Chapter 3: Multiple Linear Regression

Regression

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

where p = k + 1
p is the total number of betas (or parameters), k is the number of regressor variables.
$\epsilon \sim$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 $\sim$N(xβ , σ^2I)

Least Square Estimate for β and σ^2

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

E[$\hat{\beta}$] = β
V[$\hat{\beta}$] = $(x'x)^{-1}\sigma^2 = c\sigma^2$
V[$\hat{\beta}_j$] = $c_{jj}\sigma^2$
E[β_j] = β_j
$\hat{\beta}_j \sim$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$'$ = $x(x'x)^{-1}x'\sigma^2$ = Hσ^2
$\hat{y} \sim$N(xβ , Hσ^2)
$\hat{y}_j \sim$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 β_j is

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

$\hat{y}_j \sim N(x_j\beta , h_{jj}\sigma^2)$

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

where $\sigma^2 = MS_{res}$
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}			

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