}$

The model with the best prediction ability is the one that minimize $\sum_{i=1}^{n} |\delta_i|$ or $\sum_{i=1}^{n} \delta_i^2$

or eventually the one that maximizes

$$R^2_{pred} = 1 - \frac{\sum_{i=1}^{n} \delta_i^2}{SS_T}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{SS_T}$$

11.12 Test on correlation

X, Y random variables.

$$Y = \alpha + \beta X + \varepsilon \begin{cases} E[\varepsilon] = 0 \\ Var[\varepsilon] = \sigma^2 \end{cases}$$

$E[Y] = \alpha + \beta E[X]$ and $Cov(Y, X)$

$$= E\left\{[\beta(X - E[X]) + \varepsilon][X - E[X]]\right\} = \beta \sigma_X^2$$

such that $\rho(Y, X) = \dfrac{\beta \sigma_X^2}{\sigma_X \sigma_Y} = \dfrac{\beta \sigma_X}{\sigma_Y}$; $\rho = 0 \Leftrightarrow \beta = 0$

Simple linear regression

$$b = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

Estimate for ρ = $$r = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{b\sqrt{S_{xx}}}{\sqrt{S_{yy}}}$$

Some calculations

$$SS_E = \sum_{i=1}^{n} (y_i - a - bx_i) = \sum_{i=1}^{n} (y_i - \bar{y} - b(x_i - \bar{x}))^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2b \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x}) + b^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \sum_{i=1}^{n} (y_i - \bar{y})^2 - 2b \frac{S_{xy}}{\cancel{S_{xx}}} + b \frac{S_{xy}}{S_{xx}} \cdot S_{xx} = S_{yy} - \underbrace{b\, S_{xy}}_{SS_R}$$

This gives $r^2 = \frac{b^2 S_{xx}}{S_{yy}} = b \frac{S_{xy}}{S_{xx}} \cdot \frac{S_{xx}}{S_{yy}} = b \frac{S_{xy}}{S_{yy}} = R^2$

$H_0 : \rho = 0 \quad H_1 : \rho \neq 0 \iff H_0 : \beta = 0 \quad H_1 : \beta \neq 0$

Teststatistic for β:

$$t = \frac{b}{\dfrac{s}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}} = \frac{b}{\dfrac{s}{\sqrt{S_{xx}}}} = \frac{b\sqrt{S_{xx}}}{s} = \frac{r\sqrt{S_{yy}}\sqrt{S_{xx}}}{s\sqrt{S_{xx}}}$$

$$= \frac{r\sqrt{S_{yy}}}{s}.$$

Substituting $s = \sqrt{\frac{SS_E}{n-2}}$

We get. $$t = \frac{r\sqrt{S_{yy}}}{\sqrt{\dfrac{S_{yy} - b S_{xy}}{n-2}}} = \frac{r\sqrt{n-2}}{\sqrt{1 - b\dfrac{S_{xy}}{S_{yy}}}} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Reject if $|t| > t_{\frac{\alpha}{2}, n-2}$.

Mallows C_p

$$\hat{Y}_i - E[Y_i] = \hat{Y}_i - E[\hat{Y}_i] + E[\hat{Y}_i] - E[Y_i]$$

$$\left[\hat{Y}_i - E[Y_i]\right]^2 = \left(\hat{Y}_i - E[\hat{Y}_i]\right)^2 + \left(E[\hat{Y}_i] - E[Y_i]\right)^2$$

$$+ 2\left(\hat{Y}_i - E[\hat{Y}_i]\right)\left(E[\hat{Y}_i] - E[Y_i]\right)$$

$$E\left[\left(\hat{Y}_i - E[Y_i]\right)^2\right] = Var[\hat{Y}_i] + (Bias\ \hat{Y}_i)^2$$

Want to minimize

$$E\left[\sum_{i=1}^{n} \frac{(\hat{Y}_i - E[Y_i])^2}{\sigma^2}\right] = \underbrace{\sum_{i=1}^{n} \frac{Var[\hat{Y}_i]}{\sigma^2}}_{p} + \sum_{i=1}^{n} \frac{(Bias\ \hat{Y}_i)^2}{\sigma^2}$$

The estimate for $E\left[\sum_{i=1}^{n} \frac{(\hat{Y}_i - E[Y_i])^2}{\sigma^2}\right]$ is Mallows C_p

given by $C_p = p + \frac{(s^2 - \hat{\sigma}^2)(n-p)}{\hat{\sigma}^2}$

p is the number of parameters.

s^2 is an estimate for $Var(Y)$ in the model with

p parameters.

$\hat{\sigma}^2$ is an estimate for σ^2 (often used full model).