# Linear Statistical Models Assignment 1

Kim Seang CHY

## Question 1

**a.** For a symmetric matrix $\mathbf{A}$, if $\mathbf{A}^2 = \mathbf{A}^3$, then $\mathbf{A}$ is idempotent.

There are two cases the first is the first is matrix $\mathbf{A}$ over $\mathbb{R}$ and the second case is matrix $\mathbf{A}$ over $\mathbb{C}$.

**Case 1:** Let $\mathbf{A}$ be a symmetric matrix over $\mathbb{R}$ such that $\mathbf{A}^2 = \mathbf{A}^3$.

By eigenvalue properties, $\mathbf{A}$ is diagonalisable with real eigenvalues and orthogonal eigenvectors, thus there exist a diagonalise matrix $\mathbf{P}$ such that:

$$\mathbf{P}^T\mathbf{A}\mathbf{P} = \mathbf{D}$$

where $\mathbf{D}$=diag$(\lambda_1, \lambda_2, \ldots, \lambda_i, \ldots)$ and $\lambda_i$ is the eigenvalue of A for all $i \in \mathbb{N} \setminus \{0\}$

Since $\mathbf{A}^2 = \mathbf{A}^3$, this implied:

$$\begin{aligned}
&\mathbf{D}^2 = \mathbf{D}^3 && \text{(Diagonalise relationship)}\\
&\implies (\lambda_i)^2 = (\lambda_i)^3,\ \forall i \in \mathbb{N} \setminus \{0\}\\
&\implies \lambda_i^2(\lambda_i - 1) = 0\\
&\implies \lambda_i = 0 \text{ or } \lambda_i = 1
\end{aligned}$$

Since the eigenvalue can only take value of 0 or 1. We can inferred:

$$\begin{aligned}
&\mathbf{D} = \mathbf{D}^n,\ n = 1, 2, 3\ldots && \text{(Diagonal matrix properties )}\\
&\implies \mathbf{A} = \mathbf{A}^n && \text{(By diagonalisaiton properties)}
\end{aligned}$$

Given $\mathbf{A} = \mathbf{A}^n$. Hence,$\mathbf{A}$ is an idempotent matrix by the definition of idempotent matrix and its diagonalise properties. The statement is true if $\mathbf{A}$ is a real matrix and symmetric.

**Case 2**: Proof by counter example.

Let

$$\mathbf{A} = \begin{bmatrix} -\mathbf{i} & 1 \\ 1 & \mathbf{i} \end{bmatrix}$$

Notice that $\mathbf{A} = \mathbf{A}^T$, this implied $\mathbf{A}$ is a symmetric matrix. Now notice that:

$$\mathbf{A}^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}$$

This implied $\mathbf{A}^3 = \mathbf{0}\mathbf{A} = \mathbf{0} = \mathbf{A}^2$. However, $\mathbf{A} \neq \mathbf{A}^2$, thus $\mathbf{A}$ is not idempotent. Hence, the statement is false for a complex symmetric matrix.

Therefore, a symmetric matrix $\mathbf{A}$ such that $\mathbf{A}^2 = \mathbf{A}^3$, implied $\mathbf{A}$ is idempotent if and only if $\mathbf{A}$ has all real element.

**b.** For a symmetric matrix $\mathbf{A}$, if $\mathbf{A} = \mathbf{A}^3$, then $\mathbf{A}$ is idempotent.

**Proof:** Proof by counter example.

$$\text{Let } \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Notice that $\mathbf{A} = \mathbf{A}^T$, this implied $\mathbf{A}$ is a symmetric by definition of symmetric.

$$\mathbf{A}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{A}^3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

$\mathbf{A} \neq \mathbf{A}^2$, hence $\mathbf{A}$ is not idempotent. Since $\mathbf{A} = \mathbf{A}^3$ but $\mathbf{A}$ is not idempotent. Therefore, the statement is false.

# Question 2:

Prove for $A_1, A_2, \ldots, A_m$ set of symmetric $k \times k$ matrices and there exist an orthogonal matrix $P$ such that $P^T A_i P$ is diagonal for all $i$ then $A_i A_j = A_j A_i$ for every pair of $i, j = 1, 2, \ldots, m$.

**Proof:**
Supposed $A_1$ and $A_2$ are symmetric $k \times k$ matrices, simultaneously diagonalisable by orthogonal matrix $P$. This implied there exist:

$$A_1 = PD_1P^T$$

and

$$A_2 = PD_2P^T$$

Where $D_1$ is the diagonal matrix contain its respective eigenvalue for $A_1$ and $D_2$ is the diagonal matrix contain its respective eigenvalue for $A_2$

Hence:

$$\begin{aligned} A_1A_2 &= PD_1P^TPD_2P^T \\ &= PD_1D_2P^T && (P^TP = I \text{ by othrogal matrix properties}) \end{aligned}$$

Since diagonal matrix are commutable this implied $D_1D_2 = D_2D_1$. Therefore:

$$\begin{aligned} A_1A_2 &= PD_2D_1P^T \\ &= PD_2P^TPD_1P = A_2A_1 \end{aligned}$$

The above can expand to be show for any set of symmetric matrix that is simultaneously diagonalisable by the same orthogonal matrix. Therefore, if a set of symmetric matrices $A_1, A_2, \ldots, A_m$ is simultaneously diagonalisable by $P$ then $A_iA_j = A_jA_i$ for every pair of $i, j = 1, 2, \ldots, m$.

# Question 3:

Let $\mathbf{A}$ be an $n \times p$ matrix with rank p and all element belong to $\mathbb{R}$. This implied column vector of $\mathbf{A}$ is linearly independent. Let $\mathbf{y}$ be a $p \times 1$ matrix such that $\mathbf{y} = [y_1, y_2, \ldots, y_p]^T$.

Note that:

$$\mathbf{y}^T\mathbf{A}^T\mathbf{A}\mathbf{y} = (\mathbf{A}\mathbf{y})^T\mathbf{A}\mathbf{y} \qquad \text{(By the symmetric properties)}$$

For $\mathbf{Ay}$ matrix, it can be represented in a from $\mathbf{Ay} = \mathbf{a}_1y_1 + \mathbf{a}_2y_2 + \ldots + \mathbf{a}_py_p$ where $\mathbf{a}_i$ represent the column vectors for matrix $\mathbf{A}$ for column $i$ where $i = 1, 2, \ldots, p$.

Since the column are linearly independent, this implied there a non-zero solution to $\mathbf{Ay}$=0 expect for $\mathbf{y} = \mathbf{0}$, thus implied there are also non-zero solution to$\mathbf{y}^T\mathbf{A}^T\mathbf{Ay}$=0. Since $\mathbf{y}^T\mathbf{A}^T\mathbf{Ay} = (\mathbf{a}_1y_1)^2 + (\mathbf{a}_2y_2)^2 + \ldots + (\mathbf{a}_py_p)^2$ , this this is just a sum of squares, which implied $\mathbf{y}^T\mathbf{A}^T\mathbf{Ay} > 0$ for all $y \in \mathbb{R}\backslash\{0\}$. Therefore, $\mathbf{y}^T\mathbf{A}^T\mathbf{Ay}$ is positively defined matrix.

# Question 4:

**a.** Let $\mathbf{y}$=$A\mathbf{x}$. Find $\mathbf{A}$.

Want to find a vector A such that: $\mathbf{y} = A\mathbf{x} = \mathbf{x} - \mathbf{1}\bar{x}$ where $\mathbf{1}$ is a column vector of 1 with three rows.
Notice that:

$$\begin{aligned} \bar{x} &= \frac{1}{3}\mathbf{1}^T\mathbf{x} \\ \implies \mathbf{x} - \bar{x}\mathbf{1} &= \mathbf{x} - \frac{1}{3}\mathbf{1}\mathbf{1}^T\mathbf{x} \\ &= (\mathbf{I} - \frac{1}{3}\mathbf{1}\mathbf{1}^T)\mathbf{x} \\ \implies A &= \mathbf{I} - \frac{1}{3}\mathbf{1}\mathbf{1}^T \end{aligned}$$

$$\mathbf{I} - \frac{1}{3}\mathbf{1}\mathbf{1}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T = \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix}$$

Hence

$$A = \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix}$$

**b.** Find the rank of A.

The rank(A) = 2, since the sum of column 1 and column 2 is equal to negative of column 3 and column 1 is not a multiple of column 2.

**c.** Find E($\mathbf{y}^T\mathbf{y}$).

$$\mathrm{E}(\mathbf{y}^T\mathbf{y}) = E\left(\begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \end{bmatrix}^T \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ x_3 - \bar{x} \end{bmatrix}\right)$$

$$\begin{aligned} &= \mathrm{E}\left(\sum_{i=1}^{3}(x_i - \bar{x})^2\right) \\ &= \mathrm{E}\left[\sum_{i=1}^{3} x_i\right] - \mathrm{E}\left[3\bar{x}^2\right] \\ &= \sum_{i=1}^{3} E[x_i^2] - 3\mathrm{E}[\bar{x}^2] \end{aligned}$$

Notice that Var($x$)=E($x^2$) − (E($x$))$^2$, this implied E($x^2$) = Var($x$) + (E($x$))$^2$ = $\sigma^2 + \mu^2$. Similarly, Var($\bar{x}$)=E($\bar{x}^2$) − (E($\bar{x}$))$^2$, this implied E($\bar{x}^2$) = Var($\bar{x}$) + E($\bar{x}$)$^2$ = $\frac{\sigma^2}{3} + \mu^2$, since $x_1, x_2, x_3$ is IID Therefore,

$$\begin{aligned} \mathrm{E}(\mathbf{y}^T\mathbf{y}) &= \left[\sum_{i=1}^{3}(\sigma^2 + \mu^2) - 3\left(\frac{\sigma^2}{3} + \mu^2\right)\right] \\ &= \left[(\sigma^2 + \mu^2) - 3\left(\frac{\sigma^2}{3} + \mu^2\right)\right] = \left[2\sigma^2\right] \end{aligned}$$

**d.** Using theorem 3.5, find the distribution of $\frac{\mathbf{y}^T\mathbf{y}}{\sigma^2}$.

From part 1 we know $\mathbf{y} = A\mathbf{x}$ and $\mathbf{x} \sim MVN(\mu, \sigma^2)$ is $3 \times 1$ random vectors and

$$\begin{aligned} \mathrm{A} &= \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix} \\ \implies A^2 &= \begin{bmatrix} \frac{2}{3} & \frac{-1}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3} \\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3} \end{bmatrix} \end{aligned}$$

Notice that $A = A^T$ and $A = A^2$, this implies $A$ is a symmetric and idempotent matrix. Thus:

$$\begin{aligned}\frac{\mathbf{y}^T\mathbf{y}}{\sigma^2} &= \frac{1}{\sigma^2}(A\mathbf{x})^T(A\mathbf{x}) \\ &= \frac{1}{\sigma^2}\mathbf{x}^T A^T A\mathbf{x} \\ &= \frac{1}{\sigma^2}\mathbf{x}^T A^2\mathbf{x} \\ &= \frac{\mathbf{x}^T}{\sigma}A\frac{\mathbf{x}}{\sigma}\end{aligned}$$

Since $\sigma$ is a constant this implied $\frac{1}{\sigma}\mathbf{x} \sim MVN(\frac{1}{\sigma}\mu, \sigma^2)$. Therefore, by Theorem 3.5, implied $\frac{\mathbf{y}^T\mathbf{y}}{\sigma^2}$ has a noncentral $\chi^2$ with rank$(A) = 2$ degree of freedoms, and a noncentral parameter $\lambda = \frac{1}{2}\left(\frac{\boldsymbol{\mu}}{\sigma}\right)^T A\left(\frac{\boldsymbol{\mu}}{\sigma}\right) = \frac{1}{2\sigma^2}\boldsymbol{\mu}^T A\boldsymbol{\mu}$. In this case:

$$\lambda = \frac{1}{2\sigma^2}\begin{bmatrix}\mu\\ \mu\\ \mu\end{bmatrix}^T\begin{bmatrix}\frac{2}{3} & \frac{-1}{3} & \frac{-1}{3}\\ \frac{-1}{3} & \frac{2}{3} & \frac{-1}{3}\\ \frac{-1}{3} & \frac{-1}{3} & \frac{2}{3}\end{bmatrix}\begin{bmatrix}\mu\\ \mu\\ \mu\end{bmatrix} = [0]$$

Since $\lambda = 0$, this implied $\frac{\mathbf{y}^T\mathbf{y}}{\sigma^2}$ is regular $\chi^2$ distribution with 2 degree of freedom.

# Question 5:

All the matrix calculation and other calculation for this question will be in appendix.

**a.** Write down the matrices and vector involved in the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

Since, we are trying to model the price of fish in 1980, based on the 1970 price then the predictor variables($y_i$) is the price of fish in 1980 and the explanation variable $x_{i2}$ is the price of fish in 1970. The model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is given by:

$$\begin{bmatrix}27.3\\42.4\\38.7\\4.5\\23\\166.3\\109.7\\80.1\\150.7\\20.3\\189.7\\131.3\\404.2\\149\end{bmatrix} = \begin{bmatrix}1 & 13.1\\1 & 15.3\\1 & 25.8\\1 & 1.8\\1 & 4.9\\1 & 55.4\\1 & 39.3\\1 & 26.7\\1 & 47.5\\1 & 6.6\\1 & 94.7\\1 & 61.1\\1 & 135.6\\1 & 47.6\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} + \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_{14}\end{bmatrix}$$

Where $\beta_0, beta_1$ are the parameter and $\varepsilon_i$ are the error term in row i.

**b.** Find the least squares estimators of the parameters.

$$\text{Let } \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

By theorem 4.4, the best least estimator is give by: $\mathbf{b} = (X^TX)^{-1}X^T\mathbf{y}$.

$$X^TX = \begin{bmatrix} 14.0 & 575.4 \\ 575.4 & 42079.36 \end{bmatrix} \text{ and } (X^TX)^{-1} = \begin{bmatrix} 0.16308 & -0.0022 \\ -0.0022 & 5.43\text{e-}5 \end{bmatrix}$$

$$\begin{aligned} \implies \mathbf{b} &= (X^TX)^{-1}X^T\mathbf{y} \\ &= \begin{bmatrix} -1.2338 \\ 2.7016 \end{bmatrix} \end{aligned}$$

The least square estimator $b_0 = -1.2338$ with $\text{var}(b_0) = 0.163\sigma^2$ and $b_1 = 2.7016$ with $\text{var}(b_1) = (5.43\text{e-}5)\sigma^2$.

**c.** Find the sample variances $s^2$.

By theorem 4.6, the unbiased for $\sigma^2$ is given by:

$$s^2 = \frac{(\mathbf{y} - X\mathbf{b})^T(\mathbf{y} - X\mathbf{b})}{n - (k+1)}$$

Where n is row of $\mathbf{x}$ and $\mathbf{y}$ and $k+1$ is the column of $\mathbf{x}$. Hence for our data sets

$$\begin{aligned} s^2 &= \frac{(\mathbf{y} - X\mathbf{b})^T(\mathbf{y} - X\mathbf{b})}{14 - 2} \\ &= \frac{9325.833}{12} = 777.1528 \end{aligned}$$

The sample variance for our model is 777.1528.

Since the variance of the parameter is given by $\text{var}(\mathbf{b}) = (X^TX)^{-1}\sigma^2$ or $(X^TX)^{-1}s^2$, this implied $\text{var}(b_0) = 126.74$ and $\text{var}(b_1) = 0.04217$

**d.** Predict the price for ocean trout in 1980 given a fisher sold ocean trout for 28c/pound in 1970.

$$\text{Let } t = \begin{bmatrix} 1 \\ 28 \end{bmatrix}$$

By theorem 4.5, we can uses $\mathbf{t}^T\mathbf{b}$ we can use to predict the price as it is best linear unbiased estimator for $\mathbf{t}^T\boldsymbol{\beta}$.
Therefore, to predict the of ocean trout given fisher sold ocean trout for 28c/pound and our model:

$$\begin{aligned} \text{price}_{1980} &= -1.2338 + 2.7016 \text{ price}_{1970} \\ &= \begin{bmatrix} 1 \\ 28 \end{bmatrix}^T \begin{bmatrix} -1.2338 \\ 2.7016 \end{bmatrix} = [74.40965] \end{aligned}$$

The price for ocean trout the fisherman expected get in 1980 is \$74.41c /pound.

**e.** Calculate the standardised residual for sea scallops.

Let $H = X(X^TX)^{-1}X^T$ and $\mathbf{e} = \mathbf{y} - X\mathbf{b}$. Hence $H_{13\ 13} = 0.5560$ and $\mathbf{e}_{13\ 1} = 39.10$ since the sea scallop is in the 13th index of our data

The standardised residual is given by:

$$
\begin{aligned}
z_i &= \frac{e_{1i}}{\sqrt{s^2(1 - H_{ii})}} \\
\implies z_{13} &= \frac{39.10}{\sqrt{777.1528(1 - 0.5560)}} = 2.104999
\end{aligned}
$$

The standardised residual for sea scallops is 2.1050.

**f.** Calculate the Cook's distance for sea scallops.

The Cook's distance for sea scallops is:

$$
\begin{aligned}
D_{13} &= \frac{z_{13}^2}{1+1}\left(\frac{H_{13\ 13}}{1 - H_{13\ 13}}\right) \\
&= \frac{2.1050^2}{2}\left(\frac{0.5560}{1 - 0.5560}\right) = 2.7740
\end{aligned}
$$

**g.** Does sea scallops fit the linear model? Justify your argument.

The sea scallops Cook's distance is 2.774 which is significantly greater than 1 thus implied the sea scallops observation is a point of high leverage and high standardised residual. Such point is considered to be an influential outlier that significant impact the fit a model hence the sea scallops does not fit the linear model.

This was also reflected in best fit line of the graph below where the best fit line without sea scallops observation is steeper than the best fit line with sea scallops observation.