Chapter 6

Multiple Regression I

Instructor: Li C.X.

Outline

- Multiple regression models
- General linear regression model in matrix form
- Inference about regression parameters
- Estimation of mean response and prediction
- Diagnostic and Remedial Measures

6.1 Multiple regression models

- One of the most widely used tools in statistical analysis.
- Still have single response variable Y, but have multiple explanatory variables
- Examples: $Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_4 X_{i4} + \varepsilon_i$
  - Blood Pressure vs Age, Weight, Diet, Fitness Level
- Can include polynomial terms to allow for nonlinear relations
- Can include product terms to allow for interactions when effect of one variable depends on level of another variable
- Can include "dummy" variables for categorical predictors.

Models

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

- Have p−1 predictors $\rightarrow$ p coefficients (parameters)

- Goal: to determine effects (if any) of each predictor, controlling for others.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

$$E\left\{\varepsilon_i\right\} = 0 \quad \Rightarrow \quad E\left\{Y\right\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

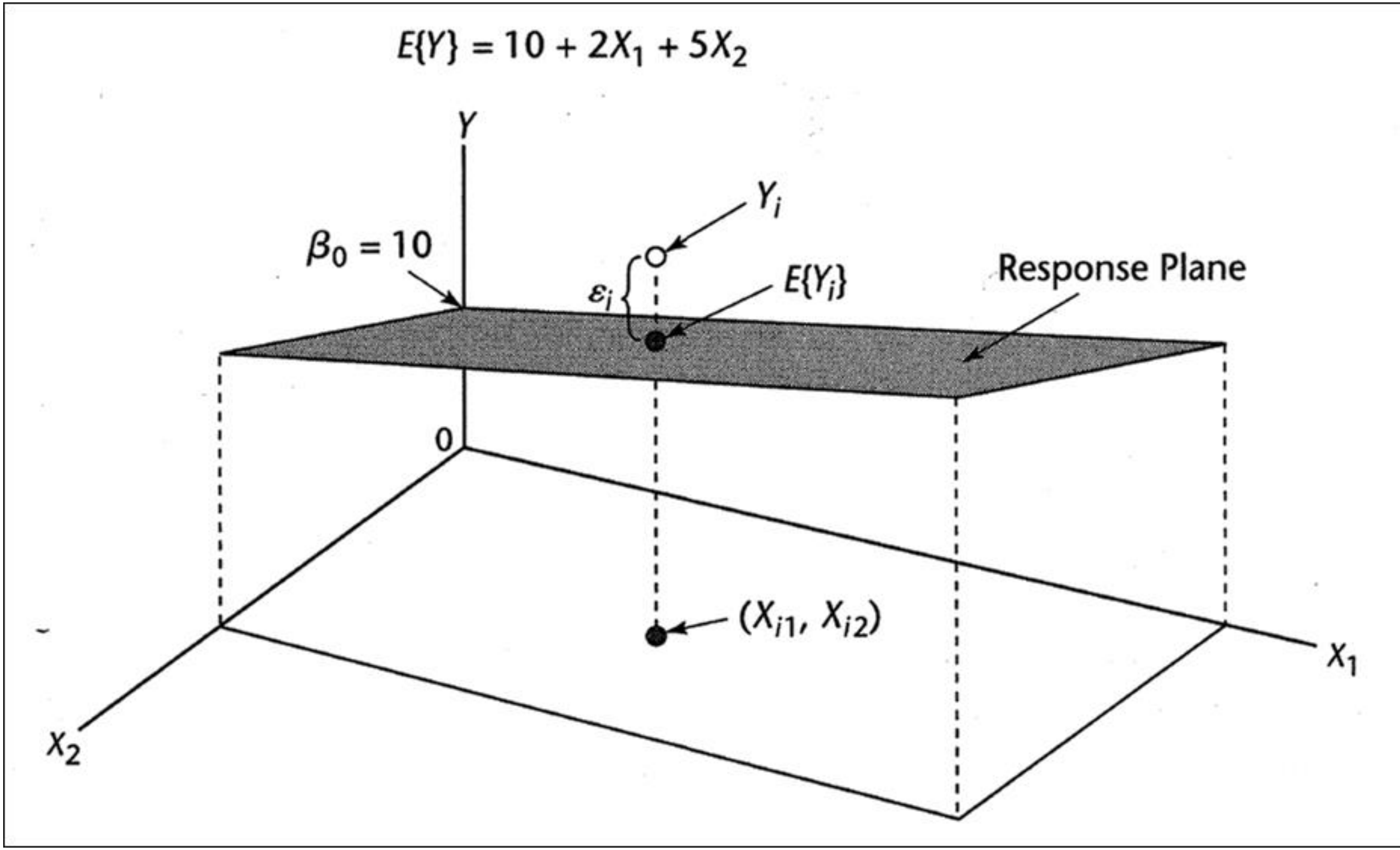
First-Order Model with 2 Numeric Predictors

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

$$E\{\varepsilon_i\} = 0 \quad \Rightarrow \quad E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

- Meaning of regression coefficients
  - β_1 describes change in mean response per unit increase in X_1 when X_2 is held constant
  - β_2 describes change in mean response per unit increase in X_2 when X_1 is held constant
- Variables X_1 and X_2 are **additive**. There is no **interaction**.
- The parameters 1 and 2 are sometimes called partial regression coefficients. They represents the partial effect of one predictor variable when the other predictor variable is included in the model and is hold constant.

The response surface is a plane.



Interaction Model

- When the effect of X_1 on the mean response does not depend on the level X_2 (and vice versa) the two predictor variables are said to have additive effects or not to interact.

  $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

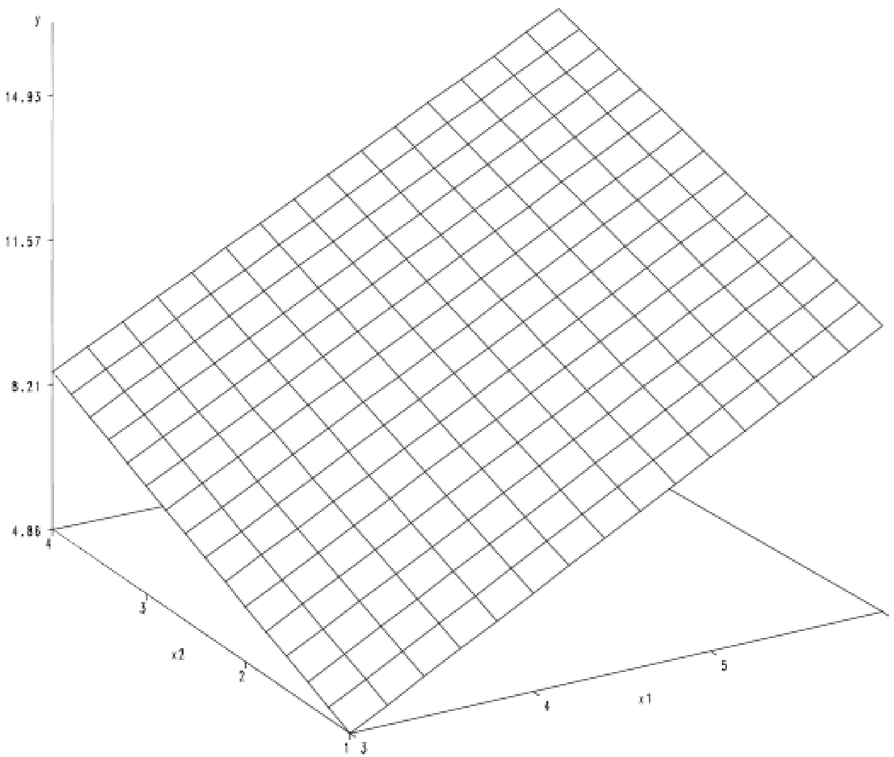
- Interaction Model:

  $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

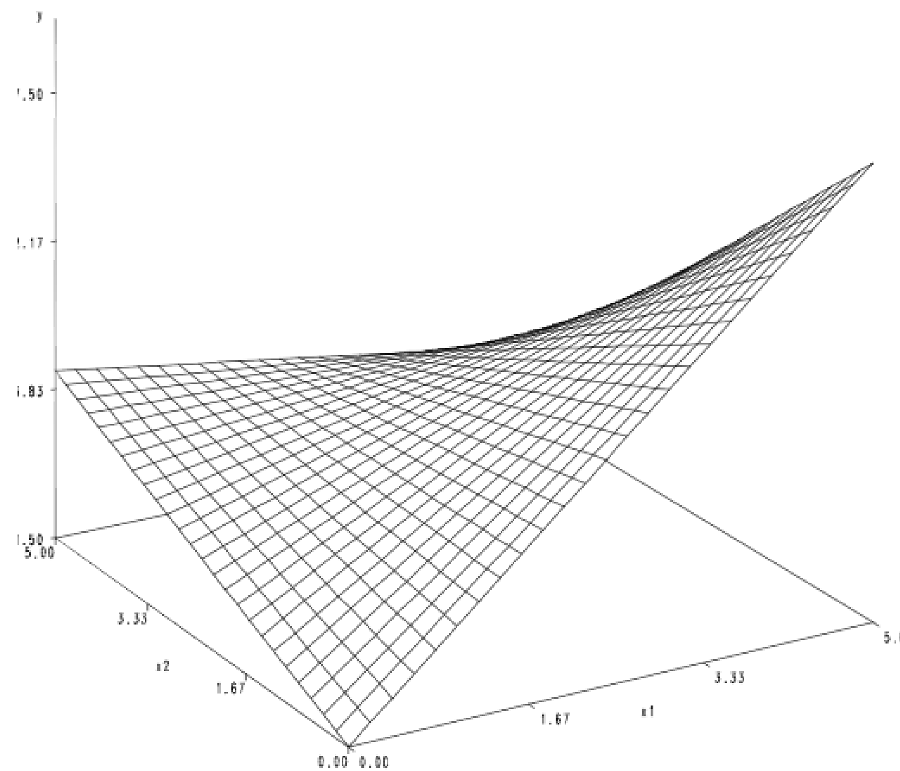
  - $E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$
  - When $X_2 = 0$: Effect of increasing X_1 by 1: $\beta_1(1) + \beta_3(1)(0) = \beta_1$
  - When $X_2 = 1$: Effect of increasing X_1 by 1: $\beta_1(1) + \beta_3(1)(1) = \beta_1 + \beta_3$
- The effect of X_1 depends on level of X_2, and vice versa

Additive and Interaction models

$$\widehat{Y}_i = -2.79 + 2.14X_{i1} + 1.21X_{i2}$$



$$\widehat{Y}_i = 1.5 + 3.2X_{i1} + 1.2X_{i2} - .75X_{i1}X_{i2}$$



General linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

$$E\{\varepsilon_i\} = 0 \quad \Rightarrow \quad E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{p-1} X_{p-1}$$

(Hyperplane in p-dimensions)

$p - 1 = 1 \quad \Rightarrow \quad$ Simple linear regression

Normality, independence, and constant variance for errors:

$$\varepsilon_i \sim NID\left(0, \sigma^2\right)$$

$$\Rightarrow \quad Y_i \sim N\left(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1}, \sigma^2\right) \quad \sigma\left\{Y_i, Y_j\right\} = 0 \quad \forall\, i \neq j$$

Special Types of Variables/Models - I

- *p*-1 distinct numeric predictors (attributes)
  - Y = Sales, X_1=Advertising, X_2=Price
- Categorical Predictors – Indicator (Dummy) variables, representing *m*-1 levels of a m level categorical variable
  - Y = Salary, X_1=Experience, X_2=1 if College Grad, 0 if Not
- Polynomial Regression with Polynomial Terms – Allow for bends in the Regression
  - MPG = $\beta_0 + \beta_1\text{Speed} + \beta_2\text{Speed}^2 + \varepsilon$
    - Y=MPG, X_1=Speed, X_2=Speed^2
  - $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^{\,2} + \beta_3 X_2 + \beta_4 X_2^{\,2} + \beta_5 X_1 X_2 + \varepsilon$
- Transformed Variables – Transformed Y variable to achieve linearity $Y^* = \ln(Y)$ $Y^* = 1/Y$

$$Y_i^* = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

Special Types of Variables/Models - II

- Interaction Effects – Effect of one predictor depends on levels of other predictors
- Y = Salary, X_1=Experience, X_2=1 if Coll Grad, 0 if Not, $X_3=X_1X_2$
- $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$
  - Non-College Grads (X_2 =0):
  - $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2(0) + \beta_3 X_1(0) = \beta_0 + \beta_1 X_1$
  - College Grads (X_2 =1):
  - $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2(1) + \beta_3 X_1(1) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1$
- Response Surface Models
  - $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \beta_5 X_1 X_2$
- Note: Although the Response Surface Model has polynomial terms, it is linear wrt Regression parameters

6.2 General Linear Regression Model in Matrix Form

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad i = 1,\ldots,n$$

$$\Rightarrow \quad Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \qquad \text{where: } X_{i0} \equiv 1$$

Matrix Form:

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad \underset{n\times p}{\mathbf{X}} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix} \qquad \underset{p\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

$$\underset{n\times 1}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \mathbf{E}\left\{\underset{n\times 1}{\boldsymbol{\varepsilon}}\right\} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \boldsymbol{\sigma}^2\left\{\underset{n\times 1}{\boldsymbol{\varepsilon}}\right\} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I}$$

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \quad \Rightarrow \quad \mathbf{E}\left\{\underset{n\times 1}{\mathbf{Y}}\right\} = \mathbf{E}\left\{\underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}}\right\} = \underset{n\times p}{\mathbf{X}}\,\underset{p\times 1}{\boldsymbol{\beta}} \qquad \boldsymbol{\sigma}^2\left\{\underset{n\times 1}{\mathbf{Y}}\right\} = \sigma^2 \mathbf{I}$$

Studio data

Y: sales of portraits of children

X_1: number of children

X_2: per capita personal income

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\mathbf{X} = \begin{bmatrix} 1 & 68.5 & 16.7 \\ 1 & 45.2 & 16.8 \\ \vdots & \vdots & \vdots \\ 1 & 52.3 & 16.0 \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} 174.4 \\ 164.4 \\ \vdots \\ 166.5 \end{bmatrix}$$

CASE	X1	X2	Y
1	68.5	16.7	174.4
2	45.2	16.8	164.4
3	91.3	18.2	244.2
4	47.8	16.3	154.6
5	46.9	17.3	181.6
6	66.1	18.2	207.5
7	49.5	15.9	152.8
8	52.0	17.2	163.2
9	48.9	16.6	145.4
10	38.4	16.0	137.2
11	87.9	18.3	241.9
12	72.8	17.1	191.1
13	88.4	17.4	232.0
14	42.9	15.8	145.3
15	52.5	17.8	161.1
16	85.7	18.4	209.7
17	41.3	16.5	146.4
18	51.7	16.3	144.0
19	89.6	18.1	232.6
20	82.7	19.1	224.1
21	52.3	16.0	166.5

6.3 Estimation of Regression Coefficients

Least Squares Estimation (LSE):

Goal: Minimize: $Q = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{i,p-1} \right)^2$

$\Rightarrow$ Obtain Estimates $b_0, b_1, \ldots, b_{p-1}$ that minimize Q

Normal Equations obtained from: $\dfrac{\partial Q}{\partial \beta_0} = 0, \ldots, \dfrac{\partial Q}{\partial \beta_{p-1}} = 0$:

$$\underset{p\times p}{\mathbf{X}'\mathbf{X}} \underset{p\times 1}{\mathbf{b}} = \underset{p\times 1}{\mathbf{X}'\mathbf{Y}} \quad \Rightarrow \quad \underset{p\times 1}{\mathbf{b}} = \left(\mathbf{X}'\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{Y}$$

Maximum Likelihood for normal error model also leads to same $\mathbf{b}$:

$$L\left(\boldsymbol{\beta}, \sigma^2 \right) = \left(2\pi\sigma^2 \right)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{i,p-1} \right)^2 \right]$$

since maximizing L involves minimizing Q.

LSE estimation of Regression Coefficients

- by matrix derivation

$$Q = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

$$= \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = \left[\frac{\partial Q}{\partial \beta_0}, \ \cdots, \frac{\partial Q}{\partial \beta_{p-1}} \right] = -2\mathbf{Y}'\mathbf{X} + 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}$$

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = 0 \quad \Rightarrow \quad \mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

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

- For the studio data, calculate

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 68.5 & 45.2 & \cdots & 52.3 \\ 16.7 & 16.8 & \cdots & 16.0 \end{bmatrix} \begin{bmatrix} 1 & 68.5 & 16.7 \\ 1 & 45.2 & 16.8 \\ \vdots & \vdots & \vdots \\ 1 & 52.3 & 16.0 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} \\ \sum X_{i2} & \sum X_{i2}X_{i1} & \sum X_{i2}^2 \end{bmatrix} = \begin{bmatrix} 21.0 & 1{,}302.4 & 360.0 \\ 1{,}302.4 & 87{,}707.9 & 22{,}609.2 \\ 360.0 & 22{,}609.2 & 6{,}190.3 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 29.7289 & .0722 & -1.9926 \\ .0722 & .00037 & -.0056 \\ -1.9926 & -.0056 & .1363 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 68.5 & 45.2 & \cdots & 52.3 \\ 16.7 & 16.8 & \cdots & 16.0 \end{bmatrix} \begin{bmatrix} 174.4 \\ 164.4 \\ \vdots \\ 166.5 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_{i1} Y_i \\ \sum X_{i2} Y_i \end{bmatrix} = \begin{bmatrix} 3{,}820 \\ 249{,}643 \\ 66{,}073 \end{bmatrix}$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 29.7289 & .0722 & -1.9926 \\ .0722 & .00037 & -.0056 \\ -1.9926 & -.0056 & .1363 \end{bmatrix} \begin{bmatrix} 3{,}820 \\ 249{,}643 \\ 66{,}073 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -68.857 \\ 1.455 \\ 9.366 \end{bmatrix} \qquad \hat{Y} = -68.857 + 1.455X_1 + 9.366X_2$$

6.4 Fitted Values and Residuals

$$\hat{Y} = -68.857 + 1.455X_1 + 9.366X_2$$

$$\text{Fitted Values: } \hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix}$$

$$\text{Residuals: } \underset{n\times 1}{\mathbf{e}} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

CASE	X1	X2	Y	FITTED	RESIDUAL
1	68.5	16.7	174.4	187.184	-12.7841
2	45.2	16.8	164.4	154.229	10.1706
3	91.3	18.2	244.2	234.396	9.8037
4	47.8	16.3	154.6	153.329	1.2715
5	46.9	17.3	181.6	161.385	20.2151
6	66.1	18.2	207.5	197.741	9.7586
7	49.5	15.9	152.8	152.055	0.7449
8	52.0	17.2	163.2	167.867	-4.6666
9	48.9	16.6	145.4	157.738	-12.3382
10	38.4	16.0	137.2	136.846	0.3540
11	87.9	18.3	241.9	230.387	11.5126
12	72.8	17.1	191.1	197.185	-6.0849
13	88.4	17.4	232.0	222.686	9.3143
14	42.9	15.8	145.3	141.518	3.7816
15	52.5	17.8	161.1	174.213	-13.1132
16	85.7	18.4	209.7	228.124	-18.4239
17	41.3	16.5	146.4	145.747	0.6530
18	51.7	16.3	144.0	159.001	-15.0013
19	89.6	18.1	232.6	230.987	1.6130
20	82.7	19.1	224.1	230.316	-6.2160
21	52.3	16.0	166.5	157.064	9.4356

Fitted Values and Residuals

$$\text{Let } \mathbf{H} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'} \quad \Rightarrow \quad \mathbf{H} = \mathbf{H}' = \mathbf{H}^2; \; \left(\mathbf{I}-\mathbf{H}\right) = \left(\mathbf{I}-\mathbf{H}\right)' = \left(\mathbf{I}-\mathbf{H}\right)^2$$

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times p}{\mathbf{X}}\;\underset{p\times 1}{\mathbf{b}} = \mathbf{X}\left(\mathbf{X'X}\right)^{-1}\mathbf{X'Y} = \mathbf{HY} \sim N\left(\mathbf{X\beta}, \sigma^2\mathbf{H}\right)$$

$$E\left\{\hat{\mathbf{Y}}\right\} = E\left\{\mathbf{HY}\right\} = \mathbf{H}E\left\{\mathbf{Y}\right\} = \mathbf{HX\beta} = \mathbf{X\beta}$$

$$\text{var}\left\{\hat{\mathbf{Y}}\right\} = \text{var}\left\{\mathbf{HY}\right\} = \mathbf{H}\,\text{var}\left\{\mathbf{Y}\right\}\mathbf{H'} = \sigma^2\mathbf{H}$$

$$\underset{n\times 1}{\mathbf{e}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times 1}{\hat{\mathbf{Y}}} = \underset{n\times 1}{\mathbf{Y}} - \underset{n\times p}{\mathbf{X}}\;\underset{p\times 1}{\mathbf{b}} = \mathbf{Y} - \mathbf{HY} = \left(\mathbf{I}-\mathbf{H}\right)\mathbf{Y} \sim N\left(\mathbf{0}, (\mathbf{I}-\mathbf{H})\sigma^2\right)$$

$$E\left\{\mathbf{e}\right\} = E\left\{\left(\mathbf{I}-\mathbf{H}\right)\mathbf{Y}\right\} = \left(\mathbf{I}-\mathbf{H}\right)E\left\{\mathbf{Y}\right\} = \left(\mathbf{I}-\mathbf{H}\right)\mathbf{X\beta} = \mathbf{X\beta} - \mathbf{X\beta} = \mathbf{0}$$

$$\text{var}\left\{\mathbf{e}\right\} = \left(\mathbf{I}-\mathbf{H}\right)\sigma^2\mathbf{I}\left(\mathbf{I}-\mathbf{H}\right)' = \sigma^2\left(\mathbf{I}-\mathbf{H}\right)$$

Hat matrix $\mathbf{H} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$

Property of hat matrix $\mathbf{H}$:

- $\mathbf{HY} = \hat{\mathbf{Y}}$; $\mathbf{HX} = \mathbf{X}$; $\mathbf{H}\hat{\mathbf{Y}} = \hat{\mathbf{Y}}$; $\mathbf{He} = \mathbf{0}$

$$\mathbf{HX} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X};\ \ \mathbf{H}\hat{\mathbf{Y}} = \mathbf{HXb} = \mathbf{Xb} = \hat{\mathbf{Y}}$$

$$\mathbf{He} = \mathbf{H}\left(\mathbf{Y} - \hat{\mathbf{Y}}\right) = \mathbf{HY} - \mathbf{H}\hat{\mathbf{Y}} = \hat{\mathbf{Y}} - \hat{\mathbf{Y}} = 0$$

- Symmetric and idempotent

$$\mathbf{H}' = \left(\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right)' = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' = \mathbf{H}$$

$$\mathbf{HH} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' = \mathbf{H}$$

- The rank is p, number of regression parameters

$$\text{rank}\left[\mathbf{H}\right] = \text{tr}\left[\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right] = \text{tr}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\right] = \text{tr}\left[\mathbf{I}_{p\times p}\right] = p$$

Hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

- Let $\mathbf{H}=(h_{ij})_{n\times n}$, then

$$h_{ii}=\sum_{i=1}^{n} h_{ij}^2; \quad \sum_{i=1}^{n} h_{ii} = \mathrm{tr}(\mathbf{H}) = p; \quad \sum_{i=1}^{n} h_{ij} = \sum_{j=1}^{n} h_{ij} = 1;$$

Proof: $\underset{n\times n}{\mathbf{H}} = \underset{n\times n}{\mathbf{H}}\,\underset{n\times n}{\mathbf{H}} \Rightarrow h_{ii}=\sum_{j=1}^{n} h_{ij}h_{ji} = \sum_{j=1}^{n} h_{ij}^2$

$$\underset{n\times n}{\mathbf{H}}\,\underset{n\times p}{\mathbf{X}} = \underset{n\times p}{\mathbf{X}} \Rightarrow \mathbf{H}\begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & \cdots & X_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p-1} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} h_{1j} & \cdots \\ \sum_{j=1}^{n} h_{2j} & \cdots \\ \vdots & \\ \sum_{j=1}^{n} h_{nj} & \cdots \end{bmatrix} = \begin{bmatrix} 1 & \cdots \\ 1 & \cdots \\ \vdots & \\ 1 & \cdots \end{bmatrix} \Rightarrow \sum_{j=1}^{n} h_{ij} = 1$$

The symmetry of H $\Rightarrow \sum_{i=1}^{n} h_{ij} = \sum_{j=1}^{n} h_{ij} = 1$

6.5 Analysis of Variance

$$\underset{n\times 1}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \mathbf{IY} \qquad \underset{n\times 1}{\hat{\mathbf{Y}}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{bmatrix} = \mathbf{Xb} = \mathbf{HY} \qquad \underset{n\times 1}{\overline{\mathbf{Y}}} = \begin{bmatrix} \overline{Y} \\ \overline{Y} \\ \vdots \\ \overline{Y} \end{bmatrix} = \frac{1}{n}\begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \left(\frac{1}{n}\right)\mathbf{JY}$$

- $\mathbf{I}-\mathbf{J}/n$, $\mathbf{H}-\mathbf{J}/n$, $\mathbf{I}-\mathbf{H}$ are idempotent and symmetric.

$$SSTO = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \left(\mathbf{Y}-\overline{\mathbf{Y}}\right)'\left(\mathbf{Y}-\overline{\mathbf{Y}}\right) = \mathbf{Y}'\left(\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{Y}$$

$$SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2 = \left(\hat{\mathbf{Y}}-\overline{\mathbf{Y}}\right)'\left(\hat{\mathbf{Y}}-\overline{\mathbf{Y}}\right) = \mathbf{Y}'\left(\mathbf{H}-\left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{Y}$$

$$SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = \left(\mathbf{Y}-\hat{\mathbf{Y}}\right)'\left(\mathbf{Y}-\hat{\mathbf{Y}}\right) = \mathbf{Y}'\left(\mathbf{I}-\mathbf{H}\right)\mathbf{Y}$$

- SSTO=SSR+SSE

ANOVA for Studio Data

$$SSTO = \mathbf{Y}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y} = \sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2$$

$$\mathbf{Y}'\mathbf{Y} = [174.4 \quad 164.4 \quad \cdots \quad 166.5]\begin{bmatrix} 174.4 \\ 164.4 \\ \vdots \\ 166.5 \end{bmatrix} = 721{,}072.40$$

$$\left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} = \frac{1}{21}[174.4 \quad 164.4 \quad \cdots \quad 166.5]\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}\begin{bmatrix} 174.4 \\ 164.4 \\ \vdots \\ 166.5 \end{bmatrix}$$

$$= \frac{(3{,}820.0)^2}{21} = 694{,}876.19$$

$$SSTO = \mathbf{Y}'\mathbf{Y} - \left(\frac{1}{n}\right)\mathbf{Y}'\mathbf{J}\mathbf{Y} = 721{,}072.40 - 694{,}876.19 = 26{,}196.21$$

$$SSE = \mathbf{Y}'\left[\mathbf{I} - \mathbf{H}\right]\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - (\mathbf{HY})'\mathbf{Y} = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$$

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$$

$$= 721{,}072.40 - \begin{bmatrix} -68.857 & 1.455 & 9.366 \end{bmatrix} \begin{bmatrix} 3{,}820 \\ 249{,}643 \\ 66{,}073 \end{bmatrix}$$

$$= 721{,}072.40 - 718{,}891.47 = 2{,}180.93$$

$$SSR = \mathbf{Y}'\left[\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y} = (\mathbf{HY})'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{JY} = \mathbf{b}'\mathbf{X}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{JY}$$

$$=718891.47-694876.19=24015.28$$

$$SSR = SSTO - SSE = 26{,}196.21 - 2{,}180.93 = 24{,}015.28$$

ANOVA in regression

$$\text{rank}\left[\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right]=n-1 \quad \text{rank}\left[\mathbf{H}-\left(\frac{1}{n}\right)\mathbf{J}\right]=p-1 \quad \text{rank}\left[\mathbf{I}-\mathbf{H}\right]=n-p$$

- For normal error models, according to Cochran's theorem,

$$SSTO=\mathbf{Y}'\left[\mathbf{I}-\left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}\sim\sigma^2\chi^2(n-1,\ \delta_{TO})$$

$$SSR=\mathbf{Y}'\left[\mathbf{H}-\left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{Y}\sim\sigma^2\chi^2(p-1,\ \delta_{R})$$

$$SSE=\mathbf{Y}'\left[\mathbf{I}-\mathbf{H}\right]\mathbf{Y}\sim\sigma^2\chi^2(n-p,\ 0) \qquad \boldsymbol{SSR\perp SSE}$$

$$\boxed{\delta_E=\frac{1}{\sigma^2}(\mathbf{X}\boldsymbol{\beta})'\left[\mathbf{I}-\mathbf{H}\right]\mathbf{X}\boldsymbol{\beta}=\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{H}\mathbf{X}\boldsymbol{\beta}=\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}=0.}$$

$$\boxed{\delta_{TO}=\delta_R=\frac{1}{\sigma^2}\sum_{k=1}^{p-1}\sum_{l=1}^{p-1}SS_{kl}\beta_k\beta_l,\quad SS_{kl}=\sum_{i=1}^{n}\left(X_{ik}-\bar{X}_k\right)\left(X_{il}-\bar{X}_l\right)}$$

$$\delta_{TO} = \delta_R = \frac{1}{\sigma^2}(\mathbf{X\beta})'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{X\beta} = \frac{1}{\sigma^2}\boldsymbol{\beta}'\mathbf{X}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J}\right]\mathbf{X\beta}$$

Let $\mathbf{A} = \mathbf{I} - \left(\frac{1}{n}\right)\mathbf{J} = \left[a_{ij}\right]_{n\times n}$, with $a_{ii} = 1 - \frac{1}{n}$, $a_{ij} = -\frac{1}{n}, i \neq j$.

$\mathbf{Q} = \mathbf{X}'\mathbf{AX} = \left[q_{kl}\right]_{p\times p}$, $\quad q_{1l} = q_{k1} = 0$, $q_{k+1,l+1} = SS_{kl}, k, l = 1, \cdots p-1$

where $SS_{kl} = \sum_{i=1}^{n}\left(X_{ik} - \bar{X}_k\right)\left(X_{il} - \bar{X}_l\right)$,

$$\delta_{TO} = \frac{1}{\sigma^2}\boldsymbol{\beta}'\mathbf{Q}\boldsymbol{\beta} = \frac{1}{\sigma^2}\sum_{k=0}^{p-1}\sum_{l=0}^{p-1} q_{k+1,l+1}\beta_k\beta_l = \frac{1}{\sigma^2}\sum_{k=1}^{p-1}\sum_{l=1}^{p-1} SS_{kl}\beta_k\beta_l$$

$$= \frac{1}{\sigma^2}\sum_{k=1}^{p-1}\beta_k^2 SS_{kk} + \frac{1}{\sigma^2}\sum_{k=1}^{p-1}\sum_{l\neq k}\beta_k\beta_l SS_{kl}$$

Mean Squares

$$(1)\xi \sim \chi^2(d,\ \delta) \Rightarrow E\xi = d + \delta$$

$$(2)E\left\{\mathbf{Y}'\mathbf{A}\mathbf{Y}\right\} = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + \mathrm{tr}(\mathbf{A}\boldsymbol{\Sigma})$$

$$MSR = \frac{SSR}{p-1} = \frac{1}{p-1}\mathbf{Y}'\left(\mathbf{H} - \left(\frac{1}{n}\right)\mathbf{J}\right)\mathbf{Y}$$

$$MSE = \frac{SSE}{n-p} = \frac{1}{n-p}\mathbf{Y}'\left(\mathbf{I} - \mathbf{H}\right)\mathbf{Y}$$

$$E\left\{MSE\right\} = \frac{1}{n-p}\left(n-p\right)\sigma^2 = \sigma^2$$

$$E\left\{MSR\right\} = \frac{1}{p-1}\left((p-1) + \delta_R\right)\sigma^2 = \sigma^2 + \frac{1}{p-1}\sum_{k=1}^{p-1}\sum_{l=1}^{p-1} SS_{kl}\beta_k\beta_l$$

$$E\left\{MSR\right\} \geq E\left\{MSE\right\} = \sigma^2$$

$$E\left\{MSR\right\} = E\left\{MSE\right\} \quad \Leftrightarrow \quad \beta_1 = \ldots = \beta_{p-1} = 0$$

F-test for regression

If $\beta_1 = \ldots = \beta_{p-1} = 0$, then $Y_i = \beta_0 + \varepsilon_i$ are iid, and $\delta_R = 0$,

$$SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2 \sim \sigma^2\chi^2(p-1)$$

$$SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \sim \sigma^2\chi^2(n-p) \qquad \boldsymbol{SSR \perp SSE}$$

Test of $H_0 : \beta_1 = \ldots = \beta_{p-1} = 0 \quad H_A :$ Not all $\beta_1, \ldots, \beta_{p-1} = 0$

Test Statistic: $F^* = \dfrac{MSR}{MSE} = \dfrac{SSR/(p-1)}{SSE/(n-p)} \overset{H_0}{\sim} F\left(p-1, n-p\right)$

Rejection Region: $F^* \geq F\left(1-\alpha; p-1, n-p\right)$

P-value$= \Pr\left\{F\left(p-1, n-p\right) \geq F^*\right\}$

ANOVA Table with p–1 predictors

Source of Variation	SS	df	MS	F	P
Regression (Model)	$SSR = \sum(\hat{Y}_i - \bar{Y})^2$	p–1	$MSR = \frac{SSR}{p-1}$	$\frac{MSR}{MSE}$	Pr(F(p–1,n–p)>F)
Error	$SSE = \sum(Y_i - \hat{Y}_i)^2$	n–p	$MSE = \frac{SSE}{n-p}$		
Total	$SSTO = \sum(Y_i - \bar{Y})^2$	n–1			

ANOVA Table for Studio data

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

H_a: not both β_1 and β_2 equal zero

Source of Variation	SS	df	MS	F	P
Regression (Model)	24015.28	2	12007.64	99.1	$1.9*10^{-10}$
Error	2180.93	18	121.1626		
Total	26196.21	20			

$$F^* = \frac{MSR}{MSE} = \frac{12{,}007.64}{121.1626} = 99.1 \; > F(.95;\, 2,\, 18) = 3.55.$$

$p<0.05$. Reject H_0.

R square and adjusted R square

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

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

- Coefficient of multiple correlation $R = \sqrt{R^2}$
  - Simple correlation of Y and $\hat{Y}$

$$R = \frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)\left(\hat{Y}_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2}}$$

- $0 \le R^2 \le 1,\ 0 \le R \le 1.$

- For the Studio example

$$R^2 = \frac{SSR}{SSTO} = \frac{24{,}015.28}{26{,}196.21} = .917$$

$$R = .957$$

R square and adjusted R square

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

- R^2 always increases when there are more variables.
- Adjusted R^2

$$R_a^2 = 1 - \frac{[SSE/(n-p)]}{[SSTO/(n-1)]} = 1 - \frac{MSE}{MSTO} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO}$$

- Adjusted R^2 may decrease when p is large.

- For the Studio example

$$R_a^2 = 1 - \frac{MSE}{MSTO} = 1 - \frac{121.1626}{26196.21/20} = 0.907$$

6.6 Inferences about Regression Parameters

Review:

(1) Covariance Matrix of Two Random Vectors

$$\operatorname{cov}\left\{\underset{m\times 1}{\mathbf{X}},\underset{n\times 1}{\mathbf{Y}}\right\}=\begin{bmatrix}\sigma_{11} & \cdots & \sigma_{1n}\\ \cdots & \cdots & \cdots\\ \sigma_{m1} & \cdots & \sigma_{mn}\end{bmatrix},\quad \text{where } \sigma_{ij}=\operatorname{cov}(X_i,Y_j)$$

(2) If **A,B** are constant matrices and **Y** is a random vector, then

$$\operatorname{cov}\{\mathbf{AY},\mathbf{BY}\}=\mathbf{A}\boldsymbol{\sigma}^2\{\mathbf{Y}\}\mathbf{B}'$$

(3) If random vector $\underset{n\times 1}{\mathbf{Y}}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma}=(\sigma_{ij})_{n\times n}$, then

$$\sigma_{ij}=0 \iff Y_i \text{ and } Y_j \text{ is independent}$$

Independence of b and SSE

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \qquad \underset{n\times 1}{\boldsymbol{\varepsilon}} \sim N\left(\underset{n\times 1}{\mathbf{0}}, \sigma^2 \underset{n\times n}{\mathbf{I}}\right)$$

$$\underset{n\times 1}{\mathbf{e}} = (\mathbf{I}-\mathbf{H})\mathbf{Y},\ \ \underset{p\times 1}{\mathbf{b}} = (\mathbf{X'X})^{-1}\mathbf{X'Y},$$

$$\text{cov}(\underset{n\times 1}{\mathbf{e}}, \underset{p\times 1}{\mathbf{b}}) = \text{cov}\left\{(\mathbf{I}-\mathbf{H})\mathbf{Y}, (\mathbf{X'X})^{-1}\mathbf{X'Y}\right\}$$

$$= (\mathbf{I}-\mathbf{H})\boldsymbol{\sigma}^2\{\mathbf{Y}\}\left[(\mathbf{X'X})^{-1}\mathbf{X'}\right]'$$

$$= \sigma^2(\mathbf{I}-\mathbf{H})\mathbf{X}(\mathbf{X'X})^{-1} = \sigma^2(\mathbf{X}-\mathbf{HX})(\mathbf{X'X})^{-1} = \underset{n\times p}{\mathbf{0}}$$

$(\mathbf{e}, \mathbf{b}) = (e_1, \ldots, e_{\mathrm{n}}, b_0, \ldots, b_{\mathrm{p-1}})$ is multiple normal distributed.

- Then **b** is independent with residual vector **e**, and SSE as well.

Inferences about Regression Parameters

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \qquad \underset{n\times 1}{\boldsymbol{\varepsilon}} \sim N\left(\underset{n\times 1}{\mathbf{0}}, \sigma^2 \underset{n\times n}{\mathbf{I}}\right)$$

- Parameter estimators

$$\underset{p\times 1}{\mathbf{b}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \sim N\left(\boldsymbol{\beta}, \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

$$\boxed{\mathbf{E}\{\mathbf{b}\} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{E}\{\mathbf{Y}\} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}}$$

$$\boxed{\begin{aligned}\boldsymbol{\sigma}^2\{\mathbf{b}\} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\sigma}^2\{\mathbf{Y}\}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right]' \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\sigma^2\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\end{aligned}}$$

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

Inferences about parameters

$$\boldsymbol{\sigma}^2\{\mathbf{b}\} = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \qquad Let\ \mathbf{A} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} = \left[a_{ij}\right]_{p\times p}$$

$$b_k \sim N(\beta_k, \sigma^2\left\{b_k\right\}), \quad \text{with } \sigma^2\left\{b_k\right\} = a_{k+1,k+1}\sigma^2$$

$$s^2\left\{b_k\right\} = a_{k+1,k+1}MSE = a_{k+1,k+1}SSE/(n-p),$$

b_k is independent with SSE. Then

$$\frac{b_k-\beta_k}{s\left\{b_k\right\}} = \frac{(b_k-\beta_k)/\sigma\left\{b_k\right\}}{s\left\{b_k\right\}/\sigma\left\{b_k\right\}} = \frac{(b_k-\beta_k)/\sigma\left\{b_k\right\}}{\sqrt{\frac{SSE}{\sigma^2}/(n-p)}} \sim t\left(n-p\right)$$

CI for parameters

$$\frac{b_k - \beta_k}{s\{b_k\}} \sim t(n-p)$$

$$P\left\{\left|\frac{b_k - \beta_k}{s\{b_k\}}\right| < t\left(1-\frac{\alpha}{2}; n-p\right)\right\} = 1-\alpha$$

$(1-\alpha)100\%$ CI for β_k : $\quad b_k \pm t\left(1-\frac{\alpha}{2}; n-p\right)s\{b_k\}$

Simultaneous $(1-\alpha)100\%$ CI^s for $g \le p$:

$$b_k \pm t\left(1-\frac{\alpha}{2g}; n-p\right)s\{b_k\}$$

Hypothesis test about parameters

Test of $H_0 : \beta_k = 0 \quad H_A : \beta_k \neq 0$

Test Statistic: $t^* = \dfrac{b_k}{s\left\{b_k\right\}} \overset{H_0}{\sim} t\left(n-p\right)$

Rejection Region: $\left|t^*\right| \geq t\left(1-\dfrac{\alpha}{2}; n-p\right)$

P-value=$2\Pr\left(t\left(n-p\right) \geq \left|t^*\right|\right)$

Inference of the parameters: Studio data

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

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} \\ \sum X_{i2} & \sum X_{i2}X_{i1} & \sum X_{i2}^2 \end{bmatrix} = \begin{bmatrix} 21.0 & 1{,}302.4 & 360.0 \\ 1{,}302.4 & 87{,}707.9 & 22{,}609.2 \\ 360.0 & 22{,}609.2 & 6{,}190.3 \end{bmatrix}$$

$$\mathbf{s}^2\{\mathbf{b}\} = 121.1626 \begin{bmatrix} 29.7289 & .0722 & -1.9926 \\ .0722 & .00037 & -.0056 \\ -1.9926 & -.0056 & .1363 \end{bmatrix}$$

$$= \begin{bmatrix} 3{,}602.0 & 8.748 & -241.43 \\ 8.748 & .0448 & -.679 \\ -241.43 & -.679 & 16.514 \end{bmatrix}$$

Inference of the parameters: Studio data

$$s^2\{b_1\} = .0448 \quad \text{or} \quad s\{b_1\} = .212$$
$$s^2\{b_2\} = 16.514 \quad \text{or} \quad s\{b_2\} = 4.06$$

- **Single and simultaneous CIs?**
- Simultaneous 90% CI for parameters

$$B = t[1 - .10/2(2); 18] = t(.975; 18) = 2.101$$

$$1.455 \pm 2.101(.212)$$

$$9.366 \pm 2.101(4.06)$$

Test of $H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0$

Test Statistic: $t^* = \dfrac{b_1}{s\{b_1\}} \overset{H_0}{\sim} t(18)$

$$t^* = \frac{b_1}{s\{b_1\}} = \frac{1.455}{0.212} > t(0.975;18) = 2.101$$

$$\text{P-value} = 2\Pr\left(t(18) \geq \frac{1.455}{0.212}\right) < 0.05$$

Test of $H_0: \beta_2 = 0 \quad H_A: \beta_2 \neq 0$

Test Statistic: $t^* = \dfrac{b_2}{s\{b_2\}} \overset{H_0}{\sim} t(18)$

$$t^* = \frac{b_2}{s\{b_2\}} = \frac{9.366}{4.06} > t(0.975;18) = 2.101$$

$$\text{P-value} = 2\Pr\left(t(18) \geq \frac{9.366}{4.06}\right) < 0.05$$

6.7 Estimating Mean Response and Prediction of New Observations

- **Estimating Mean Response at Specific X-levels**

Given set of levels of $X_1, \ldots, X_{p-1}$: $\quad X_{h1}, \ldots, X_{h,p-1}$

$$\underset{p\times 1}{\mathbf{X}_h} = \left[1, X_{h1}, \cdots, X_{h,p-1}\right]', \qquad E\left\{Y_h\right\} = \mathbf{X}_h^{'}\boldsymbol{\beta} \qquad \hat{Y}_h = \mathbf{X}_h^{'}\mathbf{b}$$

$$E\left\{\hat{Y}_h\right\} = \mathbf{X}_h^{'}\boldsymbol{\beta} \qquad \sigma^2\left\{\hat{Y}_h\right\} = \mathbf{X}_h^{'}\boldsymbol{\sigma}^2\left\{\mathbf{b}\right\}\mathbf{X}_h = \sigma^2\mathbf{X}_h^{'}\left(\mathbf{X'X}\right)^{-1}\mathbf{X}_h$$

$$s^2\left\{\hat{Y}_h\right\} = MSE\left(\mathbf{X}_h^{'}\left(\mathbf{X'X}\right)^{-1}\mathbf{X}_h\right) = \mathbf{X}_h^{'}\mathbf{s}^2\left\{\mathbf{b}\right\}\mathbf{X}_h$$

$$\hat{Y}_h = \mathbf{X}_h^{'}\mathbf{b} \sim N\left(E\left\{\hat{Y}_h\right\} = \mathbf{X}_h^{'}\boldsymbol{\beta}, \sigma^2\left\{\hat{Y}_h\right\}\right), \;\; \hat{Y}_h \perp SSE$$

$$\frac{\hat{Y}_h - E\left\{\hat{Y}_h\right\}}{s\left\{\hat{Y}_h\right\}} = \frac{\left(\hat{Y}_h - E\left\{\hat{Y}_h\right\}\right)/\sigma\left\{\hat{Y}_h\right\}}{s\left\{\hat{Y}_h\right\}/\sigma\left\{\hat{Y}_h\right\}} = \frac{\left(\hat{Y}_h - E\left\{\hat{Y}_h\right\}\right)/\sigma\left\{\hat{Y}_h\right\}}{\sqrt{\dfrac{SSE}{\sigma^2}/(n-p)}} \sim t\left(n-p\right)$$

CI for Mean Response

$$E\left\{Y_h\right\} = \mathbf{X}_h^{'}\boldsymbol{\beta}, \quad \hat{Y}_h = \mathbf{X}_h^{'}\mathbf{b}, \quad \frac{\hat{Y}_h - E\left\{\hat{Y}_h\right\}}{s\left\{\hat{Y}_h\right\}} \sim t\left(n-p\right)$$

$$P\left\{\left|\frac{\hat{Y}_h - E\left\{\hat{Y}_h\right\}}{s\left\{\hat{Y}_h\right\}}\right| < t\left(1-\frac{\alpha}{2}; n-p\right)\right\} = 1-\alpha$$

$$\left(1-\alpha\right)100\% \text{ CI for } E\left\{\hat{Y}_h\right\}: \quad \hat{Y}_h \pm t\left(1-\frac{\alpha}{2}; n-p\right)s\left\{\hat{Y}_h\right\}$$

$$\left(1-\alpha\right)100\% \text{ CI for several } (g) \;\; E\left\{\hat{Y}_h\right\}: \quad E\left\{\hat{Y}_h\right\}: \hat{Y}_h \pm B \cdot s\left\{\hat{Y}_h\right\}$$

$$\left(1-\alpha\right)100\% \text{ Confidence Region for Regression Surface: } \hat{Y}_h \pm W \cdot s\left\{\hat{Y}_h\right\}$$

$$where \;\; B = t\left(1-\frac{\alpha}{2g}; n-p\right), W = \sqrt{pF\left(1-\alpha; p, n-p\right)}$$

Prediction of New Observations

Given set of levels of new $X_1, ..., X_{p-1}$: $\underset{p\times 1}{\mathbf{X}_h} = \left[1, X_{h1}, \cdots, X_{h,p-1}\right]'$,

Predicted New Response at $\mathbf{X}_{new} = \mathbf{X}_h$:

$$Y_{h(new)} = \mathbf{X_h \beta} + \varepsilon_{h,new} \sim N(\mathbf{X_h \beta}, \sigma^2)$$

$$\hat{Y}_{h(new)} = \mathbf{X_h b} \sim N(\mathbf{X_h \beta}, \sigma^2 \mathbf{X_h'} \left(\mathbf{X'X}\right)^{-1} \mathbf{X_h})$$

Prediction error $Y_{h(new)} - \hat{Y}_{h(new)} \sim N\left(0, \sigma^2 \left(1 + \mathbf{X_h'} \left(\mathbf{X'X}\right)^{-1} \mathbf{X_h}\right)\right)$

$$s^2\left\{\text{pred}\right\} = MSE\left(1 + \mathbf{X_h'} \left(\mathbf{X'X}\right)^{-1} \mathbf{X_h}\right)$$

$$\frac{Y_{h(\text{new})} - \hat{Y}_h}{s\left\{pred\right\}} \sim t\left(n-p\right), \quad P\left\{\left|\frac{Y_{h(\text{new})} - \hat{Y}_h}{s\left\{pred\right\}}\right| < t\left(1-\left(\alpha/2\right); n-p\right)\right\} = 1-\alpha$$

$(1-\alpha)*100\%$ prediction interval of $Y_{h(\text{new})} = \mathbf{X}_h \boldsymbol{\beta} + \varepsilon_{h(\text{new})}$

$$\hat{Y}_h \pm t\left(1-\left(\alpha/2\right); n-p\right) s\left\{pred\right\}$$

Predicting New Observations

$(1-\alpha)100\%$ prediction interval for $Y_{h(\text{new})}$:

$$\hat{Y}_h \pm t\left(1-\frac{\alpha}{2};n-p\right)s\{\text{pred}\}$$

Bonferroni: $(1-\alpha)100\%$ prediction interval for several (g) $Y_{h(\text{new})}$:

$$\hat{Y}_h \pm B \cdot s\{\text{pred}\}, \qquad B = t\left(1-\frac{\alpha}{2g};n-p\right)$$

Estimate mean response for Studio data

- Mean response at X1=65.4, X2=17.6

$$\mathbf{X}_h = \begin{bmatrix} 1 \\ 65.4 \\ 17.6 \end{bmatrix}$$

$$\hat{Y}_h = \mathbf{X}_h'\mathbf{b} = [1 \quad 65.4 \quad 17.6] \begin{bmatrix} -68.857 \\ 1.455 \\ 9.366 \end{bmatrix} = 191.10$$

$$\begin{aligned} s^2\{\hat{Y}_h\} &= \mathbf{X}_h' \mathbf{s}^2\{\mathbf{b}\}\mathbf{X}_h \\ &= [1 \quad 65.4 \quad 17.6] \begin{bmatrix} 3{,}602.0 & 8.748 & -241.43 \\ 8.748 & .0448 & -.679 \\ -241.43 & -.679 & 16.514 \end{bmatrix} \begin{bmatrix} 1 \\ 65.4 \\ 17.6 \end{bmatrix} \\ &= 7.656 \end{aligned}$$

$$s\{\hat{Y}_h\} = 2.77$$

95% CI: $191.10 \pm 2.101(2.77)$ $\quad 185.3 \leq E\{Y_h\} \leq 196.9$

Prediction for Studio data

- Prediction for new observation(city) with X1=65.4, X2=17.6,

$$\mathbf{X}_h = \begin{bmatrix} 1 \\ 65.4 \\ 17.6 \end{bmatrix}$$

$$\hat{Y}_h = 191.10 \qquad s^2\{\hat{Y}_h\} = 7.656 \qquad MSE = 121.1626$$

$$s^2\{\text{pred}\} = MSE + s^2\{\hat{Y}_h\} = 121.1626 + 7.656 = 128.82$$

$$s\{\text{pred}\} = 11.35$$

95% prediction interval: $191.10 \pm 2.101\, s\{\text{pred}\}$

$$167.3 \le Y_{h(\text{new})} \le 214.9$$

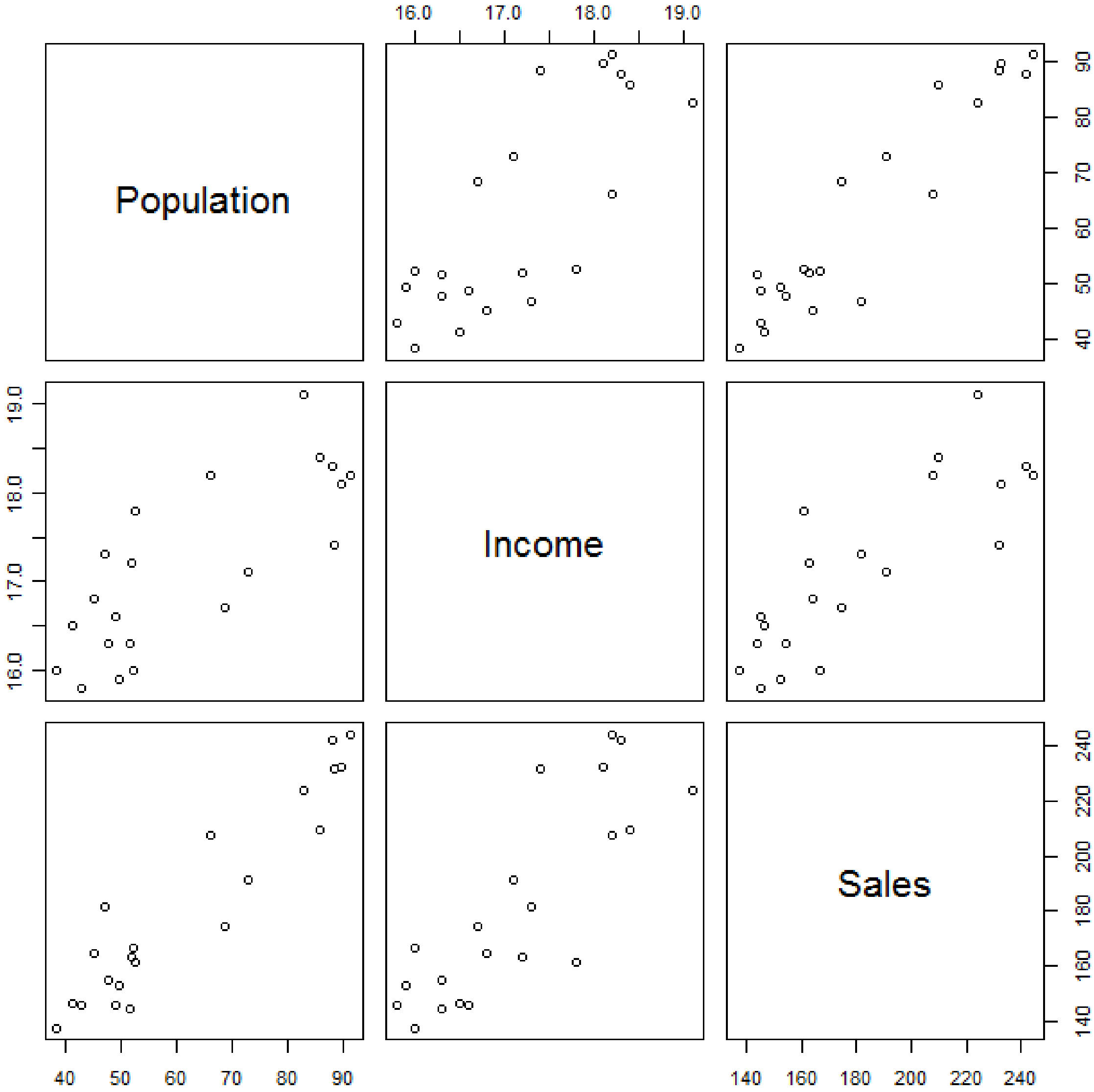
6.8 Diagnostics and Remedial Measures

- Very similar to simple linear regression.
- Only mention the difference.
- Given more than one predictor, must also consider relationship between predictors
- Scatterplot matrix summarizes bivariate relationships between Y and X_j as well as between X_j and X_k $(j, k = 1, 2, \ldots, p-1)$
  - Nature of bivariate relationships
  - Strength of bivariate relationships
  - Detection of outliers
  - Range spanned by X's

Scatter Plot Matrix for Studio data

R code:

pairs(data)



Correlation Matrix

- Displays all pairwise correlations
- R code: cor(data)

	Population	Income	Sales
Population	1.00	0.78	0.94
Income	0.78	1.00	0.84
Sales	0.94	0.84	1.00

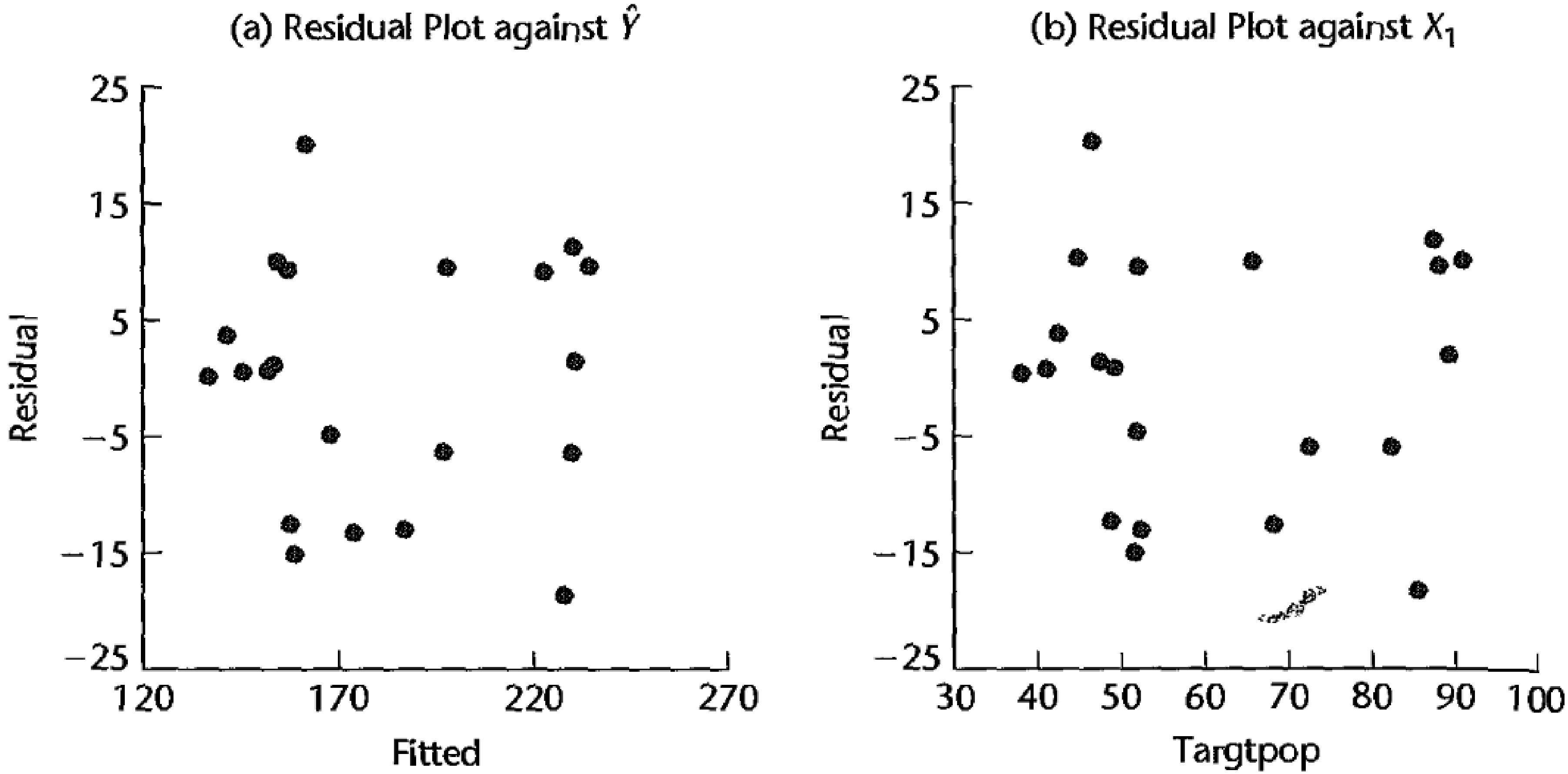
Residual Plots

- Plot e vs $\hat{Y}$ (overall)
- Plot e vs X_j (with respect to X_j)
- Plot e vs missing variable (e.g., $X_j X_k$)

Used for similar assessment of assumptions

- Linear model is correct
- Independence
- Normality
- Equal Variance
- omitted variables (including the interaction terms)?
- Outliers?

Residual Plots for Studio data



Tests for Diagnosis

- Correlation Test for Normality (Same, since it is on the residuals)
- Brown-Forsythe Test for Constancy of Error Variance (Need to find a way to divide the X space)
- Breusch-Pagan Test for Constancy of Error Variance (Same)
- F Test for Lack of Fit (Need to have replicates where all X fixed at same levels)
- Box-Cox Transformations (Same, since it is on Y)

Breusch-Pagan Test

Breusch-Pagan (aka Cook-Weisberg) Test:

$H_0:$ Equal Variance Among Errors $\sigma^2\{\varepsilon_i\} = \sigma^2 \ \forall\, i$

$H_A:$ Unequal Variance Among Errors $\sigma_i^2 = \sigma^2 h\left(\gamma_1 X_{i1} + \ldots + \gamma_k X_{ik}\right)$

1) Let $SSE = \sum_{i=1}^{n} e_i^2$ from original regression

2) Fit Regression of e_i^2 on $X_{i1}, \ldots X_{ik}$ and obtain $SS(\text{Reg}^*)$

Test Statistic: $X_{BP}^2 = \dfrac{SS(\text{Reg}^*)/2}{\left(\sum_{i=1}^{n} e_i^2 \big/ n\right)^2} \overset{H_0}{\sim} \chi_k^2$

Reject H_0 if $X_{BP}^2 \geq \chi^2(1-\alpha; k)$ k = # of predictors

Lack of Fit Test

- Compare
  - (reduced) linear model

    $$H_0: E(Y_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1}$$

  - (full) model where Y has c means (i.e. c combinations of X_i)

    $$H_a: E(Y_i) \neq \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1}$$

- In full model, there are c parameters $\hat{\mu}_j = \overline{Y}_j$,

$$SSE(F) = \sum_{j=1}^{c} \sum_{i=1}^{n_j} \left(Y_{ij} - \overline{Y}_j \right)^2 \quad df_F = n - c$$

Lack of Fit Test

$$SSE(R) = \sum_{j=1}^{c} \sum_{i=1}^{n_j} \left(Y_{ij} - \hat{Y}_{ij}\right)^2 \quad df_R = n - p$$

$$\hat{Y}_{ij} = b_0 + b_1 X_{ij1} + \cdots + b_{p-1} X_{ij,p-1}$$

$$SSE(F) = \sum_{j=1}^{c} \sum_{i=1}^{n_j} \left(Y_{ij} - \overline{Y}_j\right)^2 \quad df_F = n - c$$

- $F^* = \frac{\{SSE(R)-SSE(F)\}/\{(n-p)-(n-c)\}}{SSE(F)/(n-c)} \sim F(c-p, n-c)$ under H_0
- Reject H_0 if $F^* > F(1-\alpha, c-p, n-c)$
- If reject H_0, conclude that a more complex relationship between Y and $X_1, \ldots, X_{p-1}$ is needed

R code for studio data

```
studio = read.table('studio.txt')
names(studio)=c('X1','X2','Y')
fit = lm(Y~X1+X2, data=studio)
summary(fit);   confint(fit);  anova(fit)

newx = data.frame(X1 = 65.4, X2 = 17.6)
yhat= predict(fit)
predict(fit, newx, interval="confidence",level=.95)
predict(fit, newx, interval="prediction",level=.95)

resi = fit$resi;  plot(yhat, resi);  plot(X1,resi);  plot(X2,resi)
```

Homework

- Page 253

  6.23 6.24 6.25 6.26

- Page 253

  6.27 6.31(a)(b)(c)