北京师范大学 香港浸会大学 联合国际学院

BEIJING NORMAL UNIVERSITY · HONG KONG BAPTIST UNIVERSITY

UNITED INTERNATIONAL COLLEGE

STAT2013 Regression Analysis

Semester 1, 2018-2019

Room: T3-401-R1

Instructor: Ye, Huajun Terry

TA：Jenna Otto

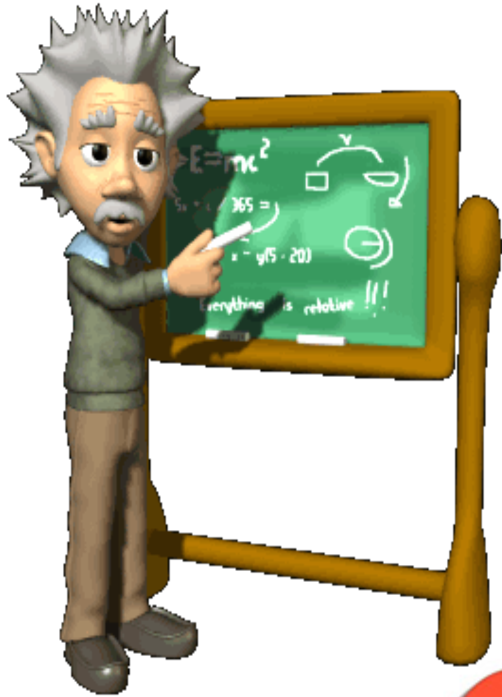
Email:hjye@uic.edu.hk

Jotto@uic.edu.hk (TA)

Chapter 2
The Simple Linear Regression Model

In this Chapter, we will know some important concepts.

- Dada, model and the least squares estimation
- Estimator Properties
- Tests of Hypothesis
- Confidence Intervals
- More Regression Models



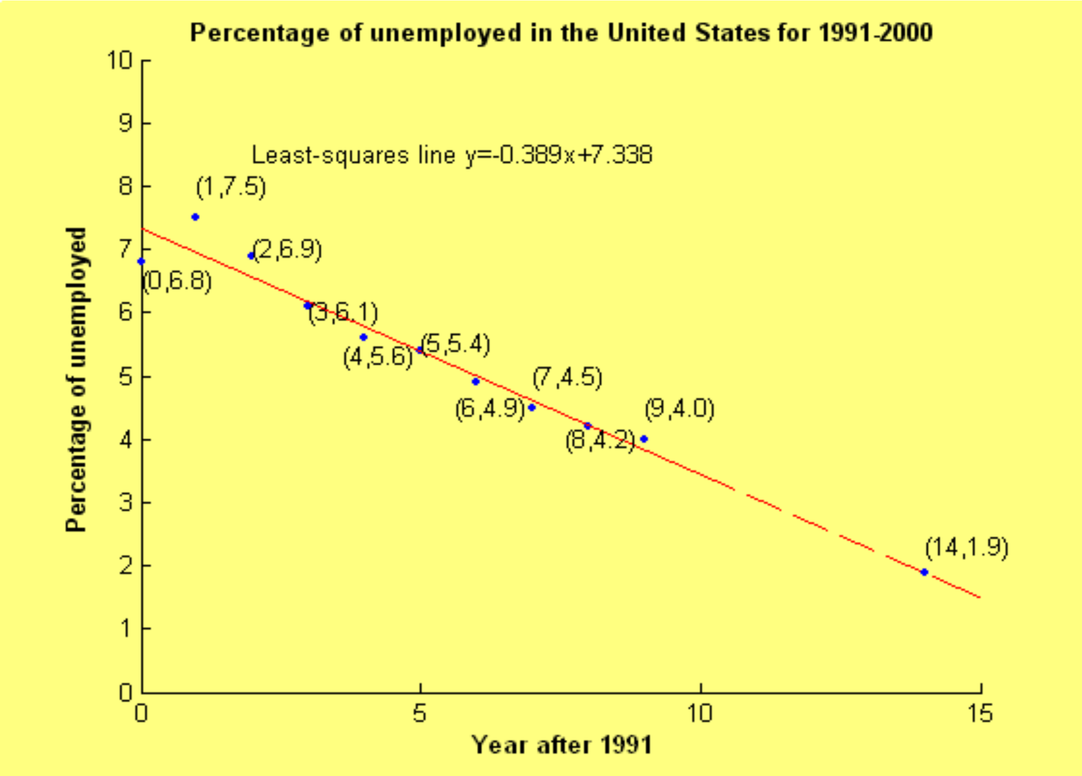
2.1

Data, model and the least squares estimation

Table 3.1 lists the percentage of the labour force that was unemployed during the decade *1991-2000*. Plot a graph with the time (years after 1991) on the x axis and percentage of unemployment on the y axis. Do the points follow a clear pattern? Based on these data, what would you expect the percentage of unemployment to be in the year *2005*?

Table 3.1 Percentage of Civilian Unemployment

Year	Number of Years from 1991	Percentage of Unemployed
1991	0	6.8
1992	1	7.5
1993	2	6.9
1994	3	6.1
1995	4	5.6
1996	5	5.4
1997	6	4.9
1998	7	4.5
1999	8	4.2
2000	9	4.0



The pattern does suggest that we may be able to get useful information by finding a line that "best fits" the data in some meaningful way. It produces the "best-fitting line".

$$y = -0.389x + 7.338$$

Based on this formula, we can attempt a prediction of the unemployment rate in the year *2005*:

$$y(14) = -0.389(14) + 7.338 = 1.892$$

Note: Care must be taken when making predictions by extrapolating from known data, especially when the data set is as small as the one in this example.

Correlation(相关) vs. Regression(回归)

- A scatter diagram (散点图) can be used to show the relationship between two variables
- Correlation (相关) analysis is used to measure strength of the association (linear relationship) between two variables
  - Correlation is only concerned with strength of the relationship
  - No causal effect (因果效应) is implied with correlation

Scatter Diagrams

- Scatter Diagrams are used to examine possible relationships between two numerical variables

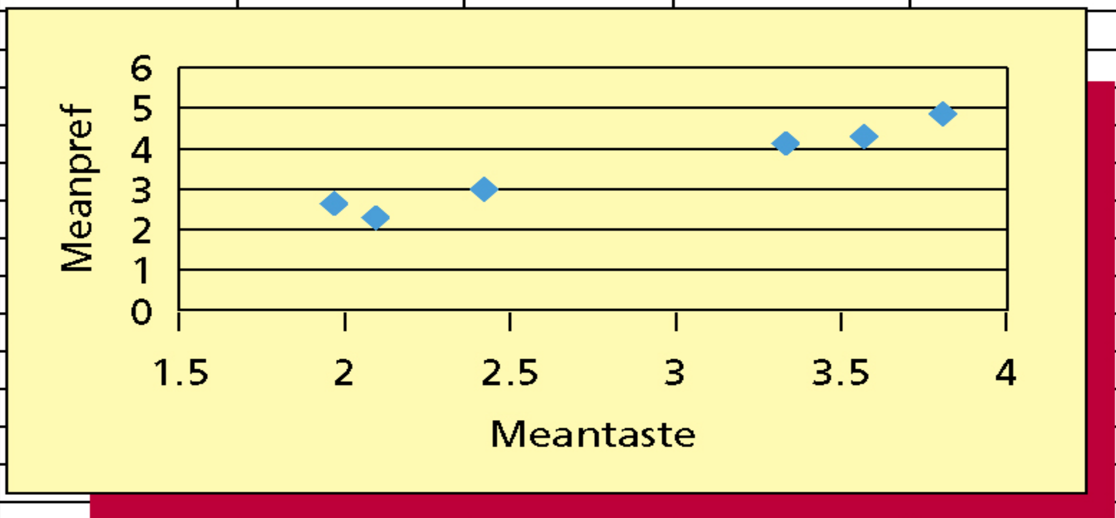
- The Scatter Diagram:
  - one variable is measured on the vertical axis and the other variable is measured on the horizontal axis

Scatter Plots(散点图)

Visualize the data to see patterns, especially “trends”

Restaurant Ratings: Mean Preference vs. Mean Taste

	A	B	C	D	E	F
1	Restaurant	Meantaste	Meanconv	Meanfam	Meanprice	Meanpref
2	Borden Burger	3.5659	2.7005	2.5282	2.9372	4.2552
3	Hardee’s	3.329	3.3483	2.7345	2.7513	4.0911
4	Burger King	2.4231	2.7377	2.3368	3.0761	3.0052
5	McDonald’s	2.0895	1.938	1.4619	2.4884	2.2429
6	Wendy’s	1.9661	2.892	2.3376	4.0814	2.5351
7	White Castle	3.8061	3.7242	2.6515	1.708	4.7812
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						



Introduction to Regression Analysis

- Regression analysis is used to:
  - Predict the value of a dependent variable based on the value of at least one independent variable
  - Explain the impact of changes in an independent variable on the dependent variable

Dependent variable: the variable we wish to predict or explain

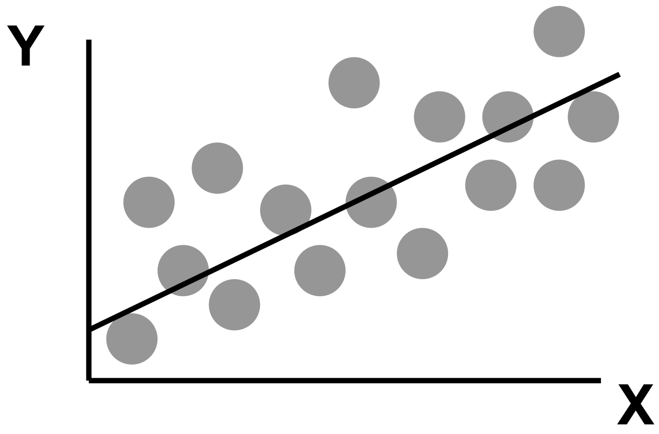
Independent variable: the variable used to explain the dependent variable

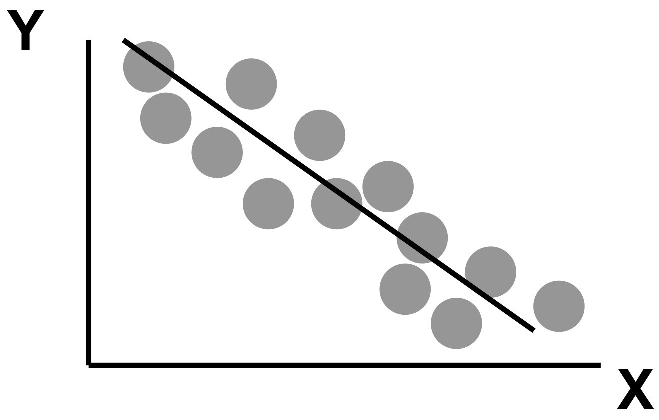
Simple Linear Regression Model

- Only **one** independent variable, X
- Relationship between X and Y is described by a linear function
- Changes in Y are assumed to be caused by changes in X

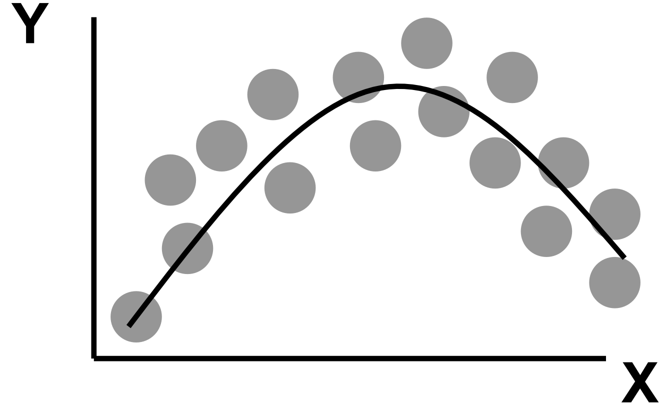
Types of Relationships

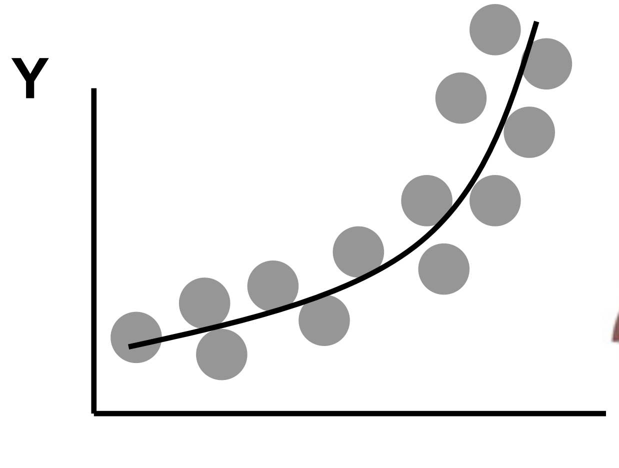
Linear relationships





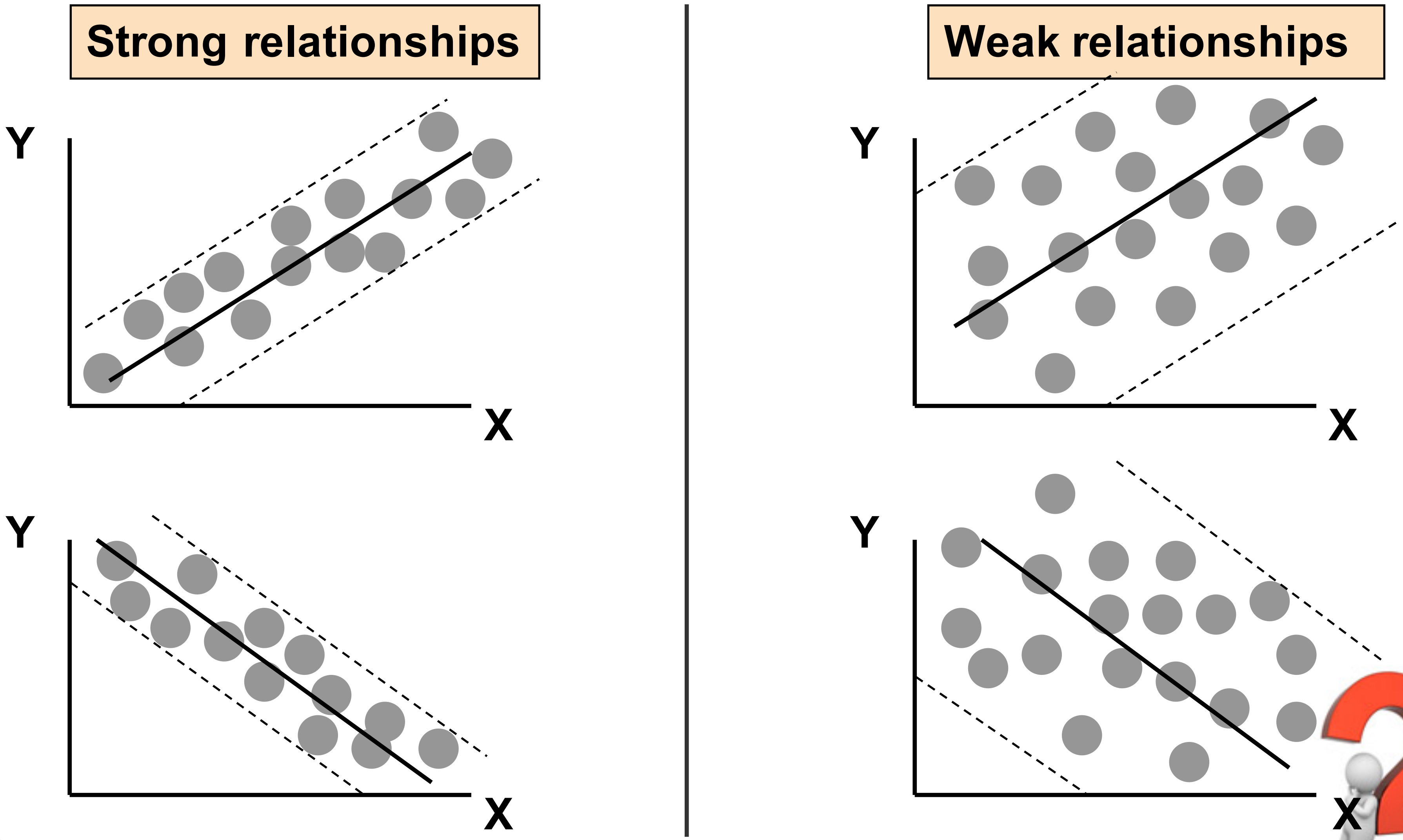
Curvilinear relationships





Types of Relationships

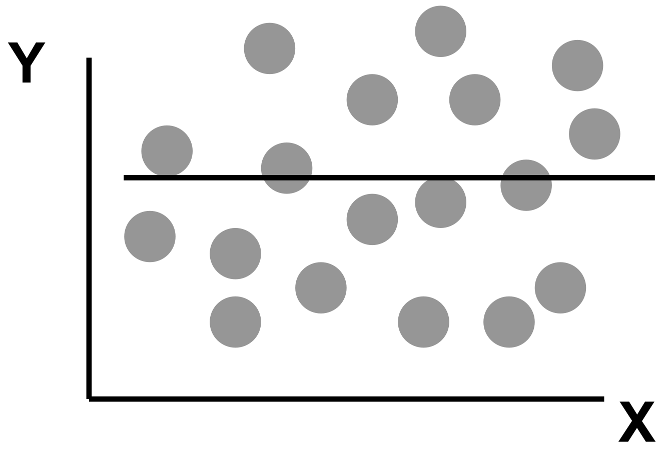
(continued)

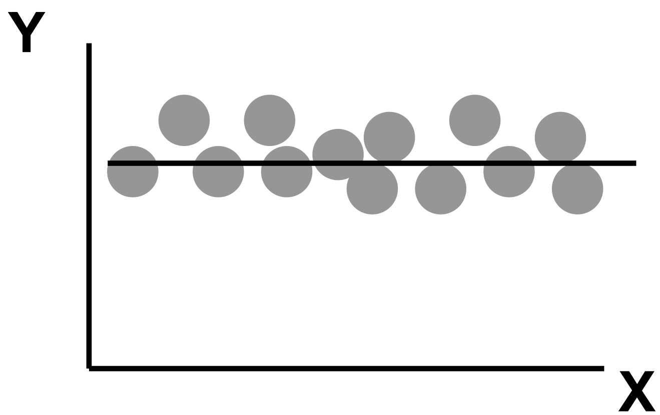


Types of Relationships

(continued)

No relationship





Simple Linear Regression Model

Dependent Variable

Population Y intercept

Population Slope Coefficient

Independent Variable

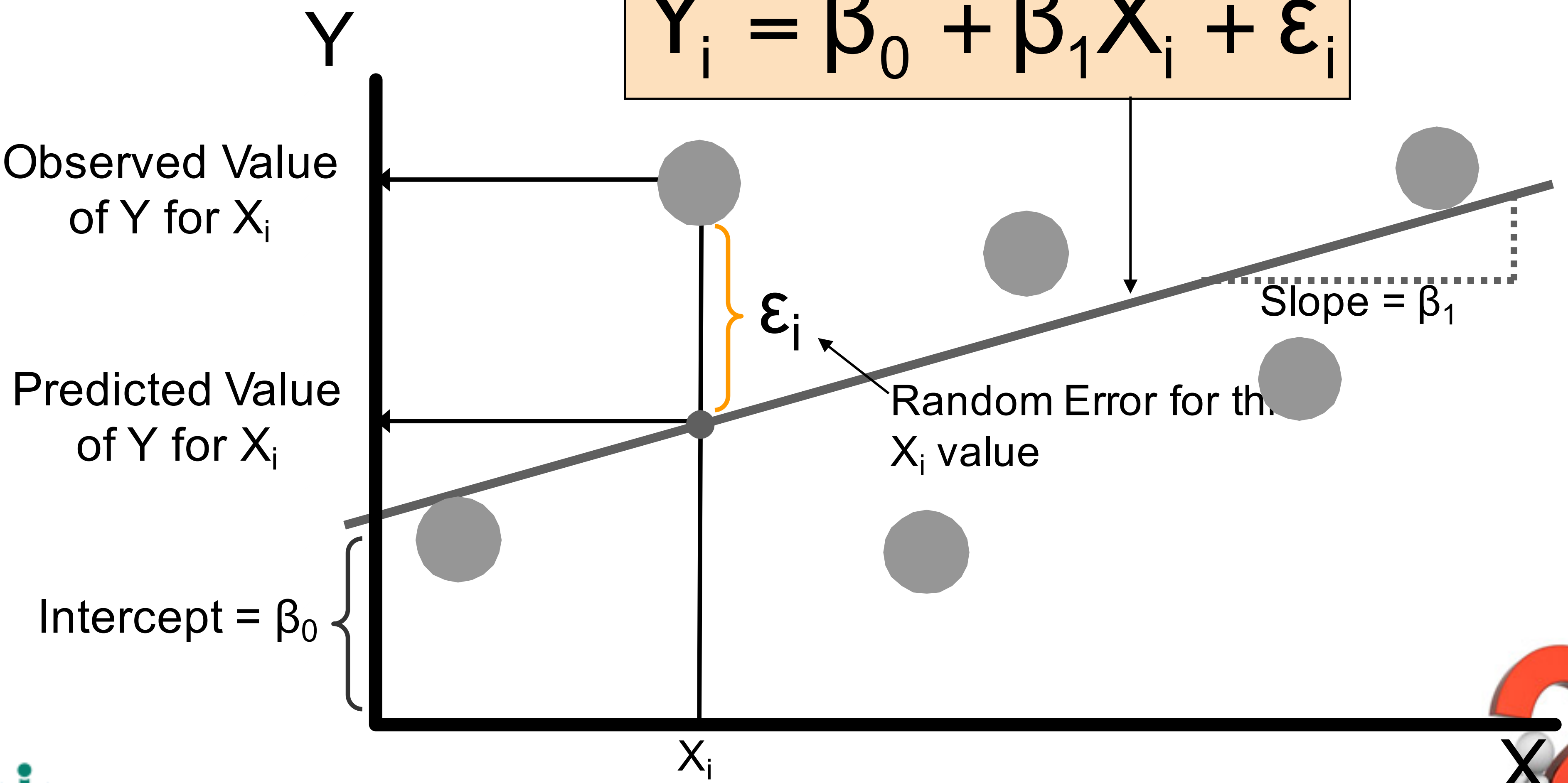
Random Error term

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Linear component}} + \underbrace{\varepsilon_i}_{\text{Random Error component}}$$

Simple Linear Regression Model

(continued)

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$



Simple Linear Regression Equation (Prediction Line)

The simple linear regression equation provides an estimate of the population regression line

Estimated (or predicted) Y value for observation i

Estimate of the regression intercept

Estimate of the regression slope

Value of X for observation i

$$\hat{Y}_i = b_0 + b_1 X_i$$

The individual random error terms e_i have a mean of zero

Least Squares Method
(最小二乘方法)

- b_0 and b_1 are obtained by finding the values of b_0 and b_1 that minimize the sum of the squared differences between Y and $\hat{Y}$:

$$\min \sum (Y_i - \hat{Y}_i)^2 = \min \sum (Y_i - (b_0 + b_1 X_i))^2$$

Interpretation of the Slope(斜率) and the Intercept(截距)

- b_0 is the estimated average value of Y when the value of X is zero

- b_1 is the estimated change in the average value of Y as a result of a one-unit change in X

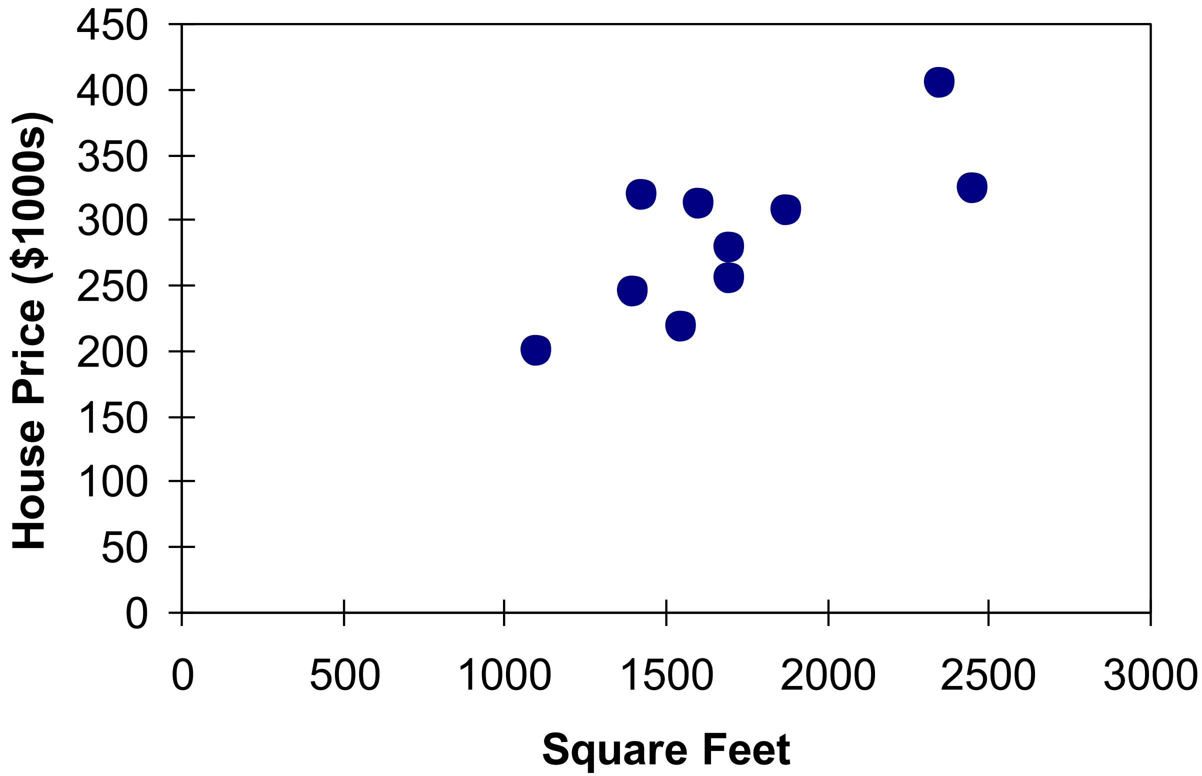
The House Price Case

- A real estate agent wishes to examine the relationship between the selling price of a home and its size (measured in square feet)

- A random sample of 10 houses is selected
  - Dependent variable (Y) = house price in $1000s
  - Independent variable (X) = square feet

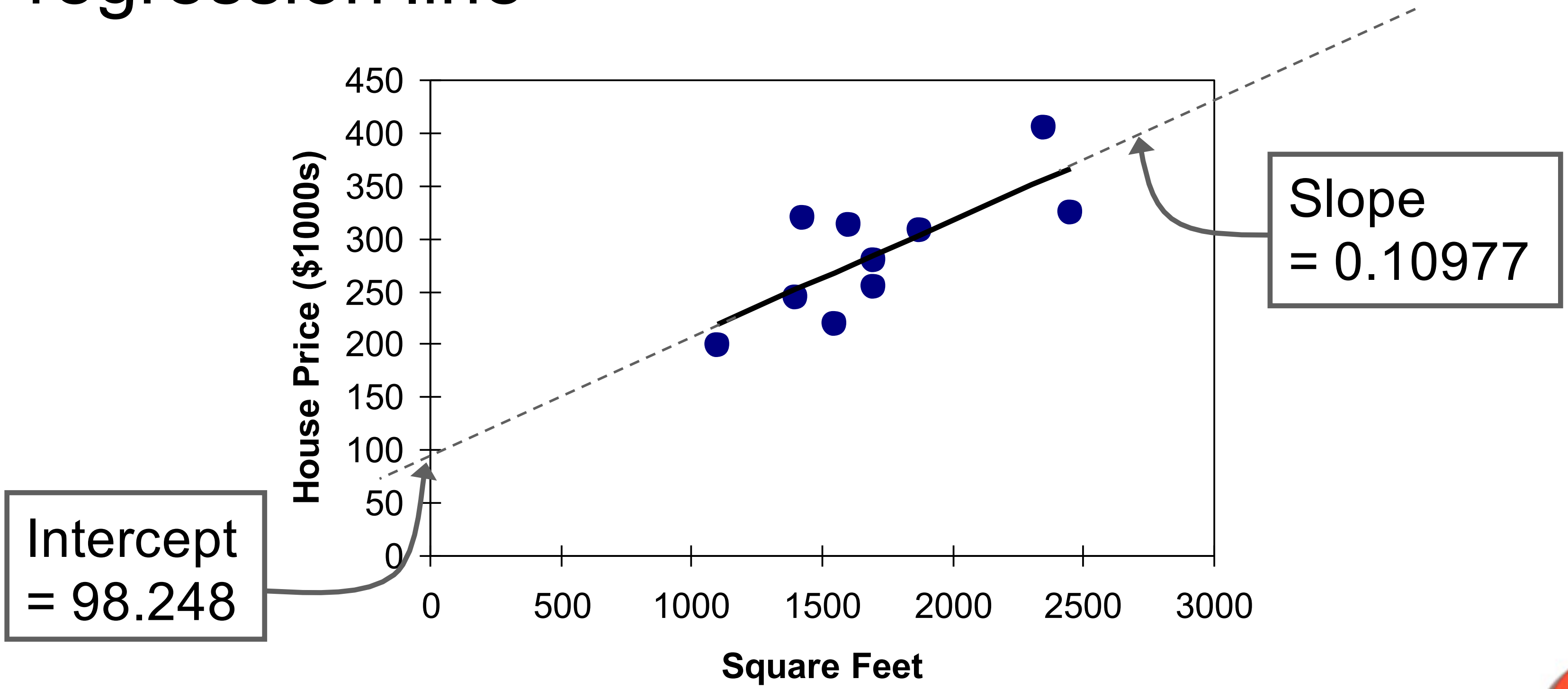
Graphical Presentation

- House price model: scatter plot



Graphical Presentation

- House price model: scatter plot and regression line



$$\widehat{\text{house price}} = 98.24833 + 0.10977\ (\text{square feet})$$

Interpretation of the Intercept, b_0

$$\widehat{\text{house price}} = \boxed{98.24833} + 0.10977\,(\text{square feet})$$

- b_0 is the estimated average value of Y when the value of X is zero (if X = 0 is in the range of observed X values)
  - Here, no houses had 0 square feet, so $b_0 = 98.24833$ just indicates that, for houses within the range of sizes observed, \$98,248.33 is the portion of the house price not explained by square feet

Interpretation of the Slope Coefficient, b_1

$$\widehat{\text{house price}} = 98.24833 + \boxed{0.10977}\,(\text{square feet})$$

- b_1 measures the estimated change in the average value of Y as a result of a one-unit change in X
  - Here, b_1 = .10977 tells us that the average value of a house increases by .10977($1000) = $109.77, on average, for each additional one square foot of size

Predictions using Regression Analysis

Predict the price for a house with 2000 square feet:

$$\widehat{\text{house price}} = 98.25 + 0.1098\,(\text{sq.ft.})$$

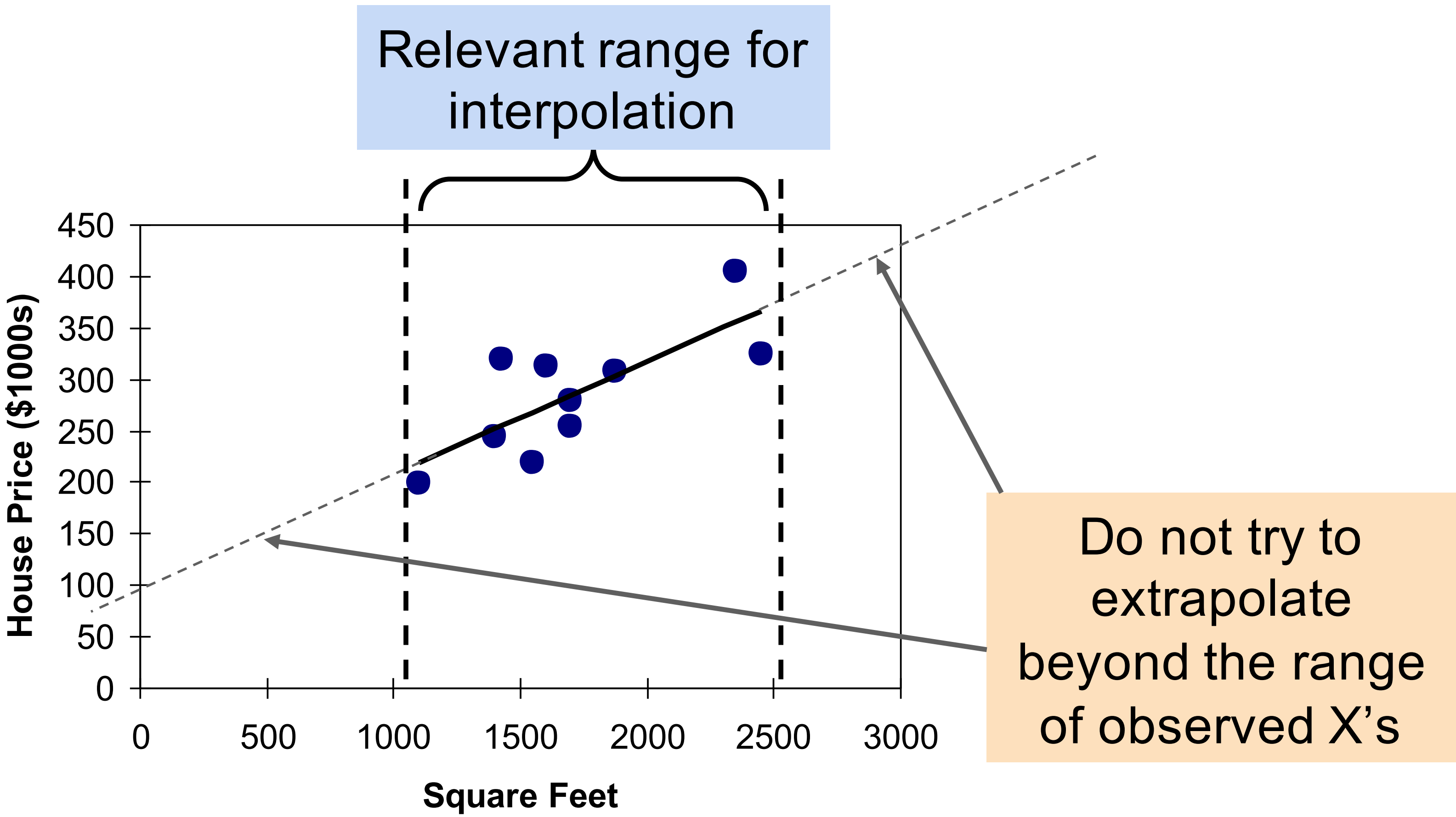
$$= 98.25 + 0.1098(2000)$$

$$= 317.85$$

The predicted price for a house with 2000 square feet is 317.85($1,000s) = $317,850

Interpolation vs. Extrapolation

- When using a regression model for prediction, only predict within the relevant range of data



The Least Square Point Estimates

Estimation/prediction equation

$$\hat{y} = b_0 + b_1 x$$

Least squares point estimate of the slope β_1

$$b_1 = \frac{S_{xy}}{S_{xx}} \qquad S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n}$$

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n} \qquad S_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}$$

Least squares point estimate of the y-intercept β_0

$$b_0 = \bar{y} - b_1 \bar{x} \qquad \bar{y} = \frac{\sum y_i}{n} \qquad \bar{x} = \frac{\sum x_i}{n}$$

Basic model with one regressor:

SLR Model Form

$$y = \beta_0 + \beta_1 x + \epsilon$$

The least squares procedure and SLR have the following assumptions:

1. x_i's are nonrandom
2. $\epsilon_i \sim N(0, \sigma^2)$ and *independent*
3. ϵ_i are uncorrelated

Thus, for any value of x_i, the SLR model says that $y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (i = 1, \ldots, n)$. The mean of the distribution of the y_i, $\hat{y}_i$, is given by

$$E[y_i] = \underbrace{\beta_0 + \beta_1 x_i}_{constants} + \underbrace{E[\epsilon_i]}_{=0} = \beta_0 + \beta_1 x_i$$

and the variance is

$$Var[y_i] = \underbrace{Var[\beta_0 + \beta_1 x_i]}_{=0} + \underbrace{Var[\epsilon_i]}_{=\sigma^2} = \sigma^2.$$

interpretation of β_0 may be foolhardy as a value of *x = 0* may not be within the range of the data, nor a feasible value.

An alternative SLR model exists by centering the data about the mean.

$$\hat{y}_i = \beta_0 + \beta_1 x_i \underbrace{- \beta_1 \overline{x} + \beta_1 \overline{x}}_{=0} = \underbrace{(\beta_0 + \beta_1 \overline{x})}_{\beta_0^*} + \beta_1 (x_i - \overline{x})$$

Alternative SLR Equation

$$\hat{y}_i = \beta_0^* + \beta_1 (x_i - \overline{x}).$$

A. Methods of Estimation

Observation: $(x_1, y_1), \cdots, (x_n, y_n)$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i , \qquad i = 1, \cdots, n$$

$$\varepsilon_1, \cdots, \varepsilon_n \text{ are i.i.d.,} \quad E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \sigma^2$$

We want to estimate β_0, β_1 and σ^2.

- If we know the distribution of ε, the *maximum likelihood estimation (MLE)* is available. Otherwise, the *least squares estimates (LSE) and other estimates* are considered.

$$e_i = y_i - \hat{y}_i \quad : \text{residual}$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \ : \text{estimated value of } y_i$$

$\hat{\beta}_0$ and $\hat{\beta}_1$ are estimates of β_0 and β_1 respectively.

→ We want to minimize all e_i's in a certain sense.

Let $b_0, b_1,$ *and* b_0^* be estimators of the unknown coefficients. i.e. $b_0 \approx \beta_0, b_1 \approx \beta_1, b_0^* \approx \beta_0^*$. Least squares says that we want to minimize the squared distance of our observations, y_i, from their corresponding *fitted values*, $\hat{y}_i$. Using the alternative formula, we want to minimize

$$\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - b_0^* - b_1(x_i - \overline{x}))^2.$$

Taking partial derivatives with respect to β_0^*, we have

$$\begin{aligned} -2\sum_{i=1}^{n}(y_i - b_0^* - b_1(x_i - \overline{x})) = -2\sum_{i=1}^{n} \epsilon_i &= 0 \\ n\overline{y} - nb_0^* &= 0 \\ b_0^* &= \overline{y} \end{aligned}$$

Taking partial derivatives with respect to β_1 , we have

$$-2\sum_{i=1}^{n}(x_i-\overline{x})(y_i-b_0^*-b_1(x_i-\overline{x})) = 0$$

$$\underbrace{\sum_{i=1}^{n}(x_i-\overline{x})y_i-\sum_{i=1}^{n}(x_i-\overline{x})b_0^*}_{=\sum_{i=1}^{n}(x_i-\overline{x})(y_i-b_0^*)}-b_1\sum_{i=1}^{n}(x_i-\overline{x})^2 = 0$$

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) \qquad b_1=\frac{s_{xy}}{s_{xx}}$$

Therefore, since we know that $b_0 = b_0^* - b_1\overline{x}$, we can solve for b_0, which is

$$b_0 = \overline{y} - b_1\overline{x}.$$

This gives the ordinary least squares estimates b_0 of β_0 and b_1 of β_1 as

$$b_0 = \overline{y} - b_1\overline{x}$$

$$b_1 = \frac{s_{xy}}{s_{xx}}$$

where

$$s_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}),\quad s_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2,\quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n}x_i,\ \overline{y} = \frac{1}{n}\sum_{i=1}^{n}y_i.$$

$$\frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0^2} = -2\sum_{i=1}^{n}(-1) = 2n,$$

$$\frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_1^2} = 2\sum_{i=1}^{n} x_i^2$$

$$\frac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0 \partial \beta_1} = 2\sum_{i=1}^{n} x_t = 2n\overline{x}.$$

The Hessian matrix which is the matrix of second order partial derivatives

$$H^* = \begin{pmatrix} \dfrac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0^2} & \dfrac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0 \partial \beta_1} \\ \dfrac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_0 \partial \beta_1} & \dfrac{\partial^2 S(\beta_0,\beta_1)}{\partial \beta_1^2} \end{pmatrix}$$

$$= 2\begin{pmatrix} n & n\overline{x} \\ n\overline{x} & \sum_{i=1}^{n} x_i^2 \end{pmatrix}$$

$$\left| H^* \right| = 2\left(n\sum_{i=1}^{n} x_i^2 - n^2\overline{x}^2 \right)$$

$$= 2n\sum_{i=1}^{n}(x_i - \overline{x})^2$$

The case when $\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0$ is not interesting because all the observations in this case are identical, i.e. $x_i = c$ (some constant). In such a case there is no relationship between x and y in the context of regression analysis. Since $\sum_{i=1}^{n}(x_i - \overline{x})^2 > 0$, therefore $|H| > 0$. So H is positive definite for any (β_0, β_1); therefore $S(\beta_0, \beta_1)$ has a global minimum at (b_0, b_1).

Reference books

方开泰， 陈敏 (2013), 统计学中的矩阵代数,

高等教育出版社, 北京.

Lay D (1994),

Linear Algebra and Its Applications,

Addison-Wesley Publishing Company.

:

$\mathbf{A} = (a_{ij}) : n \times p$

Transpose

$\mathbf{A}' : p \times n$

Square Matrix

$\mathbf{A} : n \times n$

Row Vector

$\boldsymbol{a} = (a_1, \ldots, a_p): 1 \times p$

Column Vector

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} : n \times 1$$

Symmetric

$\mathbf{A} = \mathbf{A}'$

Upper Triangle

$\mathbf{A} = (a_{ij}),\ a_{ij} = 0 \quad \text{if } i > j$

Diagonal Matrix

$$\mathbf{A} = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} = \mathrm{diag}(a_1, \cdots, a_n)$$

Lower Triangle

$\mathbf{A} = (a_{ij}),\ a_{ij} = 0 \quad \text{if } i < j$

Identical Matrix of *n*

$\mathbf{I}_n = \mathrm{diag}(1, \cdots, 1)$

The Derivatives of a Matrix

A multivariate function

$$y = f(x_1,\cdots,x_n) = f(\boldsymbol{x}),\quad \boldsymbol{x} \in \boldsymbol{R}^n$$

If there exist all the derivatives of $\partial y/\partial x_i$ ($i = 1,\cdots,n$), we write

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_1} \\ \vdots \\ \dfrac{\partial y}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f(x_1,\cdots,x_n)}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(x_1,\cdots,x_n)}{\partial x_n} \end{bmatrix}$$

and called the derivatives of y with respect to $\boldsymbol{x}$.

Example:

$$y = \boldsymbol{x}'\boldsymbol{x} = x_1^2 + \; + x_n^2 \; .$$

$$\frac{\partial y}{\partial x_i} = 2x_i, \quad i = 1, \cdots, n \qquad \Rightarrow \qquad \frac{\partial(\boldsymbol{x}'\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\boldsymbol{x}.$$

Example:

$y = \boldsymbol{x}'\mathbf{A}\boldsymbol{x}$, $\mathbf{A}=(a_{ij})$: $n\times n$, constant matrix

$$\frac{\partial y}{\partial x_i} = \sum_{j=1}^{n}(a_{ij} + a_{ji})x_j, \quad i = 1,\cdots,n$$

We have

$$\frac{\partial(\boldsymbol{x}'\mathbf{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = (\mathbf{A} + \mathbf{A}')\boldsymbol{x}.$$

In particular, we have

$$\frac{\partial(\boldsymbol{x}'\mathbf{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\mathbf{A}\boldsymbol{x}.$$

if $\mathbf{A} = \mathbf{A}'$ is symmetric.

Example:

$$y = \boldsymbol{a}'\boldsymbol{x}, \qquad \frac{\partial y}{\partial \boldsymbol{x}} = \boldsymbol{a}$$

Let $y = f(\boldsymbol{X}) = f(x_{11}, \cdots, x_{1m}, x_{21}, \cdots x_{2m}, \cdots, x_{n1}, \cdots x_{nm})$,

$\boldsymbol{X} = (x_{nm}) : n \times m$ be a scale function of x.

We can similarly define

$$\frac{\partial y}{\partial \boldsymbol{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \cdots & \frac{\partial y}{\partial x_{1m}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{n1}} & \cdots & \frac{\partial y}{\partial x_{nm}} \end{bmatrix} = \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}.$$

Matrix Algebra in Estimation for Regression Model

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

Then the linear model can be expressed as:

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

and

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

$$\frac{\partial \mathrm{Q}}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta})$$

$$= -2X'\boldsymbol{y} + 2X'\boldsymbol{X}\boldsymbol{\beta} = 0$$

$$\frac{\partial \mathrm{Q}}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta})$$

$$= -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = 0$$

The normal equation becomes

$$(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{y}$$

If $\boldsymbol{X'X}$ is not singular ($\boldsymbol{X'X}$ >0), then the least squares estimate of $\boldsymbol{\beta}$ is given by:

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

We should point out that **Q(*β*)** arrives its minimum at $\hat{\beta}$.

It is easy to see that

$$
\begin{aligned}
Q(\beta) &= (y - X\beta)'(y - X\beta) \\
&= (y - X\hat{\beta} + X\hat{\beta} - X\beta)'(y - X\hat{\beta} + X\hat{\beta} - X\beta) \\
&= (y - X\hat{\beta})'(y - X\hat{\beta}) \quad \} \quad \text{I} \\
&\quad + (\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta) \quad \} \quad \text{II} \\
&\quad + (y - X\hat{\beta})'(X\hat{\beta} - X\beta) \quad \} \quad \text{III} \\
&\quad + (X\hat{\beta} - X\beta)'(y - X\beta) \quad \} \quad \text{IV}
\end{aligned}
$$

$$
\mathrm{I}: \begin{cases} (y - X\hat{\boldsymbol{\beta}})'(y - X\hat{\boldsymbol{\beta}}) \\ = \mathrm{Q}(\hat{\boldsymbol{\beta}}) \end{cases}
$$

$$
\mathrm{II}: \begin{cases} \text{For} \quad (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' X' X (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ \text{Let} \quad \boldsymbol{u} = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ \boldsymbol{u}' X' X\boldsymbol{u} = (X\boldsymbol{u})'(X\boldsymbol{u}) \begin{cases} \geq 0 & \text{if } X \text{ is singular,} \\ > 0 & \text{if } X \text{ is non-singular, } \boldsymbol{u} \neq \mathbf{0}. \end{cases} \end{cases}
$$

$$
\mathrm{III}:\begin{cases}
(y-X\hat{\beta})'(X\hat{\beta}-X\beta) \\
=(y-X(X'X)^{-1}X'y)'X(\hat{\beta}-\beta) \\
=y'(I_n-X(X'X)^{-1}X')'X(\hat{\beta}-\beta) \\
=\boldsymbol{y'(I_n-X(X'X)^{-1}X')X(\hat{\beta}-\beta)} \\
=\boldsymbol{y'(X-X(\cancel{X'X})^{-1}(\cancel{X'X}))(\hat{\beta}-\beta)} \\
=\mathbf{0}
\end{cases}
$$

$$
\mathrm{IV}:\begin{cases}
\boldsymbol{(X\hat{\beta}-X\beta)'(y-X\hat{\beta})=0} \\
\text{(follows by III)}
\end{cases}
$$

Therefore,

$$
\begin{aligned}
\mathrm{Q}(\boldsymbol{\beta}) &= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} \\
&= \mathrm{Q}(\hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \boldsymbol{X}' \boldsymbol{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + 0 + 0 \\
&\begin{cases} = \mathrm{Q}(\hat{\boldsymbol{\beta}}) & \text{if } \boldsymbol{\beta} = \hat{\boldsymbol{\beta}} \\ \geq \mathrm{Q}(\hat{\boldsymbol{\beta}}) & \text{if } \boldsymbol{\beta} \neq \hat{\boldsymbol{\beta}} \end{cases} \quad \forall \boldsymbol{\beta} \in \boldsymbol{R}^2
\end{aligned}
$$

Estimation of $\theta = X\beta$

$$\hat{y} = X\hat{\beta} \quad \text{is an estimator for } \boldsymbol{\theta} = X\beta$$

$$\boxed{\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Hy}$$

where ***H*** = ***X***(***X'X***) $^{-1}$ ***X*** is called ***Hat Matrix***.

2.2 Estimator Properties

Unbiased Property

Note that $b_1 = \frac{s_{xy}}{s_{xx}}$ and $b_0 = \overline{y} - b_1\overline{x}$ are the linear combinations of $y_i (i = 1, \ldots, n)$.

Therefore

$$b_1 = \sum_{i=1}^{n} k_i y_i$$

where $k_i = (x_i - \overline{x}) / s_{xx}$. Note that $\sum_{i=1}^{n} k_i = 0$ and $\sum_{i=1}^{n} k_i x_i = 1$, so

$$\begin{aligned} E(b_1) &= \sum_{i=1}^{n} k_i E(y_i) \\ &= \sum_{i=1}^{n} k_i (\beta_0 + \beta_1 x_i) . \\ &= \beta_1. \end{aligned}$$

$$
\begin{aligned}
E(b_0) &= E\left[\bar{y} - b_1\bar{x}\right] \\
&= E\left[\beta_0 + \beta_1\bar{x} + \bar{\varepsilon} - b_1\bar{x}\right] \\
&= \beta_0 + \beta_1\bar{x} - \beta_1\bar{x} \\
&= \beta_0 .
\end{aligned}
$$

$$
\begin{aligned}
E[b_0^*] &= E[\bar{y}] = \frac{1}{n}E\left[\sum_{i=1}^{n} y_i\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} \beta_0^* + \beta_1(x_i - \bar{x})\right] \\
&= \frac{1}{n}\left(\underbrace{\sum_{i=1}^{n} \beta_0^*}_{n\beta_0^*} + \beta_1 \underbrace{\sum_{i=1}^{n}(x_i - \bar{x})}_{=0}\right) = \beta_0^*
\end{aligned}
$$

Variances

Using the assumption that y_i's are independently distributed, the variance of b_1 is

$$Var(b_1) = \sum_{i=1}^{n} k_i^2 Var(y_i) + \sum_i \sum_{j \neq i} k_i k_j Cov(y_i, y_j)$$

$$= \sigma^2 \frac{\sum_i (x_i - \bar{x})^2}{s_{xx}^2} \quad (Cov(y_i, y_j) = 0 \text{ as } y_1, ..., y_n \text{ are independent})$$

$$= \frac{\sigma^2 s_{xx}}{s_{xx}^2}$$

$$= \frac{\sigma^2}{s_{xx}}.$$

$$Var[b_0^*] = Var[\overline{y}] = \frac{\sigma^2}{n}.$$

The variance of b_0 is

$$Var(b_0) = Var(\overline{y}) + \overline{x}^2 Var(b_1) - 2\overline{x}Cov(\overline{y}, b_1).$$

First we find that

$$\begin{aligned}
Cov(\overline{y}, b_1) &= E\left[\{\overline{y} - E(\overline{y})\}\{b_1 - E(b_1)\}\right] \\
&= E\left[\overline{\varepsilon}(\sum_i c_i y_i - \beta_1)\right] \\
&= \frac{1}{n}E\left[(\sum_i \varepsilon_i)(\beta_0 \sum_i c_i + \beta_1 \sum_i c_i x_i + \sum_i c_i \varepsilon_i) - \beta_1 \sum_i \varepsilon_i\right] \\
&= \frac{1}{n}[0 + 0 + 0 + 0] \\
&= 0
\end{aligned}$$

$$Var(b_0) = \sigma^2\left(\frac{1}{n} + \frac{\overline{x}^2}{s_{xx}}\right).$$

Covariance

The covariance between b_0 and b_1 is

$$Cov(b_0, b_1) = Cov(\bar{y}, b_1) - \bar{x}Var(b_1)$$
$$= -\frac{\bar{x}}{s_{xx}}\sigma^2.$$

$$Cov(b_0^*, b_1) = Cov(\bar{y}, b_1) = 0$$

What about σ^2?

It is the variance of the residuals, but it's used in computing the variance of the slope and intercept.

Can we estimate it?

An unbiased estimator of the variance of the residuals is

$$\sigma^2 \approx s^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}.$$

Why *n-2*?

There are *n* residuals, but *2* parameters (i.e slope and intercept) need to be estimated in order to determine the residuals. Thus, the degrees of freedom, *df*, is *n-2*.

The minimum of Q($\boldsymbol{\beta}$) over $\boldsymbol{\beta} \in \boldsymbol{R}^2$ is

$$
\begin{aligned}
\mathrm{Q}(\hat{\boldsymbol{\beta}}) &= (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= (\boldsymbol{y} - \boldsymbol{H}\boldsymbol{y})'(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{y}) \\
&= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{EQ}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}(\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}) \\
&= \mathrm{tr}\left[(\boldsymbol{I} - \boldsymbol{H})\ \mathrm{Cov}(\boldsymbol{y})\right] + \mathrm{E}(\boldsymbol{y})'(\boldsymbol{I} - \boldsymbol{H})\mathrm{E}(\boldsymbol{y}) \\
&= \sigma^2 \mathrm{tr}(\boldsymbol{I} - \boldsymbol{H}) + \boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} \\
&= \sigma^2(n - \mathrm{tr}(\boldsymbol{H})) \\
&= \sigma^2\left(n - \mathrm{tr}[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\right) \\
&= \sigma^2\left(n - \mathrm{tr}[(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X})]\right) \\
&= \sigma^2(n - \mathrm{tr}(\boldsymbol{I}_2)) = \sigma^2(n - 2)
\end{aligned}
$$

Therefore

$$\hat{\sigma}^2 = \frac{1}{n-2} Q(\hat{\boldsymbol{\beta}}) = \frac{1}{n-2} \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$$

is an unbiased estimate of σ^2.

$$s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}}$$

Unbiased Property

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

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\beta}}) &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{y}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \mathrm{E}(\boldsymbol{\varepsilon}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \mathbf{0} = \boldsymbol{\beta}
\end{aligned}$$

$$\begin{aligned}
\hat{\boldsymbol{\sigma}}^2 &= \frac{1}{n-2}\mathbf{SS}_{\mathrm{Res}} \\
&= \frac{1}{n-2}\mathbf{Q}(\hat{\boldsymbol{\beta}}) \\
&= \frac{1}{n-2}\boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}
\end{aligned}$$

$$\mathrm{E}(\hat{\boldsymbol{\sigma}}^2) = \boldsymbol{\sigma}^2$$

$$\text{Let } \boldsymbol{\theta} = \boldsymbol{X}\boldsymbol{\beta} \qquad \hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

$$\mathrm{E}(\hat{\boldsymbol{y}}) = \boldsymbol{X}\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\theta}$$

Variance - Covariance

$$
\begin{aligned}
\operatorname{Cov}(\hat{\boldsymbol{\beta}}) &= \operatorname{Cov}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\operatorname{Cov}(\boldsymbol{y})[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']' \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\left(\sigma^2\boldsymbol{I}\right)\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \begin{pmatrix} \operatorname{Var}(\hat{\boldsymbol{\beta}}_0) & \operatorname{Cov}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1) \\ \operatorname{Cov}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1) & \operatorname{Var}(\hat{\boldsymbol{\beta}}_1) \end{pmatrix}
\end{aligned}
$$

As $\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}$

$$
\begin{aligned}
(\boldsymbol{X}'\boldsymbol{X})^{-1} &= \frac{1}{|\boldsymbol{X}'\boldsymbol{X}|}\begin{pmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{pmatrix} \\
&= \frac{1}{n\mathbf{S}_{XX}}\begin{pmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{pmatrix} = \frac{1}{\mathbf{S}_{XX}}\begin{pmatrix} \frac{1}{n}\sum X_i^2 & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix}
\end{aligned}
$$

Therefore,

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} \mathrm{Var}(\hat{\boldsymbol{\beta}}_0) & \mathrm{Cov}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1) \\ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1) & \mathrm{Var}(\hat{\boldsymbol{\beta}}_1) \end{pmatrix}$$

where

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_1) = \sigma^2 \big/ \mathbf{S_{XX}}$$

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_0) = \sigma^2 \frac{\sum X_i^2}{n\mathbf{S_{XX}}}$$

$$= \frac{\sigma^2}{n\mathbf{S_{XX}}}\left[\sum X_i^2 - n\overline{X}^2 + n\overline{X}^2\right]$$

$$= \frac{\sigma^2}{n\mathbf{S_{XX}}}[\mathbf{S_{XX}} + n\overline{X}^2] = \sigma^2\left[\frac{1}{n} + \frac{\overline{X}^2}{\mathbf{S_{XX}}}\right]$$

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_0, \hat{\boldsymbol{\beta}}_1) = -\sigma^2 \frac{\overline{X}}{\mathbf{S_{XX}}}$$

2.3

Tests of Hypotheses

Several questions may raise from the fitted regression line:

1. Does x truly influence y?
2. Is there an adequate fit of the data to the model?
3. Will the model adequately predict response?

Hypothesis: H_0: $\beta_1 = 0$, H_1: $\beta_1 \neq 0$

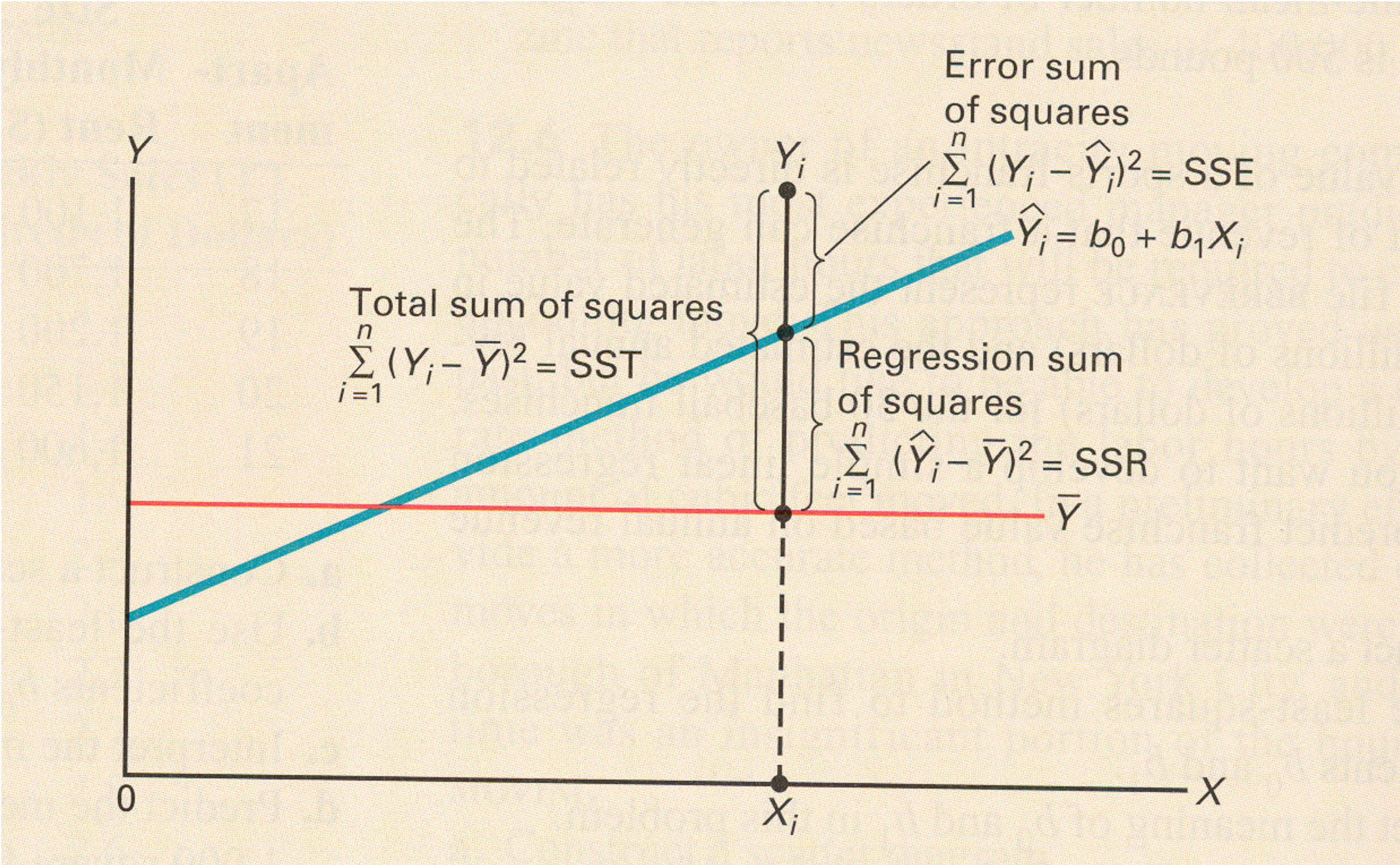
If H_0 is true,

$E(y) = \beta_0$, and the regressor variable does not influence y.

There are several ways to test this hypothesis.

ANOVA (Analysis of Variance)

- **The total sum of squares (SST):** a measure of variation of the values around their mean
- **The regression sum of squares (SSR):** which is due to the relationship between *X* and *Y*.
- **The error sum of squares (SSE):** which is due to random error.



Error sum
of squares
$\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 = SSE$
$\widehat{Y}_i = b_0 + b_1X_i$
Total sum of squares
$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = SST$
Regression sum
of squares
$\sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 = SSR$
Y
Y_i
$\overline{Y}$
0
X_i
X

Partitioning Total Variability

- Decomposition of sum of squares (variation)

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \quad (\bar{\hat{y}} = \bar{y})$$

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

$$SST = SSR + SSE$$

$$y_i - \hat{y}_i = (y_i - \bar{y}) - (\hat{y}_i - \bar{y}),$$

the sum of squared residuals is

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

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

Further consider

$$\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n} (y_i - \bar{y}) b_1 (x_i - \bar{x})$$

$$= b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

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

Partitioning Total Variability

$(x_i, y_i) \qquad i = 1, \cdots, n$

The total variation of y_i's is:

$$\boldsymbol{SS}_{Total} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Sources:

x: Regressor

Random Error

$$\boldsymbol{SS}_{Total} = \boldsymbol{y}' \boldsymbol{D}_n \boldsymbol{y}$$

where $\boldsymbol{y} = (y_1, \cdots, y_n)'$ and $\boldsymbol{D}_n = \boldsymbol{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n'$

Properties of D_n:

1 D_n is a projection matrix of rank (n-1)

2
$$D_n x = x - \bar{x}\mathbf{1}_n = \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix} \qquad \forall\ x \in R^n$$

Let $z = D_n x = (I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')x = x - \bar{x}\mathbf{1}_n$

$\bar{z} = 0$

3 $D_n x = x$ if $x \in R^n$ and $\bar{x} = 0$

D_n is a centered operator.

$$
\begin{aligned}
SS_{\text{Total}} &= \mathbf{y}'\mathbf{D_n}\mathbf{y} \\
&= (\mathbf{y}-\hat{\mathbf{y}}+\hat{\mathbf{y}})'\mathbf{D_n}(\mathbf{y}-\hat{\mathbf{y}}+\hat{\mathbf{y}}) \\
&\boxed{= (\mathbf{y}-\hat{\mathbf{y}})'\mathbf{D_n}(\mathbf{y}-\hat{\mathbf{y}})} + \hat{\mathbf{y}}'\mathbf{D_n}\hat{\mathbf{y}} + 2(\mathbf{y}-\hat{\mathbf{y}})'\mathbf{D_n}\hat{\mathbf{y}}
\end{aligned}
$$

As $\hat{y}_{\mathbf{i}} = b_0 + b_1 x_i$

$$\tfrac{1}{n}\sum \hat{y}_{\mathbf{i}} = b_0 + b_1 \tfrac{1}{n}\sum x_{\mathbf{i}}$$

$$\bar{\hat{y}} = b_0 + b_1\bar{x} = \bar{y} \qquad \Rightarrow \bar{\hat{y}} = \bar{y}$$

$$\boxed{(y-\hat{y})'\underbrace{D_n(y-\hat{y})}_{(y-\hat{y})}}$$

See Properties 3

$$= (\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}}) = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \equiv SS_{\text{Res}}$$

So

$$SS_{\text{Total}} = (y-\hat{y})'(y-\hat{y}) + \hat{y}' D_n \hat{y} + 2\boxed{(y-\hat{y})' D_n \hat{y}}$$

$$\boxed{(y-\hat{y})' D_n \hat{y}}$$

$$= [\hat{y}' D_n (y-\hat{y})]'$$

$$= [\hat{y}'(y-\hat{y})]' = (y-\hat{y})'\hat{y}$$

屏幕截图 Ctrl + Alt + A

截图时隐藏当前窗口

$$= (y-Hy)'Hy = y'(I-H)H'y$$

$$= y'(I-H)Hy$$

$$= y'(H-H^2)y$$

$$= 0$$

Hat Matrix:
Symmetric
$H' = H$

Hat Matrix:
Projection
$H^2 = H$

Therefore,

$$\begin{aligned}\mathbf{SS}_{\text{Total}} &= (\mathbf{y}-\hat{\mathbf{y}})'(\mathbf{y}-\hat{\mathbf{y}}) + \hat{\mathbf{y}}'\mathbf{D}_{\mathbf{n}}\hat{\mathbf{y}} + \mathbf{0} \\ &= \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \\ &= \mathbf{SS}_{\text{Res}} + \mathbf{SS}_{\text{Reg}}\end{aligned}$$

$$\boxed{SS_{\text{Total}} = SS_{\text{Reg}} + SS_{\text{Res}}}$$

ANOVA Table

Source	Sum of Squares	Degrees of Freedom	Mean Squares	F
Regression	SS_{Reg}	1	$MS_{Reg} = SS_{Reg} / 1$	MS_{Reg} / MS_{Res}
Residual	SS_{Res}	$n - 2$	$MS_{Res} = SS_{Res} / (n\text{-}2)$	
Total	S_{YY}	n - 1		

Example: Hospital Manpower Data

The data was taken from 17 U.S. Naval Hospitals:

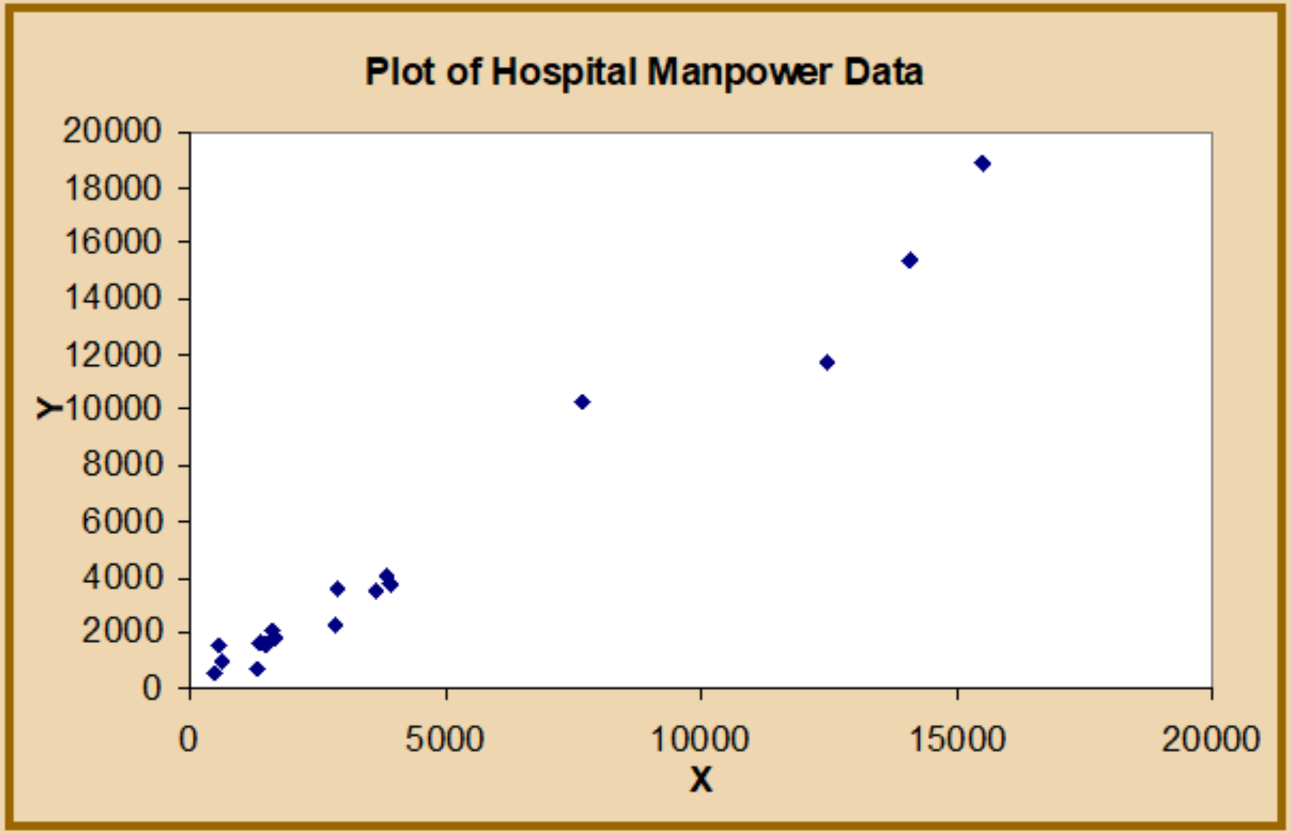
where

Y: Monthly man-hours

X: Monthly occupied bed days

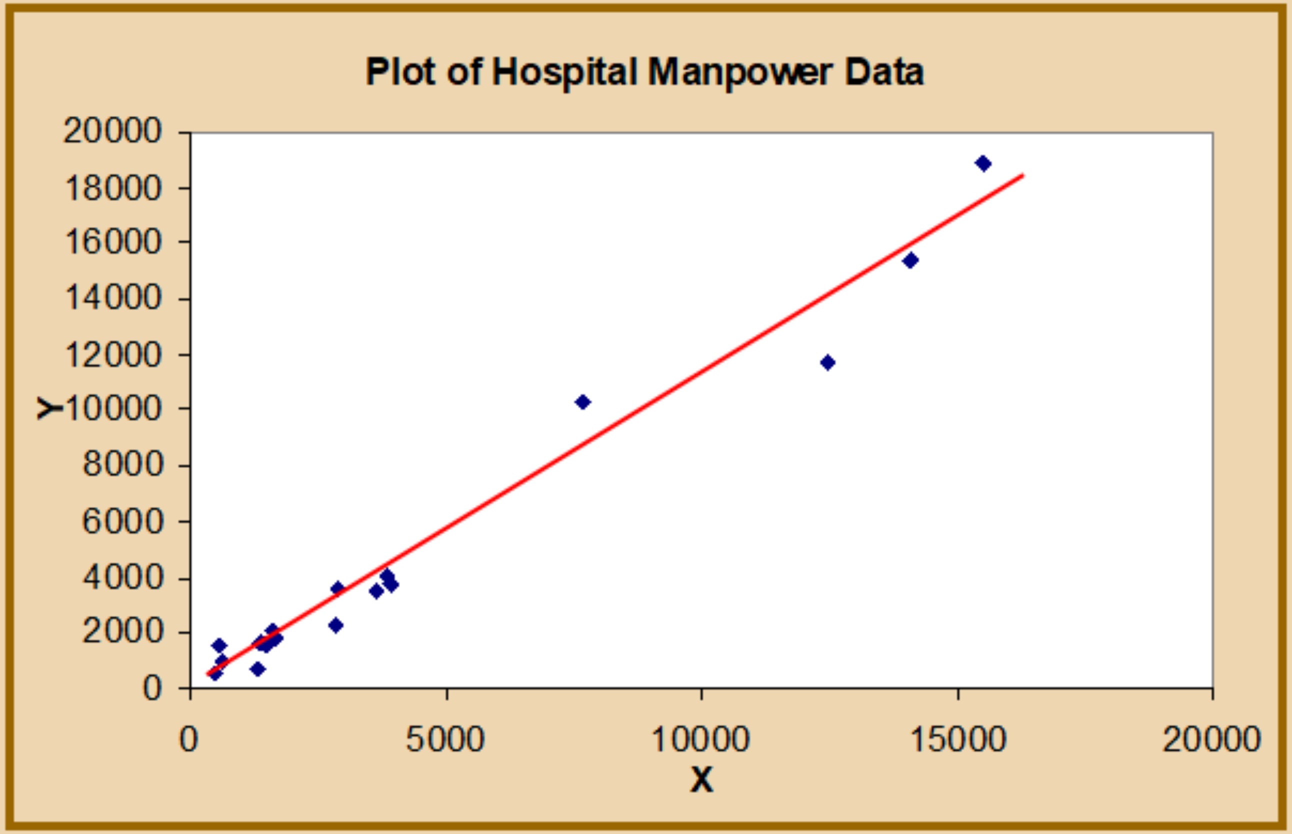
	Y	X
1	566.52	472.92
2	696.82	1339.75
3	1033.15	620.25
4	1603.62	568.33
5	1611.37	1497.60
6	1613.27	1365.83
7	1854.17	1687.00
8	2160.55	1639.92
9	2305.58	2872.33
10	3503.93	3655.08
11	3571.89	2912.00
12	3741.40	3921.00
13	4026.52	3865.67
14	10343.81	7684.10
15	11732.17	12446.33
16	15414.94	14098.40
17	18854.45	15524.00

Plot the data:



The regression equation:

$$\hat{y} = -28.1286 + 1.1174x$$



Example: Hospital Manpower Data *(cont…)*

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
Regression	1	480950232	480950232	524.2037
Residual	15	13762308.9	917487.258	
Total	16	494712540		

Quadratic Forms in Statistics (Review)

1.1. Quadratic forms

Definition 1: A quadratic form on R^n is a polynomial function Q: $R^n \to R$ of the form

$$Q(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j, \qquad (1)$$

in which each term has degree 2 and $\mathbf{x} = (x_1, \ldots, x_n)'$.

1.1. Quadratic forms - matrix presentation

Definition 2: Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ symmetric matrix with real entries and $\mathbf{x} = (x_1, \ldots, x_n)'$ be a column vector. Then

- ◇ $Q(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x}$ is said to be a quadratic form of $\mathbf{x}$,
- ◇ $\mathbf{A}$: the matrix of the quadratic form and
- ◇ rank($\mathbf{A}$): the rank of the quadratic form.

Regression Analysis: Consider the linear model

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n) \end{aligned} \tag{12}$$

where $\mathbf{X} : n \times p, \boldsymbol{\beta}'' \, p \times 1$.

◇ The least squared estimator of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{y}$,

◇ an unbiased estimator of σ^2 is given by $\hat{\sigma^2} = \frac{1}{n-p}\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}$,

where $p = \text{Rank}(\mathbf{X})$ and $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the hat matrix.

◇ Unbiased estimator of $\mathbf{y}$ is given by $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$.

Regression Analysis:

◇ The decomposition of the sum squares of $SS_{total}, SS_{res}, SS_{reg}$ is given by

$$Q(\mathbf{y}) \equiv \sum_{j=1}^{n}(y_j - \bar{y})^2 = \sum_{j=1}^{n}(y_j - \hat{y}_j)^2 + \sum_{j=1}^{n}(\hat{y}_j - \bar{y})^2 \equiv Q_1(\mathbf{y}) + Q_2(\mathbf{y})$$

That can be expressed as

$$\begin{aligned}\mathbf{y}'\mathbf{D}_n\mathbf{y} &= \mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y} + \hat{y}'\mathbf{D}_n\hat{\mathbf{y}} \\ &= \mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y} + \mathbf{y}'\mathbf{H}\mathbf{D}_n\mathbf{H}\mathbf{y}.\end{aligned}$$

by the use of the fact $\bar{y} = \bar{\hat{y}}$, where $\mathbf{D}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_N\mathbf{1}_n'$.

Regression Analysis:

◇ As $\mathbf{D}_n$ is a projection matrix with rank $n-1$ and $\mathbf{y} \sim N_n(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$ we have $Q(\mathbf{y}) = \mathbf{y}'\mathbf{D}_n\mathbf{y} \sim \chi^2_{n-1}(\lambda)$, where $\lambda = \beta'\mathbf{X}'\mathbf{D}_n\mathbf{X}\beta$.

◇ As $(\mathbf{I}-\mathbf{H})$ is a projection matrix with rank $n-2$ and $(\mathbf{I}-\mathbf{H})\mathbf{X} = \mathbf{O}$, it implies that $SS_{res} = Q_1(\mathbf{y}) = \mathbf{y}'(\mathbf{I}-\mathbf{H})\mathbf{y} \sim \chi^2_{n-p}$.

◇ It shows that SS_{res} and SS_{reg} are independent by Theorem 2 (Craig) as $(\mathbf{I}-\mathbf{H})\mathbf{H}\mathbf{D}_n\mathbf{H} = \mathbf{O}$.

◇ From Theorem 3 we have that $\mathbf{H}\mathbf{D}_n\mathbf{H}$ is a projection matrix with rank $(n-1)-(n-p) = p-1$, and $SS_{reg} = \mathbf{y}'\mathbf{H}\mathbf{D}_n\mathbf{H}\mathbf{y} \sim \chi^2_{p-1}(\lambda_2), \lambda_2 = \beta'\mathbf{X}'\mathbf{H}\mathbf{D}_n\mathbf{H}\mathbf{X}\beta = \beta'\mathbf{X}'\mathbf{D}_n\mathbf{X}\beta$.

Regression Analysis:

Applying the theory of quadratic forms to model (12) prove the following facts. If some conditions are necessary, please show the related condition.

(1) The distributions of $Q(\mathbf{y})$, $Q_1(\mathbf{y})$ and $Q_2(\mathbf{y})$;

(2) Prove $Q_1(\mathbf{y})$ and $Q_2(\mathbf{y})$ are independent;

(3) Prove $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{y}$ and $\widehat{\sigma^2}$ are independent.

Coefficient of Determinate (R^2)

$$R^2 = \frac{SS_{\text{Reg}}}{SS_{\text{Total}}} = 1 - \frac{SS_{\text{Res}}}{SS_{\text{Total}}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\begin{aligned}
SS_{\text{Reg}} &= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \\
&= \sum_{i=1}^{n}(b_0 + b_1 x_i - b_0 - b_1\bar{x})^2 \\
&= \sum_{i=1}^{n}\left[b_1(x_i - \bar{x})\right]^2 \\
&= b_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= b_1^2 S_{XX} = b_1 S_{XY} = \frac{S_{XY}^2}{S_{XX}}
\end{aligned}$$

$$\boxed{R^2 = \frac{S_{XY}^2}{S_{XX}S_{YY}} = r^2 \qquad 0 \le R^2 \le 1}$$

Example:

Hospital Manpower Data *(cont…)*

$$R^2 = 0.9722$$

Correlation Coefficient

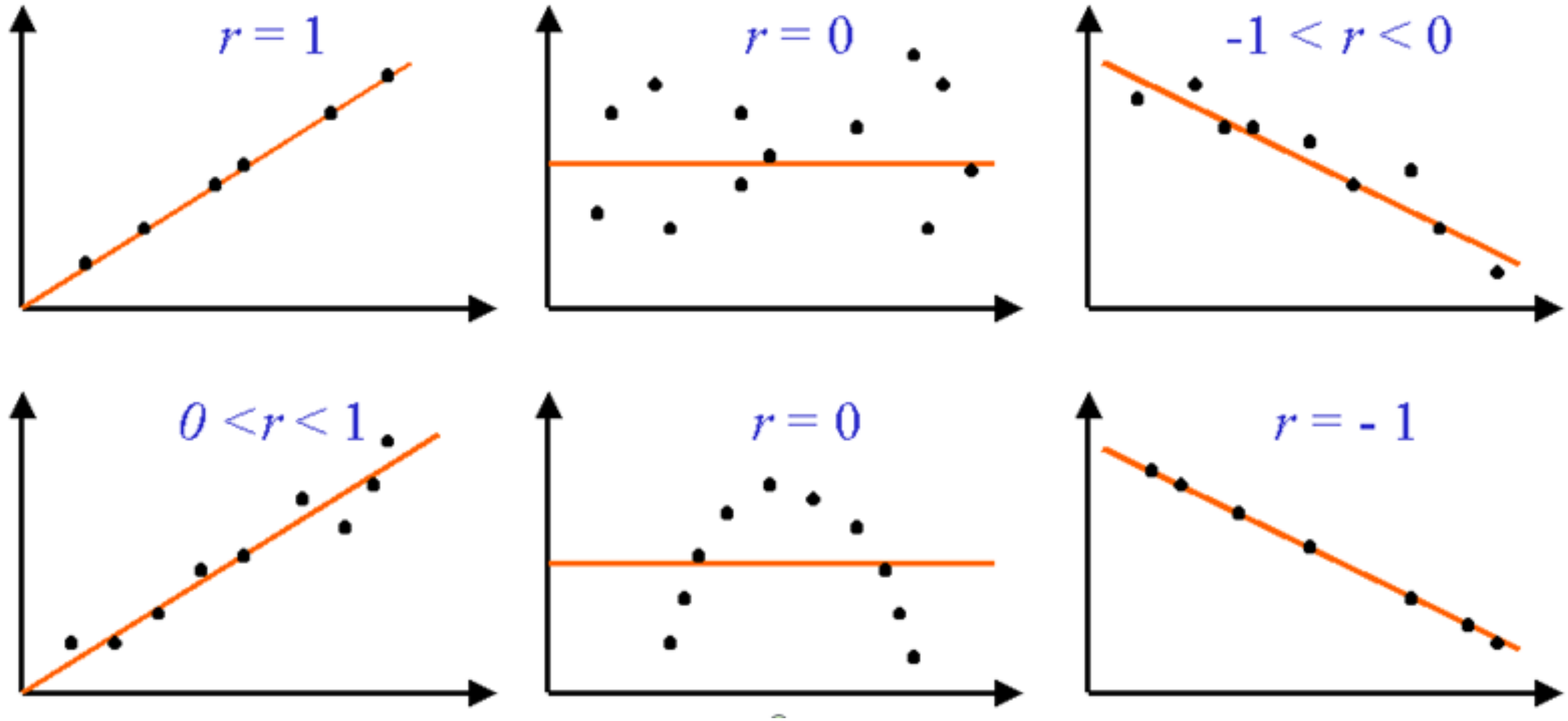
$$\boxed{\rho = \frac{\mathrm{Cov}(x,y)}{\sqrt{\mathrm{Var}(x)\,\mathrm{Var}(y)}}}$$

Hypotheses:

$$H_0: \rho = 0 \qquad \text{VS} \qquad H_1: \rho \neq 0$$

$$r = \hat{\rho} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

$$= \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} = \frac{S_{XY}}{S_{XX}} \sqrt{\frac{S_{XX}}{S_{YY}}} = b_1 \sqrt{\frac{S_{XX}}{S_{YY}}}$$



Example: Hospital Manpower Data *(cont…)*

$$r = \frac{430421977}{\sqrt{385202181 \times 494712540}} = 0.9860$$

2.4

Confidence Intervals

$$\begin{cases} y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \\ \varepsilon_1, \cdots, \varepsilon_n \quad \text{iid}, \quad N(0, \sigma^2) \end{cases}$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon, \qquad \mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \qquad \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \sim N_2(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$$

and $b_1 = \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{XX}}\right) \Rightarrow \dfrac{b_1 - \beta_1}{\sqrt{\sigma^2 / S_{XX}}} = \dfrac{b_1 - \beta_1}{\sigma} \sqrt{S_{XX}} \sim N(0,1)$

As σ is unknown, use s^2 to estimate σ^2 :

$$\frac{(b_1 - \beta_1)}{s} \sqrt{S_{XX}} \sim t_{n-2}$$

<u>Hypothesis:</u> H_0: $\beta_1 = 0$, H_1: $\beta_1 \neq 0$

$$\boxed{t = \frac{b_1\sqrt{S_{XX}}}{s}}$$

As $$SS_{\text{Reg}} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

$$= \sum_{i=1}^{n}(b_0 + b_1 x_i - b_0 - b_1\bar{x})^2$$

$$= b_1^2\sum_{i=1}^{n}(x_i - \bar{x})^2 = b_1^2 S_{XX}$$

So $$t^2 = \frac{SS_{\text{Reg}}}{s^2} = F \quad \Rightarrow$$ *t-test is equivalent to F-test.*

As

$$\frac{(b_1 - \beta_1)}{s}\sqrt{S_{XX}} \sim t_{n-2}$$

We have

$$\mathrm{P}\left\{-t_{\alpha/2,n-2} < \frac{(b_1 - \beta_1)\sqrt{S_{XX}}}{s} < t_{\alpha/2,n-2}\right\} = 1 - \alpha$$

The (1-α) confidence interval of β_1 is given by

$$b_1 \pm t_{\alpha/2,n-2}\frac{s}{\sqrt{S_{xx}}}$$

The confident intervals for the mean response (**fitting**, cf Section 2.9 of the textbook):

$$
\begin{aligned}
\hat{y}(x_0) &= b_0 + b_1 x_0 \\
\operatorname{Var}\left(\hat{y}(x_0)\right) &= \operatorname{Var}\left(b_0 + b_1 x_0\right) \\
&= \operatorname{Var}\left(y - b_1(x_0 - \bar{x})\right) \\
&= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}}\right]
\end{aligned}
$$

as $\bar{y}$ and b_1 are independent.

The (1-α) confidence interval for prediction is:

$$\hat{y}(x_0) \pm t_{\alpha/2, n-2}\ s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}}}$$

Confident interval

- For the confident intervals for **prediction**:

as $E(y_0 - \hat{y}(x_0)) = 0$

$$Var(y - \hat{y}(x_0)) = \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}} \right]$$

$$\frac{y - \hat{y}(x_0)}{\sigma \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}}}} \sim N(0,1), \text{ then } \frac{y - \hat{y}(x_0)}{s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}}}} \sim t_{n-2}$$

the $(1-\alpha)$ confidence interval for prediction is :

$$\hat{y}(x_0) \pm t_{\alpha/2, n-2} s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}}}$$

2.5
More regression models

Simple Regression Through the Origin (see p.33-36)

$$\begin{cases} y = \beta X + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

Find the LSE of β and unbiased estimate of σ^2.

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$\mathrm{Q}(\hat{\beta}) = \mathbf{y}'[\mathbf{I} - H]\mathbf{y} = \mathbf{y}'\mathbf{y} - \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\sum_{i=1}^{n} x_i^2}$$

$$H = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \frac{\mathbf{X}\mathbf{X}'}{\sum_{i=1}^{n} x_i^2}, \qquad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$