6. Regression Analysis

6.1 Simple linear regression for normal theory Gauss-Markov models.

Model 1:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where $\epsilon_i \sim NID(0, \sigma^2)$ for $i = 1, \ldots, n$.

Matrix formulation:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

The OLS estimator (b.l.u.e.) for $\boldsymbol{\beta}$ is

$$\mathbf{b} = \underline{(X^T X)}^{-1} X^T \mathbf{Y}$$

$\uparrow$ when does this exist?

Here

$$X^T X = \begin{bmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{bmatrix}, \quad X^T \mathbf{Y} = \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{bmatrix}$$

$(X^TX)^{-1}$

$$= \frac{1}{n\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{bmatrix}$$

$$= \frac{1}{n\sum_{i=1}^n (X_i - \bar{X})^2} \begin{bmatrix} \sum_{i=1}^n X_i^2 & -n\bar{X} \\ -n\bar{X} & n \end{bmatrix}$$

Then

$$
\begin{aligned}
\mathbf{b} =& (X^T X)^{-1} X^T \mathbf{Y} \\
=& \frac{1}{n \sum_{i=1}^{n} (X_i - \bar{X})^2} \left[\begin{array}{c} \left(\sum_{i=1}^{n} X_i^2\right)\left(\sum_{i=1}^{n} Y_i\right) - n\bar{X} \sum_{i=1}^{n} X_i Y_i \\ -n\bar{X} \sum_{i=1}^{n} Y_i + n \sum_{i=1}^{n} X_i Y_i \end{array} \right]
\end{aligned}
$$

and

$$
\mathbf{b} = \left[\begin{array}{c} b_0 \\ b_1 \end{array} \right] = \left[\begin{array}{c} \bar{Y} - b_1 \bar{X} \\ \dfrac{\sum_{i=1}^{n} (X_i - \bar{X}) Y_i}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \end{array} \right]
$$

Covariance matrix:

$$
\begin{aligned}
Var(\mathbf{b}) &= Var\left[(X^TX)^{-1}X^T\mathbf{Y}\right] \\
&= (X^TX)^{-1}X^T(\sigma^2 I)X(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1} \\
&= \sigma^2 \begin{bmatrix} \frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X_i-\bar{X})^2} & \frac{-\bar{X}}{\Sigma(X_i-\bar{X})^2} \\ \frac{-\bar{X}}{\Sigma(X_i-\bar{X})^2} & \frac{1}{\Sigma(X_i-\bar{X})^2} \end{bmatrix}
\end{aligned}
$$

Estimate the covariance matrix for **b** as

$$S_{\mathbf{b}} = \mathrm{MSE}\,(X^TX)^{-1}$$

where

$$\mathrm{MSE} = \mathrm{SSE}/(n-2) = \frac{1}{n-2}\mathbf{Y}^T(I - P_X)\mathbf{Y}.$$

Analysis of Variance:

$$
\begin{aligned}
\sum_{i=1}^{n} Y_i^2 &= \mathbf{Y}^T\mathbf{Y} \\
&= \mathbf{Y}^T(I - P_X + P_X - P_{\mathbf{1}} + P_{\mathbf{1}})\mathbf{Y} \\
&= \mathbf{Y}^T(I - P_X)\mathbf{Y} + \mathbf{Y}^T(P_X - P_{\mathbf{1}})\mathbf{Y} + \mathbf{Y}^T P_{\mathbf{1}}\mathbf{Y}
\end{aligned}
$$

$\nearrow$	$\nearrow$	$\uparrow$
SSE	Corrected model sum of squares	Correction for the "mean"
	$\uparrow$	$\uparrow$
	call this $R(\beta_1\|\beta_0)$	call this $R(\beta_0)$

(i) By Cochran's Theorem, these three sums of squares are multiples of independent chi-squared random variables.

(ii) By result 4.7, $\frac{1}{\sigma^2}\,\mathrm{SSE} \sim \chi^2_{(\mathrm{n}-2)}$ if the model is correctly specified.

Notation:

Reduction in residual sum of squares:

$$R(\beta_{k+1}, \ldots, \beta_{k+q} \mid \beta_0, \beta_1, \ldots, \beta_k)$$

$$= \underbrace{\mathbf{Y}^T(I - P_{X_1})\mathbf{Y}}_{\substack{\uparrow \\ \text{sum of squared} \\ \text{residuals for the} \\ \text{smaller model}}} - \underbrace{\mathbf{Y}^T(I - P_X)\mathbf{Y}}_{\substack{\uparrow \\ \text{sum of squared} \\ \text{residuals for the} \\ \text{larger model}}}$$

Here

$$X = [\ \underbrace{X_1}_{\substack{\nearrow \\ \text{columns} \\ \text{corresponding} \\ \text{to } \beta_0, \beta_1, \ldots, \beta_k}} \mid \underbrace{X_2}_{\substack{\nwarrow \\ \text{columns} \\ \text{corresponding} \\ \text{to } \beta_{k+1} \cdots \beta_{k+q}}}\ ]$$

Correction for the overall mean:

$$
\begin{aligned}
R(\beta_0) &= \mathbf{Y}^T P_{\mathbf{1}} \mathbf{Y} \\
&= \mathbf{Y}^T (I - I + P_{\mathbf{1}}) \mathbf{Y} \\
&= \mathbf{Y}^T I \mathbf{Y} - \mathbf{Y}^T (I - P_{\mathbf{1}}) \mathbf{Y} \\
&= \sum_{i=1}^{n} (Y_i - 0)^2 - \sum_{i=1}^{n} (Y_i - \bar{Y})^2
\end{aligned}
$$

$\nearrow$

Sum of squared residuals from fitting the model $Y_i = \alpha + \epsilon_i$.

The OLS estimator for $\alpha = E(Y_i)$ is

$$
\begin{aligned}
\hat{\alpha} &= (\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{Y} \\
&= (n)^{-1} \left(\sum_{i=1}^{n} Y_i \right) \\
&= \bar{Y}
\end{aligned}
$$

An alternative formula

$$
\begin{aligned}
R(\beta_0) &= \mathbf{Y}^T \mathbf{P_1} \mathbf{Y} \\
&= \mathbf{Y}^T \mathbf{1}(\mathbf{1}^T \mathbf{1})^{-1} \mathbf{1}^T \mathbf{Y} \\
&= \left(\sum_{i=1}^{n} Y_i \right) (n)^{-1} \left(\sum_{i=1}^{n} Y_i \right) \\
&= (n)^{-1} \left(\sum_{i=1}^{n} Y_i \right)^2 \\
&= n\bar{Y}^2
\end{aligned}
$$

with $df = rank(P_{\mathbf{1}}) = rank(\mathbf{1}) = 1$.

Reduction in the residual sum of squares for regression on X_1:

$$\begin{aligned} R(\beta_1|\beta_0) &= \mathbf{Y}^T(P_X - P_{\mathbf{1}})\mathbf{Y} \\ &= \mathbf{Y}^T(P_X - I + I - P_{\mathbf{1}})\mathbf{Y} \\ &= \mathbf{Y}^T(I - P_{\mathbf{1}} - (I - P_X))\mathbf{Y} \\ &= \mathbf{Y}^T(I - P_{\mathbf{1}})\mathbf{Y} - \mathbf{Y}^T(I - P_X)\mathbf{Y} \end{aligned}$$

$\nearrow$ sum of squared residuals for fitting the model $Y_i = \alpha + \epsilon_i$

$\uparrow$ sum of squared residuals for fitting the model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

ANOVA table:

Source of variation	d.f.	Sum of Squares
Regression on X	1	$R(\beta_1 \vert \beta_0) = \mathbf{Y}^T(P_X - P_{\mathbf{1}})\mathbf{Y}$
Residuals	$n-2$	$\mathbf{Y}^T(I - P_X)\mathbf{Y}$
Corrected total	$n-1$	$\mathbf{Y}^T(I - P_{\mathbf{1}})\mathbf{Y}$
Correction for the mean	1	$\mathbf{Y}^T P_{\mathbf{1}}\mathbf{Y} = n\bar{Y}^2$

F-tests

From result 4.7 we have

$$\frac{1}{\sigma^2}\ R(\beta_0) = \frac{1}{\sigma^2}\ \mathbf{Y}^T P_{\mathbf{1}} \mathbf{Y} \sim \chi_1^2(\delta^2)$$

where

$$\delta^2 = \frac{1}{\sigma^2}\ \boldsymbol{\beta}^T \mathbf{X}^T P_{\mathbf{1}} \mathbf{X} \boldsymbol{\beta} = \frac{1}{\sigma^2}\ \boldsymbol{\beta}^T \mathbf{X}^T P_{\mathbf{1}}^T P_{\mathbf{1}} X \boldsymbol{\beta}$$

$$= \frac{1}{\sigma^2} (P_{\mathbf{1}} X \boldsymbol{\beta})^T (P_{\mathbf{1}} X \boldsymbol{\beta}) = \frac{n}{\sigma^2} (\beta_0 + \beta_1 \bar{X})^2$$

Hypothesis test: Reject $H_0 : \beta_0 + \beta_1 \bar{X} = 0$ if

$$F = \frac{R(\beta_0)}{\mathrm{MSE}} > F_{(1,n-2)}, \alpha$$

Also use Result 4.7 to show that

$$\frac{1}{\sigma^2} SSE = \frac{1}{\sigma^2} \, \mathbf{Y}^T (I - P_X) \mathbf{Y} \sim \chi^2_{(n-2)}$$

and, use Result 4.8 to show that

$$SSE = \frac{1}{\sigma^2} \, \mathbf{Y}^T (I - P_X) \mathbf{Y}$$

is distributed independently of

$$R(\beta_0) = \frac{1}{\sigma^2} \, \mathbf{Y}^T P_{\mathbf{1}} \mathbf{Y} \ .$$

This follows from

$$(I - P_X) P_{\mathbf{1}} = 0 \ .$$

Consequently,

$$F = \frac{R(\beta_0)}{\mathrm{MSE}} \sim F_{(1,n-2)}(\delta^2)$$

and this becomes a central F-distribution when the null hypothesis is true.

Test the null hypothesis $H_0 : \beta_1 = 0$

$$
\begin{aligned}
F &= \frac{R(\beta_1|\beta_0)/1}{\mathrm{MSE}} \\
&= \frac{[\mathbf{Y}^T(P_X - P_{\mathbf{1}})\mathbf{Y}]/[1\sigma^2]}{[\mathbf{Y}^T(I - P_X)\mathbf{Y}]/[(n-2)\sigma^2]} \\
&\sim F_{(1,n-2)}(\delta^2)
\end{aligned}
$$

where

$$
\begin{aligned}
\delta^2 &= \frac{1}{\sigma^2}\,\boldsymbol{\beta}^T X^T (P_X - P_{\mathbf{1}}) X\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\,\boldsymbol{\beta}^T X^T (P_X - P_{\mathbf{1}})^T\, \underline{(P_X - P_{\mathbf{1}}) X\boldsymbol{\beta}}
\end{aligned}
$$

The null hypothesis is $H_0 : (P_X - P_{\mathbf{1}})X\boldsymbol{\beta} = \mathbf{0}$

Here

$$
\begin{aligned}
(P_X - P_{\mathbf{1}})X &= (P_X - P_{\mathbf{1}})[\mathbf{1}|\mathbf{X}] = \left[(P_X - P_{\mathbf{1}})\mathbf{1} \,\middle|\, (P_X - P_{\mathbf{1}})\mathbf{X}\right] \\
&= \left[P_X\mathbf{1} - P_{\mathbf{1}}\mathbf{1} \,\middle|\, P_X\mathbf{X} - P_{\mathbf{1}}\mathbf{X}\right] = \left[\mathbf{1} - \mathbf{1} \,\middle|\, \mathbf{X} - \bar{X}\mathbf{1}\right] \\
&= \begin{bmatrix} 0 & | & X_1 - \bar{X} \\ 0 & | & X_2 - \bar{X} \\ \vdots & | & \vdots \\ 0 & | & X_n - \bar{X} \end{bmatrix}
\end{aligned}
$$

If any $X_i \neq X_j$, then we cannot have both

$$X_j - \bar{X} = 0$$

and

$$X_i - \bar{X} = 0 \; .$$

Consequently, if any $X_i \neq X_j$ then

$$(P_X - P_{\mathbf{1}})X\boldsymbol{\beta} = 0$$

if and only if

$$\beta_1 = 0 \ .$$

Hence, the null hypothesis is

$$H_0 : \beta_1 = 0.$$

Note that

$$\delta^2 = \frac{1}{\sigma^2} \ \beta_1^2 \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Maximize the power of the F-test for

$$H_0 : \beta_1 = 0 \quad vs. \; H_A : \beta_1 \neq 0$$

by maximizing

$$\delta^2 = \frac{1}{\sigma^2} \; \beta_1^2 \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Reparameterize the model:

$$Y_i = \alpha + \beta_1 (X_i - \bar{X}) + \epsilon_i$$

with $\epsilon_i \sim NID(0, \sigma^2), i = 1, \ldots, n$.

Interpretation of parameters:

$\alpha = E(Y)$ when $X = \bar{X}$

β_1 is the change in $E(Y)$ when X is increased by one unit.

Matrix formulation:

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 - \bar{X} \\ \vdots & \vdots \\ 1 & X_n - \bar{X} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

or

$$Y = W\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

Clearly,

$$W = X \begin{bmatrix} 1 & -\bar{X} \\ 0 & 1 \end{bmatrix} = XF$$

$$X = W \begin{bmatrix} 1 & \bar{X} \\ 0 & 1 \end{bmatrix} = WG$$

For this reparameterization, the columns of W are orthogonal and

$$W^T W = \begin{bmatrix} n & 0 \\ 0 & \sum_{i=1}^{n}(X_i - \bar{X})^2 \end{bmatrix}$$

$$(W^T W)^{-1} = \begin{bmatrix} \frac{1}{n} & 0 \\ 0 & \frac{1}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

$$W^T \mathbf{Y} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n}(X_i - \bar{X}) Y_i \end{bmatrix}$$

Then,

$$\hat{\gamma} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = (W^T W)^{-1} W^T \mathbf{Y}$$

$$= \begin{bmatrix} \bar{Y} \\ \frac{\Sigma(X_i - \bar{X}) Y_i}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

and

$$Var(\hat{\gamma}) = \sigma^2 (W^T W)^{-1}$$

$$= \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^2}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

Hence, $\bar{Y}$ and $\hat{\beta}_1 = \frac{\Sigma(X_i - \bar{X}) Y_i}{\Sigma(X_i - \bar{X})^2}$ are uncorrelated (independent for the normal theory Gauss-Markov model).

Analysis of variance:

The reparamterization does not change the ANOVA table. Note that

$$
\begin{aligned}
P_X &= X(X^T X)^{-1} X^T \\
&= W(W^T W)^{-1} W^T = P_W
\end{aligned}
$$

and

$R(\beta_0) + R(\beta_1|\beta_0) + \text{SSE}$

$$= \mathbf{Y}^T P_{\mathbf{1}} \mathbf{Y} + \mathbf{Y}^T (P_X - P_{\mathbf{1}}) \mathbf{Y} + \mathbf{Y}^T (I - P_X) \mathbf{Y}$$

$$= \mathbf{Y}^T P_{\mathbf{1}} \mathbf{Y} + \mathbf{Y}^T (P_W - P_{\mathbf{1}}) \mathbf{Y} + \mathbf{Y}^T (I - P_W) \mathbf{Y}$$

$$= R(\alpha) + R(\beta_1|\alpha) + \text{SSE}$$

6.2 Multiple regression analysis for the normal-theory G-M model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_r X_{ri} + \epsilon_i$$

where

$$\epsilon_i \sim NID(0, \sigma^2), \quad \text{for} \quad i = 1, \ldots, n\,.$$

Matrix formulation:

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I)$$

where

$$X\boldsymbol{\beta} = \begin{array}{c} \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{r1} \\ 1 & X_{12} & X_{22} & \cdots & X_{r2} \\ 1 & \vdots & \vdots & \vdots & \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & X_{1n} & X_{2n} & \cdots & X_{rn} \end{bmatrix} \\ \begin{array}{ccccc} \uparrow & \uparrow & \uparrow & & \uparrow \\ \mathbf{1} & \mathbf{X}_1 & \mathbf{X}_2 & & \mathbf{X}_r \end{array} \end{array} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}$$

Suppose $\text{rank}(X) = r + 1$, then

(i) the OLS estimator (b.l.u.e.) for $\boldsymbol{\beta}$ is

$$\mathbf{b} = (X^T X)^{-1} X^T \mathbf{Y}$$

(ii) $Var(\mathbf{b}) = \sigma^2 (X^T X)^{-1}$

(iii) $\hat{\mathbf{Y}} = X\mathbf{b} = X(X^T X)^{-1} X^T \mathbf{Y} = P_X \mathbf{Y}$

(iv) $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (I - P_X)\mathbf{Y}$

(v) By result 4.7,

$$\begin{aligned} \frac{1}{\sigma^2}\,\mathrm{SSE} = \frac{1}{\sigma^2}\,\mathbf{e}^{\mathrm{T}}\mathbf{e} &= \frac{1}{\sigma^2}\,\mathbf{Y}^T (I - P_X)\mathbf{Y} \\ &\sim \chi^2_{(n-r-1)} \end{aligned}$$

(vi) $\text{MSE} = \frac{\mathrm{SSE}}{n-r-1}$ is an unbiased estimator of σ^2.

ANOVA

Source of variation	d.f.	Sum of squares
Model (regression on $X_1, \ldots, X_r$)	r	$R(\beta_1, \ldots, \beta_r \mid \beta_0)$
		$= \mathbf{Y}^T(P_X - P_{\mathbf{1}})\mathbf{Y}$
Error (or residuals)	$n - r - 1$	$\mathbf{Y}^T(I - P_X)\mathbf{Y}$
Corrected total	$n - 1$	$\mathbf{Y}^T(I - P_{\mathbf{1}})\mathbf{Y}$
Correction for the mean	1	$R(\beta_0) = \mathbf{Y}^T P_{\mathbf{1}} \mathbf{Y} = n\bar{Y}^2$

Reduction in the residual sum of squares obtained by regression on $X_1, X_2, \ldots, X_r$ is denoted as

$$R(\beta_1, \beta_2, \ldots, \beta_r \,\Big|\, \beta_0)$$

$$= \mathbf{Y}^T(I - P_{\mathbf{1}})\mathbf{Y} - \mathbf{Y}^T(I - P_X)\mathbf{Y}$$

$$= \mathbf{Y}^T(\mathbf{P}_X - P_{\mathbf{1}})\mathbf{Y}$$

Use Cochran's theorem or results 4.7 and 4.8 to show that SSE is distributed independently of

$$R(\beta_1, \beta_2, \ldots, \beta_r | \beta_0) = SS_{\text{model}}$$

and

$$\frac{1}{\sigma^2} SSE \sim \chi^2_{(n-r-1)}$$

Then

$$F = \frac{R(\beta_1, \ldots, \beta_r|\beta_0)/r}{\mathrm{MSE}} \sim F_{(r,n-r-1)}(\delta^2)$$

where

$$\begin{aligned}
\delta^2 &= \frac{1}{\sigma^2}\,\boldsymbol{\beta}^T \mathbf{X}^T (P_X - P_{\mathbf{1}}) X\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\,\boldsymbol{\beta}^T X^T (P_X - I + I - P_{\mathbf{1}}) X\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\,\left[\boldsymbol{\beta}^T X^T (I - P_{\mathbf{1}}) X\boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \underline{(I - P_X)} X\boldsymbol{\beta}\right] \\
&= \frac{1}{\sigma^2}\,\boldsymbol{\beta}^T X^T (I - P_{\mathbf{1}}) X\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\,\boldsymbol{\beta}^T X^T (I - P_{\mathbf{1}})(I - P_{\mathbf{1}}) X\boldsymbol{\beta} \\
&= \frac{1}{\sigma^2}\,[(I - P_{\mathbf{1}}) X\boldsymbol{\beta}]^T (I - P_{\mathbf{1}}) X\boldsymbol{\beta}
\end{aligned}$$

Note that

$$
\begin{aligned}
(I - P_{\mathbf{1}})X &= \left[(I - P_{\mathbf{1}})\mathbf{1} \middle| (I - P_{\mathbf{1}})\mathbf{X}_1 \middle| \cdots \middle| (I - P_{\mathbf{1}})\mathbf{X}_r \right] \\
&= \left[\; \mathbf{0} \; \middle| \; \mathbf{X}_1 - \bar{X}_1\mathbf{1} \; \middle| \cdots \middle| \; \mathbf{X}_r - \bar{X}_r\mathbf{1} \right]
\end{aligned}
$$

$$
\Longrightarrow (I - P_{\mathbf{1}})X\boldsymbol{\beta} = \sum_{j=1}^{r} \beta_j(\mathbf{X}_j - \bar{X}_j\mathbf{1})
$$

$$
\begin{aligned}
\Longrightarrow \delta^2 &= \frac{1}{\sigma^2} \left[\sum_{j=1}^{r} \beta_j^2 (\mathbf{X}_j - \bar{X}_j\mathbf{1})^T (\mathbf{X}_j - \bar{X}_j\mathbf{1}) \right. \\
&\qquad + \left. \sum\sum_{j \neq k} \beta_j\beta_k (\mathbf{X}_j - \bar{X}_j\mathbf{1})^T (\mathbf{X}_k - \bar{X}_k\mathbf{1}) \right]
\end{aligned}
$$

$$= \frac{1}{\sigma^2}\, \boldsymbol{\beta}_*^T \left[\sum_{i=1}^{n} (\mathbf{X}_{*j} - \bar{\mathbf{X}}_*)(\mathbf{X}_{*i} - \bar{\mathbf{X}}_*)^T \right] \boldsymbol{\beta}_*$$

where

$$\boldsymbol{\beta}_* = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} \qquad \bar{\mathbf{X}}_* = \begin{bmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_r \end{bmatrix} \qquad \mathbf{X}_{*j} = \begin{bmatrix} X_{1j} \\ \vdots \\ X_{rj} \end{bmatrix}$$

If $\sum_{j=1}^{n}(\mathbf{X}_{*j} - \bar{\mathbf{X}}_*)(\mathbf{X}_{*j} - \bar{\mathbf{X}}_*)^T$ is positive definite, then the null hypothesis corresponding to $\delta^2 = 0$ is

$$H_0 : \boldsymbol{\beta}_* = \mathbf{0} \quad (\quad \text{or} \quad \beta_1 = \beta_2 = \cdots = \beta_r = 0)$$

Reject

$$H_0 : \boldsymbol{\beta}_* = \mathbf{0}$$

if

$$F = \frac{\mathbf{Y}^T(P_X - P_{\mathbf{1}})\mathbf{Y}/r}{\mathbf{Y}^T(I - P_X)\mathbf{Y}/(n-r-1)} > F_{(r,n-r-1)}, \alpha$$

Sequential sums of squares (Type I sums of squares in PROC GLM or PROC REG in SAS).

Let

$$X_0 = \mathbf{1} \qquad P_0 = X_0(X_0^T X_0)^{-1} X_0^T$$

$$X_1 = [\mathbf{1}|\mathbf{X}_1] \qquad P_1 = X_1(X_1^T X_1)^{-} X_1^T$$

$$X_2 = [\mathbf{1}|\mathbf{X}_1|\mathbf{X}_2] \qquad P_2 = X_2(X_2^T X_2)^{-} X_2^T$$

$$\vdots \qquad \vdots$$

$$X_r = [\mathbf{1}|\mathbf{X}_1|\cdots|\mathbf{X}_r] \qquad P_r = X_r(X_r^T X_r)^{-} X_r^T$$

$$
\begin{aligned}
\mathbf{Y}^T\mathbf{Y} &= \mathbf{Y}^T P_0 \mathbf{Y} + \mathbf{Y}^T (P_1 - P_0)\mathbf{Y} + \mathbf{Y}^T (P_2 - P_1)\mathbf{Y} \\
&\quad + \cdots + \mathbf{Y}^T (P_r - P_{r-1})\mathbf{Y} + \mathbf{Y}^T (I - P_r)\mathbf{Y} \\
&= R(\beta_0) + R(\beta_1|\beta_0) + R(\beta_2|\beta_0, \beta_1) + \cdots + R(\beta_r|\beta_0, \beta_1, \ldots, \beta_{r-1}) \\
&\quad + \text{SSE}
\end{aligned}
$$

- Use Cochran's theorem to show
  - these sums of squares are distributed independently of each other.
  - Each $\frac{1}{\sigma^2}\ R(\beta_i|\beta_0, \ldots, \beta_{i-1})$ has a chi-squared distribution with one degree of freedom.
- Use Result 4.7 to show $\frac{1}{\sigma^2}$ SSE $\sim \chi^2_{(n-r-1)}$.

Then

$$F = \frac{R(\beta_j|\beta_0, \ldots, \beta_{j-1})/1}{\text{MSE}} \sim F_{1,n-r-1}(\delta^2)$$

where

$$\begin{aligned} \delta^2 &= \frac{1}{\sigma^2}\, \boldsymbol{\beta}^T X^T (P_j - P_{j-1}) X\boldsymbol{\beta} \\ &= \frac{1}{\sigma^2}\, \boldsymbol{\beta}^T X^T (P_j - P_{j-1})^T (P_j - P_{j-1}) X\boldsymbol{\beta} \\ &= \frac{1}{\sigma^2}\, [(P_j - P_{j-1}) X\boldsymbol{\beta}]^T (P_j - P_{j-1}) X\boldsymbol{\beta} \end{aligned}$$

Hence, this is a test of

$$H_0 : (P_j - P_{j-1}) X\boldsymbol{\beta} = \mathbf{0} \quad vs \quad H_a : (P_j - P_{j-1}) X\boldsymbol{\beta} \neq 0$$

Note that

$$
\begin{aligned}
&(P_j - P_{j-1})X \\
=&(P_j - P_{j-1})\left[\mathbf{1} \;\middle|\; \mathbf{X}_1 \;\middle|\; \cdots \;\middle|\; \mathbf{X}_{j-1} \;\middle|\; \mathbf{X}_j \;\middle|\; \cdots \;\middle|\; \mathbf{X}_r\right] \\
=&\left[(P_j - P_{j-1})\mathbf{1} \;\middle|\; (P_j - P_{j-1})\mathbf{X}_1 \;\middle|\; \cdots \;\middle|\; (P_j - P_{j-1})\mathbf{X}_j \;\middle|\; \cdots\right] \\
=&\left[O_{n\times j} \;\middle|\; (P_j - P_{j-1})\mathbf{X}_j \;\middle|\; \cdots \;\middle|\; (P_j - P_{j-1})\mathbf{X}_r\right]
\end{aligned}
$$

Then

$$
\begin{aligned}
(P_j - P_{j-1})X\boldsymbol{\beta} &= \sum_{k=j}^{r} \beta_k (P_j - P_{j-1})X_k \\
&= \beta_j (P_j - P_{j-1})X_j \\
&\qquad + \sum_{k=j+1}^{r} \beta_k (P_j - P_{j-1})X_k
\end{aligned}
$$

and the null hypothesis is

$$
H_0 : \mathbf{0} = \beta_j (P_j - P_{j-1})\mathbf{X}_j + \sum_{k=j+1}^{r} \beta_k (P_j - P_{j-1})\mathbf{X}_k
$$

Type II sums of squares in SAS (these are also Type III and Type IV sums of squares for regression problems).

$$R(\beta_j|\beta_0 \text{ and all other } \beta_k' s) = \mathbf{Y}^T(P_X - P_{-j})\mathbf{Y},$$

where

$$P_{-j} = X_{-j}(X_{-j}^T X_{-j})^{-} X_{-j}^T$$

and X_{-j} is obtained by deleting the $(j+1)$-th column of X.

From the previous discussion:

$$F = \frac{\mathbf{Y}^T (P_X - P_{-j})\mathbf{Y}/1}{\mathrm{MSE}} \sim F_{(1,n-r-1)}(\delta^2)$$

where

$$\begin{aligned} \delta^2 &= \frac{1}{\sigma^2}\, \boldsymbol{\beta}^T X^T (P_X - P_{-j}) X \boldsymbol{\beta} \\ &= \frac{1}{\sigma^2}\, \beta_j^2 \mathbf{X}_j^T (P_X - P_{-j}) \mathbf{X}_j \end{aligned}$$

This F-test provides a test of

$$H_0 : \beta_j = 0 \qquad \text{vs} \qquad H_A : \beta_j \neq 0$$

if $(P_X - P_{-j})\mathbf{X}_j \neq \mathbf{0}$.

Variable	Type I Sums of squares	Type II Sums of squares
X_1	$R(\beta_1 \mid \beta_0)$ $=\mathbf{Y}^T(P_1 - P_0)\mathbf{Y}$	$R(\beta_1 \mid \text{ other } \beta'\text{s})$ $=\mathbf{Y}^T(P_X - P_{-1})\mathbf{Y}$
X_2	$R(\beta_2 \mid \beta_0, \beta_1)$ $= \mathbf{Y}^T(P_2 - P_1)\mathbf{Y}$	$R(\beta_2 \mid \text{ other } \beta\text{'s})$ $= \mathbf{Y}^T(P_X - P_{-2})\mathbf{Y}$
$\vdots$	$\vdots$	$\vdots$
X_r	$R(\beta_r \mid \beta_0, \beta_1, \ldots, \beta_{r-1})$ $= \mathbf{Y}^T(P_r - P_{r-1})\mathbf{Y}$	$R(\beta_r \mid \beta_0, \ldots, \beta_{r-1})$ $= \mathbf{Y}^T(P_X - P_{-r})\mathbf{Y}$
Residuals	$\text{SSE} = \mathbf{Y}^T(I - P_X)\mathbf{Y}$	
Corrected Total	$\mathbf{Y}^T(I - P_{\mathbf{1}})\mathbf{Y}$	

When $X_1, X_2, \ldots, X_r$ are all uncorrelated, then

(i) $R(\beta_j \mid \beta_0 \text{ and any other subset of } \beta\text{'s}) = R(\beta_j|\beta_0)$ and there is only one ANOVA table.

(ii) $R(\beta_j|\beta_0) = \hat{\beta}_j^2 \sum_{i=1}^{n}(X_{ji} - \bar{X}_{j.})^2$

(iii)

$$F = \frac{R(\beta_j|\beta_0)}{\text{MSE}} \sim F_{1,n-k-1}(\delta^2)$$

where

$$\delta^2 = \frac{1}{\sigma^2}\beta_j^2 \sum_{i=1}^{n}(X_{ji} - \bar{X}_{j.})^2$$

and this F-statistic provides a test of

$$H_0 : \beta_j = 0 \quad \text{versus} \quad H_A : \beta_j \neq 0.$$

Testable Hypothesis

For any testable hypothesis, reject $H_0 : C\boldsymbol{\beta} = \mathbf{d}$ in favor of the general alternative $H_A : C\boldsymbol{\beta} \neq \mathbf{d}$ if

$$
\begin{aligned}
F &= \frac{(C\mathbf{b} - \mathbf{d})^T [C(X^T X)^- C^T]^{-1} (C\mathbf{b} - \mathbf{d})/m}{\mathbf{Y}^T (I - P_X)\mathbf{Y}/(n - \text{rank}(X))} \\
&> F_{(m, n - \text{rank}(X)), \alpha}
\end{aligned}
$$

where

$$m = \text{number of rows in } C = \text{rank}(C)$$

and

$$\mathbf{b} = (X^T X)^- X^T \mathbf{Y}$$

Confidence interval for an estimable function $\mathbf{c}^T\beta$

$$\mathbf{c}^T\mathbf{b} \pm t_{(n-rank(X))\alpha/2}\sqrt{MSE\ \mathbf{c}^T(X^TX)^{-}\mathbf{c}}$$

- Use $\mathbf{c}^T = (0\ 0\ ..\ 0\ \underset{\substack{\uparrow \\ j\text{-th position}}}{1}\ 0\ ..\ 0)$

  to construct a confidence interval for β_{j-1}

- Use $\mathbf{c}^T = (1, x_1, x_2, \ldots, x_r)$ to construct a confidence interval for

$$E(\mathbf{Y}|X_1 = x_1, \ldots, X_r = x_r) = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r$$

Prediction Intervals:

Predict a future observation at

$$X_1 = x_1, \ \ldots \ , \ X_r = x_r$$

i.e., predict

$$Y = \underline{\beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r} + \epsilon$$

estimate the conditional mean as

$$b_0 + b_1 x_1 + \cdots + b_r x_r$$

estimate this with its mean

$$E(\epsilon) = 0$$

A $(1-\alpha)\times$ 100% prediction interval is

$$(\mathbf{c}^T\mathbf{b} + 0) \pm t_{(n-rank(X)),\alpha/2}\sqrt{\text{MSE}\ [1 + \mathbf{c}^T(X^TX)^{-}\mathbf{c}]}$$

where

$$\mathbf{c}^T = (1 \ x_1 \ \cdots \ x_r)$$

Refer slide6_r1.pdf