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 3
The Multiple Linear Regression Model

In this Chapter, we will know some important concepts.

- Data and model
- Estimation
- The properties of estimators
- Linear models with constraints
- Hypothesis testing
- Data with Repeated Observations
- Multicollinearity in Multiple Regression Data

3.1 Data and Model

Example 3.1.: Hospital Manpower Data

In the last chapters, we have studied the monthly manpower problem, where we found that the **monthly man-hours (*Y*)** has a high correlation with the **monthly occupied bed days**. In fact, there are more variables that are related to the monthly man-hours, for instance:

Y:	Monthly Man-hours
X_1:	Average Daily Patient Load
X_2:	Monthly X-ray Exposures
X_3:	Monthly Occupied Bed Days
X_4:	Eligible Population in the Area /1000
X_5:	Average Length of Patients' Stay in Days

The related data is listed in Table 3.1:

Table 3.1

Obs	X1	X2	X3	X4	X5	Y
1	15.57	2463	472.92	18.0	4.45	566.52
2	44.02	2048	1339.75	9.5	6.92	696.82
3	20.42	3940	620.25	12.8	4.28	1033.15
4	18.74	6505	568.33	36.7	3.90	1603.62
5	49.20	5723	1497.60	35.7	5.50	1611.37
6	44.92	11520	1365.83	24.0	4.60	1613.27
7	55.48	5779	1687.00	43.3	5.62	1854.17
8	59.28	5969	1639.92	46.7	5.15	2160.55
9	94.39	8461	2872.33	78.7	6.18	2305.58
10	128.02	20106	3655.08	180.5	6.15	3503.93
11	96.00	13313	2912.00	60.9	5.88	3571.89
12	131.41	10771	3921.00	103.7	4.88	3741.40
13	127.21	15543	3865.67	126.8	5.50	4026.52
14	252.90	36194	7684.10	157.7	7.00	10343.81
15	409.20	34703	12446.33	167.4	10.78	11732.17
16	463.70	39204	14098.40	331.4	7.05	15414.94
17	510.22	86533	15524.00	371.6	6.35	18854.45

Fit this data by a regression model ➪

Regression Model for Hospital Manpower problem:

$$\{y_i, x_{i1}, x_{i2}, \cdots, x_{i5}, \quad i = 1, \cdots, 17\}$$

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon \\ E(\varepsilon) = 0, \qquad \mathrm{Var}(\varepsilon) = \sigma^2 \text{ unknown} \end{cases}$$

or

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i \\ \varepsilon_1, \ldots, \varepsilon_n \text{ are iid.} \quad E(\varepsilon_i) = 0, \quad \mathrm{Var}(\varepsilon_i) = \sigma^2. \end{cases}$$

where $n = 17$.

In general:

Regression Model:

$$\{y_i, x_{i1}, x_{i2}, \cdots, x_{i,p-1}, \quad i = 1, \cdots, n\}$$

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1} x_{p-1} + \varepsilon \\ E(\varepsilon) = 0, \qquad Var(\varepsilon) = \sigma^2 \text{ unknown} \end{cases}$$

or

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i,\,p-1} + \varepsilon_i \\ \varepsilon_i, \ldots, \varepsilon_n \text{ are iid.} \quad E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \sigma^2. \end{cases}$$

General Regression Model in terms of Matrix:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ 1 & x_{21} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{n,p-1} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

$$\begin{cases} \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ E(\boldsymbol{\varepsilon}) = 0, \qquad Cov(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n \end{cases} \tag{3.1}$$

where

$\boldsymbol{y}$: $n\times1$, $\boldsymbol{X}$: $n\times p$, $\boldsymbol{\beta}$: $p\times1$,

$\boldsymbol{\varepsilon}$: $n\times1$ with iid components.

What is a Linear Model?

The word "linear" refers to the coefficients, *not* the regressors. Some examples of linear models are

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \epsilon \\
y &= \beta_0 + \beta_1 x_1 x_2 + \epsilon \\
y &= \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + \epsilon \\
y &= \beta_0 + \beta_1 / x_1 + \epsilon,
\end{aligned}
$$

whereas the following are considered nonlinear models

$$
\begin{aligned}
y &= \beta_0 + \beta_1 x_1^{\beta_3} + \beta_2 x_2^{\beta_4} + \epsilon \\
y &= \frac{\beta_0}{1 + e^{-\beta_1 x_1}}.
\end{aligned}
$$

Question: Is $y = \beta_0 X^{\beta_1}$ linear or nonlinear?

Assumptions in multiple linear regression model

Some assumptions are needed in the model $y = X\beta + \varepsilon$ for drawing the statistical inferences. The following assumptions are made:

(i) $E(\varepsilon) = 0$

(ii) $E(\varepsilon\varepsilon') = \sigma^2 I_n$

(iii) $Rank(X) = k$

(iv) X is a non-stochastic matrix

(v) $\varepsilon \sim N(0, \sigma^2 I_n)$.

This model (3.1) involves many useful models:

(a) $$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_{p-1} x^{p-1} + \varepsilon$$

(b) $$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$

Example 3.2 *Quadratic Model*

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{1 \le i \le j \le k} \beta_{ij} x_i x_j + \varepsilon$$

Example 3.2 *Quadratic Model (cont…)*

When $k = 2$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 + \varepsilon$$

where

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{11}x_{12} & x_{12}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & x_{n1}x_{n2} & x_{n2}^2 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{11} \\ \beta_{12} \\ \beta_{22} \end{pmatrix}$$

Example 3.3 *Models with Non-linear functions*

For example

$$y = \beta_0 + \beta_1 x_1 + \beta_2 \log x_2 + \beta_3 \frac{1}{x_1 x_2} + \varepsilon$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \log(x_{12}) & \frac{1}{x_{11}x_{12}} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \log(x_{n2}) & \frac{1}{x_{n1}x_{n2}} \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$$

where $x_{i2} > 0$ and $x_{i1} \neq 0$, $i = 1, \ldots, n$.

3.2 Estimation

Recall the general linear regression model :

$$\begin{cases} \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \mathrm{E}(\boldsymbol{\varepsilon}) = 0, \qquad \mathrm{Cov}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n \end{cases} \tag{3.1}$$

A. The least squares estimator of $\boldsymbol{\beta}$

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

$$\frac{\partial \mathrm{Q}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}$$

$$\Rightarrow \quad \boldsymbol{b} = \hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{3.2}$$

*Does this solution minimize **Q**(β)?*

$$\begin{aligned} Q(\beta) &= (y - X\hat{\beta} + X\hat{\beta} - X\beta)'(y - X\hat{\beta} + X\hat{\beta} - X\beta) \\ &= Q(\hat{\beta}) + (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \\ &\geq Q(\hat{\beta}) \end{aligned}$$

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

$$= \mathbf{y}'\underbrace{(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X}}_{\boldsymbol{0}}(\hat{\beta} - \beta)$$

$$= 0$$

Therefore, $Q(\beta) \geq Q(\hat{\beta})$ and this equality holds $\Leftrightarrow \beta = \hat{\beta}$

(if $(X'X) > 0$)

For regression model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \qquad k = p-1$$

By the use of the inverse of partition matrix, we can show that the regular equation $(X'X)\beta = X'y$ is equivalent to:

$$\begin{pmatrix} S_{x_1x_1} & \cdots & S_{x_1x_k} \\ \vdots & \ddots & \vdots \\ S_{x_kx_1} & \cdots & S_{x_kx_k} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} S_{x_1y} \\ \vdots \\ S_{x_ky} \end{pmatrix} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \cdots - \hat{\beta}_k \bar{x}_k$$

where $\mathbf{S}_{x_ix_j} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$, $\quad i, j = 1, \cdots, k$

$$\bar{x}_j = \tfrac{1}{n} \sum_{k=1}^{n} x_{kj} \qquad \mathbf{S}_{X_iy} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(y_i - \bar{y})$$

Let's express the above regression model in terms of matrix algebra.

Denote $\boldsymbol{X} = (\boldsymbol{1}_n \boldsymbol{X}_*)$ $\qquad \boldsymbol{S}_{\mathrm{XX}} = \boldsymbol{X}_*^{'} \boldsymbol{D}_n \boldsymbol{X}_*$

$\boldsymbol{S}_{\mathrm{XY}} = \boldsymbol{X}_*^{'} \boldsymbol{D}_n \boldsymbol{y}$ $\qquad \boldsymbol{S}_{\mathrm{YY}} = \boldsymbol{y}' \boldsymbol{D}_n \boldsymbol{y}$

where $\boldsymbol{D}_n = \boldsymbol{I}_n - \frac{1}{n} \boldsymbol{1}_n \boldsymbol{1}_n{}'$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \mathbf{1}_n{}'\mathbf{X}_* \\ \mathbf{X}_*\mathbf{1}_n & \mathbf{X}_*^{'}\mathbf{X}_* \end{pmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} n & \mathbf{1}_n{}'\mathbf{X}_* \\ \mathbf{X}'_*\mathbf{1}_n & \mathbf{X}_*{}'\mathbf{X}_* \end{pmatrix}^{-1} = \begin{pmatrix} n & n\bar{x}' \\ n\bar{x} & \mathbf{X}_*^{'}\mathbf{X}_* \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \frac{1}{n} + \bar{x}'\mathbf{S}_{\mathrm{XX}}^{-1}\bar{x} & -\bar{x}'\mathbf{S}_{\mathrm{XX}}^{-1} \\ -\mathbf{S}_{\mathrm{XX}}^{-1}\bar{x} & \mathbf{S}_{\mathrm{XX}}^{-1} \end{pmatrix}$$

Review: Partition of Matrix

Partition **A** and **A**$^{-1}$ into:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \qquad \mathbf{B} \equiv \mathbf{A}^{-1} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$$

where $\mathbf{A}_{11}$ and $\mathbf{B}_{11}$ are $m \times m$.

Theorem: If $\mathbf{A}_{11}$ is non-singular, then:

$$\mathbf{A}^{-1} \equiv \mathbf{B} = \left(\begin{array}{c:c} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}_{22\cdot 1}^{-1} \\ \hdashline -\mathbf{A}_{22\cdot 1}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{A}_{22\cdot 1}^{-1} \end{array}\right)$$

where $\mathbf{A}_{22\cdot 1} = \mathbf{A}_{22} - \mathbf{A}_{21}\,\mathbf{A}_{11}^{-1}\,\mathbf{A}_{12}$.

Therefore

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_* \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n} + \bar{x}'\mathbf{S}_{\mathrm{XX}}^{-1}\bar{x} & -\bar{x}'\mathbf{S}_{\mathrm{XX}}^{-1} \\ -\mathbf{S}_{\mathrm{XX}}^{-1}\bar{x} & \mathbf{S}_{\mathrm{XX}}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{1}_n'\mathbf{y} \\ \mathbf{X}_*'\mathbf{y} \end{pmatrix}$$

$$= \begin{pmatrix} \left(\frac{1}{n} + \bar{x}'\mathbf{S}_{\mathrm{XX}}^{-1}\bar{x}\right) n\,\bar{\mathbf{y}} - \bar{x}'\mathbf{S}_{\mathrm{XX}}^{-1}\mathbf{X}_*'\mathbf{y} \\ -\mathbf{S}_{\mathrm{XX}}^{-1}\bar{x}\, n\,\bar{\mathbf{y}} + \mathbf{S}_{\mathrm{XX}}^{-1}\mathbf{X}_*'\mathbf{y} \end{pmatrix}$$

$$= \begin{pmatrix} \bar{\mathbf{y}} - \bar{x}'\mathbf{S}_{\mathrm{XX}}^{-1}(\mathbf{X}_*'\mathbf{y} - n\,\bar{x}\,\bar{\mathbf{y}}) \\ \mathbf{S}_{\mathrm{XX}}^{-1}(\mathbf{X}_*'\mathbf{y} - n\,\bar{x}\,\bar{\mathbf{y}}) \end{pmatrix} \qquad \text{as } \mathbf{S}_{\mathrm{XY}} = \mathbf{X}_*'\mathbf{y} - n\,\bar{x}\,\bar{\mathbf{y}}$$

$$= \begin{pmatrix} \bar{\mathbf{y}} - \bar{x}'\hat{\beta}_* \\ \mathbf{S}_{\mathrm{XX}}^{-1}\mathbf{S}_{\mathrm{XY}} \end{pmatrix} \qquad \Rightarrow \begin{cases} \hat{\beta}_* = \mathbf{S}_{\mathrm{XX}}^{-1}\mathbf{S}_{\mathrm{XY}} \\ \hat{\beta}_0 = \bar{\mathbf{y}} - \bar{x}'\hat{\beta}_* \end{cases}$$

$$\boldsymbol{S}_{XX}\, b_1 = \boldsymbol{S}_{XY}$$

where

$$\boldsymbol{S}_{XX} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

$$\boldsymbol{S}_{XY} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

$$\Rightarrow \begin{cases} b_1 = \mathrm{S}_{XY} / \mathrm{S}_{XX} \\ b_0 = \bar{y} - b_1 \bar{x} \end{cases}$$

B. Estimation of σ^2

Recall

$$Q(\hat{\beta}) = y'(I - H)y,$$

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

$$\mathrm{E}\left(Q(\hat{\beta})\right) = \mathrm{E}\left(y'(I-H)y\right) \qquad \boxed{\mathrm{E}(y) = X\beta \qquad \mathrm{Cov}(y) = \sigma^2 I_n}$$

$$= (X\beta)'\underbrace{(I-H)X\beta}_{0} + \mathrm{tr}[(I-H)\sigma^2 I_n]$$

$$= \sigma^2 \mathrm{tr}\left(I - H\right)$$

$$= \sigma^2\left(n - \mathrm{tr}(X(X'X)^{-1}X')\right)$$

$$= \sigma^2\left(n - \mathrm{tr}(X'X(X'X)^{-1})\right) = \sigma^2(n-p)$$

$$\boxed{s^2 = \hat{\sigma}^2 = \tfrac{1}{n-p}\mathbf{Q}(\hat{\beta}) = \tfrac{1}{n-p}\mathbf{y}'(\mathbf{I}_n - \mathbf{H})\mathbf{y} \quad \text{is an unbiased estimator.}}$$

Maximum likelihood estimation

In the model $y = X\beta + \varepsilon$, it is assumed that the errors are normally and independently distributed with constant variance σ^2 or $\varepsilon \sim N(0, \sigma^2 I)$.

The normal density function for the errors is

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\varepsilon_i^2\right] \quad i = 1, 2, \ldots, n..$$

The likelihood function is the joint density of $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ *iven as*

$$
\begin{aligned}
L(\beta, \sigma^2) &= \prod_{i=1}^{n} f(\varepsilon_i) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \varepsilon_i^2 \right] \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \varepsilon' \varepsilon \right] \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right].
\end{aligned}
$$

Since the log transformation is monotonic, so we maximize $\ln L(\beta,\sigma^2)$ instead of $L(\beta,\sigma^2)$.

$$\ln L(\beta,\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

$$\frac{\partial \ln L(\beta,\sigma^2)}{\partial \beta} = \frac{1}{2\sigma^2}2X'(y - X\beta) = 0$$

$$\frac{\partial \ln L(\beta,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(y - X\beta)'(y - X\beta).$$

Since $\text{rank}(X) = k$, so that the unique mle of β and σ^2 are obtained as

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

$$\tilde{\sigma}^2 = \frac{1}{n}(y - X\tilde{\beta})'(y - X\tilde{\beta}).$$

Further to verify that these values maximize the likelihood function, we find

$$\frac{\partial^2 \ln L(\beta, \sigma^2)}{\partial \beta^2} = -\frac{1}{\sigma^2} X'X$$

$$\frac{\partial^2 \ln L(\beta, \sigma^2)}{\partial^2 (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}(y - X\beta)'(y - X\beta)$$

$$\frac{\partial^2 \ln L(\beta, \sigma^2)}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4} X'(y - X\beta).$$

Thus the Hessian matrix of second order partial derivatives

$$\begin{pmatrix} \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial \beta^2} & \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial \beta \partial \sigma^2} \\ \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial \sigma^2 \partial \beta} & \dfrac{\partial^2 \ln L(\beta,\sigma^2)}{\partial^2 (\sigma^2)^2} \end{pmatrix}$$

which is negative definite at $\beta = \tilde{\beta}$ and $\sigma_2 = \tilde{\sigma}^2$.

Comparing with OLSEs, we find that

(i) OLSE and m.l.e. of β are same. So m.l.e. of β is also an unbiased estimator of β.

(ii) OLSE of σ^2 is s^2 which is related to m.l.e. of σ^2 as $\tilde{\sigma}^2 = \frac{n-k}{n} s^2$. So m.l.e. of σ^2 is a biased estimator of σ^2.

3.3 Properties of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$.

Theorem 3.1

Consider the model

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

Then

(i) $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n)$

(ii) $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1})$

(iii) $\frac{1}{\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\mathbf{X}'\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \sim \chi^2_p$

(iv) $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent

(v) $\mathbf{SS}_{\text{Res}} / \sigma^2 \sim \chi^2_{n-p}$

Proof: Theorem 3.1 (i) - (ii) (a linear function of a multivariate normal is also a multivariate normal)

(i) As

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$$

and a linear function of $\boldsymbol{\varepsilon}$ is still normal distributed.

$$\Rightarrow \boldsymbol{y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n) \quad \blacksquare$$

(ii) $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, $\hat{\boldsymbol{\beta}}$ has a multivariate normal

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

It follows $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$

Proof: Theorem 3.1 (iii)

(iii)

$$\boldsymbol{Z} \sim N_p(\mathbf{0}, \boldsymbol{I}_p) \quad \Rightarrow \quad \boldsymbol{Z}'\boldsymbol{Z} \sim \chi_p^2$$

$$\boldsymbol{Z} \sim N_p(\mathbf{0}, \sigma^2 \boldsymbol{I}_p) \quad \Rightarrow \quad \tfrac{1}{\sigma^2}\boldsymbol{Z}'\boldsymbol{Z} \sim \chi_p^2$$

$$\boldsymbol{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}) \quad \Rightarrow \quad \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{Z} \sim N_p(\mathbf{0}, \boldsymbol{I}_p) \quad \Rightarrow \quad \boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z} \sim \chi_p^2$$

From (ii) we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim N_p(\mathbf{0}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$$

Then (iii) follows.

■

Lemma 3.1

Suppose $\boldsymbol{y} \sim N_n(\mu, \sigma^2 \boldsymbol{I}_n)$ and let $\boldsymbol{U} = \boldsymbol{A}\boldsymbol{y}$ and $\boldsymbol{V} = \boldsymbol{B}\boldsymbol{y}$, where $\boldsymbol{A}$: $s \times n$, $\boldsymbol{B}$: $t \times n$, $s < n$ and $t < n$. Then $\boldsymbol{U}$ and $\boldsymbol{V}$ are *independent* iff $\boldsymbol{A}\boldsymbol{B}' = \boldsymbol{0}$.

Proof: Lemma 3.1

Let

$$\mathbf{Z} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{y} = \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} : (s+t) \times 1$$

$$\Rightarrow \quad \mathbf{Z} \sim N_{s+t}\left(\mathbf{0}, \sigma^2 \begin{pmatrix} \mathbf{AA'} & \mathbf{AB'} \\ \mathbf{BA'} & \mathbf{BB'} \end{pmatrix} \right)$$

$$\Rightarrow \quad \mathbf{U} \text{ and } \mathbf{V} \text{ are independent iff } \mathbf{AB'} = \mathbf{0}.$$

Lemma 3.2

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are *independent*, then $f(\boldsymbol{x})$ and $g(\boldsymbol{y})$ are *independent*, where f and g are measurable.

Example:

$$\text{If } \; x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \; \text{ and } \; \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \; \text{ are independent,}$$

$$\Rightarrow \begin{pmatrix} x_2 \\ x_1^2 + x_2^2 \end{pmatrix} \; \text{ and } \; \begin{pmatrix} \sin(y_1) \\ y_1 + 2y_2 \\ \cos(y_3) \end{pmatrix} \; \text{ are independent.}$$

Proof: Theorem 3.1 (iv)

(iv)

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

$$SS_{\text{Res}} = \mathbf{y}'(\mathbf{I}-\mathbf{H})(\mathbf{I}-\mathbf{H})\mathbf{y} = [(\mathbf{I}-\mathbf{H})\mathbf{y}]'[(\mathbf{I}-\mathbf{H})\mathbf{y}]$$

Let $\boldsymbol{z} = \underbrace{(\boldsymbol{I}-\boldsymbol{H})}_{\boldsymbol{B}}\boldsymbol{y}$ and $\hat{\boldsymbol{\beta}} = \underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'}_{\boldsymbol{A}}\boldsymbol{y}$

$$\boldsymbol{A}\boldsymbol{B}' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\underbrace{\boldsymbol{X}'(\boldsymbol{I}-\boldsymbol{H})}_{\mathbf{0}} = \mathbf{0}$$

From Lemma 3.1 and Lemma 3.2,

$\hat{\boldsymbol{\beta}}$ and $\boldsymbol{z}$ are independent

$\Rightarrow \hat{\boldsymbol{\beta}}$ and $\boldsymbol{z}'\boldsymbol{z} = SS_{\text{Res}}$ are independent

$\Rightarrow \hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2 = \frac{1}{n-p}SS_{\text{Res}}$ are independent ■

Lemma 3.3 (Cochran's Theorem)

Let $\boldsymbol{y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \boldsymbol{I}_n)$ and $\boldsymbol{A}$ be an $n \times n$ symmetric matrix, then

$$\frac{1}{\sigma^2}\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} \sim \chi^2_k \quad \Leftrightarrow \quad \begin{cases} \boldsymbol{A}\boldsymbol{\mu} = \boldsymbol{0} \\ \boldsymbol{A}^2 = \boldsymbol{A} \\ rank\ \boldsymbol{A} = k \end{cases}$$

Proof: Theorem 3.1 (v)

(v) As $\boldsymbol{SS}_{\text{Res}} = \boldsymbol{y}'\underbrace{(\boldsymbol{I} - \boldsymbol{H})}_{\boldsymbol{A}}\boldsymbol{y}$ and $\boldsymbol{y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$

$(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}$, $(\boldsymbol{I} - \boldsymbol{H})^2 = (\boldsymbol{I} - \boldsymbol{H})$ and $\text{rank}(\boldsymbol{I} - \boldsymbol{H}) = n - p$

$\Rightarrow \quad SS_{\text{Res}} / \sigma^2 \sim \chi^2_{n-p}$ ■

Let $\hat{\boldsymbol{\theta}}$ be an estimate of $\boldsymbol{\theta}$ based on a sample. We wish:

(i) $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$ unbiased

(ii) $Var(\hat{\boldsymbol{\theta}})$ is the smallest in a certain group of estimators.

Let $\boldsymbol{\theta} = \boldsymbol{X\beta}$ be the considered parameters,

$\hat{\boldsymbol{\theta}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ be the LSE of θ, unbiased.

Consider $\phi = \boldsymbol{a}'\boldsymbol{\theta}$

Look at $\hat{\phi} = \boldsymbol{a}'\hat{\boldsymbol{\theta}} = \boldsymbol{a}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{a}'\boldsymbol{H}\boldsymbol{y} \equiv \boldsymbol{c}'\boldsymbol{y}$

$$\boldsymbol{c}' = \boldsymbol{a}'\boldsymbol{H} \quad \Rightarrow \quad \boldsymbol{c} = \boldsymbol{H}\boldsymbol{a}$$

$\boldsymbol{c}$ belongs to $\Omega = \{\boldsymbol{b} : \boldsymbol{b} = \boldsymbol{H}\boldsymbol{a}, \quad \forall \boldsymbol{a} \in \boldsymbol{R}^n\}$

Define $\Phi = \{\hat{\phi} = \boldsymbol{b}'\boldsymbol{y}, \quad \boldsymbol{b} \in \Omega, \quad E\hat{\phi} = \phi\}$

Consider $\phi = \boldsymbol{a}'\boldsymbol{\theta}$

Look at $\phi = a'\theta = a'X(X'X)^{-1}X'y = a'Hy \equiv c'y$

$$c' = a'H \Rightarrow c = Ha$$

$\boldsymbol{c}$ belongs to $\Omega = \{\boldsymbol{b} : \boldsymbol{b} = \boldsymbol{Ha}, \quad \forall \boldsymbol{a} \in \boldsymbol{R}^n\}$

Define $\Phi = \{\, \hat{\phi} = \boldsymbol{b}'\boldsymbol{y}, \quad \boldsymbol{b} \in \Omega, \quad \mathrm{E}\hat{\phi} = \phi \,\}$

Gauss - Markov Theorem:

Let $\boldsymbol{\theta} = \boldsymbol{X\beta}$ and $\phi = \boldsymbol{a}'\ \boldsymbol{\theta}$, where $\boldsymbol{a} \in R^n$. Then

$$\hat{\phi} = \mathbf{a}'\hat{\boldsymbol{\theta}} = \mathbf{a}'\mathbf{Hy}$$

is the Best Linear Unbiased Estimate **(BLUE)** under the group $\boldsymbol{\Phi}$.

Refer to the textbook P. 92, where use UMVM for BLUE.

3.4 Linear Models with Constraints

$$\begin{cases} \boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, & \boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n) \\ \boldsymbol{A\beta} = \boldsymbol{c}, & \boldsymbol{A}: q \times p \ (q < p), \quad \boldsymbol{c}: q \times 1. \\ & \text{Rank}(\boldsymbol{A}) = q, \end{cases}$$

- $\hat{\boldsymbol{\beta}}_H, \hat{\sigma}_H^2$ under the constraints $\boldsymbol{A\beta} = \boldsymbol{c}$;
- $\hat{\boldsymbol{\beta}}, \hat{\sigma}^2$ estimates without constraints.

Theorem 3.2

Under the constraint model, we have

(i) $\hat{\boldsymbol{\beta}}_H = \hat{\boldsymbol{\beta}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{A}']^{-1}(\boldsymbol{c} - \boldsymbol{A}\hat{\boldsymbol{\beta}})$

(ii) $Q_H = Q + (\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})'[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{A}']^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})$

where $Q = Q(\hat{\boldsymbol{\beta}})$ and $Q_H = Q(\hat{\boldsymbol{\beta}}_H)$.

3.5 Hypothesis Testing

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i, \quad i = 1,\ldots,n \\ \varepsilon_i,\ldots,\varepsilon_n \text{ are iid.} \quad \varepsilon_i \sim N(0, \sigma^2). \end{cases}$$

The following hypotheses are often needed in the practice:

A. Testing whether the model makes the sense.

$\mathbf{H_0}$: $\beta_1 = \cdots = \beta_k = 0$ VS $\mathbf{H_a}$: Some $\beta_j \neq 0$, $1 \leq j \leq k$

B. Testing whether the variable x_j gives a significant contribution to the model.

$\mathbf{H_0}$: $\beta_j = 0$ VS $\mathbf{H_a}$: $\beta_j \neq 0$

The following hypotheses are often needed in the practice:

A. Testing whether the model makes the sense

$\mathbf{H_0}$: $\beta_1 = \cdots = \beta_k$ VS $\mathbf{H_a}$: Some $\beta_j \neq 0$, $1 \leq j \leq k$

B. Testing whether the variable x_j gives a significant contribution to the model

$\mathbf{H_0}$: $\beta_j = 0$ VS $\mathbf{H_a}$: $\beta_j \neq 0$

C. Testing whether the q ($q < k$) regressors give a significant contribution to the model

$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, $X_1 : n\times(p\text{-}q)$ $X_2 : n\times q$

$\mathbf{H_0}$: $\beta_2 = 0$ VS $\mathbf{H_a}$: $\beta_2 \neq 0$

D. General Hypothesis

$\mathbf{H_0}$: $A\beta = c$ VS $\mathbf{H_a}$: $A\beta \neq c$

Theorem 3.3

Denote $Q = Q(\hat{\boldsymbol{\beta}})$ and $Q_H = Q(\hat{\boldsymbol{\beta}}_H)$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_H$ are given in the previous section.

Under the above notations for case (D), we have:

(i) $$Q_H - Q = (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})$$

(ii) $$E(Q_H - Q) = \sigma^2 q + (\mathbf{A}\boldsymbol{\beta} - \mathbf{c})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{A}\boldsymbol{\beta} - \mathbf{c})$$

(iii) When $\boldsymbol{H_0}$ is true, the statistics

$$F = \frac{(Q_H - Q)/q}{Q/(n-p)} = \frac{Q_H - Q}{qs^2} \sim F_{q,n-p}$$

The proof can refer to Seber (1977) and the Textbook, p.105.

Example 3.4: (Case **A**)

$\mathbf{H_0}$: $\beta_1 = \cdots = \beta_k = 0$ VS $\mathbf{H_a}$: Some $\beta_j \neq 0$, $1 \leq j \leq k$

Taking $\boldsymbol{A} = (0, \boldsymbol{I}_k)$, $\boldsymbol{c} = 0$, we have:

$$\mathrm{A}\hat{\boldsymbol{\beta}} - \mathrm{c} = \hat{\boldsymbol{\beta}}_*, \quad \text{where } \hat{\boldsymbol{\beta}}_* = (\hat{\beta}_1, \cdots \hat{\beta}_k)'$$

From theorem 3.3, the testing statistic becomes:

$$F = \frac{\hat{\boldsymbol{\beta}}_*' \boldsymbol{S}_{XX} \hat{\boldsymbol{\beta}}_*}{ks^2} = \frac{\hat{\boldsymbol{\beta}}_*' \boldsymbol{S}_{XX} \hat{\boldsymbol{\beta}}_*}{(p-1)s^2} \sim F_{p-1,n-p}$$

As $$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} \frac{1}{n} + \bar{x}'\boldsymbol{S}_{XX}^{-1}\bar{x} & -\bar{x}'\boldsymbol{S}_{XX}^{-1} \\ -\boldsymbol{S}_{XX}^{-1}\bar{x} & \boldsymbol{S}_{XX}^{-1} \end{bmatrix}$$

and $$[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}']^{-1} = [\boldsymbol{S}_{XX}^{-1}]^{-1} = \boldsymbol{S}_{XX}.$$

Example 3.4: (*cont...*)

This test is usually expressed as a ANOVA table:

$$\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{\boldsymbol{SS}_{\text{Total}}}=\underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{\boldsymbol{SS}_{\text{Res}}}+\underbrace{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}_{\boldsymbol{SS}_{\text{Reg}}}$$

$$Q_H-Q=\sum_{i=1}^{n}(y_i-\bar{y})^2-\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\boldsymbol{SS}_{\text{Reg}}$$

So

$$F=\frac{\boldsymbol{SS}_{\text{Reg}}}{(p-1)s^2},\qquad R^2=\frac{\boldsymbol{SS}_{\text{Reg}}}{\boldsymbol{SS}_{\text{Total}}}$$

Coefficient of determination

Example 3.4: (*cont...*)

It is easy to verify

$$F = \frac{n-p}{p-1}\frac{R^2}{1-R^2} \sim F_{k,n-p}$$

F and R^2 are equivalent as $\frac{x}{1-x}$ is a monotone function.

ANOVA Table: ($k = p - 1$)

Source	DF	SS	MS	F	p
Regression	k	SS_{Reg}	$MS_{Reg} = SS_{Reg}/k$	$F = \frac{MS_{Reg}}{MS_{Res}}$	
Error	$n - p$	SS_{Res}	$MS_{Res} = SS_{Res}/n\text{-}p$		
Total	$n - 1$	SS_{Total}			

Example 3.5: (Case **B**) Refer to the textbook, p 98.

$$\mathbf{H_0}: \beta_j = 0 \quad \text{VS} \quad \mathbf{H_a}: \beta_j \neq 0$$

Take $\boldsymbol{A} = (0,\ldots,0,1,0,\ldots,0) \equiv \boldsymbol{e}_{j+1}$, where '1' is in the $j+1$ th position, $\boldsymbol{c} = 0$. Then $\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c} = \hat{\boldsymbol{\beta}}_j$

Denote $$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \mathbf{C} = (c_{ij})$$

$$[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}']^{-1} = c_{j+1,,j+1}^{-1}$$

$$Q_H - Q = \hat{\boldsymbol{\beta}}_j c_{ij}^{-1} \hat{\boldsymbol{\beta}}_j = \hat{\boldsymbol{\beta}}_j^2 / c_{j+1,j+1}$$

$$F = \frac{\hat{\boldsymbol{\beta}}_j^2}{c_{j+1,j+1}\, s^2} \sim F_{1,n-p}$$

or $$t = \frac{\hat{\boldsymbol{\beta}}_j}{\sqrt{c_{j+1,j+1}}\, s} \sim t_{n-p}$$

Example 3.1 (Hospital Manpower problem) *cont...*

If we choose x_2, x_3 and x_5 as the regressors, the following are the results:

The Model Equation:

$$Y = 1523.39 + 0.0530\, x_2 + 0.9785\, x_3 - 320.951\, x_5$$

ANOVA:

Source	DF	*SS*	*MS*	*F*	*p*
Model	3	489799142	163266381	431.9745	0.0001
Error	13	4913398.50	377953.731		
Total	16	494712540			

Type III Tests:

Source	DF	*SS*	*MS*	*F*	*p*
X2	1	2628687.60	2628687.60	6.9551	0.0205
X3	1	32726194.90	32726194.90	86.5878	0.0001
X5	1	1658984.04	1658984.04	4.3894	0.0563

with $R^2 = 0.9901$. Since x_2 and x_3 are significant at $\alpha = 0.05$, and x_5 is nearly significant at $\alpha = 0.05$, the model makes sense.

Example 3.6: (Case **C**) Refer to the textbook, pp 96-98

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon, \qquad \beta_2 : q\times 1$$

$$\mathbf{H_0}: \beta_2 = \mathbf{0} \quad \text{VS} \quad \mathbf{H_a}: \beta_2 \neq \mathbf{0}$$

Take $\boldsymbol{A} = [\mathbf{0}, \boldsymbol{I}_q]$, $\boldsymbol{c} = \mathbf{0}$

$$\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c} = \hat{\boldsymbol{\beta}}_2$$

$$[\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c}]'[\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}']^{-1}[\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c}] = \hat{\boldsymbol{\beta}}_2'\boldsymbol{C}_{22}^{-1}\hat{\boldsymbol{\beta}}_2$$

where $(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{pmatrix}, \qquad \boldsymbol{C}_{22} : q\times q$

Then

$$F = \frac{\hat{\boldsymbol{\beta}}_2'\boldsymbol{C}_{22}^{-1}\hat{\boldsymbol{\beta}}_2}{qs^2} \sim F_{q,n-p}$$

Example 3.7: (Case **D**) Refer to the textbook, 104-108, 110-111.

$\mathbf{H_0}$: $A\beta = c$ VS $\mathbf{H_a}$: $A\beta \neq c$

Set 1	Set 2
$y_1 \quad x_1$	$y_{n+1} \quad x_{n+1}$
$\vdots \quad \vdots$	$\vdots \quad \vdots$
$y_n \quad x_n$	$y_{n+m} \quad y_{n+m}$

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ $\qquad$ $y_i = \gamma_0 + \gamma_1 x_i + \varepsilon_i$

$i = 1, \ldots, n$ $\qquad$ $i = n+1, \ldots, n+m$

Testing: If two sets of data have the same situations?

$\mathbf{H_0}$: $\beta_0 = \gamma_0$ & $\beta_1 = \gamma_1$

$\mathbf{H_a}$: Some $\beta_i \neq \gamma_i$ for $i = 0, 1$.

Example 3.7: (*cont...*)

$$\mathbf{H_0}: \ \beta_0 = \gamma_0 \ \ \& \ \ \beta_1 = \gamma_1$$

$$\mathbf{H_a}: \ \text{Some } \beta_i \neq \gamma_i \ \ \text{for } i = 0, 1.$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \\ y_{n+1} \\ \vdots \\ y_{n+m} \end{pmatrix} = \begin{pmatrix} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & 0 & 0 \\ 0 & 0 & 1 & x_{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{n+m} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \gamma_0 \\ \gamma_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \\ \varepsilon_{n+1} \\ \vdots \\ \varepsilon_{n+m} \end{pmatrix}$$

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

If we put into the form: $\mathbf{H_0}$: $\boldsymbol{A\beta} = \boldsymbol{c}$

Example 3.7: (*cont...*)

$\mathbf{H_0}: \boldsymbol{A\beta} = \boldsymbol{c}$

where $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}, \qquad \boldsymbol{c} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

so $$\boldsymbol{A\beta} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \gamma_0 \\ \gamma_1 \end{pmatrix} = \begin{pmatrix} \beta_0 - \gamma_0 \\ \beta_1 - \gamma_1 \end{pmatrix} = \boldsymbol{c} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which is the same as testing: $\mathbf{H_0}$: $\beta_0 = \gamma_0$ & $\beta_1 = \gamma_1$

If we only test for one β: $\mathbf{H_0}$: $\beta_1 = \gamma_1$ VS $\mathbf{H_a}$: $\beta_1 \neq \gamma_1$

Then $\boldsymbol{A} = (0 \;\; 1 \;\; 0 \;\; -1)$ and $\boldsymbol{c} = (\,0\,)$.

3.6 Data with Repeated Observations

(Section 3.6, 116-8)

Obs	*X1*	*X2*	*X3*	*Y*
1	225	46	0.5	6.6
2	285	46	0.5	6.9
3	225	64	0.5	7.9
4	285	64	0.5	6.1
5	225	46	1.7	9.2
6	285	46	1.7	6.8
7	225	64	1.7	10.4
8	285	64	1.7	7.3
9	204.5	55	1.1	9.8
10	305.5	55	1.1	5.0
11	255	39.9	1.1	6.9
12	255	70.1	1.1	6.3
13	255	55	0.09	4.0
14	255	55	2.11	8.6
15	255	55	1.1	10.1
16	255	55	1.1	9.9
17	255	55	1.1	12.2
18	255	55	1.1	9.7
19	255	55	1.1	9.7
20	255	55	1.1	9.6

Example 3.8: (Breadwrapper Data , p.118)

To determine the between the seal strength (y) in grams per inch of a breadwrapper stock and regressor variables: sealing temperature (x_1), cooling bar temperature (x_2) and percent polyethylene in the stock (x_3).

We can apply the simple linear model into the multiple linear model.

For example, there is a repeated data:

x_1	$\cdots$	x_1	x_2	$\cdots$	x_2	$\cdots$	$\cdots$	x_m	$\cdots$	x_m
y_{11}	$\cdots$	y_{1n_1}	y_{21}	$\cdots$	y_{2n_2}	$\cdots$	$\cdots$	y_{m1}	$\cdots$	y_{mn_m}

$$n = n_1 + n_2 + \ldots + n_m$$

We want to fit the data by a simple linear model:

$$y = \beta_0 + \beta_1 x + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2)$$

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \beta_0 - \beta_1 x_i)^2 \to \min$$

$$\Rightarrow \hat{\beta}_1 = \frac{S_{X\overline{Y}}}{S_{XX}}, \quad \text{and} \quad \hat{\beta}_0 = \overline{\overline{y}} - \hat{\beta}_1 \overline{x}$$

where $S_{X\overline{Y}} = \sum_{i=1}^{m} (x_i - \overline{x})(\overline{y}_i - \overline{\overline{y}})$; $\quad \overline{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$;

$$\overline{\overline{y}} = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{ij} , \qquad \overline{x} = \frac{1}{m} \sum_{i=1}^{m} x_i .$$

i.e. $\hat{\beta}_0$ and $\hat{\beta}_1$ are the LSE's of β_0 and β_1 based on the data $\{(x_i, \overline{y}_i) : i = 1, \cdots, m\}$.

$$SS_{Total} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2$$

$$SS_{Reg} = \sum_{i=1}^{m} n_i(\hat{y}_i - \bar{\bar{y}})^2$$

$$SS_{Res} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \hat{y}_{ij})^2 = \boxed{\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2} + \boxed{\sum_{i=1}^{m} n_i(\bar{y}_i - \hat{\bar{y}}_i)^2}$$

Measures variation due to repeated observations

Degree of freedom:

$$\sum_{i=1}^{m}(n_i - 1) = n - m$$

Lack of fit sum of squares

Degree of freedom: m - 2

Two estimators of σ^2:

$$\hat{\sigma}^2 = \frac{1}{n-m}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 \text{ : pure random error}$$

$$\hat{\sigma}_l^2 = \frac{1}{m-2}\sum_{i=1}^{m} n_i(\bar{y}_i-\hat{\bar{y}}_i)^2 \equiv MS_{\text{Lof}} \text{ : another unbiased estimator of } \sigma^2.$$

and the *testing statistic for model adequacy*:

$$F = \frac{\hat{\sigma}_l^2}{\hat{\sigma}^2} \sim F(m-2, n-m)$$

Test Process

- Step 1. Do lack of fit test
  - Accepted: go to Step 2
  - Rejected: Go to Step 3
- Step 2. Merge DF and SS of "Lack of fit" and "Pure error", recalculate MS, do *F*-test
  - Accepted: the current model is not useful
  - Rejected: the current model is good
- Step 3. Calculate *F*=MS_reg/MS_pure and do test
  - Accepted: the current model is not useful
  - Rejected: the current model is useful, but it can be improved

Example 3.8: (Breadwrapper Data , p.118)

We fit the multiple regression model as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{33} x_{3i}^2 + \beta_{12} x_{1i} x_{2i} + \beta_{13} x_{1i} x_{3i} + \beta_{23} x_{2i} x_{3i} + \varepsilon_i$$

The X matrix can then be used to compute $X'X$ and $(X'X)^{-1}$. The coeficients are then computed from $\hat{\beta} = (X'X)^{-1} X'y$. The fitted least squares regression equation is as follows :

$$\begin{aligned}\hat{y} = &-104.8568 + 0.4948x_1 + 1.7303x_2 + 14.2620x_3 \\ &- 0.0008425x_1^2 - 0.01291x_2^2 - 3.1846x_3^2 \\ &- 0.0013x_1x_2 - 0.02778x_1x_3 + 0.02778x_2x_3\end{aligned}$$

with the residual sum of squares

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = 11.8678$$

and the residual estimate of variance, s^2 (10 residual degrees of freedom)

$$s^2 = \frac{11.8678}{10} = 1.18678$$

Example 3.8: (Breadwrapper Data)

Source	DF	*SS*	*MS*	*F*	*P*
Regression	9	70.3022	7.8113	—	—
Lack of fit	5	6.9078	1.3816	1.39	0.3633
Pure error	5	4.9600	0.9920	—	—

The F-statistic is not significant, and thus we infer that *lack of fit is insignificant*. Therefore, we combine the lack of fit and pure error together and perform the ANOVA test again.

Source	DF	*SS*	*MS*	*F*	*p*
Model	9	70.3022	7.8113	6.582	0.0034
Error	10	11.8678	1.18678		
Total	19	82.17			

3.7 Multicollinearity in Multiple Regression Data

(Sec. 3.8, 123-133)

Consider two somewhat different scenarios both containing two regressor variables.

Case 1:

x1	10	10	10	10	15	15	15	15
x2	10	10	15	15	10	10	15	15

Case 2:

x1	10.0	11.0	11.9	12.7	13.3	14.2	14.7	15.0
x2	10.0	11.4	12.2	12.5	13.2	13.9	14.4	15.0

- ✓ For Case 1, correlation coefficient r=0, and r=0.99215 for Case 2;

- ✓ For Case 1,

$$X^{*'}X^{*} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \qquad (X^{*'}X^{*})^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- ✓ For Case 2,

$$X^{*'}X^{*} = \begin{bmatrix} 1 & 0.99215 \\ 0.99215 & 1 \end{bmatrix}; \qquad (X^{*'}X^{*})^{-1} = \begin{bmatrix} 63.94 & -63.44 \\ -63.44 & 63.94 \end{bmatrix}$$

- ✓ Notice how the variances of the coefficients were inflated due to collinearity.

Recall

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

and

$$\text{Conditional number of } \boldsymbol{A} = \frac{\lambda_{\max}(\boldsymbol{A})}{\lambda_{\min}(\boldsymbol{A})}$$

If $\boldsymbol{X}'\boldsymbol{X}$ is nearly singular, the LSE of $\boldsymbol{\beta}$ has a poor performance.

Example 3.9: Polynomial Regression Model

When k is large,

$$y = \beta_0 + \beta_1 x + \ldots + \beta_k x^k + \varepsilon$$

and suppose that x is taken 1, 2, …, 10.

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 4 & \cdots & 2^k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 10 & 100 & \cdots & 10^k \end{pmatrix}$$

The conditional number of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is

k	1	2	3	4	5
Cond. Num	187.11	40847	11583988	4398349265	2.15170E+12

How to solve the multicollinearity problem?

- Use the ***Principle Component Regression***

- ***Ridge Regression***

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

$$\text{Ridge Regression}: \quad \hat{\boldsymbol{\beta}}_r = (\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y} \qquad \lambda \in [0, \infty)$$

- Use of ***Orthogonal Polynomials*** for polynomial regression models.

- ***Selection of Variables***

3.8 Categorical or Indicator Variables

(Sec. 3.10)

1. ONE-FACTOR EXPERIMENT

Level	1	2	$\cdots$	q
	y_{11}	y_{21}	$\cdots$	y_{q1}
	$\vdots$	$\vdots$		$\vdots$
	y_{1n_1}	y_{2n_2}	$\cdots$	y_{qn_q}
	$\bar{y}_1$	$\bar{y}_2$	$\cdots$	$\bar{y}_q$

$$n = n_1 + \cdots + n_q$$

Model: $y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

$\varepsilon \sim N(0, \sigma^2)$ $\qquad \alpha_1 + \ldots + \alpha_q = 0.$

α_i : the main effect of level i ;

μ : the overall mean of y ;

μ_i : the mean at level i ;

y_{ij} : the jth response at level i ;

ε_{ij} : error at the jth experiment at level i .

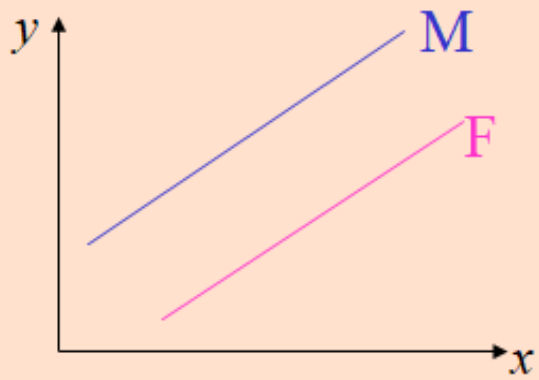
Model: $y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

$$
\begin{array}{c}
y_{ij} \\
\begin{pmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{q1} \\ \vdots \\ y_{qn_q} \end{pmatrix}
\end{array}
=
\begin{array}{c}
\begin{array}{ccccc} \mu & \alpha_1 & \alpha_2 & \cdots & \alpha_{q-1} \end{array} \\
\begin{pmatrix}
1 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & 0 & \cdots & 0 \\
1 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & -1 & -1 & \cdots & -1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & -1 & -1 & \cdots & -1
\end{pmatrix}
\end{array}
\begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_{q-1} \end{pmatrix}
+
\begin{array}{c}
\varepsilon_{ij} \\
\begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \vdots \\ \varepsilon_{q1} \\ \vdots \\ \varepsilon_{qn_q} \end{pmatrix}
\end{array}
$$

$$
\boxed{
\begin{aligned}
&\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \\
&\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
&\hat{\mu} = \bar{\bar{y}} \\
&\hat{\alpha}_i = \bar{y}_i - \bar{\bar{y}}
\end{aligned}
}
$$

2. SINGLE CATEGORICAL VARIABLE WITH TWO CATEGORIES

Consider a single quantitative regressor variable x, in a situation where there are two categories, for example, gender (Male or Female).



Define a categorical variable z as:

$$z = \begin{cases} 0, & \text{if in the first category;} \\ 1, & \text{if in the second category.} \end{cases}$$

The model becomes:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_{n_1} & 0 \\ 1 & x_{n_1+1} & 1 \\ \vdots & \vdots & \vdots \\ 1 & x_{n_1+n_2} & 1 \end{pmatrix} \begin{array}{l} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{Category 1} \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \text{Category 2} \end{array}$$

As a result, for the first category:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1,\ldots,n_1$$

and for the second category:

$$y_i = (\beta_0 + \beta_2) + \beta_1 x_i + \varepsilon_i, \quad i = n_1+1,\ldots,n_1+n_2$$

3. ONE CATEGORICAL VARIABLE WITH MULTIPLE LEVELS

z have l levels: $z_1,\ldots,z_l$, say.

We define l - 1 dummy variables, and the model becomes:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \gamma_1 z_1 + \ldots + \gamma_{l-1} z_{l-1} + \varepsilon$$

$$
\boldsymbol{X} = \begin{pmatrix}
1 & x_{11} & \cdots & x_{1k} & 1 & 0 & \cdots & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \cdots & \cdots & \vdots \\
1 & x_{n_1 1} & \cdots & x_{n_1 k} & 1 & 0 & \cdots & \cdots & 0 \\
1 & x_{n_1+1,1} & \cdots & x_{n_1+1,k} & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{n_1+n_2,1} & \cdots & x_{n_1+n_2,k} & 0 & 1 & 0 & \cdots & 0 \\
\vdots & & \vdots & & & & \vdots & & \\
1 & \vdots & \vdots & \vdots & 0 & \cdots & \cdots & 0 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & & & \vdots & \vdots \\
1 & \vdots & \vdots & \vdots & 0 & \cdots & \cdots & 0 & 1 \\
1 & \vdots & \vdots & \vdots & 0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \ddots & & \vdots \\
1 & \vdots & \vdots & \vdots & 0 & \cdots & \cdots & \cdots & 0
\end{pmatrix}
\begin{matrix}
\left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} z_1 \\
\left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} z_2 \\
\\
\left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} z_{l\text{-}1} \\
\left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} z_l
\end{matrix}
\qquad
\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \\ \gamma_1 \\ \vdots \\ \gamma_{l-1} \end{pmatrix}
$$

z have l levels: $z_1, \ldots, z_l$, say.

We define l - 1 dummy variables, and the model becomes:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \gamma_1 z_1 + \ldots + \gamma_{l-1} z_{l-1} + \varepsilon$$

As a result, we have the following results:

$$y = \begin{cases} (\beta_0 + \gamma_1) + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon & \text{for Category 1} \\ (\beta_0 + \gamma_2) + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon & \text{for Category 2} \\ \vdots & \vdots \\ (\beta_0 + \gamma_{l-1}) + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon & \text{for Category } l-1 \\ (\beta_0) + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon & \text{for Category } l \end{cases}$$

4. EQUALITY OF REGRESSOR EFFECTS ACROSS CLASSES

For model discussed in case 2, we have:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i,$$

$$i = 1, \ldots, n_1$$

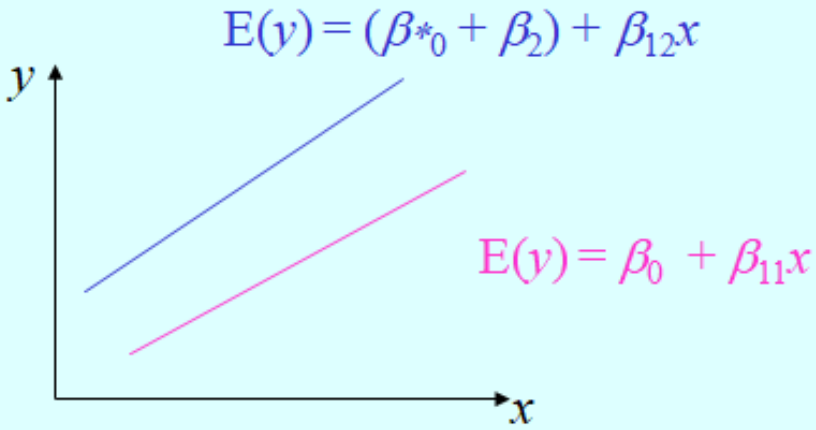
$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_{n_1} & 0 \\ 1 & x_{n_1+1} & 1 \\ \vdots & \vdots & \vdots \\ 1 & x_{n_1+n_2} & 1 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\ \vdots \\ 1\end{matrix}}\right\} \text{Category 1} \\ \left.\vphantom{\begin{matrix}1\\ \vdots \\ 1\end{matrix}}\right\} \text{Category 2} \end{matrix}$$

$$E(y_i) = \begin{cases} \beta_0 + \beta_{11} x_i & i = 1, \cdots, n_1 \\ \beta_0^* + \beta_{12} x_i + \beta_2 z_i & i = n_1 + 1, \cdots, n_1 + n_2 \end{cases}$$

$$E(y_i) = \begin{cases} \beta_0 + \beta_{11} x_i & i = 1, \cdots, n_1 \\ \beta_0^* + \beta_{12} x_i + \beta_2 z_i & i = n_1 + 1, \cdots, n_1 + n_2 \end{cases}$$

For testing additivity: (parallel or not)

$$\mathbf{H_0}: \beta_{11} = \beta_{12} \quad \text{VS} \quad \mathbf{H_a}: \beta_{11} \neq \beta_{12}$$



Parallel $\Rightarrow$ Accept

Otherwise $\Rightarrow$ Reject

Another Approach: Testing the interaction between x and z.

Consider the model:

$$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz + \varepsilon$$

This leads to:

$$E(y) = \begin{cases} \beta_0 + \beta_1 x & Category\,1 \\ \beta_0 + \beta_2 + (\beta_1 + \beta_3)x & Category\,2 \end{cases}$$

If $\beta_3 = 0$, two models have the same slope. Therefore, the test additivity is equivalent to test:

$$\mathbf{H_0}: \beta_3 = 0 \quad \text{VS} \quad \mathbf{H_a}: \beta_3 \neq 0$$