Practice problems for midterm

April 5, 2024

1. Suppose we have a population $\{1, 2, 3, 4, 5\}$. Suppose that we draw a random sample $X_1, X_2, \cdots, X_n$ from this population with replacement. How much n should be so that the variance of the sample mean is equal to or smaller than 0.1 ?

2. Suppose we want to transform the left dataset(D) into the right dataset($D2$), where the unit of height is not centimeters but meters. **Which matrix** should we multiply to D and on **which side** (left or right) of D should we multiply it?

```
D.head()
```

	height	weight
0	158.64	48.00
1	156.59	45.78
2	172.70	62.65
3	154.18	43.89
4	178.39	66.24

```
D2.head()
```

	height	weight
0	1.5864	48.00
1	1.5659	45.78
2	1.7270	62.65
3	1.5418	43.89
4	1.7839	66.24

3. Suppose we have random variables $X_1, X_2, \cdots, X_n$ which are identically distributed with expectation μ and variance σ^2. ($E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$ and $Var(X_1) = Var(X_2) = \cdots = Var(X_n) = \sigma^2$.) Suppose however that $X_1, X_2, \cdots, X_n$ are *not* independent and $Cov(X_i, X_j) = \rho^2 > 0$ for any $i \neq j$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) What is $E(\bar{X})$? Show your derivation.

(b) What is $Var(\bar{X})$? Show your derivation.

(c) What is the meaning of (b) ?

4. In the following code,

(a) express the expected outcome using the cumulative distribution function of the standard normal distribution, $\Phi()$.

(b) explain how `sum(reject)/len(reject)` will change according to the value of `sigma` and `n`.

```
In [1]: import numpy as np
        import scipy.stats as stats
```

```
In [2]: sigma=1
        n=100
```

```
In [3]: reject=[]

        for i in range(100000):
            samp=np.random.normal(loc=-0.1,scale=sigma,size=n)

            Z=(np.mean(samp)-0)/sigma*np.sqrt(len(samp))

            reject.append(Z<stats.norm.ppf(0.05))
```

```
In [4]: sum(reject)/len(reject)
```

5. Consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (i = 1, 2, \cdots, N)$$

where ε_i's are i.i.d. with mean 0. Consider the following, slightly different model

$$y_i = \alpha_0 + \alpha_1(x_i - \bar{x}) + \varepsilon_i \quad (i = 1, 2, \cdots, N)$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. Suppose $\hat{\beta}$ and $\hat{\alpha}$ are least-squares estimates. Prove that $\hat{\alpha}_1 = \hat{\beta}_1$.

6. Suppose we have data $(x_1, y_1), (x_2, y_2), \cdots, (x_{10}, y_{10})$ which satisfy

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (1 \le i \le 10)$$

where $\varepsilon_1, \cdots, \varepsilon_{10}$ are independent with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = 2$ for every i. Suppose the values of x_i can be selected anywhere between 0 and 2. What is the smallest value of $Var(\hat{\beta}_1)$ possible? ($\hat{\beta}_1$ is the estimate of β_1 obtained by Least-Square Method.) Explain the response.

7. Suppose we conduct linear regression on outcome variable (y) and explanatory variable (x). We posit the following model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

where ε_i's are i.i.d. $\mathcal{N}(0, \sigma^2)$.

Consider the following, slightly different model

$$y_i = \alpha_0 + \alpha_1 c x_i + \epsilon_i \tag{2}$$

where ϵ_i's are i.i.d. $\mathcal{N}(0, \sigma^2)$.

In (2), we scaled the variable x_i by $c (\neq 0)$. (For example, when x_i is the height in cm, we can change it to height in meter by multiplying $c = 0.01$.)

Prove that the test results for

$$H_0 : \beta_1 = 0 \qquad vs \qquad H_1 : \beta_1 \neq 0$$

and

$$H_0 : \alpha_1 = 0 \qquad vs \qquad H_1 : \alpha_1 \neq 0$$

are identical. (Assume we know the value of σ^2.)

8. Let $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1N} & x_{2N} \end{bmatrix}$, $\mathcal{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$. Suppose we have

$$Y = X\beta + \mathcal{E}$$

where $\varepsilon_1, \cdots, \varepsilon_N$ are independent with $E[\varepsilon_i] = 0$ and $Var[\varepsilon_i] = \sigma^2$ for every i. Also, suppose $X^T X = \begin{bmatrix} 4 & -2 & 6 \\ -2 & 2 & -5 \\ 6 & -5 & 29 \end{bmatrix}$ and $X^T Y = \begin{bmatrix} 22 \\ -14 \\ 71 \end{bmatrix}$.

(a) Conduct Cholesky decomposition of $X^T X$.

(b) What is the least-square estimate of β? (Do not use the inverse formula of 3×3 matrix.)

9. Let $M = \begin{bmatrix} y_1 & x_{11} & x_{21} \\ y_2 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ y_N & x_{1N} & x_{2N} \end{bmatrix}$. We posit the following model

$$y_i = \beta_0 + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + \varepsilon_i$$

where ε_i's are i.i.d. with mean 0 and $\bar{x}_1 = \frac{1}{N}\sum_{i=1}^N x_{1i}$, $\bar{x}_2 = \frac{1}{N}\sum_{i=1}^N x_{2i}$. Suppose $N = 15$, $\sum_{i=1}^N y_i = 45$, and suppose we have

$$M^T \left(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T \right) M = \begin{bmatrix} 5 & 30 & -30 \\ 30 & 10 & 0 \\ -30 & 0 & 15 \end{bmatrix},$$

where I is $N \times N$ identity matrix and $\mathbf{1}$ is a $N \times 1$ vector with all elements equal to 1. What are the least-squares estimates of β_0, β_1 and β_2?

10. Suppose we conduct linear regression on outcome variable (y) and explanatory variables (x_1, x_2, $\cdots$, x_p). We posit the following model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

where ε_i's are i.i.d. with mean 0.

Consider the following, slightly different model

$$y_i = \alpha_0 + \alpha_1 \underline{c_1 x_{i1}} + \alpha_2 \underline{c_2 x_{i2}} + \cdots + \alpha_p \underline{c_p x_{ip}} + \varepsilon_i$$

where $c_1, c_2, \cdots, c_p \neq 0$. (For example, when x_{i1} is the height in cm, we can change it to height in meter by multiplying $c_1 = 0.01$.)

Suppose $\hat{\beta}$ and $\hat{\alpha}$ are least-squares estimates. Using matrix operations, show that $\hat{\alpha}_j = \frac{1}{c_j}\hat{\beta}_j$ for $j = 1, 2, \cdots, p$. (*Hint: see Problem 2*)

1. Suppose we have a population $\{1, 2, 3, 4, 5\}$. Suppose that we draw a random sample $X_1, X_2, \cdots, X_n$ from this population with replacement. How much n should be so that the variance of the sample mean is equal to or smaller than 0.1 ?

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{4^2+1^2+1^2+4^2}{5} = 2.$$

Note that $Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{2}{n} \le 0.1 \Rightarrow n \geq 20.$

2. Suppose we want to transform the left dataset(D) into the right dataset($D2$), where the unit of height is not centimeters but meters. **Which matrix** should we multiply to D and on **which side** (left or right) of D should we multiply it?

```
D.head()
```

	height	weight
0	158.64	48.00
1	156.59	45.78
2	172.70	62.65
3	154.18	43.89
4	178.39	66.24

```
D2.head()
```

	height	weight
0	1.5864	48.00
1	1.5659	45.78
2	1.7270	62.65
3	1.5418	43.89
4	1.7839	66.24

$$D_2 = D\begin{bmatrix} 1/100 & 0 \\ 0 & 1 \end{bmatrix}$$

3. Suppose we have random variables $X_1, X_2, \cdots, X_n$ which are identically distributed with expectation μ and variance σ^2. ($E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$ and $Var(X_1) = Var(X_2) = \cdots = Var(X_n) = \sigma^2$.) Suppose however that $X_1, X_2, \cdots, X_n$ are *not* independent and $Cov(X_i, X_j) = \rho^2 > 0$ for any $i \neq j$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) What is $E(\bar{X})$? Show your derivation.

(b) What is $Var(\bar{X})$? Show your derivation.

(c) What is the meaning of (b) ?

(a) $E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n} n\mu = \mu.$

(b) $Var(\bar{X}) = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} Cov\left[\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)$

$= \frac{1}{n^2}\left[\sum_{i=1}^{n} Cov(X_i, X_i) + \sum_{i \neq j} Cov(X_i, X_j)\right]$

$= \frac{1}{n^2}\left[\sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} \rho^2\right]$

$= \frac{1}{n^2}\left[\sum_{i=1}^{n} \sigma^2 + (n^2 - n)\rho^2\right]$

$= \frac{1}{n^2}\left[n\sigma^2 + n(n-1)\rho^2\right]$

$= \frac{1}{n}\sigma^2 + \left(1 - \frac{1}{n}\right)\rho^2.$

(c) Since $\left(1-\frac{1}{n}\right)\rho^2 > 0$, $Var(\bar{X})$ is always larger than when $X_i \perp X_j$ $(i \neq j)$

Also, $\lim_{n\to\infty}\left[\frac{1}{n}\sigma^2 + \left(1-\frac{1}{n}\right)\rho^2\right] = 0 + \rho^2 = \rho^2 > 0.$

$\hookrightarrow$ $Var(\bar{X})$ cannot converge to 0 no matter how large n is.

4. In the following code,

(a) express the expected outcome using the cumulative distribution function of the standard normal distribution, $\Phi()$.

(b) explain how `sum(reject)/len(reject)` will change according to the value of `sigma` and `n`.

In [1]:
```
import numpy as np
import scipy.stats as stats
```

In [2]:
```
sigma=1
n=100
```

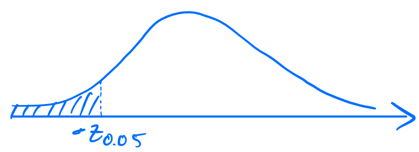
In [3]:
```
reject=[]

for i in range(100000):
    samp=np.random.normal(loc=-0.1,scale=sigma,size=n)

    Z=(np.mean(samp)-0)/sigma*np.sqrt(len(samp))

    reject.append(Z<stats.norm.ppf(0.05))
```

In [4]:
```
sum(reject)/len(reject)
```

(a) ⊗ let samp $= \{X_1, X_2, \dots, X_n\}$ where $n = 100$, $X_i \sim N(-0.1, 1^2)$

⊗ The Z-statistic: $Z = \dfrac{\bar{X} - 0}{\sigma/\sqrt{n}} = \dfrac{\bar{X}\sqrt{100}}{1} = 10\bar{X} \sim N(-1, 1)$

⊗ The hypothesis: $\begin{cases} H_0: \mu \geq 0 \\ H_1: \mu < 0. \end{cases}$



⊗ $\dfrac{\text{sum(reject)}}{\text{len(reject)}}$ = the ratio of #of rejects over the #of trials.

$= P[Z < -z_{0.05}] = P[Z + 1 < 1 - z_{0.05}]$

$= P[Z' < 1 - z_{0.05}]$ $(Z' = Z + 1 \sim N(0,1))$

$= \Phi(1 - z_{0.05}) \approx \Phi(1 - 1.645) = \Phi(-0.645)$

(b) Note that $Z = \dfrac{\bar{X} - 0}{\sigma/\sqrt{n}} = \dfrac{\bar{X}}{\sigma/\sqrt{n}} \sim N\left(\dfrac{-0.1}{\sigma/\sqrt{n}}, 1\right)$

$\Rightarrow$ Result $= P[Z < -z_{0.05}] = P\left[Z + \dfrac{0.1}{\sigma/\sqrt{n}} < \dfrac{0.1}{\sigma/\sqrt{n}} - z_{0.05}\right]$

$= P\left[Z' < \dfrac{0.1}{\sigma/\sqrt{n}} - z_{0.05}\right]$ $\left(Z' = Z + \dfrac{0.1}{\sigma/\sqrt{n}} \sim N(0,1)\right)$

$= \Phi\left(\dfrac{0.1}{\sigma/\sqrt{n}} - z_{0.05}\right)$

⊗ When $\sigma \uparrow$, $\dfrac{0.1}{\sigma/\sqrt{n}} \downarrow \Rightarrow$ result $\downarrow$.

$\sigma \downarrow$, $\dfrac{0.1}{\sigma/\sqrt{n}} \uparrow \Rightarrow$ result $\uparrow$

when $n \uparrow$, $\dfrac{0.1}{\sigma/\sqrt{n}} \uparrow \Rightarrow$ result $\uparrow$

$n \downarrow$, $\dfrac{0.1}{\sigma/\sqrt{n}} \downarrow \Rightarrow$ result $\downarrow$.

5. Consider a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (i = 1, 2, \cdots, N)$$

where ε_i's are i.i.d. with mean 0. Consider the following, slightly different model

$$y_i = \alpha_0 + \alpha_1 (x_i - \bar{x}) + \varepsilon_i \quad (i = 1, 2, \cdots, N)$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$. Suppose $\hat{\beta}$ and $\hat{\alpha}$ are least-squares estimates. Prove that $\hat{\alpha}_1 = \hat{\beta}_1$.

⊗ We have: $\overline{(x_i - \bar{x})} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}) = \frac{1}{N}\sum_{i=1}^{N} x_i - \frac{1}{N}\sum_{i=1}^{N}\bar{x} = \frac{1}{N}N\bar{x} - \frac{1}{N}N\bar{x} = 0.$

⊗ Since $\hat{\beta}$ and $\hat{\alpha}$ are least-square estimates,

$$\begin{cases} \hat{\beta}_1 = \dfrac{\sum_{i=1}^{N}(x_i - \bar{x})y_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \\ \hat{\alpha}_1 = \dfrac{\sum_{i=1}^{N}\left[x_i - \bar{x} - \overline{(x_i - \bar{x})}\right]y_i}{\sum_{i=1}^{N}\left[(x_i - \bar{x}) - \overline{(x_i - \bar{x})}\right]^2} = \dfrac{\sum_{i=1}^{N}(x_i - \bar{x})y_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} = \hat{\beta}_1 \end{cases}$$

6. Suppose we have data $(x_1, y_1), (x_2, y_2), \cdots, (x_{10}, y_{10})$ which satisfy

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (1 \le i \le 10)$$

where $\varepsilon_1, \cdots, \varepsilon_{10}$ are independent with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = 2$ for every i. Suppose the values of x_i can be selected anywhere between 0 and 2. What is the smallest value of $Var(\hat{\beta}_1)$ possible? ($\hat{\beta}_1$ is the estimate of β_1 obtained by Least-Square Method.) Explain the response.

⊗ Note that: $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} \Rightarrow Var(\hat{\beta}_1) = Var\left(\dfrac{S_{xy}}{S_{xx}}\right) = \dfrac{1}{S_{xx}^2} Var\left(\sum_{i=1}^{10}(x_i - \bar{x})y_i\right) = \dfrac{1}{S_{xx}^2}\sum_{i=1}^{10}(x_i - \bar{x})^2 Var(y_i)$

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

⊗ To minimize $Var(\hat{\beta}_1)$, we need to maximize $S_{xx} = \sum_{i=1}^{10}(x_i - \bar{x})^2$.

One way to maximize S_{xx}: $\begin{cases} x_1 = x_2 = x_3 = x_4 = x_5 = 0 \\ x_6 = x_7 = x_8 = x_9 = x_{10} = 2 \end{cases} \Rightarrow \bar{x} = 1 \Rightarrow S_{xx} = \sum_{i=1}^{10}(x_i - \bar{x})^2 = 10 \times 1 = 10.$

$\Rightarrow \min_{x_1, x_2, \dots, x_{10}} \{Var(\hat{\beta}_1)\} = \dfrac{2}{10} = \dfrac{1}{5}$ when $\begin{cases} x_1 = x_2 = x_3 = x_4 = x_5 = 0 \\ x_6 = x_7 = x_8 = x_9 = x_{10} = 2 \end{cases}.$

7. Suppose we conduct linear regression on outcome variable (y) and explanatory variable (x). We posit the following model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

where ε_i's are i.i.d. $\mathcal{N}(0, \sigma^2)$.

Consider the following, slightly different model

$$y_i = \alpha_0 + \alpha_1 c x_i + \epsilon_i \tag{2}$$

where ϵ_i's are i.i.d. $\mathcal{N}(0, \sigma^2)$.

In (2), we scaled the variable x_i by $c (\neq 0)$. (For example, when x_i is the height in cm, we can change it to height in meter by multiplying $c = 0.01$.)

Prove that the test results for

$$H_0 : \beta_1 = 0 \quad vs \quad H_1 : \beta_1 \neq 0$$

and

$$H_0 : \alpha_1 = 0 \quad vs \quad H_1 : \alpha_1 \neq 0$$

are identical. (Assume we know the value of σ^2.)

⊗ Test statistic for β_1: $Z_{\beta_1} = \dfrac{\beta_1 - 0}{\sigma/\sqrt{S_{xx\beta}}} = \dfrac{\beta_1}{\sigma/\sqrt{S_{xx\beta}}}$

$\alpha_1 = \dfrac{\sum(cx_i - c\bar{x}_i)y_i}{\sum(cx_i - c\bar{x}_i)^2} = \dfrac{c\sum(x_i-\bar{x})y_i}{c^2\sum(x_i-\bar{x})^2} = \dfrac{1}{c}\beta_1$

⊗ Test statistic for α_1: $Z_{\alpha_1} = \dfrac{\alpha_1 - 0}{\sigma/\sqrt{S_{xx\alpha}}} = \dfrac{\frac{1}{c}\beta_1}{\sigma/c\sqrt{S_{xx\beta}}} = \dfrac{\beta_1}{\sigma/\sqrt{S_{xx\beta}}}$

$S_{xx_\alpha} = c^2 S_{xx\beta}$

⊗ The 2 statistics are the same, hence, the test results are the same

8. Let $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$, $X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1N} & x_{2N} \end{bmatrix}$, $\mathcal{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$. Suppose we have

$$Y = X\beta + \mathcal{E}$$

where $\varepsilon_1, \cdots, \varepsilon_N$ are independent with $E[\varepsilon_i] = 0$ and $Var[\varepsilon_i] = \sigma^2$ for every i.
Also, suppose $X^T X = \begin{bmatrix} 4 & -2 & 6 \\ -2 & 2 & -5 \\ 6 & -5 & 29 \end{bmatrix}$ and $X^T Y = \begin{bmatrix} 22 \\ -14 \\ 71 \end{bmatrix}$.

(a) Conduct Cholesky decomposition of $X^T X$.

(b) What is the least-square estimate of β? (Do not use the inverse formula of 3×3 matrix.)

(a) $X^T X = \begin{bmatrix} 4 & -2 & 6 \\ -2 & 2 & -5 \\ 6 & -5 & 29 \end{bmatrix}$. Cholesky decomposition: $X^T X = (L^*)(L^*)^T$

$\Rightarrow \begin{bmatrix} 4 & -2 & 6 \\ -2 & 2 & -5 \\ 6 & -5 & 29 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$

$$\Rightarrow \begin{cases} l_{11}^2 = 4 \Rightarrow l_{11} = 2. \\ l_{21} l_{11} = -2 \Rightarrow l_{21} = -1. \\ l_{21}^2 + l_{22}^2 = 2 \Rightarrow l_{22} = \sqrt{2-1} = 1. \\ l_{31} l_{11} = 6 \Rightarrow l_{31} = 3 \\ l_{31} l_{21} + l_{32} l_{22} = -5 \Rightarrow 3\times(-1) + l_{32} = -5 \Rightarrow l_{32} = -2 \\ l_{33}^2 + l_{32}^2 + l_{31}^2 = 29 \Rightarrow l_{33} = \sqrt{29-9-4} = \sqrt{16} = 4. \end{cases}$$

$\Rightarrow L^* = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & -2 & 4 \end{bmatrix}$

Thus, $X^T X = L^*(L^*)^T = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & -2 & 4 \end{bmatrix} \begin{bmatrix} 2 & -1 & 3 \\ 0 & 1 & -2 \\ 0 & 0 & 4 \end{bmatrix}$

(b) ⊗ Note that $X^T X\beta = X^T Y \Rightarrow L^*(L^*)^T\beta = \begin{bmatrix} 22 \\ -14 \\ 71 \end{bmatrix}$. Let $\gamma = (L^*)^T\beta$, we have: $L^*\gamma = \begin{bmatrix} 22 \\ -14 \\ 71 \end{bmatrix}$

$\Rightarrow \begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 3 & -2 & 4 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} = \begin{bmatrix} 22 \\ -14 \\ 71 \end{bmatrix}$

$$\Rightarrow \begin{cases} 2\gamma_1 = 22 \\ -\gamma_1 + \gamma_2 = -14 \\ 3\gamma_1 - 2\gamma_2 + 4\gamma_3 = 71 \end{cases}$$

$$\Rightarrow \begin{cases} \gamma_1 = 11 \\ \gamma_2 = -3 \\ \gamma_3 = \dfrac{71 - 33 - 6}{4} = 8 \end{cases}$$

$\Rightarrow \gamma = \begin{bmatrix} 11 \\ -3 \\ 8 \end{bmatrix}$

⊗ Now, $(L^*)^T\beta = \gamma \Rightarrow \begin{bmatrix} 2 & -1 & 3 \\ 0 & 1 & -2 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 11 \\ -3 \\ 8 \end{bmatrix}$

$$\Rightarrow \begin{cases} 2\beta_0 - \beta_1 + 3\beta_2 = 11 \\ \beta_1 - 2\beta_2 = -3 \\ 4\beta_2 = 8 \end{cases}$$

$$\Rightarrow \begin{cases} \beta_0 = 3 \\ \beta_1 = 1 \\ \beta_2 = 2 \end{cases} \Rightarrow \beta = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$$

9. Let $M = \begin{bmatrix} y_1 & x_{11} & x_{21} \\ y_2 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ y_N & x_{1N} & x_{2N} \end{bmatrix}$. We posit the following model

$$y_i = \beta_0 + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + \varepsilon_i$$

where ε_i's are i.i.d. with mean 0 and $\bar{x}_1 = \frac{1}{N}\sum_{i=1}^N x_{1i}$, $\bar{x}_2 = \frac{1}{N}\sum_{i=1}^N x_{2i}$. Suppose $N = 15$, $\sum_{i=1}^N y_i = 45$, and suppose we have

$$M^T\left(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T\right)M = \begin{bmatrix} 5 & 30 & -30 \\ 30 & 10 & 0 \\ -30 & 0 & 15 \end{bmatrix},$$

where I is $N \times N$ identity matrix and $\mathbf{1}$ is a $N \times 1$ vector with all elements equal to 1. What are the least-squares estimates of β_0, β_1 and β_2?

least-squares: $Y = X\beta + \varepsilon$.

Minimize : $(Y - X\beta)^T(Y - X\beta) \Rightarrow \hat{\beta} = \underset{\beta \in \mathbb{R}^{3\times 1}}{\operatorname{argmin}}\{(Y - X\beta)^T(Y - X\beta)\}$

$$\boxed{\bar{x} = \frac{1}{n}1^T x}$$

$\hookrightarrow n\bar{x} = 1^T x$

$$X = \begin{bmatrix} 1 & x_{11} - \bar{x}_1 & x_{21} - \bar{x}_2 \\ 1 & x_{12} - \bar{x}_1 & x_{22} - \bar{x}_2 \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} - \bar{x}_1 & x_{2n} - \bar{x}_2 \end{bmatrix} \quad X^T X\beta = X^T Y$$

We have $M^T(I_n - \frac{1}{n}11^T)M = M^T M - \frac{1}{n}M^T 1 1^T M = M^T M - \frac{1}{n}(M^T 1)(M^T 1)^T$

$$M^T M = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{bmatrix} \begin{bmatrix} y_1 & x_{11} & x_{21} \\ y_2 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ y_n & x_{1n} & x_{2n} \end{bmatrix} = \begin{bmatrix} | & | & | \\ M^T y & M^T x_1 & M^T x_2 \\ | & | & | \end{bmatrix}$$

$$M^T 1 = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i} \\ \sum x_{2i} \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ n\bar{x}_1 \\ n\bar{x}_2 \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ 15\bar{x}_1 \\ 15\bar{x}_2 \end{bmatrix} = n\begin{bmatrix} \bar{y} \\ \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}.$$

$\Rightarrow (M^T 1)^T = n[\bar{y} \;\; \bar{x}_1 \;\; \bar{x}_2]$

$$\Rightarrow \frac{1}{n}(M^T 1)(M^T 1)^T = n\begin{bmatrix} \bar{y} \\ \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}[\bar{y} \;\; \bar{x}_1 \;\; \bar{x}_2] = n\begin{bmatrix} \bar{y}^2 & \bar{y}\bar{x}_1 & \bar{y}\bar{x}_2 \\ \bar{y}\bar{x}_1 & \bar{y}\bar{x}_1^2 & \bar{x}_1\bar{x}_2 \\ \bar{y}\bar{x}_2 & \bar{x}_1\bar{x}_2 & \bar{x}_2^2 \end{bmatrix} = \frac{1}{n}\begin{bmatrix} 1^T y 1^T y & 1^T y 1^T x_1 & 1^T y 1^T x_2 \\ 1^T y 1^T x_1 & 1^T x_1 1^T x_1 & 1^T x_1 1^T x_2 \\ 1^T y 1^T x_2 & 1^T x_1 1^T x_2 & 1^T x_2 1^T x_2 \end{bmatrix}$$

Thus, $$M^T M - \frac{1}{n}(M^T 1)(M^T 1)^T = \begin{bmatrix} y^T y - \frac{1}{n}1^T y 1^T y & y^T x_1 - \frac{1}{n}1^T y 1^T x_1 & y^T x_2 - \frac{1}{n}1^T y 1^T x_2 \\ x_1^T y - \frac{1}{n}1^T y 1^T x_1 & x_1^T x_1 - \frac{1}{n}1^T x_1 1^T x_1 & x_1^T x_2 - \frac{1}{n}1^T x_1 1^T x_2 \\ x_2^T y - \frac{1}{n}1^T y 1^T x_2 & x_2^T x_1 - \frac{1}{n}1^T x_2 1^T x_1 & x_2^T x_2 - \frac{1}{n}1^T x_2 1^T x_2 \end{bmatrix} = \begin{bmatrix} 5 & 30 & -30 \\ 30 & 10 & 0 \\ -30 & 0 & 15 \end{bmatrix}$$

Let $S_{ab} = a^T b - \frac{1}{n}1^T a 1^T b$. Then : $\begin{bmatrix} S_{yy} & S_{yx_1} & S_{yx_1} \\ S_{x_1 y} & S_{x_1 x_1} & S_{x_1 x_2} \\ S_{x_2 y} & S_{x_2 x_1} & S_{x_2 x_2} \end{bmatrix} = \begin{bmatrix} 5 & 30 & -30 \\ 30 & 10 & 0 \\ -30 & 0 & 15 \end{bmatrix} \Rightarrow \begin{cases} S_{x_1 x_1} = 10 \\ S_{x_1 x_2} = S_{x_2 x_1} = 0 \\ S_{x_2 x_2} = 15 \\ S_{YY} = 5 \\ S_{x_1 Y} = S_{Y x_1} = 30 \\ S_{x_2 Y} = S_{Y x_2} = -30 \end{cases}$

$= \sum a_i b_i - n\bar{a}\bar{b}$

$= \sum a_i b_i - \bar{a}\sum b_i$

$= \sum(a_i b_i - \bar{a} b_i)$

$= \sum(a_i - \bar{a}) b_i$

$= \sum(a_i - \bar{a})(b_i - \bar{b})$

⊛ Now, $$X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} - \bar{x}_1 & x_{12} - \bar{x}_1 & \cdots & x_{1n} - \bar{x}_1 \\ x_{21} - \bar{x}_2 & x_{22} - \bar{x}_2 & \cdots & x_{2n} - \bar{x}_2 \end{bmatrix} \begin{bmatrix} 1 & x_{11} - \bar{x}_1 & x_{21} - \bar{x}_2 \\ 1 & x_{12} - \bar{x}_1 & x_{22} - \bar{x}_2 \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} - \bar{x}_1 & x_{2n} - \bar{x}_2 \end{bmatrix} = \begin{bmatrix} n & \sum(x_{1i} - \bar{x}_{1i}) & \sum(x_{2i} - \bar{x}_2) \\ \sum(x_{1i} - \bar{x}_1) & \sum(x_{1i} - \bar{x}_1)^2 & \sum(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \\ \sum(x_{2i} - \bar{x}_2) & \sum(x_{2i} - \bar{x}_2)(x_{1i} - \bar{x}_1) & \sum(x_{2i} - \bar{x}_2)^2 \end{bmatrix}$$

$$\Rightarrow X^T X = \begin{bmatrix} n & 0 & 0 \\ 0 & S_{x_1x_1} & S_{x_1x_2} \\ 0 & S_{x_2x_1} & S_{x_2x_2} \end{bmatrix} = \begin{bmatrix} 15 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 15 \end{bmatrix}$$

⊗ Note that $X^T Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11}-\bar{x}_1 & x_{12}-\bar{x}_1 & \cdots & x_{1n}-\bar{x}_1 \\ x_{21}-\bar{x}_2 & x_{22}-\bar{x}_2 & \cdots & x_{2n}-\bar{x}_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum (x_{1i}-\bar{x}_1) y_i \\ \sum (x_{2i}-\bar{x}_2) y_i \end{bmatrix} = \begin{bmatrix} 45 \\ S_{x_1Y} \\ S_{x_2Y} \end{bmatrix} = \begin{bmatrix} 45 \\ 30 \\ -30 \end{bmatrix}$

⊗ The least-square estimate must satisfy: $X^T X \beta = X^T Y \Rightarrow \beta = (X^T X)^{-1} X^T Y$

$$= \begin{bmatrix} 15 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 15 \end{bmatrix}^{-1} \begin{bmatrix} 45 \\ 30 \\ -30 \end{bmatrix}$$

$$= \begin{bmatrix} 1/15 & 0 & 0 \\ 0 & 1/10 & 0 \\ 0 & 0 & 1/15 \end{bmatrix} \begin{bmatrix} 45 \\ 30 \\ -30 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ 3 \\ -2 \end{bmatrix}$$

⊗ Therefore, $\begin{cases} \beta_0 = 3 \\ \beta_1 = 3 \\ \beta_2 = -2 \end{cases}$.

10. Suppose we conduct linear regression on outcome variable (y) and explanatory variables ($x_1, x_2, \cdots, x_p$). We posit the following model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

where ε_i's are i.i.d. with mean 0.

Consider the following, slightly different model

$$y_i = \alpha_0 + \alpha_1 \underline{c_1 x_{i1}} + \alpha_2 \underline{c_2 x_{i2}} + \cdots + \alpha_p \underline{c_p x_{ip}} + \varepsilon_i$$

where $c_1, c_2, \cdots, c_p \neq 0$. (For example, when x_{i1} is the height in cm, we can change it to height in meter by multiplying $c_1 = 0.01$.)

Suppose $\hat{\beta}$ and $\hat{\alpha}$ are least-squares estimates. Using matrix operations, show that $\hat{\alpha}_j = \frac{1}{c_j}\hat{\beta}_j$ for $j = 1, 2, \cdots, p$. (*Hint: see Problem 2*)

⊗ Let the explanatory variables matrix of the original data: $X_1 = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix}$.

Scaling the features of X_1 by $c_1, c_2, \ldots, c_p$, we obtained the newly scaled explanatory variables matrix:

$$X_2 = \begin{bmatrix} 1 & c_1x_{11} & c_2x_{21} & \cdots & c_px_{p1} \\ 1 & c_1x_{12} & c_2x_{22} & \cdots & c_px_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & c_1x_{1n} & c_2x_{2n} & \cdots & c_px_{pn} \end{bmatrix} = \begin{bmatrix} | & | & | & & | \\ 1 & c_1x_1 & c_2x_2 & \cdots & c_px_p \\ | & | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} | & | & | & & | \\ 1 & x_1 & x_2 & \cdots & x_p \\ | & | & | & & | \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & c_1 & 0 & \cdots & 0 \\ 0 & 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & c_p \end{bmatrix}$$

$$= X_1 \begin{bmatrix} 1 & & & & 0 \\ & c_1 & & & \\ & & c_2 & & \\ & & & \ddots & \\ 0 & & & & c_p \end{bmatrix} = X_1 C \quad \left(C = \begin{bmatrix} 1 & & & & 0 \\ & c_1 & & & \\ & & c_2 & & \\ & & & \ddots & \\ 0 & & & & c_p \end{bmatrix} \right)$$

⊗ In the original data, the least-square estimate is:

$$X_1^T X_1 \hat{\beta} = X_1^T Y \Rightarrow \hat{\beta} = (X_1^T X_1)^{-1} X_1^T Y.$$

⊗ In the scaled data, the least-square estimate is:

$$X_2^T X_2 \hat{\alpha} = X_2^T Y \Rightarrow \hat{\alpha} = (X_2^T X_2)^{-1} X_2^T Y.$$

⊗ Simplifying $\hat{\alpha}$, we obtain:

$$\hat{\alpha} = (X_2^T X_2)^{-1} X_2^T Y = (C^T X_1^T X_1 C)^{-1} C^T X_1^T Y = (C X_1^T X_1 C)^{-1} C X_1^T X_1 \hat{\beta}$$

$$= C^{-1} (C X_1^T X_1)^{-1} C X_1^T X_1 \hat{\beta}$$

$$= C^{-1} (X_1^T X_1)^{-1} C^{-1} C X_1^T X_1 \hat{\beta}$$

$$= C^{-1} (X_1^T X_1)^{-1} X_1^T X_1 \beta$$

$$= C^{-1} \beta$$

$$\Rightarrow \hat{\alpha} = C^{-1}\beta$$

⊛ Note that $C^{-1} = \begin{bmatrix} 1 & & & & 0 \\ & c_1 & & & \\ & & c_2 & & \\ & & & \ddots & \\ 0 & & & & c_p \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & & & 0 \\ & \frac{1}{c_1} & & & \\ & & \frac{1}{c_2} & & \\ & & & \ddots & \\ 0 & & & & \frac{1}{c_p} \end{bmatrix}$.

Thus, the least-square estimate of the scaled data is:

$$\begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_p \end{bmatrix} = \hat{\alpha} = \begin{bmatrix} 1 & & & & 0 \\ & \frac{1}{c_1} & & & \\ & & \frac{1}{c_2} & & \\ & & & \ddots & \\ 0 & & & & \frac{1}{c_p} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \frac{1}{c_1}\beta_1 \\ \frac{1}{c_2}\beta_2 \\ \vdots \\ \frac{1}{c_p}\beta_p \end{bmatrix}$$

$$\Rightarrow \hat{\alpha}_j = \frac{1}{c_j}\beta_j \quad \text{for } j = 1, 2, \dots, p. \quad \square.$$

Proof of the configuration in problem 6.

"Maximize: $\sum_{i=1}^{10}(x_i-\bar{x})^2$

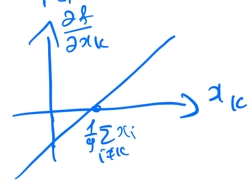
Constraint: $0 \le x_i \le 2 \ \forall i$"

⊛ Let $f(x_1, x_2, \dots, x_{10}) = \sum_{i=1}^{10}(x_i-\bar{x})^2$.

Note that $\sum_{i=1}^{10}(x_i-\bar{x})^2 = \sum_{i=1}^{10} x_i^2 - 2\bar{x}\sum_{i=1}^{10} x_i + \sum_{i=1}^{10}\bar{x}^2 = \sum_{i=1}^{10} x_i^2 - 20\bar{x}^2 + 10\bar{x}^2 = \sum_{i=1}^{10} x_i^2 - 10\bar{x}^2$

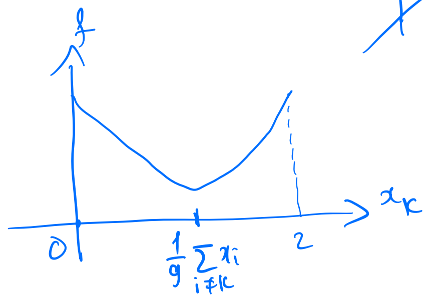
$= \sum_{i=1}^{10} x_i^2 - \frac{1}{10}\left(\sum_{i=1}^{10} x_i\right)^2$.

⊛ $\frac{\partial f}{\partial x_k} = 2x_k - \frac{1}{5}\sum_{i=1}^{10} x_i = 0 \rightarrow \frac{9}{5}x_k = \frac{1}{5}\sum_{i\ne k} x_i \Rightarrow x_k = \frac{1}{9}\sum_{i \ne k} x_i$



Min when $x_1 = x_2 = \dots = x_{10}$.

x_k	0	$\frac{1}{9}\sum_{i\ne k} x_i$	2
$\frac{\partial f}{\partial x_k}$	$-$	0	$+$
f	A $\searrow$	m	$\nearrow$ B



$\Rightarrow f$ is maximized when $x_k = 0$ or 2.

⊛ Let $g(x'_1, x'_2, \dots, x'_{10}) = \sum_{i=1}^{10}(x'_i - \bar{x}')^2$ where $x_i \in \{0, 2\}$

$\Rightarrow \max_{0\le x_i \le 2}\{f(x_1, x_2, \dots, x_{10})\} = \max_{x_i\in\{0,2\}} g(x_1, x_2, \dots, x_{10})$.

⊛ Note that $\begin{cases} \bar{x} = \frac{1}{10}(k_0 \times 0 + k_2 \times 2) = \frac{1}{5}k_2 \\ \sum_{i=1}^{10} x_i^2 = k_0 \times 0^2 + k_2 \times 2^2 = 4k_2 \end{cases}$, where k_0 and k_2 is the # of 0's and 2's respectively.

$\Rightarrow g(x_1, x_2, \dots, x_{10}) = 4k_2 - 10\left(\frac{1}{5}k_2\right)^2 = 4k_2 - \frac{2}{5}k_2^2$.

⊛ $4k_2 - \frac{2}{5}k_2^2$ achieve its maximum value when $k_2 = \frac{-4}{2\left(\frac{-2}{5}\right)} = \frac{-4}{-4/5} = 5 \Rightarrow k_0 = 10 - 5 = 5$.

⊛ Therefore, $g(x_1, x_2, \dots, x_{10})$ is maximized when there are 5 0's and 5 2's.

Thus, $f(x_1, x_2, \dots, x_{10})$ is also maximized when there are 5 0's and 5 2's.