3. Linear Models

- Unified approach
  - regression analysis
  - analysis of variance
  - balanced factorial experiments
- Extensions
  - analysis of unbalanced experiments
  - models with correlated responses
    - split plot designs
    - repeated measures
  - models with fixed and random components (mixed models)

Regression Analysis

Example 3.1: Yield of a chemical process

Yield (%) Y	Temperature (^{o}F) X_1	Time (hr) X_2
77	160	1
82	165	3
84	165	2
89	170	1
94	175	2

Simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$$

$i = 1, 2, 3, 4, 5$

Matrix formulation:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \epsilon_1 \\ \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \epsilon_2 \\ \beta_0 + \beta_1 X_{13} + \beta_2 X_{23} + \epsilon_3 \\ \beta_0 + \beta_1 X_{14} + \beta_2 X_{24} + \epsilon_4 \\ \beta_0 + \beta_1 X_{15} + \beta_2 X_{25} + \epsilon_5 \end{bmatrix}$$

$$= \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} \\ \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} \\ \beta_0 + \beta_1 X_{13} + \beta_2 X_{23} \\ \beta_0 + \beta_1 X_{14} + \beta_2 X_{24} \\ \beta_0 + \beta_1 X_{15} + \beta_2 X_{25} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & X_{11} & X_{21} \\ 1 & X_{12} & X_{22} \\ 1 & X_{13} & X_{23} \\ 1 & X_{14} & X_{24} \\ 1 & X_{15} & X_{25} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \end{bmatrix}$$

Analysis of Variance (ANOVA)

Source of Variation	d.f.	Sums of Squares	Mean Squares
Model	2	$\sum_{i=1}^{5}(\hat{Y}_i - \bar{Y}_.)^2$	$\frac{1}{2} SS_{\text{model}}$
Error	2	$\sum_{i=1}^{5}(Y_i - \hat{Y}_i)^2$	$\frac{1}{2} SS_{\text{error}}$
C. total	4	$\sum_{i=1}^{5}(Y_i - \bar{Y}_.)^2$	

where

$$\bar{Y}_. = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

n = total number of observations

Example 3.2. Blood coagulation times (in seconds) for blood samples from six different rats. Each rat was fed one of three diets.

Diet 1	Diet 2	Diet 3
$Y_{11} = 62$	$Y_{21} = 71$	$Y_{31} = 72$
$Y_{12} = 60$		$Y_{32} = 68$
		$Y_{33} = 67$

"Means" model

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

observed time for the j-th rat fed the i-th diet

mean time for rats given the i-th diet

random error with $E(\epsilon_{ij}) = 0$

You can express this model as

$$\underset{\mathbf{Y}}{\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix}} = \underset{X}{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}} \underset{\beta}{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}} + \underset{\epsilon}{\begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}}$$

Assuming that $E(\epsilon_{ij}) = 0$ for all (i, j), this is a linear model with

$$E(\mathbf{Y}) = X\beta \quad \text{and} \quad Var(\mathbf{Y}) = \Sigma$$

An "effects" model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

This can be expressed as

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

or

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

This is a linear model with

$$E(\mathbf{Y}) = X\beta \quad \text{and} \quad Var(\mathbf{Y}) = \Sigma$$

You could add the assumptions

- independent errors
- homogeneous variance, i.e.,

$$Var(\epsilon_{ij}) = \sigma^2$$

to obtain a linear model

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

with

$$E(\mathbf{Y}) = X\beta$$
$$Var(\mathbf{Y}) = Var(\epsilon) = \sigma^2 I$$

Analysis of Variance (ANOVA)

Source of Variation	d.f.	Sums of Squares	Mean Squares
Diets	3 - 1 = 2	$\sum_{i=1}^{3} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$\frac{1}{2} SS_{\text{diets}}$
Error	$\sum_{i=1}^{3}(n_i - 1) = 3$	$\sum_{i=1}^{3} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$	$\frac{1}{3} SS_{\text{error}}$
C. total	$\sum_{i=1}^{3}(n_i - 1) = 5$	$\sum_{i=1}^{3} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$	

where

$$n_i = \text{number of rats fed the i-th diet}$$

$$\bar{Y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$$

$$\bar{Y}_{..} = \frac{1}{n_.} \sum_{i=1}^{3} \sum_{j=1}^{n_i} Y_{ij}$$

$$n_. = \sum_{i=1}^{3} n_i$$

$$= \text{total number of observations}$$

Example 3.3. A 2×2 factorial experiment

Experimental units: 8 plots with 5 trees per plot.

Factor 1: Variety (A or B)

Factor 2: Fungicide use (new or old)

Response: Percentage of apples with spots

Percentage of apples with spots	Variety	Fungicide use
$Y_{111} = 4.6$	A	new
$Y_{112} = 7.4$	A	new
$Y_{121} = 18.3$	A	old
$Y_{122} = 15.7$	A	old
$Y_{211} = 9.8$	B	new
$Y_{212} = 14.2$	B	new
$Y_{211} = 21.1$	B	old
$Y_{222} = 18.9$	B	old

Analysis of Variance (ANOVA)

Source of Variation	d.f.	Sums of Squares
Varieties	$2 - 1 = 1$	$4 \sum_{i=1}^{2} (\bar{Y}_{i..} - \bar{Y}_{...})^2$
Fungicide use	$2 - 1 = 1$	$4 \sum_{j=1}^{2} (\bar{Y}_{.j.} - \bar{Y}_{...})^2$
Variety $\times$ Fung. use interaction	$(2-1)(2-1)$ =1	$2 \sum_{i=1}^{2} \sum_{j=1}^{2} (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$
Error	$4(2-1) = 4$	$\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$
Corrected total	$8 - 1 = 7$	$\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2$

Linear model:

$$\underset{\substack{\uparrow\\\text{percent with spots}}}{Y_{ijk}} = \mu + \underset{\substack{\uparrow\\\text{variety effects}\\(i=1,2)}}{\alpha_i} + \underset{\substack{\uparrow\\\text{fung. use}\\(j=1,2)}}{\gamma_j} + \underset{\substack{\uparrow\\\text{interaction}}}{\delta_{ij}} + \underset{\substack{\uparrow\\\text{random error}}}{\epsilon_{ijk}}$$

Here we use 9 parameters

$$\beta^T = (\mu\ \alpha_1\ \alpha_2\ \gamma_1\ \gamma_2\ \delta_{11}\ \delta_{12}\ \delta_{21}\ \delta_{22})$$

to represent the 4 response means,

$$E(Y_{ijk}) = \mu_{ij}, \quad i = 1, 2, \text{ and } j = 1, 2,$$

corresponding to the 4 combinations of levels of the two factors.

Write this model in the form

$$\mathrm{Y} = X\,\beta + \epsilon$$

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{224} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \gamma_1 \\ \gamma_2 \\ \delta_{11} \\ \delta_{12} \\ \delta_{21} \\ \delta_{22} \end{bmatrix}$$

$$+ \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}$$

"Means" model

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

$$\nearrow$$

$\mu_{ij} = E(Y_{ijk}) =$ mean percentage of apples with spots

This linear model can be written in the form $\mathrm{Y} = X\,\beta + \epsilon$, that is,

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}$$

General Linear Model

Any linear model can be written as

$$\mathrm{Y} = X\beta + \epsilon$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$\uparrow$ observed responses

$\uparrow$ the elements of X are known (non-random) values

$\uparrow$ random errors are not observed

$\mathrm{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$ is a random vector with

(1) $E(\mathrm{Y}) = X\,\beta$

for some known $n \times k$ matrix X of constants and unknown $k \times 1$ parameter vector β

(2) Complete the model by specifying a probability distribution for the possible values of Y or ϵ

Sometimes we will only specify the covariance matrix

$$Var(\mathrm{Y}) = \Sigma$$

Since

$$\mathrm{Y} = X\beta + \epsilon$$

we have

$$\epsilon = \mathrm{Y} - X\beta = \mathrm{Y} - E(\mathrm{Y})$$

and

$$\begin{aligned} E(\epsilon) &= 0 \\ Var(\epsilon) &= Var(\mathrm{Y}) = \Sigma \end{aligned}$$

Gauss-Markov Model

Defn 3.1: The linear model

$$\mathrm{Y} = X\beta + \epsilon$$

is a Gauss-Markov model if

$$Var(\mathrm{Y}) = Var(\epsilon) = \sigma^2 I$$

for an unknown constant σ^2.

Notation: $\underset{\substack{\uparrow \\ \text{distributed} \\ \text{as}}}{\mathrm{Y} \sim} \ (\underset{\substack{\uparrow \\ E(\mathrm{Y})}}{X\beta} \ , \ \underset{\substack{\uparrow \\ Var(\mathrm{Y})}}{\sigma^2 I} \ )$

The distribution of Y is not completely specified.

Normal Theory Gauss-Markov Model

Defn 3.2: A normal theory Gauss-Markov model is a Gauss-Markov model in which Y (or ϵ) has a multivariate normal distribution.

$$\underset{\substack{\nearrow \\ \text{distr.} \\ \text{as}}}{\mathrm{Y} \sim} \ \underset{\substack{\uparrow \\ \text{multivar.} \\ \text{normal} \\ \text{distr.}}}{N} \ (\underset{\substack{\nwarrow \\ E(\mathrm{Y})}}{X\beta} \ , \ \underset{\substack{\nwarrow \\ Var(\mathrm{Y})}}{\sigma^2 I})$$

The additional assumption of a normal distribution is

(1) not needed for some estimation results

(2) useful in creating
- confidence intervals
- tests of hypotheses

Objectives

(i) Develop estimation procedures

- Estimate β?
- Estimate $E(\mathrm{Y}) = X\,\beta$
- Estimable functions of β.

(ii) Quantify uncertainty in estimates

- variances, standard deviations
- distributions
- confidence intervals

(iii) Analysis of Variance (ANOVA)

(iv) Tests of hypotheses

- Distributions of quadratic forms
- F-tests
- power

(v) sample size determination

Least Squares Estimation

For the linear model with

$$E(\mathrm{Y}) = X\,\beta \text{ and } Var(\mathrm{Y}) = \Sigma$$

we have

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{k1} \\ X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

and

$$\begin{aligned} Y_i &= \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i \\ &= \mathrm{X}_i^T \beta + \epsilon_i \end{aligned}$$

where $\mathrm{X}_i^T = (X_{1i}\, X_{2i} \cdots X_{ki})$ is the i-th row of the model matrix X.

OLS Estimator

Defn 3.3: For a linear model with $E(\mathrm{Y}) = X\beta$, any vector b that minimizes the sum of squared residuals

$$\begin{aligned} Q(\mathrm{b}) &= \sum_{i=1}^{n} (Y_i - \mathrm{X}_i^T \mathrm{b})^2 \\ &= (\mathrm{Y} - X\mathrm{b})^T (\mathrm{Y} - X\mathrm{b}) \end{aligned}$$

is an ordinary least squares (OLS) estimator for β.

OLS Estimating Equations

For $j = 1, 2, \ldots, k$, solve

$$0 = \frac{\partial Q(\mathrm{b})}{\partial b_j} = 2 \sum_{i=1}^{n} (Y_i - \mathrm{X}_i^T \mathrm{b})\, X_{ij}$$

These equations are expressed in matrix form as

$$\begin{aligned} 0 &= X^T(\mathrm{Y} - X\mathrm{b}) \\ &= X^T\mathrm{Y} - X^T X\mathrm{b} \end{aligned}$$

or

$$X^T X\mathrm{b} = X^T\mathrm{Y}$$

These are called the "normal" equations.

If $X_{n\times k}$ has full column rank, i.e., rank$(X) = k$, then

(i) $X^T X$ is non-singular

(ii) $(X^T X)^{-1}$ exists and is unique

Consequently,

$$(X^T X)^{-1}(X^T X)\mathrm{b} = (X^T X)^{-1} X^T \mathrm{y}$$

and

$$\mathrm{b} = (X^T X)^{-1} X^T \mathrm{Y}$$

is the unique solution to the normal equations.

If rank$(X) < k$, then

(i) there are infinitely many solutions to the normal equations

(ii) if $(X^T X)^-$ is a generalized inverse of $X^T X$, then

$$\mathrm{b} = (X^T X)^- X^T \mathrm{Y}$$

is a solution of the normal equations.

Generalized Inverse

Defn 3.4: For a given $m \times n$ matrix A, any $n \times m$ matrix G that satisfies

$$AGA = A$$

is a generalized inverse of A.

Comments

- We will often use A^- to denote a generalized inverse of A.
- There may be infinitely many generalized inverses.
- If A is an $m \times m$ nonsingular matrix, then $G = A^{-1}$ is the unique generalized inverse for A.

Example 3.5.

$$A = \begin{bmatrix} 16 & -6 & -10 \\ -6 & 21 & -15 \\ -10 & -15 & 25 \end{bmatrix}$$

with rank$(A) = 2$

A generalized inverse is

$$G = \begin{bmatrix} \frac{1}{20} & 0 & 0 \\ 0 & \frac{1}{30} & 0 \\ 0 & 0 & \frac{1}{50} \end{bmatrix}$$

Check that $AGA = A$.

Another generalized inverse is

$$G = \begin{bmatrix} \begin{bmatrix} 16 & -6 \\ -6 & 21 \end{bmatrix}^{-1} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & 0 \end{bmatrix} = \begin{bmatrix} \frac{21}{300} & \frac{6}{300} & 0 \\ \frac{6}{300} & \frac{16}{300} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Example 3.2. "means" model

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

For this model

$$X^T X = \begin{bmatrix} n_1 & 0 & 0 \\ 0 & n_2 & 0 \\ 0 & 0 & n_3 \end{bmatrix}$$

$$X^T \mathrm{Y} = \begin{bmatrix} Y_{11} + Y_{12} \\ Y_{21} \\ Y_{31} + Y_{32} + Y_{33} \end{bmatrix}$$

and the unique OLS estimator for $\beta = (\mu_1 \ \mu_2 \ \mu_3)^T$ is

$$\begin{aligned} \mathrm{b} &= (X^T X)^{-1} X^T \mathrm{Y} \\ &= \begin{bmatrix} \frac{1}{n_1} & 0 & 0 \\ 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & \frac{1}{n_3} \end{bmatrix} \begin{bmatrix} Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \end{bmatrix} \end{aligned}$$

Example 3.2. "effects" model

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \\ Y_{33} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \end{bmatrix}$$

Here

$$X^T X = \begin{bmatrix} n_. & n_1 & n_2 & n_3 \\ n_1 & n_1 & 0 & 0 \\ n_2 & 0 & n_2 & 0 \\ n_3 & 0 & 0 & n_3 \end{bmatrix}$$

$$X^T \mathrm{Y} = \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix}$$

Solution A:

$$(X^TX)^- = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{n_1} & 0 & 0 \\ 0 & 0 & \frac{1}{n_2} & 0 \\ 0 & 0 & 0 & \frac{1}{n_3} \end{bmatrix}$$

and a solution to the normal equations is

$$\begin{aligned} \mathrm{b} &= (X^TX)^- X^T \mathrm{Y} \\ &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & n_1^{-1} & 0 & 0 \\ 0 & 0 & n_2^{-1} & 0 \\ 0 & 0 & 0 & n_3^{-1} \end{bmatrix} \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \end{bmatrix} \end{aligned}$$

Solution B: Another generalized inverse for X^TX is

$$(X^TX)^- = \begin{bmatrix} \begin{bmatrix} n_. & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{bmatrix}^{-1} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & 0 \end{bmatrix}$$

Let

$$A = \begin{bmatrix} n_. & n_1 & n_2 \\ n_1 & n_1 & 0 \\ n_2 & 0 & n_2 \end{bmatrix}$$

and use result 1.4(ii) to compute

$$\begin{aligned} A^{-1} &= \frac{1}{|A|} \begin{bmatrix} |A_{11}| & -|A_{21}| & |A_{31}| \\ -|A_{12}| & |A_{22}| & -|A_{32}| \\ |A_{13}| & -|A_{23}| & |A_{33}| \end{bmatrix} \\ &= \frac{1}{n_1 n_2 n_3} \begin{bmatrix} n_1n_2 & -n_1n_2 & -n_1n_2 \\ -n_1n_2 & n_2(n_1+n_3) & n_1n_2 \\ -n_1n_2 & n_1n_2 & n_1(n_2+n_3) \end{bmatrix} \end{aligned}$$

Then

$$A^{-1} = \frac{1}{n_3} \begin{bmatrix} 1 & -1 & -1 \\ -1 & \frac{n_1+n_3}{n_1} & 1 \\ -1 & 1 & \frac{n_2+n_3}{n_2} \end{bmatrix}$$

and a solution to the normal equations is

$$\begin{aligned} \mathrm{b} &= (X^TX)^- X^T \mathrm{Y} \\ &= \frac{1}{n_3} \begin{bmatrix} 1 & -1 & -1 & 0 \\ -1 & \frac{n_1+n_3}{n_1} & 1 & 0 \\ -1 & 1 & \frac{n_2+n_3}{n_2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} \\ &= \frac{1}{n_3} \begin{bmatrix} Y_{..} - Y_{1.} - Y_{2.} \\ -Y_{..} + (\frac{n_1+n_3}{n_1})Y_{1.} + Y_{2.} \\ -Y_{..} + Y_{1.} + (\frac{n_2+n_3}{n_2})Y_{2.} \\ 0 \end{bmatrix} \end{aligned}$$

Then

$$\mathrm{b} = \begin{bmatrix} \bar{Y}_{3.} \\ \bar{Y}_{1.} - \bar{Y}_{3.} \\ \bar{Y}_{2.} - \bar{Y}_{3.} \\ 0 \end{bmatrix}$$

This is the OLS estimator for

$$\beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

reported by PROC GLM in the SAS package, but it is not the only possible solution to the normal equations.

<u>Solution C</u>: Another generalized inverse for X^TX is

$$(X^TX)^- = \frac{1}{n_1n_2n_3}\begin{bmatrix} n_2n_3 & 0 & -n_2n_3 & -n_2n_3 \\ 0 & 0 & 0 & 0 \\ -n_2n_3 & 0 & n_3(n_1+n_2) & n_2n_3 \\ -n_2n_3 & 0 & n_2n_3 & n_2(n_1+n_3) \end{bmatrix}$$

The corresponding solution to the normal equations is

$$\begin{aligned} \mathrm{b} &= (X^TX)^- X^T \mathrm{Y} \\ &= \begin{bmatrix} \bar{Y}_{1.} \\ 0 \\ \bar{Y}_{2.} - \bar{Y}_{1.} \\ \bar{Y}_{3.} - \bar{Y}_{1.} \end{bmatrix} \end{aligned}$$

<u>Solution D</u>: Another generalized inverse for X^TX is

$$(X^TX)^- = \begin{bmatrix} \frac{2}{n} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & \frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n} & 0 & \frac{1}{n_2} & 0 \\ -\frac{1}{n} & 0 & 0 & \frac{1}{n_3} \end{bmatrix}$$

The corresponding solution to the normal equations is

$$\begin{aligned} \mathrm{b} &= (X^TX)^- X^T \mathrm{Y} \\ &= \begin{bmatrix} \frac{2}{n} & -\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} \\ -\frac{1}{n} & \frac{1}{n_1} & 0 & 0 \\ -\frac{1}{n} & 0 & \frac{1}{n_2} & 0 \\ -\frac{1}{n} & 0 & 0 & \frac{1}{n_3} \end{bmatrix} \begin{bmatrix} Y_{..} \\ Y_{1.} \\ Y_{2.} \\ Y_{3.} \end{bmatrix} = \begin{bmatrix} \bar{Y}_{..} \\ \bar{Y}_{1.} - \bar{Y}_{..} \\ \bar{Y}_{2.} - \bar{Y}_{..} \\ \bar{Y}_{3.} - \bar{Y}_{..} \end{bmatrix} \end{aligned}$$

Evaluating Generalized Inverses

<u>Algorithm 3.1:</u>

(i) Find any $r \times r$ nonsingular submatrix of A where r=rank(A). Call this matrix W.

(ii) Invert and transpose W, ie., compute $(W^{-1})^T$.

(iii) Replace each element of W in A with the corresponding element of $(W^{-1})^T$

(iv) Replace all other elements in A with zeros.

(v) Transpose the resulting matrix to obtain G, a generalized inverse for A.

<u>Example 3.6.</u>

$$A = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}$$

$$W = \begin{bmatrix} 1 & 5 \\ 1 & 3 \end{bmatrix}$$

$$(W^{-1})^T = \begin{bmatrix} -3/2 & 1/2 \\ 5/2 & -1/2 \end{bmatrix}$$

$$\begin{aligned} G &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -\frac{3}{2} & \frac{1}{2} & 0 \\ 0 & \frac{5}{2} & -\frac{1}{2} & 0 \end{bmatrix}^T \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\frac{3}{2} & \frac{5}{2} \\ 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

Check: $AGA = A$

Another solution

$$A = \begin{bmatrix} 4 & 1 & 2 & 0 \\ 1 & 1 & 5 & 15 \\ 3 & 1 & 3 & 5 \end{bmatrix}$$

$$W = \begin{bmatrix} 4 & 0 \\ 3 & 5 \end{bmatrix}$$

$$(W^{-1})^T = \begin{bmatrix} -\frac{5}{20} & -\frac{3}{20} \\ \frac{0}{20} & \frac{4}{20} \end{bmatrix}$$

$$G = \begin{bmatrix} \frac{5}{20} & 0 & 0 & -\frac{3}{20} \\ 0 & 0 & 0 & 0 \\ \frac{0}{20} & 0 & 0 & \frac{4}{20} \end{bmatrix}^T$$

$$= \begin{bmatrix} \frac{5}{20} & 0 & \frac{0}{20} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{3}{20} & 0 & \frac{4}{20} \end{bmatrix}$$

Algorithm 3.2

For any $m \times n$ matrix A with rank$(A) = r$,

(i) compute a singular value decomposition of A (see result 1.14) to obtain

$$PAQ = \begin{bmatrix} D_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}$$

where
P is an $m \times m$ orthogonal matrix
Q is an $n \times n$ orthogonal matrix
D is an $r \times r$ matrix of singular values

(ii) $G = Q \begin{bmatrix} D^{-1} & F_1 \\ F_2 & F_3 \end{bmatrix} P$ is a generalized inverse for A for any choice of F_1, F_2, F_3.

Proof: Check if $AGA = A$.

$$AGA = P^T \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \underbrace{Q^T Q}_{\substack{\uparrow \\ n\times n \\ \text{identity} \\ \text{matrix}}} \begin{bmatrix} D^{-1} & F_1 \\ F_2 & F_3 \end{bmatrix} \underbrace{P\, P^T}_{\substack{\uparrow \\ m\times m \\ \text{identity} \\ \text{matrix}}} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

$$= P^T \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} D^{-1} & F_1 \\ F_2 & F_3 \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

$$= P^T \begin{bmatrix} I & DF_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

$$= P^T \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^T$$

$$= A$$

Moore-Penrose Inverse

Defn 3.5: For any matrix A there is a unique matrix M, called the Moore-Penrose inverse, that satisfies

(i) $AMA = A$

(ii) $MAM = M$

(iii) AM is symmetric

(iv) MA is symmetric

Result 3.1

$$M = Q \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} P$$

is the Moore-Penrose inverse of A, where

$$PAQ = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

is a singular value decomposition of A.

Properties of generalized inverses of X^TX

Normal equations: $(X^TX)\mathrm{b} = X^T\mathrm{Y}$

Result 3.3 If G is a generalized inverse of X^TX, then

(i) G^T is a generalized inverse of X^TX.

(ii) $XGX^TX = X$, i.e., GX^T is a generalized inverse of X.

(iii) XGX^T is invariant with respect to the choice of G.

(iv) XGX^T is symmetric.

Proof:

(i) Since G is a generalized inverse of (X^TX), $(X^TX)G(X^TX) = X^TX$.

Taking the transpose of both sides

$$\begin{aligned} \left[X^TX\right]^T &= \left[(X^TX)G(X^TX)\right]^T \\ &= (X^TX)^T G^T (X^TX)^T \end{aligned}$$

But $(X^TX)^T = X^T(X^T)^T = X^TX$,

hence $(X^TX)G^T(X^TX) = (X^TX)$

(ii) From (i) $\underbrace{(X^TX)G^T}_{\text{Call this } B}(X^TX) = (X^TX)$

Then

$$\begin{aligned} 0 &= BX^TX - X^TX \\ &= (BX^TX - X^TX)(B^T - I) \\ &= BX^TXB^T - X^TXB^T - BX^TX - X^TX \\ &= (BX^T - X^T)(BX^T - X^T)^T \end{aligned}$$

Hence, $0 = BX^T - X^T$

$$\begin{aligned} &\Rightarrow \quad BX^T = X^T \\ &\Rightarrow \quad X^TXG^TX^T = X^T \end{aligned}$$

Taking the transpose

$$\begin{aligned} X &= (X^TXG^TX^T)^T \\ &= X\underline{GX^T}X \end{aligned}$$

Hence, GX^T is a generalized inverse for X.

(iii) Suppose F and G are generalized inverses for X^TX. Then, from (ii)

$$XGX^TX = X$$

and

$$XFX^TX = X$$

It follows that

$$
\begin{aligned}
0 &= X - X \\
&= (XGX^TX - XFX^TX) \\
&= (XGX^TX - XFX^TX)(G^TX^T - F^TX^T) \\
&= (XGX^T - XFX^T)X(G^TX^T - F^TX^T) \\
&= (XGX^T - XFX^T)(XG^TX^T - XF^TX^T) \\
&= (XGX^T - XFX^T)(XGX^T - XFX^T)^T
\end{aligned}
$$

Since the (i,i) diagonal element of the result of multiplying a matrix by its transpose is the sum of the squared entries in the *i-th* row of the matrix, the diagonal elements of the product are all zero only if all entries are zero in every row of the matrix. Consequently,

$$(XGX^T - XFX^T) = 0$$

(iv) For any generalized inverse G,

$$T = GX^TXG^T$$

is a symmetric generalized inverse. Then

$$XTX^T$$

is symmetric and from (iii),

$$XGX^T = XTX^T.$$

Estimation of the Mean Vector

$$E(\mathrm{Y}) = X\beta$$

For any solution to the normal equations, say

$$\mathrm{b} = (X^TX)^-X^T\mathrm{Y} \ ,$$

the OLS estimator for $E(\mathrm{Y}) = X\beta$ is

$$
\begin{aligned}
\hat{\mathrm{Y}} = X\mathrm{b} &= X(X^TX)^-X^T\mathrm{Y} \\
&= P_X\mathrm{Y}
\end{aligned}
$$

- The matrix $P_X = X(X^TX)^-X^T$ is called an "orthogonal projection matrix".
- $\hat{\mathrm{Y}} = P_X\mathrm{Y}$ is the projection of Y onto the space spanned by the columns of X.

<u>Result 3.4</u> Properties of a projection matrix

$$P_X = X(X^TX)^-X^T$$

(i) P_X is invariant to the choice of $(X^TX)^-$. For any solution

$$\mathrm{b} = (X^TX)^-X^T\mathrm{Y}$$

to the normal equations

$$\hat{\mathrm{Y}} = X\mathrm{b} = P_X\mathrm{Y}$$

is the same.
(from Result 3.3 (iii))

(ii) P_X is symmetric (from Result 3.3 (iv))

(iii) P_X is idempodent $(P_X P_X = P_X)$

(iv) $P_X X = X$ (from Result 3.3 (ii))

(v) Partition X as

$$X = [X_1|X_2|\cdots|X_k],$$

then $P_X X_j = X_j$

<u>Residuals:</u>

$$e_i = Y_i - \hat{Y}_i \quad i = 1, \ldots, n$$

The vector of residuals is

$$\begin{aligned} \mathrm{e} &= \mathrm{Y} - \hat{\mathrm{Y}} \\ &= \mathrm{Y} - X\mathrm{b} \\ &= \mathrm{Y} - P_X \mathrm{Y} \\ &= (I - P_X)\mathrm{Y} \end{aligned}$$

<u>Comment:</u> $I - P_X$ is a projection matrix that projects Y onto the space orthogonal to the space spanned by the columns of X.

<u>Result 3.5</u> Properties of $I - P_X$

(i) $I - P_X$ is symmetric

(ii) $I - P_X$ is idempodent

$$(I - P_X)(I - P_X) = I - P_X$$

(iii)

$$\begin{aligned} (I - P_X)P_X &= P_X - P_X P_X \\ &= P_X - P_X = 0 \end{aligned}$$

(iv)

$$\begin{aligned} (I - P_X)X &= X - P_X X \\ &= X - X = 0 \end{aligned}$$

(v) Partition X as $[X_1|X_2|\cdots|X_k]$ then

$$(I - P_X)\mathrm{X}_j = 0$$

(vi) Residuals are invariant with respect to the choice of $(X^T X)^-$, so

$$\mathrm{e} - \mathrm{Y} - X\mathrm{b} = (I - P_X)\mathrm{Y}$$

is the same for any solution

$$\mathrm{b} = (X^T X)^- X^T \mathrm{Y}$$

to the normal equations