Reresja wieloraka

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1} + \xi$$

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \xi_{n\times 1}$$

p - liczba "bet" n - liczba obserwacji

$$\xi \sim N(0, \sigma^2 I)$$

$$Y \sim N(X\beta, \sigma^2 I)$$

Rozwiązania MNK:

$$b = (X'X)^{-1}X'Y$$

$$\hat{Y} = Xb = HY$$

$$H = X(X'X)^{-1}X'$$

Reszty:

$$e = Y - \hat{Y} = Y - HY$$

Macierz kowariancji reszt

$$Cov(e) = \sigma^2(I - H)$$

$$Var(e_i) = \sigma^2(1 - h_{ii})$$

$$Cov(e_i, e_j) = -sigma^2 h_{ij}$$

estymator σ^2

$$S^2 = \frac{e'e}{n-p} = MSE$$

estymacja b

$$b \sim N(\beta, \sigma^2(X'X)^{-1}$$

estymacja wariancji:

$$s^2(X'X)^{-1}$$

$$R^2 = \frac{SSM}{SST} = cos\alpha$$

$SST = \sum(Y_i - \bar{Y})^2$ $MST = \frac{SST}{dfT}$
$SSM = \sum(\hat{Y}_i - \bar{Y})^2$, $MSM = \frac{SSM}{dfM}$
$SSE = \sum(Y_i - \hat{Y}_i)^2$, $MSE = \frac{SSE}{dfE}$

$$SSt = SSM + SSE$$

$F = \frac{MSM}{MSE} \sim F(dfM, dfE) = F(1, n-2)$

$$dfM = p - 1$$

$$dfE = n - p$$
$$dfT = n - 1$$

Teoria $E(Y_h)$ CI

$$X_h = (1, X_{h1}, \ldots, X_{hp-1})$$
$$\hat{\mu}_h = X_h' b$$
$$\sigma^2(\hat{\mu}_h) = \sigma^2 X_h'(X'X)^{-1} X_h$$
$$s^2(\hat{\mu}_h) = s^2 X_h'(X'X)^{-1} X_h$$

CI : $\hat{\mu}_h \pm s(\hat{\mu}_h) t_{0.975,n-p}$

$$Y_h = X_h' \beta + \xi$$
$$\hat{Y}_h = \hat{\mu}_h = X_h' b$$
$$\sigma^2(pred) = Var(Y_h - \hat{Y}_h) = \sigma^2(1 + X_h'(X'X)^{-1} X_h)$$
$$s^2(pred) = s^2(1 + X_h'(X'X)^{-1} X_h)$$

PI : $\hat{\mu}_h \pm s(pred) t_{0.975,n-p}$

$$F = \frac{\frac{SSE(R) - SSE(F)}{dfE(R) - dfE(F)}}{MSE(F)}$$

$$F(liczbadodatkowychzmiennych, dfE(model\text{"}wikszy\text{"}))$$

C_p $B_i = E(\hat{Y}_i) - E(Y_i)$ - bias C-p to estymator

$$\sum_i^n B_i^2 / \sigma^2$$

$$C_p = \frac{SSE_p}{MSE(F)} - (n - 2p)$$

AIC

$$nlog\left(\frac{SSE_p}{n}\right) + 2p$$

SBC

$$nlog\left(\frac{SSE_p}{n}\right) + plog(n)$$

Studentyzowane reszty

$$\frac{e_i}{\sqrt{MSE(1 - h_{ii})}}$$

DFFITS 1 i $2\sqrt{p/n}$ SFBETAS 1 i $\sqrt{n}$

$$VIF_k = (1 - R_k^2)^{-1}$$

$$tol = 1/VIF$$