Chapter 13. Multiple Linear Regression and Nonlinear Regression

Multiple Linear Regression

- **Multiple linear regression** model is an extension of a simple linear regression model that allows the dependent variable y (response variable) to be modeled as a linear function of more than one input variable x_i

Multiple linear regression: 다중선형회귀분석

Multiple Linear Regression

- *Example*: (Weight of students)
  - y: Weight
  - x_1: Height
  - x_2: Age
  - x_3: Daily sleep

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \epsilon_i$$

13.1 Introduction to Multiple Linear Regression

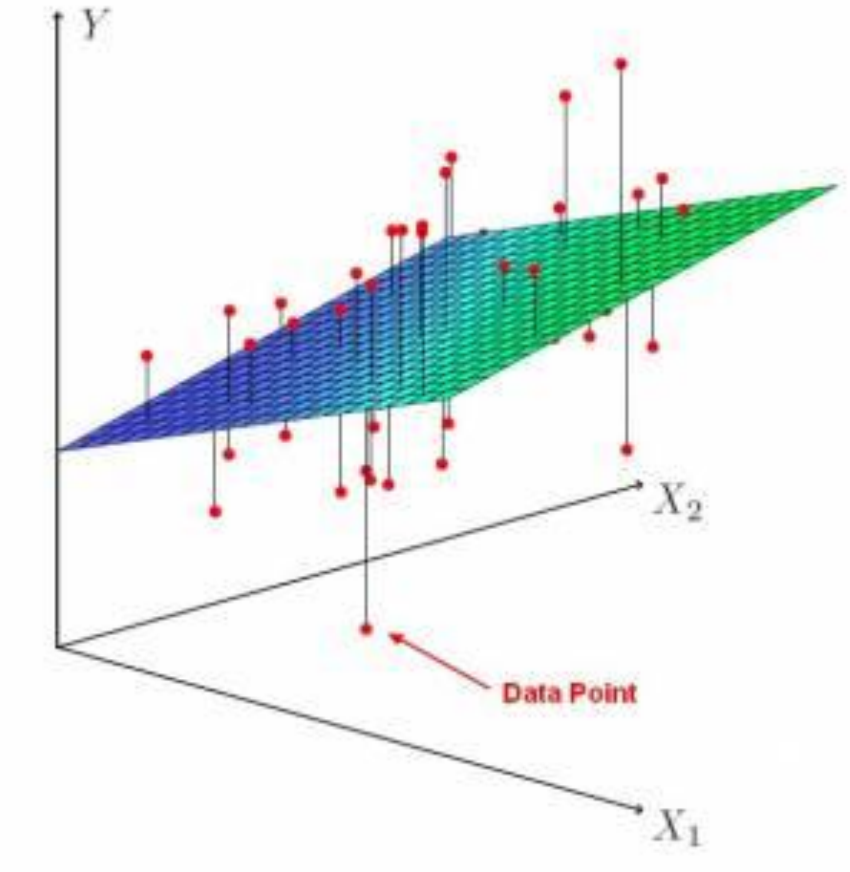
Observations

- Consider a data set consisting of n sets of values

$$(y_1, x_{11}, x_{21}, \ldots, x_{k1})$$
$$\vdots$$
$$(y_n, x_{1n}, x_{2n}, \ldots, x_{kn})$$

- Thus, y_i is the value taken by the response variable y for the ith observation, which is obtained with values $x_{1i}, x_{2i}, \ldots, x_{ki}$ of the k input variables $x_1, x_2, \ldots, x_k$

Multiple Linear Regression



Multiple Linear Regression

- In multiple linear regression, the response variable y_i is modeled as
$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i$$
which consists of a linear function $\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki}$ together with an error term ϵ_i
- The error terms $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ are generally taken to be independent observations from a $N(0, \sigma^2)$ distribution, for some error variance σ^2
- When $k = 1$, the model simplifies to a simple linear regression

Multiple Linear Regression

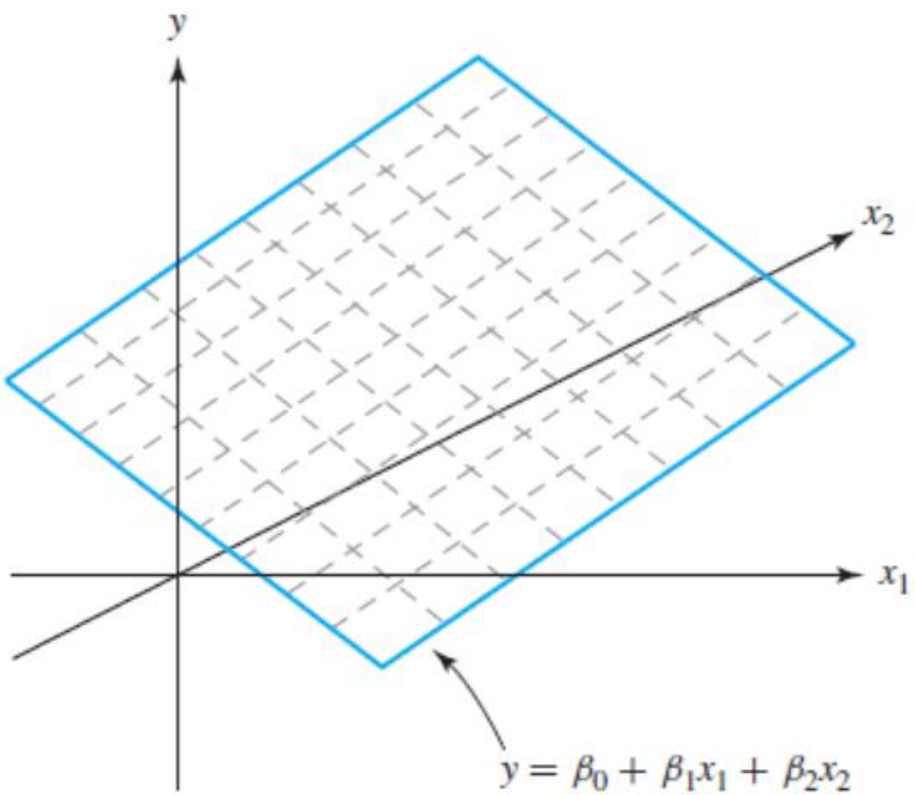
- Expected value of the response variable at $x = (x_1, x_2, \dots, x_k)$ is

$$\bar{y} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Multiple Linear Regression

- For example, with $k = 2$, the expected values of the response variable lie on the plane

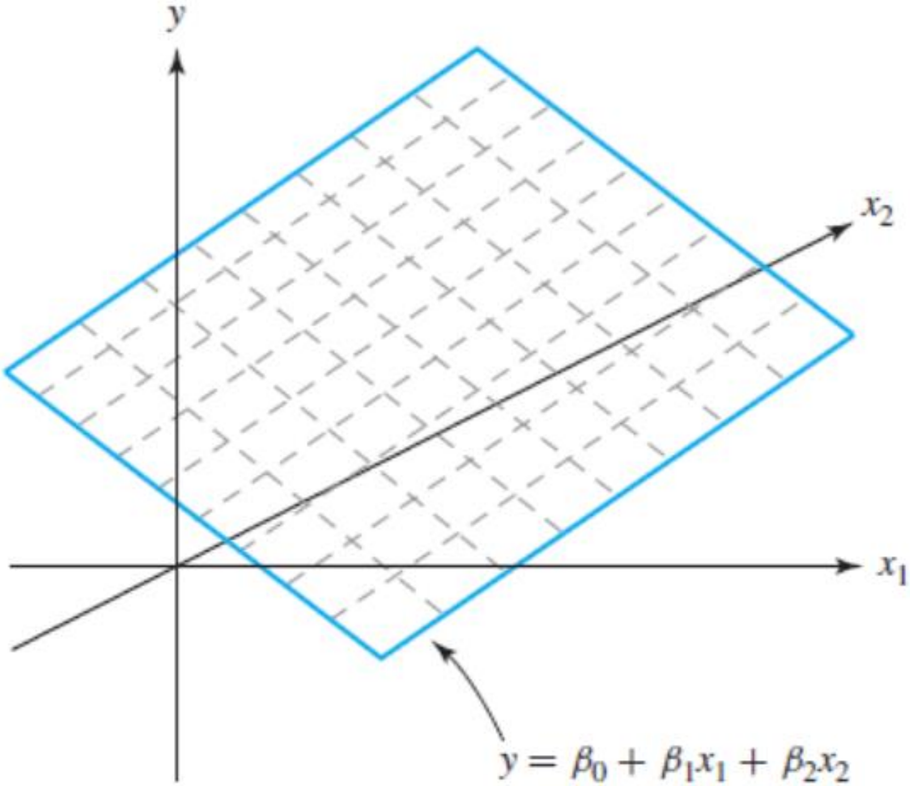
$$\bar{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

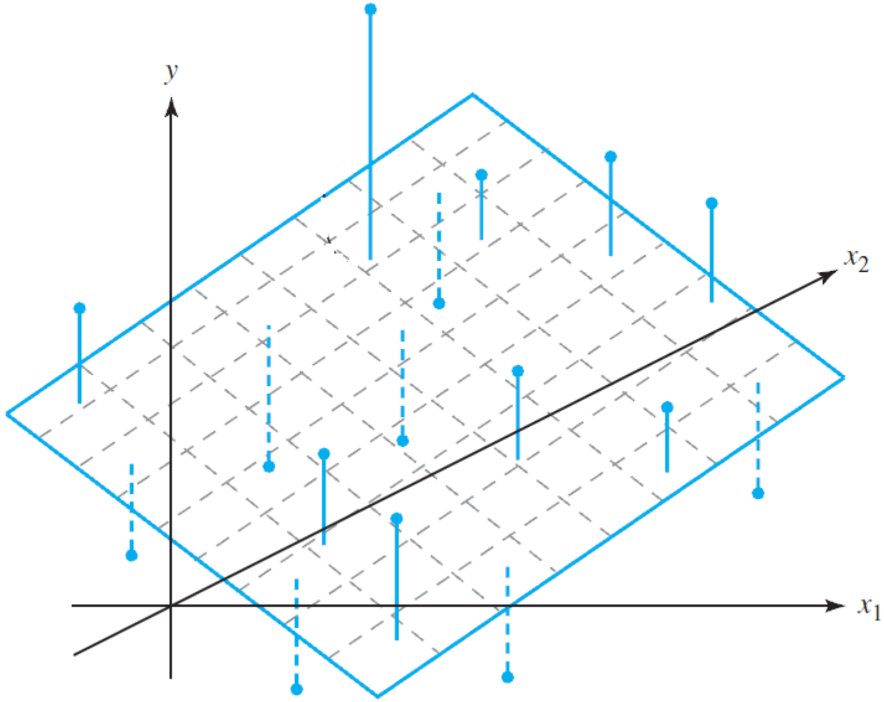


Multiple Linear Regression

- For example, with $k = 2$, the expected values of the response variable lie on the plane

$$\bar{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$





Multiple Linear Regression

- β_0 is the intercept (same as simple linear regression)
- β_i determines how x_i influences the response variable when the other input variables are kept fixed
  - If $\beta_i > 0$, then the expected value of the response variable increases as x_i increases
  - If $\beta_i < 0$, then the expected value of the response variable decreases as x_i increases
  - If $\beta_i = 0$, then the dependent variable is not influenced by changes in x_i

Fitting the Linear Regression Model

- Similar to simple linear regression, the parameters $\beta_0, \beta_1, \ldots, \beta_k$ can be estimated by minimizing error
- In other words, minimize the vertical distances between the data observations y_i and their fitted values

$$\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}$$

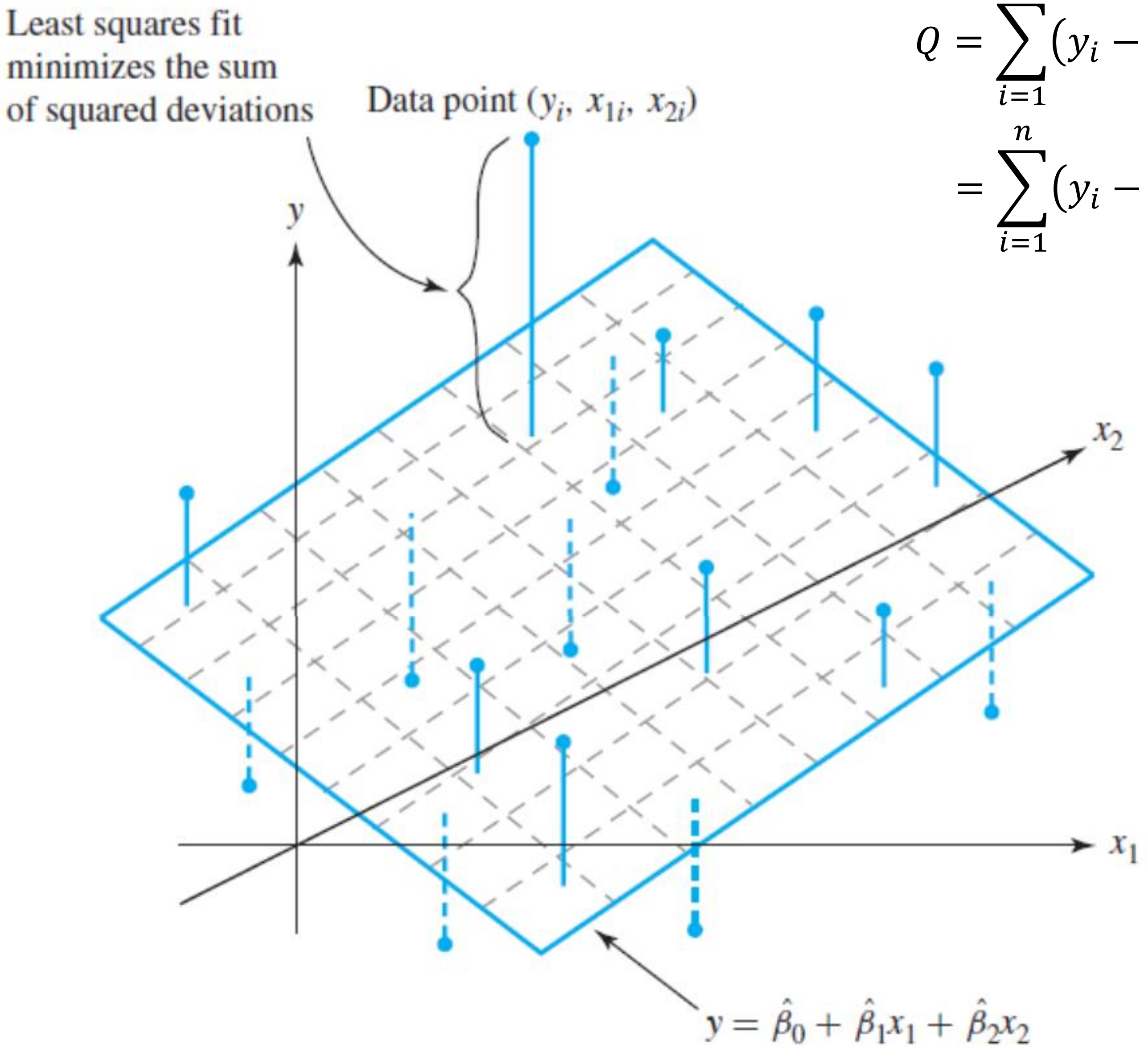
Fitting the Linear Regression Model

- *Recall*: For simple linear regression, the total distance between observations and a straight line was calculated as

$$Q = \sum_{i=1}^{n} \big(y_i - (\beta_0 + \beta_1 x_i)\big)^2$$

- For multiple linear regression, the following expresses the sum of square distances

$$Q = \sum_{i=1}^{n} \big(y_i - (\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})\big)^2$$



$$Q = \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})\right)^2$$

$$= \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})\right)^2$$

Fitting the Linear Regression Model

- How can we estimate the parameters $\beta_0, \beta_1, \dots, \beta_k$?
  - Similar to simple linear regression, we find the values of $\beta_0, \beta_1, \dots, \beta_k$ that minimize Q

Fitting the Linear Regression Model

- For finding the minimum, take the derivative of Q with respect to each of $\beta_0, \beta_1, \ldots, \beta_k$

$$Q = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki}))^2$$

$$\left.\begin{aligned} \frac{\partial Q}{\partial \beta_0} &= -2 \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})) \\ \frac{\partial Q}{\partial \beta_j} &= -2 \sum_{i=1}^{n} x_{ji} (y_i - (\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})) \end{aligned}\right\} \text{Set these to zero}$$

Fitting the Linear Regression Model

- Then, we have $k+1$ equations

$$\sum_{i=1}^{n} y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_{1i} + \beta_2 \sum_{i=1}^{n} x_{2i} + \cdots + \beta_k \sum_{i=1}^{n} x_{ki}$$

$$\sum_{i=1}^{n} y_i x_{1i} = \beta_0 \sum_{i=1}^{n} x_{1i} + \beta_1 \sum_{i=1}^{n} x_{1i}^2 + \beta_2 \sum_{i=1}^{n} x_{2i} x_{1i} + \cdots + \beta_k \sum_{i=1}^{n} x_{ki} x_{1i}$$

$$\vdots$$

$$\sum_{i=1}^{n} y_i x_{ki} = \beta_0 \sum_{i=1}^{n} x_{ki} + \beta_1 \sum_{i=1}^{n} x_{1i} x_{ki} + \beta_2 \sum_{i=1}^{n} x_{2i} x_{ki} + \cdots + \beta_k \sum_{i=1}^{n} x_{ki}^2$$

The parameter estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are the solutions to these equations

These are generally not solved by hand

Fitting the Linear Regression Model

- We will revisit the calculation for estimating $\beta_0, \beta_1, \dots, \beta_k$ later in this chapter

Analysis of the Fitted Model

- Suppose we found the estimates for $\beta_0, \beta_1, \ldots, \beta_k$ (i.e., we found the linear model)

$$\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}$$

Analysis of the Fitted Model

- Suppose we found the estimates for $\beta_0, \beta_1, \dots, \beta_k$ (i.e., we found the linear model)

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}$$

- Then the *i*th residual (or error) is $e_i = y_i - \hat{y}_i$

As in simple linear regression, the sum of squares for error is defined to be

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

Analysis of the Fitted Model

- Similar to simple linear regression, the estimate of the error variance σ^2 is

$$\text{MSE} = \hat{\sigma}^2 = \frac{\text{SSE}}{n-k-1}$$

where $n-k-1$ comes from

$$n - (number\ of\ estimated\ parameters) = n - (k+1)$$

ANOVA Table

- Also, similar to simple linear regression, we can create an ANOVA table for the following null hypothesis

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

(with the alternative hypothesis that at least one of these β_i is nonzero)

- If the null hypothesis were true, then the response variable is not related to any of the *k* variables

ANOVA Table

- The relationship between SST, SSR, and SSE still holds

SST = SSR + SSE

- Sum of squares calculation:

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

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

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

ANOVA Table

Source	Degrees of freedom	Sum of squares	Mean square	F-statistic	p-value
Regression Error	k $n-k-1$	SSR SSE	$\text{MSR} = \text{SSR}/k$ $\text{MSE} = \text{SSE}/(n-k-1)$	$F = \text{MSR}/\text{MSE}$	$P(F_{k,n-k-1} > F)$
Total	$n-1$	SST			

- F-statistic:

$$F = \frac{MSR}{MSE} = \frac{MSR}{\hat{\sigma}^2} = \frac{SSR/k}{\hat{\sigma}^2} = \frac{SSR}{k\hat{\sigma}^2}$$

- Use $F_{k,n-k-1}$ distribution to calculate p-value

- p-value shows whether we should accept or reject the null hypothesis

Small p-value indicates that the response variable is related to at least one of the input variables

Analysis of Variance

Analysis of Variance Table for Multiple Linear Regression Problem

The analysis of variance table for a multiple linear regression problem provides a test of the null hypothesis

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

which implies that the response variable is not related to any of the k input variables. The p-value is

$$p\text{-value} = P(X > F)$$

where the random variable X has an F-distribution with degrees of freedom k and $n - k - 1$, and the F-statistic is

$$F = \frac{\text{SSR}}{k\hat{\sigma}^2} = \frac{(n-k-1)R^2}{k(1-R^2)}$$

ANOVA Table

- Simple linear regression:

Source	Degrees of freedom	Sum of squares	Mean squares	F-statistic	p-value
Regression Error	1 $n-2$	SSR SSE	MSR = SSR $\hat{\sigma}^2 = \text{MSE} = \text{SSE}/(n-2)$	$F = \text{MSR/MSE}$	$P(F_{1,n-2} > F)$
Total	$n-1$	SST			

- Multiple linear regression:

Source	Degrees of freedom	Sum of squares	Mean square	F-statistic	p-value
Regression Error	k $n-k-1$	SSR SSE	$\text{MSR} = \text{SSR}/k$ $\text{MSE} = \text{SSE}/(n-k-1)$	$F = \text{MSR/MSE}$	$P(F_{k,n-k-1} > F)$
Total	$n-1$	SST			

Coefficient of Determination

- Coefficient of (multiple) determination

$$R^2 = \frac{SSR}{SST}$$

  - Takes values between 0 and 1
  - Indicates the amount of the total variability in the values of the response variable that is accounted for by the fitted regression model

Model Fitting

- It is possible that a subset of the k input variables is better than using all k input variables
- We can perform hypothesis tests for input variables individually:

$$H_0\colon \beta_i = 0 \quad \text{versus} \quad H_A\colon \beta_i \neq 0$$

- If the null hypothesis is accepted, then there is no evidence that the response variable is directly related to the input variable x_i

➔ Thus, x_i can be dropped from the model

Model Fitting

- These type of hypothesis tests:

$$H_0\colon \beta_i = 0 \quad \text{versus} \quad H_A\colon \beta_i \neq 0$$

are tested with the *t*-statistics (use *t*-distributions)

$$t = \frac{\hat{\beta}_i}{\text{s.e.}(\hat{\beta}_i)}$$

Use the *t*-distribution with $n - k - 1$ degrees of freedom

- So the *p*-value is therefore,

$$p\text{-value} = 2 \times P(X > |t|)$$

where the random variable X has a t-distribution with $n - k - 1$ degrees of freedom

(The calculation of standard error is discussed in Chapter 13.3)

Model Fitting

- For example, suppose we have the following model:

$$\bar{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}$$

- We perform the following test and accept the null hypothesis

$$H_0\colon \beta_2 = 0 \quad \text{versus} \quad H_A\colon \beta_2 \neq 0$$

$$\Downarrow$$

- Then, the model can be improved as

$$\bar{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_3 x_{3i}$$

Model Fitting

- *p*-value for the intercept is usually not important
- For other parameters, *p*-value larger than 0.1 usually indicates that the corresponding input variable can be dropped from the model
- It should be noted that when one or more variables are removed from the model, the *p*-values of the remaining variables may change when the reduced model is fitted

Model Fitting

- Model fitting is performed by finding which subset of the k input variables is required to model the dependent variable y in the best manner
- The final model that the experimenter uses for inference problems should consist of input variables that each have p-values no larger than 10%
- It is important to note that it is best to remove only one variable from the model at a time (begin by removing the one with the largest p-value and then fit the model with $k - 1$ input variables)

Variable Selection *(Not included in exam)*

- Backward elimination
- Forward selection
- Ridge regression
- Lasso regression
- …

$$\min_{\beta_0,\beta}\left\{\frac{1}{N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2\right\} \text{ subject to } \sum_{j=1}^{p}|\beta_j|\le t$$

13.2 Example of Multiple Linear Regression

Multiple Linear Regression

- *Example*: (Physical fitness test)
  - A physical test consists of 2 minutes of pushups followed by a 2-mile run
  - Data set of 84 fitness tests are collected

Number of pushups	Two-mile run time (seconds)
60	847
53	887
60	879
55	919
60	816
78	814
74	814
70	855
46	980
50	954
50	1078
59	1001
62	766
64	916

Multiple Linear Regression

- *Example*: (Physical fitness test)
  - While we only considered pushups and 2-mile run, suppose the fitness test also included situps

  - Let us see if 2-mile run time can be better explained when using both pushups and situps data

2-mile run time (seconds)	Number of pushups	Number of situps
847	60	83
887	53	67
879	60	70
919	55	60
816	60	71
814	78	83
814	74	70
855	70	69
980	46	48
954	50	55
1078	50	48
1001	59	61
766	62	71
916	64	65
798	51	62
782	66	64

Multiple Linear Regression

- *Example*: (Physical fitness test)
  - The multiple linear regression model becomes

    $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

    where

    y : run time

    x_1: number of pushups

    x_2: number of situps

Multiple Linear Regression

- *Example*: (Physical fitness test)
  - Suppose the model is fitted and we found the linear model (estimated $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$)

$$\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

  - Then, in order to check if both x_1 and x_2 are important, we perform the following two hypothesis tests:

    (1) $H_0\colon \beta_1 = 0$ versus $H_A\colon \beta_1 \neq 0$

    (2) $H_0\colon \beta_2 = 0$ versus $H_A\colon \beta_2 \neq 0$

Multiple Linear Regression

- *Example*: (Physical fitness test)
  - We find (from computer programs) that the *p*-value for each hypothesis test is:

    (1) H_0: $\beta_1 = 0$ versus H_A: $\beta_1 \neq 0$ → *p*-value = 0.0028

    (2) H_0: $\beta_2 = 0$ versus H_A: $\beta_2 \neq 0$ → *p*-value = 0.3448

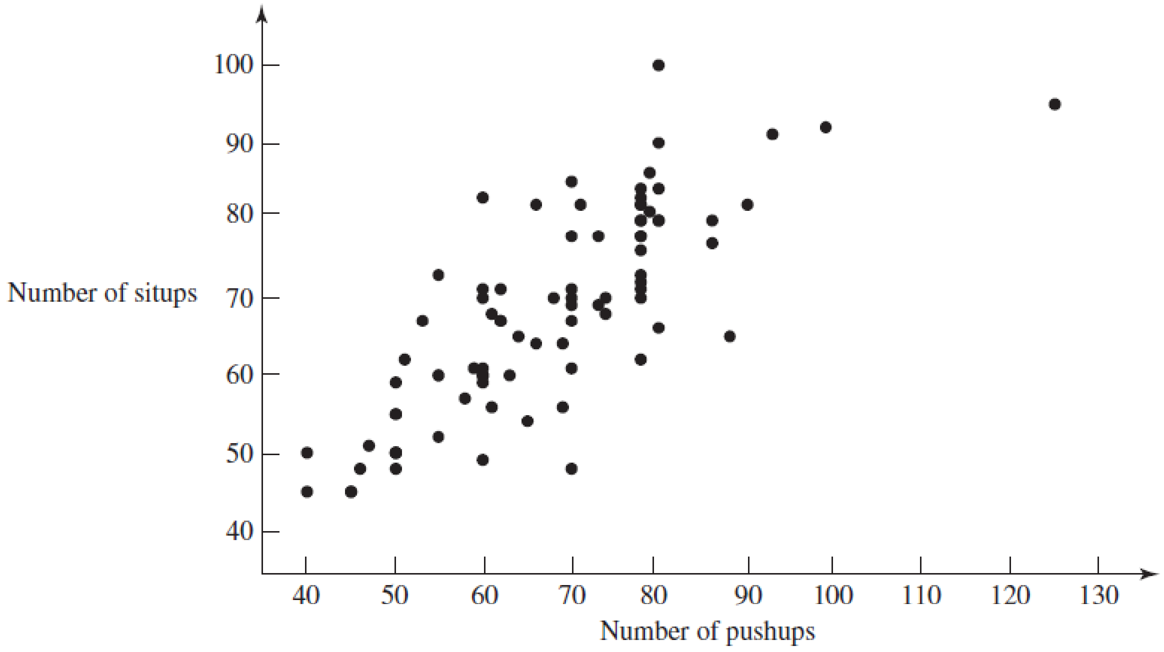
This implies that situps should be dropped from the model

(run time should be predicted from the number of pushups using simple linear regression model)

Multiple Linear Regression

- *Example*: (Physical fitness test)
  - Why is situps dropped? (Isn't more information better?)

- This is because the variable situps is correlated with the variable pushups



Multiple Linear Regression

- *Example*: (Physical fitness test)
  - Why is situps dropped? (Isn't more information better?)
  - Correlation between run time and pushups is –0.57
  - Correlation between run time and situps is –0.50

So pushups is a marginally more effective predictor of run time than situps

13.3 Matrix Algebra Formulation of Multiple Linear Regression

Matrix

- *Recall*:
  - Matrices and vectors
  - Adding/multiplying a constant
  - Addition
  - Transpose
  - Multiplication
  - Inverse

Matrix Formulation

- Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

- We can express sample data using the above equation

2-mile run time (seconds)	Number of pushups	Number of situps
847	60	83
887	53	67
879	60	70
919	55	60
816	60	71
814	78	83
814	74	70
855	70	69
980	46	48
954	50	55
1078	50	48
1001	59	61
766	62	71
916	64	65
798	51	62
782	66	64

$$847 = \beta_0 + \beta_1 60 + \beta_2 83 + \epsilon_1$$

$$887 = \beta_0 + \beta_1 53 + \beta_2 67 + \epsilon_2$$

$$879 = \beta_0 + \beta_1 60 + \beta_2 70 + \epsilon_3$$

$$919 = \beta_0 + \beta_1 55 + \beta_2 60 + \epsilon_4$$

$$\vdots$$

Matrix Formulation

- Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

- We can use matrices to express the sample data

Matrix Formulation: *Y*

- Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

- Begin with *y*:

Y is the $n \times 1$ vector of observed values of the response variable

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}$$

2-mile run time (seconds)	Number of pushups	Number of situps
847	60	83
887	53	67
879	60	70
919	55	60
816	60	71
814	78	83
814	74	70
855	70	69
980	46	48
954	50	55
1078	50	48
1001	59	61
766	62	71
916	64	65
798	51	62
782	66	64

Matrix Formulation: X

- Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

- Input variables:

design matrix X is the $n \times (k+1)$ matrix containing the values of the input variables

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$$

2-mile run time (seconds)	Number of pushups	Number of situps
1	60	83
1	53	67
1	60	70
1	55	60
1	60	71
1	78	83
1	74	70
1	70	69
1	46	48
1	50	55
1	50	48
1	59	61
1	62	71
1	64	65
1	51	62

Matrix Formulation: Error

- Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

- Error term:

$\boldsymbol{\epsilon}$ is the $n \times 1$ vector containing the error terms

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

Matrix Formulation

- Model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \epsilon_i$$

- In matrix form:

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

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

Matrix Formulation

- In matrix form: $$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$
- Expectation: $$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$$

$$\begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \bar{y}_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix}$$

Matrix Formulation

- *Example*: (Car plant electricity usage)
  - Multiple linear regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

	Electricity usage (million kWh)	Production ($ million)	Cooling degree days
January	2.48	4.51	0
February	2.26	3.58	0
March	2.47	4.31	13
April	2.77	5.06	56
May	2.99	5.64	117
June	3.05	4.99	306
July	3.18	5.29	358
August	3.46	5.83	330
September	3.03	4.70	187
October	3.26	5.61	94
November	2.67	4.90	23
December	2.53	4.20	0

where the response variable y is the electricity usage, x_1 is the production level, and x_2 is CDD

CDD: total daily temperature above 65 degrees (expecting air conditioner usage)

Matrix Formulation

- *Example*: (Car plant electricity usage)

	Electricity usage (million kWh)	Production ($ million)	Cooling degree days
January	2.48	4.51	0
February	2.26	3.58	0
March	2.47	4.31	13
April	2.77	5.06	56
May	2.99	5.64	117
June	3.05	4.99	306
July	3.18	5.29	358
August	3.46	5.83	330
September	3.03	4.70	187
October	3.26	5.61	94
November	2.67	4.90	23
December	2.53	4.20	0

$$\mathbf{Y} = \begin{pmatrix} 2.48 \\ 2.26 \\ 2.47 \\ 2.77 \\ 2.99 \\ 3.05 \\ 3.18 \\ 3.46 \\ 3.03 \\ 3.26 \\ 2.67 \\ 2.53 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 4.51 & 0 \\ 1 & 3.58 & 0 \\ 1 & 4.31 & 13 \\ 1 & 5.06 & 56 \\ 1 & 5.64 & 117 \\ 1 & 4.99 & 306 \\ 1 & 5.29 & 358 \\ 1 & 5.83 & 330 \\ 1 & 4.70 & 187 \\ 1 & 5.61 & 94 \\ 1 & 4.90 & 23 \\ 1 & 4.20 & 0 \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

Estimating Beta

- In matrix form: $$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{pmatrix}$$

How can we use the matrix formulation to estimate the betas?

Estimating Beta

- Difference between observed y_i and $\hat{y}_i$:

$$y_i - \hat{y}_i \quad \text{for each } i$$

$$\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix}$$

Estimating Beta

- Thus, Q becomes: $$Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\Downarrow$$

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

Estimating Beta

- Take the derivative with respect to each of β_i:

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

$$\mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

Set this equal to zero to find the estimates for β_i

Estimating Beta

- Therefore, need to solve:

$$\boldsymbol{X'Y} - \boldsymbol{X'X\beta} = 0$$

$$\boldsymbol{X'Y} = \boldsymbol{X'X\beta}$$

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

Estimating Beta

- Therefore, need to solve: $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \cdots & \sum_{i=1}^{n} x_{ki} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} & \cdots & \sum_{i=1}^{n} x_{1i}x_{ki} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{2i}^2 & \cdots & \sum_{i=1}^{n} x_{2i}x_{ki} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \sum_{i=1}^{n} x_{ki} & \sum_{i=1}^{n} x_{1i}x_{ki} & \sum_{i=1}^{n} x_{2i}x_{ki} & \cdots & \sum_{i=1}^{n} x_{ki}^2 \end{pmatrix} \qquad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_{1i} \\ \cdot \\ \cdot \\ \cdot \\ \sum_{i=1}^{n} y_i x_{ki} \end{pmatrix}$$

$$\Downarrow$$

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

SSE

- Once we have $\widehat{\boldsymbol{\beta}}$, we calculate estimated values $\hat{y}$:

$$\hat{\mathbf{Y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{y}_n \end{pmatrix} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

- Also, we can calculate SSE:

$$\mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix} = \mathbf{Y} - \hat{\mathbf{Y}} = \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n - \hat{y}_n \end{pmatrix}$$

$$\text{SSE} = \mathbf{e}'\mathbf{e} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Matrix Formulation

- *Example*: (Car plant electricity usage)

  - Multiple linear regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

	Electricity usage (million kWh)	Production ($ million)	Cooling degree days
January	2.48	4.51	0
February	2.26	3.58	0
March	2.47	4.31	13
April	2.77	5.06	56
May	2.99	5.64	117
June	3.05	4.99	306
July	3.18	5.29	358
August	3.46	5.83	330
September	3.03	4.70	187
October	3.26	5.61	94
November	2.67	4.90	23
December	2.53	4.20	0

where the response variable y is the electricity usage, x_1 is the production level, and x_2 is CDD

Matrix Formulation

- *Example*: (Car plant electricity usage)

	Electricity usage (million kWh)	Production ($ million)	Cooling degree days
January	2.48	4.51	0
February	2.26	3.58	0
March	2.47	4.31	13
April	2.77	5.06	56
May	2.99	5.64	117
June	3.05	4.99	306
July	3.18	5.29	358
August	3.46	5.83	330
September	3.03	4.70	187
October	3.26	5.61	94
November	2.67	4.90	23
December	2.53	4.20	0

$$\mathbf{Y} = \begin{pmatrix} 2.48 \\ 2.26 \\ 2.47 \\ 2.77 \\ 2.99 \\ 3.05 \\ 3.18 \\ 3.46 \\ 3.03 \\ 3.26 \\ 2.67 \\ 2.53 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 4.51 & 0 \\ 1 & 3.58 & 0 \\ 1 & 4.31 & 13 \\ 1 & 5.06 & 56 \\ 1 & 5.64 & 117 \\ 1 & 4.99 & 306 \\ 1 & 5.29 & 358 \\ 1 & 5.83 & 330 \\ 1 & 4.70 & 187 \\ 1 & 5.61 & 94 \\ 1 & 4.90 & 23 \\ 1 & 4.20 & 0 \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

Matrix Formulation

- *Example*: (Car plant electricity usage)
  - Since we know how to estimate $\hat{\boldsymbol{\beta}}$

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

  - First need to find $\boldsymbol{X'X}$

	Electricity usage (million kWh)	Production ($ million)	Cooling degree days
January	2.48	4.51	0
February	2.26	3.58	0
March	2.47	4.31	13
April	2.77	5.06	56
May	2.99	5.64	117
June	3.05	4.99	306
July	3.18	5.29	358
August	3.46	5.83	330
September	3.03	4.70	187
October	3.26	5.61	94
November	2.67	4.90	23
December	2.53	4.20	0

Matrix Formulation

- *Example*: (Car plant electricity usage)

	Electricity usage (million kWh)	Production ($ million)	Cooling degree days
January	2.48	4.51	0
February	2.26	3.58	0
March	2.47	4.31	13
April	2.77	5.06	56
May	2.99	5.64	117
June	3.05	4.99	306
July	3.18	5.29	358
August	3.46	5.83	330
September	3.03	4.70	187
October	3.26	5.61	94
November	2.67	4.90	23
December	2.53	4.20	0

$$\sum_{i=1}^{12} x_{1i} = 4.51 + \cdots + 4.20 = 58.62$$

$$\sum_{i=1}^{12} x_{1i}^2 = 4.51^2 + \cdots + 4.20^2 = 291.231$$

$$\sum_{i=1}^{12} x_{2i} = 0 + 0 + 13 + \cdots + 23 + 0 = 1484$$

$$\sum_{i=1}^{12} x_{2i}^2 = 0^2 + 0^2 + 13^2 + \cdots + 23^2 + 0 = 392{,}028$$

$$\sum_{i=1}^{12} x_{1i}x_{2i} = (4.51 \times 0) + \cdots + (4.20 \times 0) = 7862.87$$

Matrix Formulation

- *Example*: (Car plant electricity usage)

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{2i}^2 \end{pmatrix} = \begin{pmatrix} 12.0 & 58.6 & 1484.0 \\ 58.6 & 291.2 & 7862.8 \\ 1484.0 & 7862.8 & 392{,}028.0 \end{pmatrix}$$

This result gives

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 6.82134 & -1.47412 & 3.74529 \times 10^{-3} \\ -1.47412 & 0.32605 & -9.5962 \times 10^{-4} \\ 3.74529 \times 10^{-3} & -9.5962 \times 10^{-4} & 7.6207 \times 10^{-6} \end{pmatrix}$$

Matrix Formulation

- *Example*: (Car plant electricity usage)

$$\sum_{i=1}^{12} y_i = 2.48 + \cdots + 2.53 = 34.15$$

$$\sum_{i=1}^{12} y_i x_{1i} = (2.48 \times 4.51) + \cdots + (2.53 \times 4.20) = 169.2532$$

$$\sum_{i=1}^{12} y_i x_{2i} = (2.48 \times 0) + \cdots + (2.53 \times 0) = 4685.06$$

so that

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_{1i} \\ \sum_{i=1}^{n} y_i x_{2i} \end{pmatrix} = \begin{pmatrix} 34.15 \\ 169.2532 \\ 4685.06 \end{pmatrix}$$

Matrix Formulation

- *Example*: (Car plant electricity usage)

The parameter estimates are therefore

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$= \begin{pmatrix} 6.82134 & -1.47412 & 3.74529 \times 10^{-3} \\ -1.47412 & 0.32605 & -9.5962 \times 10^{-4} \\ 3.74529 \times 10^{-3} & -9.5962 \times 10^{-4} & 7.6207 \times 10^{-6} \end{pmatrix} \begin{pmatrix} 34.15 \\ 169.2532 \\ 4685.06 \end{pmatrix}$$

$$= \begin{pmatrix} 0.99 \\ 0.35 \\ 0.0012 \end{pmatrix}$$

so that the fitted model is

$$y = 0.99 + 0.35x_1 + 0.0012x_2$$

Matrix Formulation

- *Example*: (Car plant electricity usage)

The vector of fitted values is

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \begin{pmatrix} 1 & 4.51 & 0 \\ 1 & 3.58 & 0 \\ 1 & 4.31 & 13 \\ 1 & 5.06 & 56 \\ 1 & 5.64 & 117 \\ 1 & 4.99 & 306 \\ 1 & 5.29 & 358 \\ 1 & 5.83 & 330 \\ 1 & 4.70 & 187 \\ 1 & 5.61 & 94 \\ 1 & 4.90 & 23 \\ 1 & 4.20 & 0 \end{pmatrix} \begin{pmatrix} 0.99 \\ 0.35 \\ 0.0012 \end{pmatrix} = \begin{pmatrix} 2.568 \\ 2.243 \\ 2.514 \\ 2.827 \\ 3.102 \\ 3.099 \\ 3.265 \\ 3.421 \\ 2.856 \\ 3.064 \\ 2.732 \\ 2.460 \end{pmatrix}$$

Matrix Formulation

- *Example*: (Car plant electricity usage)

The residuals are then

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \begin{pmatrix} 2.48 \\ 2.26 \\ 2.47 \\ 2.77 \\ 2.99 \\ 3.05 \\ 3.18 \\ 3.46 \\ 3.03 \\ 3.26 \\ 2.67 \\ 2.53 \end{pmatrix} - \begin{pmatrix} 2.568 \\ 2.243 \\ 2.514 \\ 2.827 \\ 3.102 \\ 3.099 \\ 3.265 \\ 3.421 \\ 2.856 \\ 3.064 \\ 2.732 \\ 2.460 \end{pmatrix} = \begin{pmatrix} -0.088 \\ 0.017 \\ -0.044 \\ -0.057 \\ -0.112 \\ -0.049 \\ -0.085 \\ 0.039 \\ 0.174 \\ 0.196 \\ -0.062 \\ 0.070 \end{pmatrix}$$

Matrix Formulation

- *Example*: (Car plant electricity usage)

and the sum of squares for error is

$$\text{SSE} = \mathbf{e}'\mathbf{e} = (-0.088)^2 + \cdots + 0.070^2 = 0.1142$$

The estimate of the error variance is therefore

$$\hat{\sigma}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{0.1142}{9} = 0.0127$$

with $\hat{\sigma} = \sqrt{0.0127} = 0.113$

Matrix

- Matrix algebra using Excel:
  - Three steps:
    (1) Select all cells where the result will be printed
    (2) Enter the appropriate function and inputs
    (3) Press **Ctrl+Shift+Enter**
  - Matrix addition/subtraction → + or –
  - Matrix multiplication → **mmult(*array1*, *array2*)**
  - Matrix transpose → **transpose(*array*)**
  - Matrix inverse → **minverse(*array*)**

Matrix Formulation

- *Example*: (Car plant electricity usage)
  - Excel

	Electricity usage (million kWh)	Production ($ million)	Cooling degree days
January	2.48	4.51	0
February	2.26	3.58	0
March	2.47	4.31	13
April	2.77	5.06	56
May	2.99	5.64	117
June	3.05	4.99	306
July	3.18	5.29	358
August	3.46	5.83	330
September	3.03	4.70	187
October	3.26	5.61	94
November	2.67	4.90	23
December	2.53	4.20	0

Model Fitting

- If the hypothesis test is rejected,

$$H_0\colon \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

- This shows that at least one independent variable is useful
- But which one?

  ➔ One approach is to check the independent variables one by one

Model Fitting

- We can perform hypothesis tests for input variables individually:

$$H_0\colon \beta_i = 0 \quad \text{versus} \quad H_A\colon \beta_i \neq 0$$

- If the null hypothesis is accepted, then there is no evidence that the response variable is directly related to the input variable x_i

  ➔ Thus, x_i can be dropped from the model

Model Fitting

- For example, suppose we have the following model:

$$\bar{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}$$

- We perform the following tests:

① $H_0\colon \beta_1 = 0$ versus $H_A\colon \beta_1 \neq 0$ ➔ Reject

② $H_0\colon \beta_2 = 0$ versus $H_A\colon \beta_2 \neq 0$ ➔ Accept

③ $H_0\colon \beta_3 = 0$ versus $H_A\colon \beta_3 \neq 0$ ➔ Reject

- Then, the model can be improved as

$$\bar{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_3 x_{3i}$$

Model Fitting

- Individual hypothesis test:

$$H_0\colon \beta_i = 0 \quad \text{versus} \quad H_A\colon \beta_i \neq 0$$

- The test here is different from the test in Chapter 12
- Here, we want to check if x_i is useful when there are $x_1, \dots, x_k$
- Therefore, a *t*-distribution with $n - k - 1$ degrees of freedom is used

Model Fitting

- Individual hypothesis test:

$$H_0\colon \beta_i = 0 \quad \text{versus} \quad H_A\colon \beta_i \neq 0$$

- The test here is different from the test in Chapter 12
- Here, we want to check if x_i is useful when there are $x_1, \ldots, x_k$
- Therefore, a *t*-distribution with $n - k - 1$ degrees of freedom is used

- Chapter 12: use t_{n-2}
- Chapter 13: use t_{n-k-1}

Model Fitting

- Individual hypothesis test:

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

- Calculation:

t-statistic: $$t = \frac{\hat{\beta}_i}{\text{s.e.}(\hat{\beta}_i)}$$

p-value: $$2 \times P(t_{n-k-1} > |t|)$$

Model Fitting

- *Example*:

$$\bar{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}$$

- We perform the following tests:

① $H_0: \beta_1 = 0$ versus $H_A: \beta_1 \neq 0$ ➔ p-value = 0.101

② $H_0: \beta_2 = 0$ versus $H_A: \beta_2 \neq 0$ ➔ p-value = 0.003

③ $H_0: \beta_3 = 0$ versus $H_A: \beta_3 \neq 0$ ➔ p-value = 0.238

Which ones should we remove?

Model Fitting

- *Example*:

$$\bar{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} \cancel{+ \hat{\beta}_3 x_{3i}}$$

- We perform the following tests:

① H_0: $\beta_1 = 0$ versus H_A: $\beta_1 \neq 0$ ➔ *p*-value = 0.101

② H_0: $\beta_2 = 0$ versus H_A: $\beta_2 \neq 0$ ➔ *p*-value = 0.003

~~③ H_0: $\beta_3 = 0$ versus H_A: $\beta_3 \neq 0$ ➔ *p*-value = 0.238~~

If we remove x_3 and perform individual hypothesis tests again,

the *p*-values will be different

Model Fitting

- *Example*:

$$\bar{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$$

- We perform the following tests:

① $H_0: \beta_1 = 0$ versus $H_A: \beta_1 \neq 0$ ➔ *p*-value = 0.096

② $H_0: \beta_2 = 0$ versus $H_A: \beta_2 \neq 0$ ➔ *p*-value = 0.002

If we remove x_3 and perform individual hypothesis tests again,

the *p*-values will be different

Model Fitting

- Backwards elimination:

  (1) For all i, perform:

  $$H_0\colon \beta_i = 0 \quad \text{versus} \quad H_A\colon \beta_i \neq 0$$

  (2) If the null hypothesis is rejected in all cases, finish

  (3) Otherwise, remove x_i with the highest p-value from the model

  (4) Repeat (1) with the current model

Model Fitting

- Model fitting is performed by finding which subset of the k input variables is required to model the dependent variable y in the best manner
- The final model that the experimenter uses for inference problems should consist of input variables that each have p-values no larger than 10% (this significance level may change)
- It is important to note that it is best to remove only one variable from the model at a time (begin by removing the one with the largest p-value and then fit the model with $k - 1$ input variables)

Model Fitting

- We still need to discuss how to find s.e.$\left(\hat{\beta}_i\right)$
- We need two things:
  1) $\hat{\sigma}$ (this is $\sqrt{\text{MSE}}$)
  2) Diagonal elements of $(X'X)^{-1}$

➔ s.e.$\left(\hat{\beta}_i\right) = \hat{\sigma} \times \sqrt{\text{diagonal element of } (X'X)^{-1} \text{ for the } i^{\text{th}} \text{ variable}}$

Summary of Chapter 13

- Multiple linear regression
  - Estimating betas
  - Matrix representation
  - ANOVA Table
  - Model fitting (which betas are useful?)