# Multiple Linear Regression Model – Part IV

Johns Hopkins Engineering

**625.461 Statistical Models and Regression**

Module 4 – Lecture 4B

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

# Testing for Model Adequacy and Regressors

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

What is the overall adequacy of the model?

Which specific regressors seem important?

# Testing for Significance of Regressors

$$H_0 : \beta_1 = \cdots = \beta_k = 0$$

$$H_1 : \text{at least one } \beta_j \neq 0$$

Generalization of ANOVA used in simple linear regression

$$SS_T = SS_R + SS_{Res}$$

# ANOVA Table for Testing $H_0$

**TABLE 3.5 Analysis of Variance for Significance of Regression in Multiple Regression**

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F_0$ |
|---|---|---|---|---|
| Regression | $SS_R$ | $k$ | $MS_R$ | $MS_R/MS_{Res}$ |
| Residual | $SS_{Res}$ | $n-k-1$ | $MS_{Res}$ | |
| Total | $SS_T$ | $n-1$ | | |

Regression SS: $$SS_R = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

Residual SS: $$SS_{Res} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$$

Total SS: $$SS_T = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

# The $F$ test for $H_0$

$$F_0 = \frac{SS_R/k}{SS_{Res}/(n-k-1)} = \frac{MS_R}{MS_{Res}}$$

$$\sim F_{k,n-k-1} \text{ under } H_0$$

If $F_0 > F_{\alpha,k,n-k-1}$, then reject $H_0$.

# Other Ways to Assess Overall Adequacy of Model

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$$

$$R^2_{adj} = 1 - \frac{SS_{\mathrm{Re}\,s}/(n-k-1)}{SS_T/(n-1)}$$

# The Delivery Time Data (Ex 3.3, page 86 of Textbook)

$$SS_T = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$= 18{,}310.6290 - \frac{(559.60)^2}{25} = 5784.5426$$

$$SS_R = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$= 18{,}076.9030 - \frac{(559.60)^2}{25} = 5550.8166$$

# The Delivery Time Data (Ex 3.3, page 86 of Textbook)

$$SS_{\text{Res}} = SS_{\text{T}} - SS_{\text{R}}$$
$$= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 233.7260$$

$$F_0 = \frac{MS_{\text{R}}}{MS_{\text{Res}}} = \frac{2775.4083}{10.6239} = 261.24$$

which is very large and rejects
$H_0$: $\beta_1 = \beta_2 = 0$ (see ANOVA Table 3.6)

$$R^2 = 0.96$$

# Test a Subset of Regressors in Regression Model

How to test the joint impacts of $r < p$ regressors?

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim \mathbf{NIND}(\mathbf{0}, \sigma^2\mathbf{I})$$

# Test a Subset of Regressors in Regression Model

Arrange the $p$ regressors into two blocks: the $r$ regressors to be tested are placed in the second block, the remaining regressors in the first block.

$$\mathbf{X} = [\ \mathbf{X}_1 \mid \mathbf{X}_2\ ] \qquad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \hline \boldsymbol{\beta}_2 \end{bmatrix}$$

$\boldsymbol{\beta}_1$ is $(p-r) \times 1$

$\boldsymbol{\beta}_2$ is $r \times 1$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

# Test a Subset of Regressors in Regression Model

Wish to test

$$H_0\colon \boldsymbol{\beta}_2 = \mathbf{0}, \quad H_1\colon \boldsymbol{\beta}_2 \neq \mathbf{0}$$

Full model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$

Reduced model: $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$

# Test a Subset of Regressors in Regression Model

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}$$

$$SS_R(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \ (p \text{ degrees of freedom})$$

$$SS_R(\boldsymbol{\beta}_1) = \hat{\boldsymbol{\beta}}_1'\mathbf{X}_1'\mathbf{y} \ (p - r \text{ degrees of freedom})$$

# Test a Subset of Regressors in Regression Model

Regression SS due to $\boldsymbol{\beta}_2$ given $\boldsymbol{\beta}_1$ already in the model:

$$SS_R(\boldsymbol{\beta}_2 | \boldsymbol{\beta}_1) = SS_R(\boldsymbol{\beta}) - SS_R(\boldsymbol{\beta}_1)$$

with $p - (p - r) = r$ degrees of freedom

# $F$ test for $H_0$

$$F_0 = \frac{SS_R(\beta_2|\beta_1)/r}{MS_{Res}}$$

$\sim F_{r,n-k-1}$ under $H_0$  $p = k + 1$

If $F_0 > F_{\alpha,r,n-k-1}$, then reject $H_0$.

# The Delivery Time Data (Ex 3.5, page 92 of Textbook)

Suppose that we wish to assess the value of the $x_2$ (distance) given that the the $x_1$ (cases) is in the model.

$$H_0\colon \beta_2 = 0, \quad H_1\colon \beta_2 \neq 0$$

$$\begin{aligned} SS_R(\beta_2 \mid \beta_1, \beta_0) &= SS_R(\beta_1, \beta_2, \beta_0) - SS_R(\beta_1, \beta_0) \\ &= SS_R(\beta_1, \beta_2 \mid \beta_0) - SS_R(\beta_1 \mid \beta_0) \end{aligned}$$

# The Delivery Time Data (Ex 3.5, page 92 of Textbook)

$$SS_R(\beta_1, \beta_2 | \beta_0) = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = 5550.8166 \text{ (2 degrees of freedom)}$$

From Ex 2.9, the reduced model is

$$\hat{y} = 3.3208 + 2.1762x_1$$

# The Delivery Time Data (Ex 3.5, page 92 of Textbook)

$$SS_R(\beta_1|\beta_0) = \hat{\beta}_1 S_{xy} = (2.1762)(2473.3440)$$
$$= 5382.4077 \quad \text{(1 degree of freedom)}$$

$$SS_R(\beta_2|\beta_1, \beta_0) = 5550.8166 - 5382.4088$$
$$= 168.4078 \quad \text{(1 degree of freedom)}$$

$$F_0 = \frac{SS_R(\beta_2|\beta_1, \beta_0)/1}{MS_{Res}} = \frac{168.4078/1}{10.6239} = 15.85$$

$$> F_{0.05,1,22} = 4.30, \text{ we reject } H_0$$

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING