$\overrightarrow{Y} = X\overrightarrow{\beta} + \varepsilon,\ \min S(\overrightarrow{\beta}) = \sum \varepsilon^2 = (\overrightarrow{Y} - X\overrightarrow{\beta})^T(\overrightarrow{Y} - X\overrightarrow{\beta})$

normal equation $X^T X\overrightarrow{\beta} = X^T\overrightarrow{Y} \Rightarrow \widehat{\overrightarrow{\beta}} = (X^T X)^{-1}X^T\overrightarrow{Y}.$

geometric 的解釋：$\overrightarrow{Y} \in \mathbb{R}^n, L(X)$ 是其 subspace, $\widehat{\overrightarrow{Y}} \in L(X), \vec{e} \perp L(X)$

$E[\widehat{\overrightarrow{\beta}}] = \beta, V(\widehat{\overrightarrow{\beta}}) = (X^T X)^{-1}\sigma^2$

fitted values $\widehat{\overrightarrow{Y}} = X\widehat{\overrightarrow{\beta}}.E[\widehat{\overrightarrow{Y}}] = X\overrightarrow{\beta}$ hat matrix $H = X(X^T X)^{-1}X^T, V(\widehat{\overrightarrow{Y}}) = H\sigma^2$

residual $\vec{e} = \overrightarrow{Y} - X\widehat{\overrightarrow{\beta}}.E[\vec{e}] = 0, V(\vec{e}) = (I - H)\sigma^2$

σ^2 的 least squares estimator $s^2 = \dfrac{\vec{e}^{\,T}\vec{e}}{n-p-1}$ (MSE), $E[s^2] = \sigma^2$

令 v_{ii} 是 $(X^T X)^{-1}$ 的第 i 個 diagonal element. $T = \dfrac{\widehat{\beta_i} - \overline{\beta_i}}{s\sqrt{v_{ii}}} \sim T(n-p-1)$

H 和 $I-H$ 是 idempotent.(i.e., $H^2 = H, (I-H)^2 = I-H$)

mean response $\vec{x}_h : \hat{Y}_h = \vec{x}_h^{\,T}\widehat{\vec{\beta}}, V(\hat{Y}_h) = \sigma^2\vec{x}_h^{\,T}(X^T X)^{-1}\vec{x}_h$

new observation 的 predition $\vec{x}_{h(new)} : \hat{Y}_{h(new)} = \vec{x}_{h(new)}^{\,T}\widehat{\vec{\beta}}$

$V(\hat{Y}_{h(new)} - \hat{Y}_h) = \sigma^2(1 + \vec{x}_h^{\,T}(X^T X)^{-1}\vec{x}_h)$

Source of Variation	SS	df	MS
Regression	$\sum\sum(\hat{Y}_{ij} - \overline{Y})^2$	p	$\dfrac{SS(R)}{p}$
Error	$\sum\sum(Y_{ij} - \hat{Y}_{ij})^2$	$n-1-p$	$\dfrac{SS(E)}{n-1-p}$
Lack of fit	$\sum\sum(\overline{Y}_j - \hat{Y}_{ij})^2$	$c-1-p$	$\dfrac{SS(LF)}{c-1-p}$
Pure Error	$\sum\sum(Y_{ij} - \overline{Y}_j)^2$	$n-c$	$\dfrac{SS(LF)}{n-c}$
Total	$\sum(Y_i - \overline{Y})^2$	$n-1$	

當 $\overrightarrow{\beta} = \vec{0}, E(MS(R)) = \sigma^2$; 若不是的話 $E(MS(R)) > \sigma^2$

$H_0:\bar{\beta}=\bar{0}$ vs $H_1:\bar{\beta}\neq\bar{0}, F=\frac{MS(R)}{MS(E)}\overset{H_0}{\sim}F(p,n-1-p)$

Coefficient of Multiple Determination($R^2=\frac{SS(R)}{SS(TO)}=1-\frac{SS(E)}{SS(TO)}, 0\le R^2\le 1$)

adjusted $R^2=1-\frac{\frac{SS(E)}{n-1-p}}{\frac{SS(TO)}{n-1}}$ 有重複的觀測值，而且我們可以將他分為 c 個類別

$F_{obs}=\frac{MS(LF)}{MS(PE)}$，若 $F_{obs}>F(\alpha,c-2,n-c)$ 代表其無 linear 的關係

leverage:若 $h_{ii}>2\bar{h}=\frac{2(p+1)}{n}$ 我們通常認為是 high-leverage

Cook's Distance $D_i=\frac{(\hat{\beta}_{(i)}-\hat{\beta})^T(X^TX)(\hat{\beta}_{(i)}-\hat{\beta})}{(p+1)s^2}=\frac{(\hat{y}_{(i)}-\hat{y})^T(\hat{y}_{(i)}-\hat{y})}{(p+1)s^2}$,

這裡 $p+1=rank(X)$ 且 (i) 代表“將第 i 個點去掉的情形”。

$(DFFITS_i)^2=\frac{(\hat{\beta}_{(i)}-\hat{\beta})^T(X^TX)(\hat{\beta}_{(i)}-\hat{\beta})}{s_{(i)}^2}, DFFITS_i=\frac{\hat{Y}_{(i)}-\hat{Y}}{s_{(i)}\sqrt{h_{ii}}}(2\sqrt{\frac{p}{n}})$

$DFBETAS_{ij}=\frac{\hat{\beta}_j-\hat{\beta}_{j(i)}}{s_{(i)}\sqrt{v_{ii}}}(2,\frac{2}{\sqrt{n}})$

extra sum of square

Source of Variation	SS	df	MS
Regression	$SSR(x_1,x_2,x_3)$	3	$MSR(x_1,x_2,x_3)$
x_1	$SSR(x_1)$	1	$MSR(x_1)$
$x_2 \mid x_1$	$SSR(x_2 \mid x_1)$	1	$MSR(x_2 \mid x_1)$
$x_3 \mid x_1,x_2$	$SSR(x_3 \mid x_1,x_2)$	1	$MSR(x_3 \mid x_1,x_2)$
Error	$SSE(x_1,x_2,x_3)$	$n-4$	$MSE(x_1,x_2,x_3)$
Total	$SSTO$	$n-1$	

$SSTO=SSR(x_1,x_2,x_3)+SSE(x_1,x_2,x_3), SSR(x_1,x_2,x_3)=SSR(x_1)+SSR(x_2\mid x_1)+SSR(x_3\mid x_1,x_2)$

$r_{01,2}^2=\frac{SSR(x_1\mid x_2)}{SSE(x_2)}$

$$H_0:(RM),H_1:(FM),F^*=\frac{\frac{SSE(RM)-SSE(FM)}{df_{RM}-df_{FM}}}{\frac{SSE(FM)}{df_{FM}}}\overset{H_0}{\sim}F(df_{RM}-df_{FM},df_{FM})$$

若 $F^*>F(\alpha,df_{RM}-df_{FM},df_{FM})$，不 reject H_0

若 $F^* < F(\alpha, df_{RM} - df_{FM}, df_{FM})$, reject H_0

$$\text{MLE:} \min l(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 x_i)^2, \hat{\sigma}^2 = \frac{S(\vec{\beta})}{n}$$