# Outline

1. Multiple Linear Regression (Estimation, Inference, Diagnostics and Remedial Measures)
2. Special Topics for Multiple Regression
   - Extra Sums of Squares

# General Regression Model in Matrix Terms

$$Y = \begin{pmatrix} Y_1 \\ . \\ . \\ . \\ Y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,p-1} \\ \dots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{pmatrix}$$

$$\beta = \begin{pmatrix} \beta_0 \\ . \\ . \\ . \\ \beta_{p-1} \end{pmatrix} \qquad \epsilon = \begin{pmatrix} \epsilon_1 \\ . \\ . \\ . \\ \epsilon_n \end{pmatrix}$$

# General Linear Regression in Matrix Terms

$$Y = X\beta + \epsilon$$

With $E(\epsilon) = 0$ and

$$\sigma^2(\epsilon) = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \dots & & & \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix}$$

We have $E(Y) = X\beta$ and $\sigma^2\{\mathbf{y}\} = \sigma^2\mathbf{I}$

# Least Square Solution

The matrix normal equations can be derived directly from the minimization of

$$Q = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

w.r.t to $\beta$

# Least Square Solution

We can solve this equation

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$$

(if the inverse of $\mathbf{X}'\mathbf{X}$ exists) by the following

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

and since

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$$

we have

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

# Fitted Values and Residuals

Let the vector of the fitted values are

$$\hat{\mathbf{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ . \\ . \\ . \\ \hat{y}_n \end{pmatrix}$$

in matrix notation we then have $\hat{\mathbf{y}} = \mathbf{X}\mathbf{b}$

# Hat Matrix-Puts hat on $y$

We can also directly express the fitted values in terms of $\mathbf{X}$ and $\mathbf{y}$ matrices

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

and we can further define H, the "hat matrix"

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \qquad \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

The hat matrix plans an important role in diagnostics for regression analysis.

# Hat Matrix Properties

1. the hat matrix is symmetric
2. the hat matrix is idempotent, i.e. $\mathbf{HH} = \mathbf{H}$

# Residuals

The residuals, like the fitted value $\hat{\mathbf{y}}$ can be expressed as linear combinations of the response variable observations $Y_i$

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

also, remember

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\mathbf{b}$$

these are equivalent.

# Covariance of Residuals

Starting with

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

we see that

$$\sigma^2\{\mathbf{e}\} = (\mathbf{I} - \mathbf{H})\sigma^2\{\mathbf{y}\}(\mathbf{I} - \mathbf{H})'$$

but

$$\sigma^2\{\mathbf{y}\} = \sigma^2\{\epsilon\} = \sigma^2\mathbf{I}$$

which means that

$$\sigma^2\{\mathbf{e}\} = \sigma^2(\mathbf{I} - \mathbf{H})\mathbf{I}(\mathbf{I} - \mathbf{H}) = \sigma^2(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H})$$

and since $\mathbf{I} - \mathbf{H}$ is idempotent, we have $\sigma^2\{\mathbf{e}\} = \sigma^2(\mathbf{I} - \mathbf{H})$

# Quadratic Forms

- In general, a quadratic form is defined by

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \sum_i \sum_j a_{ij} Y_i Y_j \text{ where } a_{ij} = a_{ji}$$

  with $\mathbf{A}$ the matrix of the quadratic form.
- The ANOVA sums SSTO,SSE and SSR can all be arranged into quadratic forms.

$$SSTO = \mathbf{y}'(\mathbf{I} - \frac{1}{n}\mathbf{J})\mathbf{y}$$

$$SSE = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$$

$$SSR = \mathbf{y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{y}$$

# Inference

We can derive the sampling variance of the $\beta$ vector estimator by remembering that $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{A}\mathbf{y}$

where $\mathbf{A}$ is a constant matrix

$$\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \qquad \mathbf{A}' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Using the standard matrix covariance operator we see that

$$\sigma^2\{\mathbf{b}\} = \mathbf{A}\sigma^2\{\mathbf{y}\}\mathbf{A}'$$

# Variance of b

Since $\sigma^2\{\mathbf{y}\} = \sigma^2\mathbf{I}$ we can write

$$\begin{aligned}\sigma^2\{\mathbf{b}\} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{I}\\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Of course

$$\mathbb{E}(\mathbf{b}) = \mathbb{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathbb{E}(\mathbf{y}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta = \beta$$

# F-test for regression

- $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$
- $H_a$: no all $\beta_k, (k = 1, \cdots, p-1)$ equal zero

Test statistic:

$$F^* = \frac{MSR}{MSE}$$

Decision Rule:

- if $F^* \leq F(1-\alpha; p-1, n-p)$, conclude $H_0$
- if $F^* > F(1-\alpha; p-1, n-p)$, conclude $H_a$

# $R^2$ and adjusted $R^2$

- The coefficient of multiple determination $R^2$ is defined as:

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

- $0 \leq R^2 \leq 1$
- $R^2$ always increases when there are more variables.
- Therefore, adjusted $R^2$:

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO}$$

- $R_a^2$ may decrease when $p$ is large.
- Coefficient of multiple correlation:

$$R = \sqrt{R^2}$$

Always positive square root!

# Inferences

The estimated variance-covariance matrix

$$s^2\{\mathbf{b}\} = MSE(\mathbf{X}'\mathbf{X})^{-1}$$

Then, we have

$$\frac{b_k - \beta_k}{s\{b_k\}} \sim t(n-p), k = 0, 1, \cdots, p-1$$

- $1-\alpha$ confidence intervals:

$$b_k \pm t(1-\alpha/2; n-p)s\{b_k\}$$

# t test

- Tests for $\beta_k$:
  - $H_0 : \beta_k = 0$
  - $H_1 : \beta_k \neq 0$
- Test Statistic:

$$t^* = \frac{b_k}{s\{b_k\}}$$

- Decision Rule:
  - $|t^*| \leq t(1 - \alpha/2; n - p)$; conclude $H_0$
  - Otherwise, conclude $H_a$

# Outline

1. Multiple Linear Regression (Estimation, Inference, Diagnostics and Remedial Measures)
2. Special Topics for Multiple Regression
   - Extra Sums of Squares

# Extra Sums of Squares

- A topic unique to multiple regression
- An extra sum of squares measures the marginal decrease in the error sum of squares when one or several predictor variables are added to the regression model, given that other variables are already in the model.
- Equivalently-one can view an extra sum of squares as measuring the marginal increase in the regression sum of squares

# Definitions

- Definition
  $-SSR(X_1|X_2) = SSE(X_2) - SSE(X_1, X_2)$
- Equivalently
  $-SSR(X_1|X_2) = SSR(X_1, X_2) - SSR(X_2)$
- We can switch the order of $X_1$ and $X_2$ in these expressions
- We can easily generalize these definitions for more than two variables
  $-SSR(X_3|X_1, X_2) = SSE(X_1, X_2) - SSE(X_1, X_2, X_3)$
  $-SSR(X_3|X_1, X_2) = SSR(X_1, X_2, X_3) - SSR(X_1, X_2)$

# ANOVA Table

Various software packages can provide extra sums of squares for regression analysis. These are usually provided in the order in which the input variables are provided to the system, for instance

Figure:

| 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 | $SSTO$ | $n-1$ | |

# Test whether a single $\beta_k = 0$

- Does $X_k$ provide statistically significant improvement to the regression model fit?
- We can use the general linear test approach
- Example
  –First order model with three predictor variables
  $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$
  –We want to answer the following hypothesis test

$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$

# Test whether a single $\beta_k = 0$

- For the full model we have $SSE(F) = SSE(X_1, X_2, X_3)$
- The reduced model is $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$
- And for this model we have $SSE(R) = SSE(X_1, X_2)$
- Where there are $df_r = n - 3$ degrees of freedom associated with the reduced model

# Test whether a single $\beta_k = 0$

The general linear test statistics is

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} / \frac{SSE(F)}{df_F}$$

which becomes

$$F^* = \frac{SSE(X_1, X_2) - SSE(X_1, X_2, X_3)}{(n-3) - (n-4)} / \frac{SSE(X_1, X_2, X_3)}{n-4}$$

but $SSE(X_1, X_2) - SSE(X_1, X_2, X_3) = SSR(X_3 | X_1, X_2)$

# Test whether a single $\beta_k = 0$

The general linear test statistics is

$$F^* = \frac{SSR(X_3|X_1,X_2)}{1} \Big/ \frac{SSE(X_1,X_2,X_3)}{n-4} = \frac{MSR(X_3|X_1,X_2)}{MSE(X_1,X_2,X_3)}$$

Extra sum of squares has one associated degree of freedom.

# Test whether $\beta_k = 0$

Another example
$H_0 : \beta_2 = \beta_3 = 0$
$H_1$ : not both $\beta_2$ and $\beta_3$ are zero
The general linear test can be used again

$$F^* = \frac{SSE(X_1)-SSE(X_1,X_2,X_3)}{(n-2)-(n-4)} \Big/ \frac{SSE(X_1,X_2,X_3)}{n-4}$$

But $SSE(X_1) - SSE(X_1, X_2, X_3) = SSR(X_2, X_3|X_1)$
so the expression can be simplified.

# Tests concerning regression coefficients

Summary:
– General linear test can be used to determine whether or not a predictor variable( or sets of variables) should be included in the model
– The ANOVA SSE's can be used to compute $F^*$ test statistics

# Summary of Tests Concerning Regression Coefficients

- Test whether all $\beta_k = 0$
- Test whether a single $\beta_k = 0$
- Test whether some $\beta_k = 0$
- Test involving relationships among coefficients, for example,
  - $H_0 : \beta_1 = \beta_2$ vs. $H_a : \beta_1 \neq \beta_2$
  - $H_0 : \beta_1 = 3, \beta_2 = 5$ vs. $H_a$ : otherwise
- Key point in all tests: form the full model and the reduced model