29 Sept

$$\text{Res S.S.} = \mathbf{Y}^T \underbrace{\boxed{(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)}}_{\mathbf{A}} \mathbf{Y}$$

$$\begin{aligned}
\text{Reg S.S.} &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \\
&= (\hat{\mathbf{Y}} - \bar{\mathbf{Y}})^T (\hat{\mathbf{Y}} - \bar{\mathbf{Y}}) \\
&= \mathbf{Y}^T \boxed{(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \tfrac{1}{n}\mathbf{J})^T (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \tfrac{1}{n}\mathbf{J})} \mathbf{Y} \\
&= \mathbf{Y}^T \underbrace{\boxed{(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \tfrac{1}{n}\mathbf{J})}}_{\mathbf{B}} \mathbf{Y}
\end{aligned}$$

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \qquad \bar{\mathbf{Y}} = \frac{1}{n}\mathbf{J}\mathbf{Y}$$

$\cancel{(\mathbf{I}}$ $\cancel{A}$

$$\begin{aligned}
&(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \tfrac{1}{n}\mathbf{J}) \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \tfrac{1}{n}\mathbf{J}) \\
&= \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T - \tfrac{1}{n}\mathbf{J} - \underline{\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\boxed{\mathbf{X}^T\mathbf{X}}}\,\underline{\boxed{(\mathbf{X}^T\mathbf{X})^{-1}}\mathbf{X}^T} \\
&\qquad + \tfrac{1}{n}\mathbf{X}\boxed{(\mathbf{X}^T\mathbf{X})^{-1}}\boxed{\mathbf{X}^T\mathbf{J}}
\end{aligned}$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{n} & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & (\mathbf{X}_c^T\mathbf{X}_c)^{-1} & \\ 0 & & & \end{pmatrix} \qquad \mathbf{X}^T\mathbf{J} = \begin{pmatrix} n & \cdots & n \\ 0 & & \mathbf{0} \\ \vdots & & \vdots \\ 0 & \mathbf{0} & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & x_{11}-\bar{x}_1 & \cdots & x_{1p}-\bar{x}_p \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1}-\bar{x}_1 & \cdots & x_{np}-\bar{x}_p \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ 0 & & \mathbf{0} & 0 \\ \vdots & \mathbf{0} & & \vdots \\ 0 & & & \end{pmatrix} = \mathbf{J}$$

$$= 0$$

$\Rightarrow$

indep

$$\Rightarrow \quad \frac{\text{Total S.S.}}{\sigma^2} = \frac{\text{Res S.S.}}{\sigma^2} + \frac{\text{Reg S.S}}{\sigma^2}$$

$$\sim \chi^2_{(n-1,\lambda)} \qquad \sim \chi^2_{(n-p')} \qquad \overset{\text{By Theorem 4.2}}{\Longrightarrow} \quad \frac{\text{Reg S.S.}}{\sigma^2} \sim \chi^2_{(p,\lambda)}$$

where $\lambda = \frac{1}{\sigma^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \beta_i \beta_j S_{x_i x_j}$

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ (all reg. coeff. $= 0$)

Define $F = \dfrac{\text{Reg S.S.}/p}{\text{Res S.S.}/n-p'} \sim F(p, n-p', \lambda)$

Non-centrality constant in Reg S.S.

Under H_0, $\dfrac{\text{Reg S.S}/p}{\text{Res S.S}/n-p'} \sim F(p, n-p')$

$\dfrac{\text{Reg S.S}}{\sigma^2} \sim \chi^2(p,\lambda)$

$\Rightarrow E\left(\dfrac{\text{Reg S.S.}}{\sigma^2}\right) = p + \lambda$

indep.

$\dfrac{\text{Res S.S.}}{\sigma^2} \sim \chi^2(n-p')$

$\Rightarrow E\left(\dfrac{\text{Res S.S.}}{\sigma^2}\right) = n - p'$

MATH 2421

$E\left(\frac{u}{v}\right) = E(u) * E\left(\frac{1}{v}\right)$

if u & v are indep.

Since $F = \dfrac{\text{Reg S.S.}/p}{\text{Res S.S.}/(n-p')}$

$$E(F) \approx \frac{E(\text{Reg S.S.}/\sigma^2 p)}{E(\text{Res S.S}/\sigma^2(n-p'))}$$

$$= \frac{\frac{1}{p}(p+\lambda)}{\frac{1}{n-p'}(n-p')} = 1 + \frac{\lambda}{p}$$

where $\lambda = \frac{1}{\sigma^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \beta_i \beta_j S_{x_i x_j}$

Under H_0 : $\lambda = 0 \Rightarrow E(F) \approx 1$

$\Rightarrow$ F is not a ~~large~~ value much greater than 1

Under H_1 : $E(F) > 1$

$\Rightarrow F \gg 1 \Rightarrow$ Reject H_0

$\Rightarrow$ at least one of the regression coeff. is not equal to zero.

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

$$F = \frac{\text{Reg S.S.}/p}{\text{Res S.S.}/(n-p')} > \underline{F_{(p,\, n-p',\, \alpha)}}$$

↑ dist. of test stat. under H_0

$\Rightarrow$ Reject H_0

β_0 unknown

$$\text{Res S.S.} = \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{\text{total S.S.}} - \underbrace{(\hat{\beta}_1 S_{x_1y} + \cdots + \hat{\beta}_p S_{x_py})}_{\text{Reg S.S.}}$$

e.g. Example in simple linear regression (p. 3) $p = 1$

$H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ (two-sided alternative)

$$F = \frac{\text{Reg S.S.}/1}{\text{Res S.S.}/(n-2)} = \frac{\hat{\beta}_1 S_{x_1y}}{\hat{\sigma}^2} = \frac{\hat{\beta}_1^2 S_{x_1x_1}}{\hat{\sigma}^2} = \frac{\hat{\beta}_1^2}{\hat{\sigma}^2/S_{x_1x_1}} = \underline{\left(\frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{S_{x_1x_1}}}\right)^2}$$

↑ $\hat{\beta}_1 = \dfrac{S_{x_1y}}{S_{x_1x_1}}$

$\sim F(1, n-2)$

$$\boxed{F(1, v) = t_v^2}$$

$\hat{\beta}_1 = 2.9303$

$\hat{\sigma} = 0.538$

$S_{x_1x_1} = 13.10$

test stat. for testing $H_0 : \beta_1 = 0$ by t-test

$$t = \frac{2.9303}{0.538/\sqrt{13.10}} = \quad t_{\alpha/2,\, n-2} \qquad n=9$$

$$F = \leftarrow \text{Reg S.S} = \hat\beta_1 S_{x_1 y}$$

$$= \hat\beta_1^2 S_{x_1 x_1}$$

$$= (2.9303)^2 * 13.10 = 112.48$$

$$F = \frac{\text{Reg S.S}/1}{\hat\sigma^2} = \frac{112.48}{(0.538)^2}$$

For any p $\quad H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

$$F = \frac{(\hat\beta_1 S_{x_1 y} + \cdots + \hat\beta_p S_{x_p y})/p}{\hat\sigma^2}$$

ANOVA table
(p.26)

$\hat\beta_1 S_{x_1 y} + \cdots + \hat\beta_p S_{x_p y}$

Reg S.S. = total S.S. − Res S.S.

Source	S.S.	~~MS~~ df	M S	F ~~val~~	p-value
Regression	$\sum_{i=1}^{n}(\hat y_i - \bar y)^2$	p	Reg S.S. / p	$\frac{\text{Reg S.S}/p}{\text{Res S.S}/(n-p')}$	
Residual	$\sum_{i=1}^{n}(y_i - \hat y_i)^2$	$n - p'$	Res S.S. / $(n-p')$		
Total	$\sum_{i=1}^{n}(y_i - \bar y)^2$	$n-1$	*		

mean squares = $\frac{\text{S.S.}}{\text{d.f.}}$

$\cancel{\underset{\sim}{Y}^T \underset{\sim}{Y} - \underset{\sim}{X}}$

$$\underset{\sim}{Y}^T(\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)\underset{\sim}{Y}$$

$$\| $$

$$\underset{\sim}{Y}^T\underset{\sim}{Y} - \hat{\underset{\sim}{\beta}}^T \underset{\sim}{X}^T \underset{\sim}{Y}$$

\# of indep. variables or \# of β's (not include intercept)

$p' = p + 1$

intercept

(*) Reg M.S. + Res M.S. $\neq$ Total M.S.

Critic value $= F_{\alpha,1,7}$

Reject H_0, $\Rightarrow$ $\beta_1 \neq 0$

Example 5: Example in Multiple Linear Regression (cont.)

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

$$\begin{aligned} \text{Res.S.S.} &= S_{yy} - \hat{\beta}_1 S_{x_1 y} - \ldots - \hat{\beta}_p S_{x_p y} \\ \Rightarrow \text{Reg.S.S.} &= \hat{\beta}_1 S_{x_1 y} + \ldots + \hat{\beta}_p S_{x_p y} \end{aligned}$$

Reg S.S. $= \hat{\beta}_1 S_{x_1y} + \cdots + \hat{\beta}_p S_{x_py}$

Or

OR Reg S.S. = total S.S. − Res S.S.

$$\text{Res.S.S.} = \sum_{i=1}^n {y_i}^2 - \hat{\beta}_0 \sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n x_{i1} y_i - \ldots - \hat{\beta}_p \sum_{i=1}^n x_{ip} y_i$$

$$\Rightarrow \text{Reg.S.S.} = \hat{\beta}_0 \sum_{i=1}^n y_i + \hat{\beta}_1 \sum_{i=1}^n x_{i1} y_i + \ldots + \hat{\beta}_p \sum_{i=1}^n x_{ip} y_i - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

Source	Sum of squares (S.S.)	d.f.	Mean Squares (M.S.)	F
Regression	399.45437	12-9=3	399.45437/3 = 133.15146	$F = \frac{133.15146}{4.29738} = 30.98$
Residual	38.6764	9	38.6764/9 = 4.29738	
Total	438.13	12		

From p.17

> 3.86

Reject Ho

$n - p' = 13 - 4 = 9$

$F_\alpha(3,9) = 3.86$

when $\alpha = 0.05$

Another view

Under $H_0 : \beta_1 = \ldots = \beta_p = 0$, find β_0 s.t. $\sum_{i=1}^n (y_i - \beta_0)^2$ is minimized.

$$\begin{aligned} \frac{\partial \sum_{i=1}^n (y_i - \beta_0)^2}{\partial \beta_0} &= 2 \sum_{i=1}^n (y_i - \beta_0)(-1) \\ &= 0 \\ &\Rightarrow \hat{\beta}_0 = \bar{y} \\ \text{fitted value} &= \hat{\beta}_0 = \bar{y} \qquad \forall i \end{aligned}$$

$$\Rightarrow \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}|_{H_0})^2 = \text{Res S.S. under } H_0$$

$$\Rightarrow \textit{Reg. S.S.} = \textit{Res S.S.}|_{H_0} - \textit{Res S.S.}$$

Example 4: Intercept is known (cont.)

For $p = 1$

Under the model :

$$y_i' = \beta_1 x_{i1} + e_i \qquad \text{where } y_i' = y_i - \beta_0 \quad \text{and} \quad \hat{y}_i' = \hat{y}_i - \beta_0$$

RSS for the model :

$$\sum_{i=1}^n (y_i' - \hat{y}_i')^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

5

β_0 known

Model : $\underbrace{y_i - \beta_0}_{y_i'} = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$

$$\text{Res S.S.} = \boxed{\sum_{i=1}^{n} y_i'^2} - (\hat\beta_1 \sum_{i=1}^{n} x_{i1} y_i' + \cdots + \hat\beta_p \sum_{i=1}^{n} x_{ip} y_i')$$

Another view on total S.S. (Why is $\sum_{i=1}^{n} y_i'^2$ total S.S. when β_0 is known ?)

For β_0 is unknown

$$\text{total S.S.} = \sum_{i=1}^{n} (y_i - \boxed{\bar y})^2 \quad \text{(corrected total S.S.)}$$

Under the model with all regression coeff = 0

Model : $y_i = \beta_0 + e_i \qquad i = 1 —, n$

$\hat\beta_0 = \boxed{\bar y} \Rightarrow \hat y_i = \hat\beta_0 = \bar y$ for $i = 1, —, n$

$$\text{total S.S.} = \sum_{i=1}^{n} (y_i - \hat y_i |_{\text{all reg. coeff} = 0})^2$$

For β_0 is known

Under the model with all regression coeff. = 0

Model : $y_i' = e_i \qquad i = 1, —, n$

$\hat y_i' = 0$ because $E(e_i) = 0$

$$\text{total S.S.} = \sum_{i=1}^{n} (y_i' - \hat y_i' |_{\text{all reg. coefficient} = 0})^2$$

$$= \sum_{i=1}^{n} y_i'^2 \quad \text{(uncorrected total S.S)}$$

where $y_i' = y_i - \beta_0$

Since

$$\text{Res S.S.} = \underbrace{\boxed{\sum_{i=1}^{n} y_i'^2}}_{\text{total S.S.}} - (\underbrace{\hat\beta_1 \sum_{i=1}^{n} x_{i1} y_i' + \cdots + \hat\beta_p \sum_{i=1}^{n} x_{ip} y_i'}_{\text{Reg S.S.}})$$

$$\text{Res S.S.} = \underset{\sim}{Y}^T (\underset{\sim}{I} - \underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}\underset{\sim}{X}^T)\underset{\sim}{Y}$$

$$= \underset{\sim}{Y}^T\underset{\sim}{Y} - \boxed{\underset{\sim}{Y}^T\underset{\sim}{X}(\underset{\sim}{X}^T\underset{\sim}{X})^{-1}}\underset{\sim}{X}^T\underset{\sim}{Y}$$

$$= \underbrace{\boxed{\underset{\sim}{Y}^T\underset{\sim}{Y}}}_{\text{total S.S}} - \underbrace{\boxed{\hat{\underset{\sim}{\beta}}^T\underset{\sim}{X}^T\underset{\sim}{Y}}}_{\text{Reg S.S.}}$$

~~ANOVA~~ ~~F~~

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$

$$F = \frac{\text{Reg S.S.}/p}{\text{Res S.S.}/(n-p)} > F_\alpha(p, n-p)$$

Reject H_0

$$\text{Res S.S.} = \sum_{i=1}^{n} y_i'^2 - \left(\hat{\beta}_1 \sum_{i=1}^{n} x_{i1} y_i' + \cdots + \hat{\beta}_p \sum_{i=1}^{n} x_{ip} y_i'\right)$$

$$\text{total S.S.} = \sum_{i=1}^{n} y_i'^2$$

$$\text{Reg S.S.} = \text{total S.S.} - \text{Res S.S.}$$

ANOVA p.28

Source	S.S.	df	M.S	F	p-value
Regression	$\hat{\underset{\sim}{\beta}}^T\underset{\sim}{X}^T\underset{\sim}{Y}$	p	$\hat{\underset{\sim}{\beta}}^T\underset{\sim}{X}^T\underset{\sim}{Y}/p$		
Residual	$\underset{\sim}{Y}^T\underset{\sim}{Y} - \hat{\underset{\sim}{\beta}}^T\underset{\sim}{X}^T\underset{\sim}{Y}$	$n-p$	$\dfrac{\underset{\sim}{Y}^T\underset{\sim}{Y} - \hat{\underset{\sim}{\beta}}^T\underset{\sim}{X}^T\underset{\sim}{Y}}{n-p}$		
total	$\underset{\sim}{Y}^T\underset{\sim}{Y}$	n			

Model

Section 4.2.2 Subset of regression coefficients

e.g. X_1, X_2, X_3, X_4

$H_0 = \beta_3 = \beta_4 = 0$ $\qquad\qquad R(\beta_1, \beta_2 \mid \beta_0)$

Model under H_0 : Intercept, X_1, X_2 $\qquad$ Res S.S. $\mid H_0$ $\qquad$ Reg S.S. $\mid H_0$

$+ X_3, X_4$ $\qquad\qquad\qquad\qquad\qquad\qquad \vee \qquad\qquad\qquad \wedge$

under H_1 = Intercept, X_1, X_2, $\boxed{X_3, X_4}$ $\qquad$ Res S.S. $\mid H_1$ $\qquad$ Reg S.S. $\mid H_1$

$R(\beta_1, \beta_2, \beta_3, \beta_4 \mid \beta_0)$

$\uparrow \uparrow \leftarrow$ more indep. variables $\qquad$ Res S.S. $\downarrow$ $\qquad$ Reg S.S. $\uparrow$

Increase in Reg S.S. = Reg S.S. $\mid H_1$ − Reg S.S. $\mid H_0$

① Reg S.S.

$$= R(\beta_1, \beta_2, \beta_3, \beta_4 \mid \beta_0) - R(\beta_1, \beta_2 \mid \beta_0)$$

②

$$= R(\beta_3, \beta_4 \mid \beta_0, \beta_1, \beta_2)$$

increase in Reg S.S.

Distribution of Increase in Reg S.S.

Full model : $\underset{\sim}{Y} = (\underset{\sim}{j}, \underset{\sim}{X_r}, \underset{\sim}{X_s}) \begin{pmatrix} \beta_0 \\ \underset{\sim}{\beta_r} \\ \underset{\sim}{\beta_s} \end{pmatrix} + \underset{\sim}{e}$

(model under H_1)

$$\underset{\sim}{j} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \underset{\sim}{X_r} = \begin{pmatrix} X'_{13} & X'_{14} \\ \vdots & \vdots \\ X'_{n3} & X'_{n4} \end{pmatrix}, \quad \underset{\sim}{X_s} = \begin{pmatrix} X'_{11} & X'_{12} \\ \vdots & \vdots \\ X'_{n1} & X'_{n2} \end{pmatrix}$$

$$\underset{\sim}{\beta_r} = \begin{pmatrix} \beta_3 \\ \beta_4 \end{pmatrix}, \quad \underset{\sim}{\beta_s} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

where $X'_{ij} = X_{ij} - \bar{X}_j$

$i = 1, \dots, n$

$j = 1, 2, 3, 4$

$$= \beta_0 \underset{\sim}{j} + \underset{\sim}{X_r}\underset{\sim}{\beta_r} + \underset{\sim}{X_s}\underset{\sim}{\beta_s} + \underset{\sim}{e}$$

Reduced model : $\underset{\sim}{Y} = (\underset{\sim}{j}, \underset{\sim}{X_s}) \begin{pmatrix} \beta_0^* \\ \underset{\sim}{\beta_s^*} \end{pmatrix} + \underset{\sim}{e}^*$

(model under H_0)

$$= \beta_0^* \underset{\sim}{j} + \underset{\sim}{X_s}\underset{\sim}{\beta_s^*} + \underset{\sim}{e}^*$$

$H_0 = \beta_3 = \beta_4 = 0$

$\Rightarrow H_0 = \underset{\sim}{\beta_r} = \underset{\sim}{0}$

Increase in Reg S.S. = Reg S.S. (full) − Reg S.S. (reduced)

$$= \underset{\sim}{Y}^T \left(\underset{\sim}{X_f} (\underset{\sim}{X_f}^T \underset{\sim}{X_f})^{-1} \underset{\sim}{X_f}^T - \underset{\sim}{X_s} (\underset{\sim}{X_s}^T \underset{\sim}{X_s})^{-1} \underset{\sim}{X_s}^T \right) \underset{\sim}{Y}$$

$$= - - - - - -$$

$$\underset{\sim}{X_f} = (\underset{\sim}{X_r}, \underset{\sim}{X_s})$$

$$= \underset{\sim}{Y}^T \underset{\sim}{A} \underset{\sim}{Y}$$

Reg S.S. | full model = Increase in Reg S.S. + Reg S.S. (reduced)

$\underset{\sim}{Y}^T \underset{\sim}{A} \underset{\sim}{Y}$ $\qquad$ $\underset{\sim}{Y}^T \underset{\sim}{B} \underset{\sim}{Y}$

$$\underset{\sim}{B} = \underset{\sim}{X_s} (\underset{\sim}{X_s}^T \underset{\sim}{X_s})^{-1} \underset{\sim}{X_s}^T$$

Prove : $\underset{\sim}{A}^T \underset{\sim}{B} = \underset{\sim}{0}$

$\Rightarrow$ Increase in Reg S.S. & Reg S.S. (reduced) are indep.

$\sim \chi^2(4, \lambda)$ $\qquad$ $\sim \chi^2(2, \lambda_1)$

# of indep. variables in full model X_1, X_2, X_3, X_4

# of indep. variables in reduced model X_1, X_2

$$\sim \chi^2(4-2, \lambda - \lambda_1)$$

$H_0 : \underset{\sim}{\beta_r} = \underset{\sim}{0}$

$$F = \frac{\text{Increase in Reg S.S.} / r}{\text{Res S.S.} / (n-p')} \sim F(r, n-p', \lambda - \lambda_1)$$

indep