Chapter 3: Multiple Linear Regression

Regression

$$\underset{n\times 1}{y} = \underset{n\times(k+1)}{x} \times \underset{(k+1)\times 1}{\beta} + \underset{n\times 1}{\epsilon}$$

$$or$$

$$\underset{n\times 1}{y} = \underset{n\times p}{x} \times \underset{p\times 1}{\beta} + \underset{n\times 1}{\epsilon}$$

ϵ ~N(0, σ^2I) where I is the identity matrix of whatever size it needs to be.
E[y] = xβ
V[y] = V[ϵ] = σ^2I
y ~N(xβ , σ^2I)

Least Square Estimate for β and σ^2

$$\hat{\beta} = (x'x)^{-1}x'y$$

E[$\hat{\beta}$] = β
V[$\hat{\beta}$] = (x′x)$^{-1}\sigma^2$ = cσ^2
V[$\hat{\beta}_j$] = c$_{jj}\sigma^2$
E[β $_j$] = β $_j$
$\hat{\beta}_j$ ~N(β $_j$, c$_{jj}\sigma^2$)

$$\hat{y} = x\hat{\beta} = x(x'x)^{-1}x'y = Hy$$

E[$\hat{y}$] = E[x$\hat{\beta}$] = xβ
V[$\hat{y}$] = V[x$\hat{\beta}$] = xV[$\hat{\beta}$]x′ = x(x′x)$^{-1}$x′σ^2 = Hσ^2
$\hat{y}$ ~N(xβ , Hσ^2)
$\hat{y}_j$ ~N(x$_j\beta$, h$_{jj}\sigma^2$) where

$$h_{jj} = x'_j(x'x)^{-1}x_j$$

$$x_j = [x_{j0}, x_{j1}, ...x_{jk}]$$

and

$$\hat{\epsilon} = y - \hat{y} = y - Hy = (I - H)y$$

where $\hat{\epsilon}$ is the estimated error.

$$\hat{\sigma^2} = \frac{SS_{res}}{n-p} = MS_{res}$$

where p = k + 1 = the number of parameters (i.e. β 's: β $_0$, β $_1$, ... β $_k$)

Confidence Intervals

$$\hat{\beta}_j \sim N(\beta_j, c_{jj}\sigma^2) \longrightarrow \frac{\hat{\beta}_j - \beta_j}{\sqrt{c_{jj}\sigma^2}} \sim N(0,1)$$

or, if variance is unknown,

$$\hat{\beta}_j \sim N(\beta_j, c_{jj}MS_{res}) \longrightarrow \frac{\hat{\beta}_j - \beta_j}{\sqrt{c_{jj}MS_{res}}} or \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \sim t_{n-p}$$

variance:

$$\hat{\sigma^2} = MS_{res} = \frac{SS_{res}}{n-p} \sim \chi^2_{n-p}$$

So, the (1 - α) confidence interval for $\beta\ _j$ is

$$\hat{\beta}_j \pm t_{\frac{\alpha}{2},n-p} se(\hat{\beta}_j)$$

$\hat{y}_j$ ~N(x$_j\beta$, h$_{jj}\sigma^2$)

$$\frac{\hat{y}_j - x_j\beta}{\sqrt{h_{jj}\sigma^2}} \sim N(0,1) \longrightarrow \frac{\hat{y}_j - x_j\beta}{\sqrt{h_{jj}MS_{res}}} \sim t_{n-p}$$

where σ^2 = MS$_{res}$
A 1 - α confidence interval for E[y$_0$|x$_0$] is:

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-p}\sqrt{x_0'(x'x)^{-1}x_0\hat{\sigma^2}}$$

or

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-p}\sqrt{x_0'(x'x)^{-1}x_0 MS_{res}}$$

Analysis of Variance (ANOVA)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Regression	p - 1 (or k)	SS$_{Reg}$	MS$_{Reg}$	$\frac{MS_{Reg}}{MS_{res}}$	n/a
Residuals	n - p or n - (k + 1)	SS$_{res}$	MS$_{res}$		
Total	n - 1	SS$_{total}$			

R^2 = $\frac{SS_{Reg}}{SS_{Tot}}$
R$^2{}_{adj}$ = ?