(1) 4/15/16 (F) : make-up. (tentative) – still look for a room

2 – 3:45pm

(2) – Examples of the General Linear Model.

– Simple Regression : Example

(3) – One-Way ANOVA

– Chapter 2 : LSE & Linear Algebra.

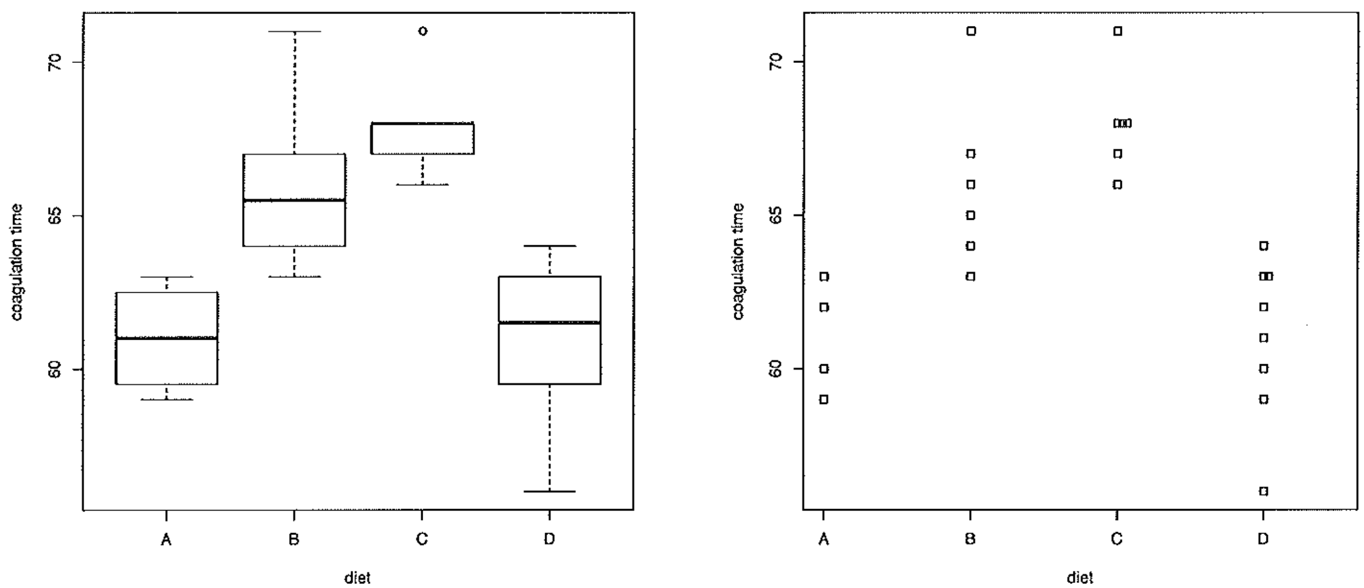
One-Way ANOVA

in Seconds

Dataset comes from a study of blood coagulation times. 24 animals were randomly assigned to four different diets and the samples were taken in a random order (taken from Linear Models with R, page 182)

```
> rm(list=ls(all=TRUE))
> library(faraway)
> data(coagulation)
>
> plot(coag~diet, coagulation, ylab="coagulation time")
> with(coagulation, stripchart(coag ~ diet, vertical=TRUE, metho
```

One-Way ANOVA



One-Way ANOVA – parameterization ~~3~~ 1

```
> options(contrasts=c("contr.sum", "contr.poly"))
> g2 <- lm(coag ~ diet, coagulation)
> summary(g2)
Call:
lm(formula = coag ~ diet, data = coagulation)
Residuals:
   Min     1Q Median     3Q    Max
 -5.00  -1.25   0.00   1.25   5.00
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.0000     0.4979 128.537  < 2e-16 ***
diet1        -3.0000     0.9736  -3.081 0.005889 **
diet2         2.0000     0.8453   2.366 0.028195 *
diet3         4.0000     0.8453   4.732 0.000128 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 2.366 on 20 degrees of freedom
Multiple R-squared:  0.6706, Adjusted R-squared:  0.6212
F-statistic: 13.57 on 3 and 20 DF,  p-value: 4.658e-05
```

$\hat{\mu}$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, $\hat{\alpha}_3$

$\hat{\alpha}_4 = -\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\alpha}_3 = +3 - 2 - 4 = -3$

$y_{ij} = \mu + \alpha_i + e_{ij}$

$i = 1, \ldots, 4$

$\Sigma \alpha_i = 0$

$\hat{\mu}$ = estimated overall mean = 64

the estimated mean response for A = 64 − 3 = 61

One-Way ANOVA – parameterization ~~3~~ 1

$\mu \quad \alpha_1 \quad \alpha_2 \quad \alpha_3$

$\alpha_4 = -\alpha_1 - \alpha_2 - \alpha_3$

```
> model.matrix(g2)
   (Intercept) diet1 diet2 diet3
1            1     1     0     0
2            1     1     0     0
3            1     1     0     0
4            1     1     0     0
5            1     0     1     0
6            1     0     1     0
7            1     0     1     0
8            1     0     1     0
9            1     0     1     0
10           1     0     1     0
11           1     0     0     1
12           1     0     0     1
13           1     0     0     1
14           1     0     0     1
15           1     0     0     1
16           1     0     0     1
17           1    -1    -1    -1
18           1    -1    -1    -1
19           1    -1    -1    -1
20           1    -1    -1    -1
21           1    -1    -1    -1
22           1    -1    -1    -1
23           1    -1    -1    -1
24           1    -1    -1    -1
attr(,"assign")
[1] 0 1 1 1
attr(,"contrasts")
attr(,"contrasts")$diet
[1] "contr.sum"
```

One-Way ANOVA – parameterization ~~1~~ 2

$y_{ij} = \mu + \alpha_i + e_{ij}$

$\alpha_1 = 0$

$E(y_{1j}) = E(\mu + \alpha_1 + e_{1j})$

$= \mu + 0 + 0$

$= \mu$

```
> g <- lm(coag ~ diet, coagulation)
> summary(g)
Call:
lm(formula = coag ~ diet, data = coagulation)

Residuals:
   Min     1Q Median     3Q    Max
 -5.00  -1.25   0.00   1.25   5.00

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.100e+01  1.183e+00  51.554  < 2e-16 ***
dietB       5.000e+00  1.528e+00   3.273 0.003803 **
dietC       7.000e+00  1.528e+00   4.583 0.000181 ***
dietD       2.991e-15  1.449e+00   0.000 1.000000
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.366 on 20 degrees of freedom
Multiple R-squared:  0.6706, Adjusted R-squared:  0.6212
F-statistic: 13.57 on 3 and 20 DF,  p-value: 4.658e-05
```

One-Way ANOVA – parameterization ~~1~~ 2

μ α_2 α_3 α_4

```
> model.matrix(g)
   (Intercept) dietB dietC dietD
1            1     0     0     0
2            1     0     0     0
3            1     0     0     0
4            1     0     0     0
5            1     1     0     0
6            1     1     0     0
7            1     1     0     0
8            1     1     0     0
9            1     1     0     0
10           1     1     0     0
11           1     0     1     0
12           1     0     1     0
13           1     0     1     0
14           1     0     1     0
15           1     0     1     0
16           1     0     1     0
17           1     0     0     1
18           1     0     0     1
19           1     0     0     1
20           1     0     0     1
21           1     0     0     1
22           1     0     0     1
23           1     0     0     1
24           1     0     0     1
attr(,"assign")
[1] 0 1 1 1
attr(,"contrasts")
attr(,"contrasts")$diet
[1] "contr.treatment"
```

$Y = X\beta + e$

One-Way ANOVA – parameterization $\cancel{2}$ 3

$y_{ij} = \theta_i + e_{ij}$

```
> g1 <- lm(coag ~ diet - 1, coagulation)
> summary(g1)
Call:
lm(formula = coag ~ diet - 1, data = coagulation)

Residuals:
   Min     1Q Median     3Q    Max
 -5.00  -1.25   0.00   1.25   5.00

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
dietA  61.0000     1.1832   51.55   <2e-16 ***
dietB  66.0000     0.9661   68.32   <2e-16 ***
dietC  68.0000     0.9661   70.39   <2e-16 ***
dietD  61.0000     0.8367   72.91   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.366 on 20 degrees of freedom
Multiple R-squared:  0.9989, Adjusted R-squared:  0.9986
F-statistic:  4399 on 4 and 20 DF,  p-value: < 2.2e-16
```

One-Way ANOVA – parameterization ~~2~~ 3

θ_1 θ_2 θ_3 θ_4

```
> model.matrix(g1)
   dietA dietB dietC dietD
1      1     0     0     0
2      1     0     0     0
3      1     0     0     0
4      1     0     0     0
5      0     1     0     0
6      0     1     0     0
7      0     1     0     0
8      0     1     0     0
9      0     1     0     0
10     0     1     0     0
11     0     0     1     0
12     0     0     1     0
13     0     0     1     0
14     0     0     1     0
15     0     0     1     0
16     0     0     1     0
17     0     0     0     1
18     0     0     0     1
19     0     0     0     1
20     0     0     0     1
21     0     0     0     1
22     0     0     0     1
23     0     0     0     1
24     0     0     0     1
attr(,"assign")
[1] 1 1 1 1
attr(,"contrasts")
attr(,"contrasts")$diet
[1] "contr.treatment"
```

AMS 256
Chapter 2: The Linear Least Squares Problem

Spring 2016

† Notation

1. Vectors: boldface lowercase.
   e.g $\mathbf{a} = [a_1, \dots, a_n]^T \in \mathbb{R}^n$, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, $\mathbf{1}$, $\mathbf{0}$

2. Matrices: boldface uppercase.
   e.g. $\underbrace{\mathbf{A}}_{\text{n} \times \text{m}}$

   $\mathbf{A}_{.j} \in \mathbb{R}^n$: the j^{th} column

3. Transpose: $\mathbf{a}^T$, $\mathbf{A}^T$

† Recall that the form of linear models is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

- $\mathbf{y}$: $n \times 1$ vector of observations (random)
- $\mathbf{X}$: $n \times p$ matrix of known constants (*design* matrix)
- $\boldsymbol{\beta}$: $p \times 1$ vector of unobservable parameters
- $\mathbf{e}$: $n \times 1$ vector of unobservable random errors

† Goal: Inference about $\boldsymbol{\beta}$ and σ^2

- point estimates

  √ Least squares estimates (**Chapters 2 & 3**):
  ⇝ Find the best approximation of $\mathbf{y}$ as a linear function of columns of $\mathbf{X}$

  √ Best linear unbiased estimator (**Chapter 4**):
  ⇝ Assume that $\mathrm{E}(\mathbf{e}) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{e}) = \sigma^2 I$ where σ^2 is some unknown parameter (will be generalized to $\mathrm{Cov}(\mathbf{e}) = \sigma^2\mathbf{V}$ where $\mathbf{V}$ is a known positive definite matrix)

  √ MLE (**Chapter 5**)
  ⇝ Assume a distribution for the $\mathbf{e}$

† Goal: Inference about $\boldsymbol{\beta}$ and σ^2 (Contd)

- tests
- confidence regions
- checking the assumptions, model selection

† Applications:

- Regression Analysis: $\mathbf{X}^T\mathbf{X}$ is nonsingular (or close to singular)
- Analysis of Variance: $\mathbf{X}^T\mathbf{X}$ is singular

$*$ Example 1: Simple linear regression
Consider the model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \qquad i = 1, \ldots, n.$$

Write it in the form of linear models.

$$Y = X\beta + e$$

$$\Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

∗ Example 2: Balanced One-Way ANOVA
Consider the model

$$y_{ij} = \mu + \alpha_i + e_{ij}, \qquad i = 1, \ldots, a \text{ and } j = 1, \ldots, n.$$

Write it in the form of linear models.

$$Y = X\beta + e$$

$$\Rightarrow \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \\ \vdots \\ y_{an} \end{bmatrix} = \underbrace{\begin{bmatrix} 1_n & 1_n & 0_n & \cdots & 0_n \\ 1_n & 0_n & 1_n & \cdots & 0_n \\ \vdots & \vdots & \vdots & & \vdots \\ 1_n & 0_n & 0_n & & 1_n \end{bmatrix}}_{an \times (a+1)} \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{an} \end{bmatrix}$$

$*$ Road map:

- Discuss least squares estimates (LSE) and its uniqueness
- Discuss the geometry associated with LSE

† Def

- (Length of a vector: Euclidean norm) The *length* of a vector $\mathbf{x}$ is $||\mathbf{x}|| \equiv \sqrt{\mathbf{x}^T\mathbf{x}} = (\sum x_i^2)^{1/2}$

- (Distance between two vectors) The *distance* of two vectors $\mathbf{x}$ and $\mathbf{y}$ is a length of their difference, i.e. $||\mathbf{x} - \mathbf{y}||$.

† Least squares estimate: Find the closest (best) approximation $\mathbf{X}\beta$ to the observed vector $\mathbf{y}$ in the Euclidean manner,

$$Q(\beta) = ||\mathbf{y} - \mathbf{X}\beta||^2 = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

$$\Rightarrow \hat{\beta} = arg \min_{\beta}(\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

The Least Squares Approach-1/3

★ How to find such β? (least squares solution: $\hat{\beta}$)

$$Q(\beta) = (\mathbf{y} - \mathbf{X}\beta)^T(\mathbf{y} - \mathbf{X}\beta)$$

$$= (y^T - \beta^T X^T)(y - X\beta)$$

$$= y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta$$

⋆ Find the gradient vector

⋆ Set the gradient to zero

The Least Squares Approach-2/3

$\star$ How to find the gradient vector (Result 2.1)

Let $\mathbf{a}$ and $\mathbf{b}$ be $p \times 1$ vectors and $\mathbf{A}$ be a $p \times p$ matrix of constants. Then

- $$\frac{\partial \mathbf{b}^T\mathbf{a}}{\partial \mathbf{b}} = \frac{\partial \mathbf{a}^T\mathbf{b}}{\partial \mathbf{b}} = \mathbf{a}.$$
- $$\frac{\partial \mathbf{b}^T\mathbf{A}\mathbf{b}}{\partial \mathbf{b}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{b}.$$

Recall

$$Q(\boldsymbol{\beta}) = \mathbf{y}^T\mathbf{y} - \underline{\mathbf{y}^T\mathbf{X}}\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}.$$

$$\frac{\partial Q(\beta)}{\partial \beta} = -(y^Tx)^T - x^Ty + (x^Tx + (x^Tx)^T)\beta$$

$$= -2x^Ty + 2x^Tx\beta$$

The Least Squares Approach-3/3

$$Q(\beta) = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}\beta - \beta^T\mathbf{X}^T\mathbf{y} + \beta^T\mathbf{X}^T\mathbf{X}\beta.$$

⋆ Find the gradient vector

⋆ Set the gradient to zero

$$\cancel{2}X^TX\beta = \cancel{2}X^Ty$$
$$X^TX\beta = X^Ty$$

$$\frac{\partial Q}{\partial \beta} = -2\mathbf{X}^T\mathbf{y} + 2\mathbf{X}^T\mathbf{X}\beta = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\beta) = 0$$

$$\Rightarrow \mathbf{X}^T\mathbf{X}\beta = \mathbf{X}^T\mathbf{y} \qquad \text{Normal Equations (NEs).}$$

Then solve the NEs for least squares solutions $\hat{\beta}$!

$*$ Example 1 (contd): Simple linear regression
Consider the model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \qquad i = 1, \ldots, n.$$

Find the NEs.

$$(X^T X)\beta = X^T y$$

① $$X^T X = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix}$$

② $$X^T y = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}$$

$\Rightarrow$ NEs $$\begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}$$

$*$ Example 2 (contd): Balanced One-Way ANOVA
Consider the model

$$y_{ij} = \mu + \alpha_i + e_{ij}, \qquad i = 1, \ldots, a \text{ and } j = 1, \ldots, n.$$

Find the NEs.

NEs $\quad X^T X \beta = X^T y$

① $$X^T X = \begin{bmatrix} 1_n^T & 1_n^T & \cdots & 1_n^T \\ 1_n^T & 0_n^T & \cdots & 0_n^T \\ 0_n^T & 1_n^T & \cdots & 0_n^T \\ \vdots & & & \\ 0_n^T & 0_n^T & \cdots & 1_n^T \end{bmatrix} \begin{array}{c} \begin{array}{ccccc} \mu & \alpha_1 & \alpha_2 & \cdots & \alpha_a \end{array} \\ \begin{bmatrix} 1_n & 1_n & 0_n & \cdots & 0_n \\ 1_n & 0_n & 1_n & \cdots & 0_n \\ 1_n & 0_n & 0_n & \cdots & 0_n \\ \vdots & & & \ddots & \\ 1_n & 0_n & 0_n & \cdots & 1_n \end{bmatrix} \end{array}$$

$$= \begin{bmatrix} na & n & n & \cdots & n \\ n & n & 0 & \cdots & 0 \\ n & 0 & n & \cdots & 0 \\ \vdots & \vdots & \vdots & & \\ n & 0 & 0 & \cdots & n \end{bmatrix}_{(a+1)\times(a+1)}$$

② $$X^T y = \begin{bmatrix} 1_n^T & 1_n^T & \cdots & 1_n^T \\ 1_n^T & 0_n^T & \cdots & 0_n^T \\ \vdots & \vdots & & \vdots \\ 0_n^T & 0_n^T & \cdots & 1_n^T \end{bmatrix} \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{an} \end{bmatrix} = \begin{bmatrix} \sum_{ij} y_{ij} \\ \sum y_{1j} \\ \vdots \\ \sum y_{aj} \end{bmatrix}_{(a+1)\times 1}$$

$$\begin{bmatrix} na & n & \cdots & n \\ n & n & \cdots & 0 \\ \vdots & & & \vdots \\ n & 0 & \cdots & n \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_a \end{bmatrix} = \begin{bmatrix} \sum_{ij} y_{ij} \\ \sum y_{1j} \\ \vdots \\ \sum y_{aj} \end{bmatrix}$$

- Normal Equations (NEs)

$$\mathbf{X}^T\mathbf{X}\beta = \mathbf{X}^T\mathbf{y}.$$

$\leftrightarrow$ **Solving a system of equations, $\mathbf{A}\mathbf{x} = \mathbf{c}$.**

- If $\mathbf{A}$ is nonsingular and square, $\mathbf{A}^{-1}$ exists and and $\mathbf{A}^{-1}\mathbf{c}$ is the unique solution.

- If $\mathbf{A}$ is singular or not square, then may not have a solution or infinitely many solutions.

$\rightarrow$ **Questions**

- $\exists$ a solution?

- How to find a solution??

† Def **[Vector Space]:** A set $\mathcal{S} \subset \mathbb{R}^n$ is a vector space if for any $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in \mathbb{R}^n$ and scalar α, β, operations of vector addition and scalar multiplication are defined such that

1. $(\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{x}+(\mathbf{y}+\mathbf{z})$
2. $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$
3. There exists a vector $\mathbf{0} \in \mathcal{S}$ such that $\mathbf{x}+\mathbf{0} = \mathbf{x} = \mathbf{0}+\mathbf{x}$ for any $\mathbf{x} \in \mathcal{S}$
4. For any $\mathbf{x} \in \mathcal{S}$, there exists $\mathbf{y} \equiv -\mathbf{x}$ such that $\mathbf{x}+\mathbf{y} = \mathbf{0} = \mathbf{y}+\mathbf{x}$
5. $\alpha(\mathbf{x}+\mathbf{y}) = \alpha\mathbf{x}+\alpha\mathbf{y}$
6. $(\alpha+\beta)\mathbf{x} = \alpha\mathbf{x}+\beta\mathbf{x}$
7. $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$
8. There exits a scalar ξ such that $\xi\mathbf{x} = \mathbf{x}$. (Typically $\xi = 1$)

† Def **[Vector Space]:** in layman's terms,

- A vector space $\mathcal{S} \subset \mathbb{R}^n$ is a set of vectors
- closed under addition and scalar multiplication, that is,

$$\text{if } \mathbf{x} \text{ and } \mathbf{y} \in \mathcal{S}, \text{ then } \alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{S}$$

- Contain the vector $\mathbf{0}$.

- Ex: $\mathbb{R}^3 = \{[x, y, z]^T \mid x, y, z \in \mathbb{R}\}$

$Y = X\beta + e$
$n\times1 \quad n\times p \quad p\times1 \quad n\times1$

† Def **[Subspace]:** Let $\mathcal{S}$ be a vector space, and let $\mathcal{M}$ be a set with $\mathcal{M} \subset \mathcal{S}$. $\mathcal{M}$ is a *subspace* of $\mathcal{S}$ if and only if $\mathcal{M}$ is a vector space.

M=

- Ex: $\{[x, y, 0]^T \mid x, y \in \mathbb{R}\}$: subspace of $\mathbb{R}^3$ consisting of the $x - y$ plane

♠ Note: Let $\mathcal{S}$ be a vector space and let $\mathbf{x}_1, \ldots, \mathbf{x}_r \in \mathcal{S}$. The set of all linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_r$ is a subspace of $\mathcal{S}$;

$$\mathcal{M} = \{\mathbf{y} \mid \mathbf{y} = \sum c_j \mathbf{x}_j, c_1, \ldots, c_r \text{ coefficients}\} \subset \mathcal{S}.$$

The space spanned by the vectors $X_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $X_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

consists of vectors $\begin{bmatrix} a \\ b \\ b \end{bmatrix}$, $a, b \in \mathbb{R}$

† Def **[Spanning]:** The set of all linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_r \in S$ is called the space *spanned* by $\mathbf{x}_1, \ldots, \mathbf{x}_r$. If $\mathcal{M}$ is a subpsce of $\mathcal{S}$ and $\mathcal{M}$ equals the space *spanned* by $\mathbf{x}_1, \ldots, \mathbf{x}_r$, then $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is called a *spanning set* for $\mathcal{M}$.

† Def **[Column Space]:** The column space of a $m \times n$ matrix $\mathbf{A}$, denoted by $C(\mathbf{A})$ is the vector space spanned by the columns of the matrix, that is,

$X\beta$

$$C(\mathbf{A}) = \{\mathbf{x} : \text{ there exists a vector } \mathbf{c} \text{ such that } \mathbf{x} = \mathbf{Ac}\}.$$

In layman's terms, if $\mathbf{x} \in C(\mathbf{A})$, then $\mathbf{x}$ is a linear combination of columns of $\mathbf{A}$ (consisting of all vectors (dimension m) formed by multiplying $\mathbf{A}$ by any vector).

$A = \begin{bmatrix} 1 & -2 \\ 2 & 4 \end{bmatrix}$

① $X = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$

$\begin{bmatrix} 3 \\ 6 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 4 \end{bmatrix} = Ac$

$c_1 - 2c_2 = 3$

$2c_1 + 4c_2 = 6$

$\Rightarrow c_1 = 3$ & $c_2 = 0$

$\Rightarrow X \in C(A)$

② $X = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$ $X \notin C(A)$

- Ex: Recall $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$.

$$\mathbf{X} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$\mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} \beta_0 + 2\beta_1 \\ \beta_0 + 3\beta_1 \\ \beta_0 + 4\beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \beta_0 + \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \beta_1$$

$\Rightarrow$ Observe $\mathbf{X}\boldsymbol{\beta} \in C(\mathbf{X})$.

† Def A system of equations $\mathbf{Ax} = \mathbf{c}$ is consistent iff there exists a solution $\mathbf{x}^*$ such that $\mathbf{Ax}^* = \mathbf{c}$.

★ Result: A system of equations $\mathbf{Ax} = \mathbf{c}$ is consistent iff $\mathbf{c} \in C(\mathbf{A})$.

♠ Recall NE;

$$\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$$

- $\mathbf{X}^T\mathbf{y} \in C(\mathbf{X}^T)$ and $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \in C(\mathbf{X}^T\mathbf{X})$
- Result: for any matrix $\mathbf{X}$, $C(\mathbf{X}^T\mathbf{X}) = C(\mathbf{X}^T)$ (Result 2.2)

$\Rightarrow$ **The normal equation is consistent (that is, $\exists$ a solution)**

- **However, still do not know how to find a solution $\hat{\boldsymbol{\beta}}$?**

ex $x_1 = \begin{bmatrix} 1 \\ -1 \\ 3 \end{bmatrix}$ $x_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ $\left.\begin{array}{l} \alpha_1 + \alpha_2 = 0 \\ -\alpha_1 + \alpha_2 = 0 \\ 3\alpha_1 + \alpha_2 = 0 \end{array}\right) \Leftrightarrow \alpha_1 = \alpha_2 = 0$

$\Rightarrow$ x_1 & x_2 are linearly indep.

ex $x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ $x_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $x_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ $\begin{array}{l} \alpha_1 + \alpha_2 + 2\alpha_3 = 0. \\ -\alpha_1 + 2\alpha_2 + \alpha_3 = 0 \end{array}$

Consider $\alpha_1 = 1$, $\alpha_2 = 1$ & $\alpha_3 = -1$ $\Rightarrow$ x_1, x_2 & x_3 are linearly dep.

† Def **[Linearly dependent]:** Let $\mathbf{x}_1, \ldots, \mathbf{x}_r$ be vectors in $\mathcal{S}$. If there exists scalars $\alpha_1, \ldots, \alpha_r$ not all zero so that $\sum \alpha_i \mathbf{x}_i = 0$, then $\mathbf{x}_1, \ldots, \mathbf{x}_r$ are *linearly dependent.*

If such α_is do not exist, $\mathbf{x}_1, \ldots, \mathbf{x}_r$ are *linearly independent.*

† Def **[Basis]:** If $\mathcal{M}$ is a subspace of $\mathcal{S}$ and if $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is a linearly independent spanning set for $\mathcal{M}$, then $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is called a *basis* for $\mathcal{M}$.

† Recall that for a $m \times n$ matrix $\mathbf{A}$, $r(\mathbf{A}) = \#$ of linearly independent rows or columns.

If $r(\mathbf{A}) = m (\leq n)$, $\mathbf{A}$ has full-row rank.

If $r(\mathbf{A}) = n (\leq m)$, $\mathbf{A}$ has full-column rank.

† Def **[Singular]:** Let $\mathbf{A}$ be an $n \times n$ matrix.

square

- $\mathbf{A}$ is *nonsingular* if there exists a matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}^{-1}\mathbf{A} = I = \mathbf{A}\mathbf{A}^{-1}$.
- If no such matrix exists, then $\mathbf{A}$ is *singular*.
- If $\mathbf{A}^{-1}$ exists, it is called the inverse of $\mathbf{A}$.

♣ Th: An $n \times n$ matrix $\mathbf{A}$ is nonsingular if and only if $r(\mathbf{A}) = n$ (full-row & full-column rank), i.e., the columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.

★ Note For $n \times n$ $\mathbf{A}$, $r(\mathbf{A}) = n \Leftrightarrow \mathbf{A}$ is nonsingluar. $|A| \neq 0$

$p \times p$

★ Result: for any matrix $\mathbf{X}$ $(n \times p)$, $r(\mathbf{X}^T\mathbf{X}) = r(\mathbf{X})$

★ Result: If $\mathbf{X}(n \times p)$ has $r(\mathbf{X}) = p$, then the $p \times p$ matrix $\mathbf{X}^T\mathbf{X}$ is nonsingular.

♠ Recall the NEs;

$$\mathbf{X}^T\mathbf{X}\beta = \mathbf{X}^T\mathbf{y}$$

$\Rightarrow$ If $r(\mathbf{X}) = p$, $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$: unique solution

Recall the NEs

$$\begin{bmatrix} \overset{a}{n} & \overset{b}{\Sigma x_i} \\ \overset{c}{\Sigma x_i} & \overset{d}{\Sigma x_i^2} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}$$

If $r(X^TX) = 2$, $\exists\ (X^TX)^{-1}$

nonsingular $\Rightarrow$ $|X^TX| = n\Sigma x_i^2 - (\Sigma x_i)^2$

$= n\Sigma x_i^2 - (n\bar{x})^2$

$= n(\Sigma x_i^2 - n\bar{x}^2)$

$= n\Sigma(x_i - \bar{x})^2 > 0$

∗ Example 1 (contd): Simple linear regression

Consider the model

$$y_i = \beta_0 + \beta_1 x_i + e_i, \qquad i = 1, \ldots, n.$$

Find $\hat{\boldsymbol{\beta}}$.

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (X^TX)^{-1}X^Ty = \frac{1}{n\Sigma(x_i-\bar{x})^2}\begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix}\begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}$$

$$= \frac{1}{n\Sigma(x_i-\bar{x})^2} \cdot \begin{bmatrix} (\Sigma x_i^2)(\Sigma y_i) - (\Sigma x_i)(\Sigma x_i y_i) \\ -(\Sigma x_i)(\Sigma y_i) + n\Sigma x_i y_i \end{bmatrix}$$

① β_1

$$\beta_1 = \frac{\cancel{n}\Sigma x_i y_i - \underbrace{(\Sigma x_i)}_{n\bar{x}}(\Sigma y_i)}{\cancel{n}\Sigma(x_i-\bar{x})^2} = \frac{\Sigma(x_i-\bar{x})y_i}{\Sigma(x_i-\bar{x})^2} = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}$$

$$\equiv \frac{S_{xy}}{S_{xx}}$$

Observe ~~Consider~~ $\Sigma(x_i-\bar{x})(y_i-\bar{y}) = \Sigma(x_i-\bar{x})y_i - \underbrace{\cancel{\Sigma(x_i-\bar{x})\bar{y}}}_{0} = \Sigma(x_i-\bar{x})y_i$

$= \Sigma x_i y_i - n\bar{x}\bar{y}$

② β_0

$$\beta_0 = \frac{\sum x_i^2 \overset{= \cancel{n}\bar{y}}{(\sum y_i)} - \overset{= \cancel{n}\bar{x}}{(\sum x_i)} \sum x_i y_i}{\cancel{n} \sum (x_i - \bar{x})^2} = \frac{\bar{y}(\sum x_i^2 - n\bar{x}^2) + n\bar{y}\bar{x}^2 - \bar{x}\sum x_i y_i}{\sum (x_i - \bar{x})^2}$$

$$= \bar{y} - \bar{x}\,\frac{(\sum x_i y_i - \overset{n\bar{x}\bar{y}}{\cancel{\;\;}})}{\sum (x_i - \bar{x})^2} = \bar{y} - \bar{x}\,\frac{S_{xy}}{S_{xx}} = \bar{y} - \bar{x}\beta_1$$

∗ Example 2 (contd): Balanced One-Way ANOVA
Consider the model

$$y_{ij} = \mu + \alpha_i + e_{ij}, \qquad i = 1, \ldots, a \text{ and } j = 1, \ldots, n.$$

Find $r(\mathbf{X}) = r(\underbrace{\mathbf{X}^T\mathbf{X}}_{(a+1)\times(a+1)})$. $= a$