PS2

Sam Olson

Problem 1

Suppose $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}^T = [1 \quad 2 \quad 3]$

$$\boldsymbol{\mu}^T = [1 \quad 2 \quad 3] \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 3 \end{bmatrix}$$

Further, define a 3×3 matrix A and a 2×3 matrix B as follows

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix}$$

a)

Determine the distribution of $u = \mathbf{1}_3^T \mathbf{y}$.

Mean of u:

$$E[u] = \mathbf{1}_3^T \boldsymbol{\mu} = [1, 1, 1] \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 + 2 + 3 = 6$$

Variance of u:

$$\text{Var}(u) = \mathbf{1}_3^T \boldsymbol{\Sigma} \mathbf{1}_3 = [1, 1, 1] \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = [1, 1, 1] \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} = 2 + 4 + 3 = 9$$

Since u is a linear combination of normally distributed variables, it follows a normal distribution with mean 6 and variance 9, i.e. the distribution of u as defined is:

$$u \sim \mathcal{N}(6, 9)$$

b)

Determine the distribution of $\mathbf{v} = \mathbf{A}\mathbf{y}$.

As defined, we start by substituting the givens, specifically using $\mathbf{v} = \mathbf{A}\mathbf{y}$:

$$\mathbf{v} \sim \mathcal{N}\left(\begin{bmatrix} 9 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 13 & 3 & -1 \\ 3 & 5 & -2 \\ -1 & -2 & 6 \end{bmatrix}\right)$$

Mean of $\mathbf{v}$:

$$E[\mathbf{v}] = \mathbf{A}\boldsymbol{\mu} = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ -2 \\ -1 \end{bmatrix}$$

Covariance of $\mathbf{v}$:

$$\text{Cov}(\mathbf{v}) = \mathbf{A\Sigma A}^T$$

Taking the first part of this expression and evaluating $\mathbf{A\Sigma}$:

$$\mathbf{A\Sigma} = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 7 & 3 \\ 3 & 0 & -4 \\ -2 & 1 & -2 \end{bmatrix}$$

Then, we take that matrix to get $\mathbf{A\Sigma A}^T$:

$$\mathbf{A\Sigma A}^T = \begin{bmatrix} 5 & 7 & 3 \\ 3 & 0 & -4 \\ -2 & 1 & -2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 0 \\ 2 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 27 & 2 & 4 \\ 2 & 7 & 4 \\ 4 & 4 & 3 \end{bmatrix}$$

Since $\mathbf{v}$ is a linear transformation of $\mathbf{y}$, it follows a multivariate normal distribution with the above mean and covariance, i.e. we may describe the distribution of $\mathbf{V}$ as:

$$\mathbf{v} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A}^T)$$

c)

Determine the distribution of $\mathbf{w}$, where $\mathbf{w}^T = [\mathbf{Ay} \quad \mathbf{By}]$.

We start by using the given information, specifically:

$$\mathbf{w} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{A}\boldsymbol{\mu} \\ \mathbf{B}\boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \mathbf{A\Sigma A}^T & \mathbf{A\Sigma B}^T \\ \mathbf{B\Sigma A}^T & \mathbf{B\Sigma B}^T \end{bmatrix}\right)$$

We just need to calculate some unknown quantities, the mean and covariance matrices of $\mathbf{w}$. To that end, we note:

The mean of $\mathbf{w}$ can be taken from part (b), $\mathbb{E}[\mathbf{Ay}] = \begin{bmatrix} 9 \\ -2 \\ -1 \end{bmatrix}$.

We then compute $\mathbb{E}[\mathbf{By}] = \mathbf{B}\boldsymbol{\mu}$:

$$\mathbf{B}\boldsymbol{\mu} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$$

Taken together this gives us:

$$E[\mathbf{w}] = \begin{bmatrix} \mathbf{A}\boldsymbol{\mu} \\ \mathbf{B}\boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} 9 \\ -2 \\ -1 \\ 6 \\ 1 \end{bmatrix}$$

We then calculate the covariance of $\mathbf{w}$:

Again, taking information from part (b), we already know $\text{Cov}(\mathbf{Ay}) = \mathbf{A\Sigma A}^T = \begin{bmatrix} 27 & 2 & 4 \\ 2 & 7 & 4 \\ 4 & 4 & 3 \end{bmatrix}$.

Compute $\text{Cov}(\mathbf{By}) = \mathbf{B\Sigma B}^T$:

$$\mathbf{B\Sigma} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 3 \\ -1 & 1 & 0 \end{bmatrix}$$

Using this, we then have:

$$\mathbf{B\Sigma B}^T = \begin{bmatrix} 2 & 4 & 3 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 9 & 2 \\ 2 & 2 \end{bmatrix}$$

We then compute $\text{Cov}(\mathbf{Ay}, \mathbf{By}) = \mathbf{A\Sigma B}^T$:

```
A <- matrix(c(5, 7, 3,
              3, 0, -4,
              2, 1, -2),
            nrow = 3, byrow = TRUE)

B <- matrix(c(1, 1, 1,
              -1, 1, 0),
            nrow = 2, byrow = TRUE)

ASigmaBT <- A %*% t(B)
ASigmaBT
```

```
##      [,1] [,2]
## [1,]   15    2
## [2,]   -1   -3
## [3,]    1   -1
```

$$\mathbf{A\Sigma B}^T = \begin{bmatrix} 5 & 7 & 3 \\ 3 & 0 & -4 \\ 2 & 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 15 & 2 \\ -1 & -3 \\ 1 & -1 \end{bmatrix}$$

The full covariance matrix is then given by:

$$
\text{Cov}(\mathbf{w}) = \begin{bmatrix} \mathbf{A\Sigma A}^T & \mathbf{A\Sigma B}^T \\ \mathbf{B\Sigma A}^T & \mathbf{B\Sigma B}^T \end{bmatrix} = \begin{bmatrix} 27 & 2 & 4 & 15 & 2 \\ 2 & 7 & 4 & -1 & -3 \\ 4 & 4 & 3 & 1 & -1 \\ 15 & -1 & 1 & 9 & 2 \\ 2 & -3 & -1 & 2 & 2 \end{bmatrix}
$$

Since $\mathbf{w}$ is a joint linear transformation of $\mathbf{y}$, it follows a multivariate normal distribution with the derived mean and covariance.

Overall, this gives us the distribution of $\mathbf{w}$:

$$
\mathbf{w} \sim \mathcal{N}\left(\begin{bmatrix} 9 \\ -2 \\ -1 \\ 6 \\ 1 \end{bmatrix}, \begin{bmatrix} 27 & 2 & 4 & 15 & 2 \\ 2 & 7 & 4 & -1 & -3 \\ 4 & 4 & 3 & 1 & -1 \\ 15 & -1 & 1 & 9 & 2 \\ 2 & -3 & -1 & 2 & 2 \end{bmatrix} \right)
$$

d)

Which of the distributions obtained in (a)–(c) are singular distributions? Recall that a distribution is singular if $\mathbf{\Sigma}$ is non-negative definite. Note that there are many algebraic properties of $\mathbf{\Sigma}$ that can be used to show that $\mathbf{\Sigma}$ is singular/nonsingular.

```r
Sigma_a <- matrix(9, nrow = 1, ncol = 1)

Sigma_b <- matrix(c(13, 3, -1,
                    3, 5, -2,
                    -1, -2, 6), nrow = 3, byrow = TRUE)

Sigma_c <- matrix(c(27, 2, 4, 15, 2,
                    2, 7, 4, -1, -3,
                    4, 4, 3, 1, -1,
                    15, -1, 1, 9, 2,
                    2, -3, -1, 2, 2), nrow = 5, byrow = TRUE)

det_a <- det(Sigma_a)
det_b <- det(Sigma_b)
det_c <- det(Sigma_c)

det_a
```

```
## [1] 9
```

```r
det_b
```

```
## [1] 291
```

```r
det_c
```

```
## [1] -1.068891e-29
```

```r
eigen_a <- eigen(Sigma_a)$values
eigen_b <- eigen(Sigma_b)$values
eigen_c <- eigen(Sigma_c)$values

eigen_a == 0
```

```
## [1] FALSE
```

```r
eigen_a > 0
```

```
## [1] TRUE
```

```r
eigen_b == 0
```

```
## [1] FALSE FALSE FALSE
```

```r
eigen_b > 0
```

```
## [1] TRUE TRUE TRUE
```

```r
eigen_c == 0
```

```
## [1] FALSE FALSE FALSE FALSE FALSE
```

```r
eigen_c > 0
```

```
## [1]  TRUE  TRUE  TRUE FALSE FALSE
```

Distribution in a):

$u \sim \mathcal{N}(6, 9)$.

The covariance matrix is a positive scalar, meaning it is positive definite, which implies non-negative definite. Thus, u is singular.

Given the distribution in b):

$$\mathbf{v} \sim \mathcal{N}\left(\begin{bmatrix} 9 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 13 & 3 & -1 \\ 3 & 5 & -2 \\ -1 & -2 & 6 \end{bmatrix}\right)$$

The covariance matrix is positive definite as all eigenvalues are strictly positive (and non-zero), implying non-negative definite. Thus, v is singular.

Given the distribution in c):

$$\mathbf{w} \sim \mathcal{N}\left(\begin{bmatrix} 9 \\ -2 \\ -1 \\ 6 \\ 1 \end{bmatrix}, \begin{bmatrix} 27 & 2 & 4 & 15 & 2 \\ 2 & 7 & 4 & -1 & -3 \\ 4 & 4 & 3 & 1 & -1 \\ 15 & -1 & 1 & 9 & 2 \\ 2 & -3 & -1 & 2 & 2 \end{bmatrix}\right)$$

The determinant is zero, meaning $\mathbf{\Sigma}$ is singular. Additionally, eigenvalues of the covariance matrix contain negative values, so the matrix is *not* non-negative definite, making w non-singular.

Summary:

All the distributions a) and b) are singular distributions, and c) is non-singular.

Problem 2

Suppose $\mathbf{X}$ and $\mathbf{W}$ are any two matrices with n rows for which $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{W})$. Show that $\mathbf{P_X} = \mathbf{P_W}$.

I'm unsure which of these is preferred, and generally apprehensive about how solid the first approach is, so I have both a Linear Algebra proof and also a more analytic algebraic proof. To that end:

Approach 1

The projection matrix $\mathbf{P_X}$ projects any vector onto the column space $\mathcal{C}(\mathbf{X})$.

Similarly, $\mathbf{P_W}$ projects any vector onto the column space $\mathcal{C}(\mathbf{W})$.

$\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{W})$, meaning the column spaces of $\mathbf{X}$ and $\mathbf{W}$ are identical.

Since $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{W})$, the projection matrices $\mathbf{P_X}$ and $\mathbf{P_W}$ must project onto the same subspace.

By the uniqueness property of projection matrices, $\mathbf{P_X} = \mathbf{P_W}$.

Approach 2 (The "better" way?)

The projection matrix $\mathbf{P_X}$ is given by:

$$\mathbf{P_X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

Similarly, $\mathbf{P_W}$ is:

$$\mathbf{P_W} = \mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T$$

Since $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{W})$, there exists a nonsingular matrix $\mathbf{C}$ such that $\mathbf{W} = \mathbf{XC}$.

So, given this, we may substitute $\mathbf{W} = \mathbf{XC}$ into $\mathbf{P_W}$:

$$\mathbf{P_W} = \mathbf{XC}\left((\mathbf{XC})^T(\mathbf{XC})\right)^{-1}(\mathbf{XC})^T$$

Simpifying gives us:

$$\mathbf{P_W} = \mathbf{XC}\left(\mathbf{C}^T\mathbf{X}^T\mathbf{XC}\right)^{-1}\mathbf{C}^T\mathbf{X}^T$$

Using the property of inverses, $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$ when $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are invertible (which we assume under the premise), we then have:

$$\mathbf{P_W} = \mathbf{XCC}^{-1}(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{C}^T)^{-1}\mathbf{C}^T\mathbf{X}^T$$

Since $\mathbf{CC}^{-1} = \mathbf{I}$ and $\mathbf{C}^T(\mathbf{C}^T)^{-1} = \mathbf{I}$, it follows:

$$\mathbf{P_W} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \mathbf{P_X}$$

Conclusion

Regardless of approach, but with preference to [Approach 2], suffice to say $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{W})$, then $\mathbf{P_X} = \mathbf{P_W}$.

Problem 3

Consider a competition among 5 table tennis players labeled 1 through 5. For $1 \leq i < j \leq 5$, define y_{ij} to be the score for player i minus the score for player j when player i plays a game against player j. Suppose for $1 \leq i < j \leq 5$,

$$y_{ij} = \beta_i - \beta_j + \epsilon_{ij},$$

where $\beta_1, \ldots, \beta_5$ are unknown parameters and the ϵ_{ij} terms are random errors with mean 0. Suppose four games will be played that will allow us to observe y_{12}, y_{34}, y_{25}, and y_{15}. Let

$$\mathbf{y} = \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{12} \\ \epsilon_{34} \\ \epsilon_{25} \\ \epsilon_{15} \end{bmatrix}$$

a)

Define a model matrix $\mathbf{X}$ so that model (1) may be written as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

For our the observed games y_{12}, y_{34}, y_{25}, and y_{15}, we model for each game with the form:

$$y_{ij} = \beta_i - \beta_j + \epsilon_{ij}$$

Each game is denoted y_{ij}, the corresponding row of $\mathbf{X}$ will have a 1 in the i-th column (for β_i), a -1 in the j-th column (for β_j), and 0 otherwise.

The model matrix $\mathbf{X}$ will have 4 rows (one for each game) and 5 columns (one for each player's parameter $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$). The rows of $\mathbf{X}$ are constructed as:

Observation 1, for y_{12}:

β_1 has a coefficient of 1, β_2 has a coefficient of -1, $\beta_3, \beta_4, \beta_5$ have coefficients of 0.

The row in the matrix $\mathbf{X}$ is $[1, -1, 0, 0, 0]$.

Observation 2, for y_{34}:

β_3 has a coefficient of 1, β_4 has a coefficient of -1, $\beta_1, \beta_2, \beta_5$ have coefficients of 0.

The row in the matrix $\mathbf{X}$ is $[0, 0, 1, -1, 0]$.

Observation 3, for y_{25}:

β_2 has a coefficient of 1, β_5 has a coefficient of -1, $\beta_1, \beta_3, \beta_4$ have coefficients of 0.

The row in the matrix $\mathbf{X}$ is $[0, 1, 0, 0, -1]$.

Observation 4, for y_{15}:

β_1 has a coefficient of 1, β_5 has a coefficient of -1, $\beta_2, \beta_3, \beta_4$ have coefficients of 0.

The row in the matrix $\mathbf{X}$ is $[1, 0, 0, 0, -1]$.

Assembling the rows defined above, we have our overall model matrix $\mathbf{X}$ as:

$$
\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}
$$

The model can now be written as:

$$
\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}
$$

where:

$$
\mathbf{y} = \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{12} \\ \epsilon_{34} \\ \epsilon_{25} \\ \epsilon_{15} \end{bmatrix}
$$

And the model matrix $\mathbf{X}$ is:

$$
\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}
$$

b)

Is $\beta_1 - \beta_2$ estimable? Prove that your answer is correct.

To determine whether $\beta_1 - \beta_2$ is estimable, we need to check if the vector $\mathbf{c} = [1, -1, 0, 0, 0]^\top$ lies in the row space of the model matrix $\mathbf{X}$. A linear function $\mathbf{c}^\top\boldsymbol{\beta}$ is estimable if and only if $\mathbf{c}$ can be expressed as a linear combination of the rows of $\mathbf{X}$.

From part a), the model matrix $\mathbf{X}$ is:

$$
\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}
$$

The vector $\mathbf{c}$ corresponding to $\beta_1 - \beta_2$ is:

$$
\mathbf{c} = [1, -1, 0, 0, 0]^\top
$$

We need to determine if $\mathbf{c}$ can be written as a linear combination of the rows of $\mathbf{X}$. That is, we need to find scalars a_1, a_2, a_3, a_4 such that:

$$
a_1[1, -1, 0, 0, 0] + a_2[0, 0, 1, -1, 0] + a_3[0, 1, 0, 0, -1] + a_4[1, 0, 0, 0, -1] = [1, -1, 0, 0, 0]
$$

This gives the system of equations:

1. $a_1 + a_4 = 1$ (for β_1),

2. $-a_1 + a_3 = -1$ (for β_2),
3. $a_2 = 0$ (for β_3),
4. $-a_2 = 0$ (for β_4),
5. $-a_3 - a_4 = 0$ (for β_5).

From equation 1: $a_1 + a_4 = 1$. From equation 2: $-a_1 + a_3 = -1$. From equation 3: $a_2 = 0$. From equation 4: $-a_2 = 0$, which is consistent with equation 3. From equation 5: $-a_3 - a_4 = 0$, which implies $a_3 = -a_4$.

Solving the system of equations, let $a_3 = -a_4$ into equation 2, giving:

$$-a_1 + (-a_4) = -1 \to -a_1 - a_4 = -1 \to a_1 + a_4 = 1$$

This is consistent with equation 1. Thus, the system has infinitely many solutions.

For example: Let $a_4 = 0$. Then $a_1 = 1$ and $a_3 = 0$. Let $a_4 = 1$. Then $a_1 = 0$ and $a_3 = -1$.

In either case, $\mathbf{c}$ can be expressed as a linear combination of the rows of $\mathbf{X}$.

Since $\mathbf{c}$ lies in the row space of $\mathbf{X}$, and the linear function $\beta_1 - \beta_2$ is estimable.

c)

Is $\beta_1 - \beta_3$ estimable? Prove that your answer is correct.

To determine whether $\beta_1 - \beta_3$ is estimable, we need to check if there exists a linear combination of the observed data $y_{12}, y_{34}, y_{25}, y_{15}$ that can express $\beta_1 - \beta_3$.

The model is given as it has previously, i.e., by:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

And with the design matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

By definition, a linear combination $\mathbf{c}^T\boldsymbol{\beta}$ is estimable if there exists a vector $\mathbf{a}$ such that:

$$\mathbf{c}^T = \mathbf{a}^T\mathbf{X}$$

For $\beta_1 - \beta_3$, the vector $\mathbf{c}$ is:

$$\mathbf{c} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

Such that we must identify/find a vector $\mathbf{a}$ such that:

$$\mathbf{c}^T = \mathbf{a}^T\mathbf{X}$$

To that end, we end up solving the system of equations given by:

$$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

This gives us the following equations:

1. $a_1 + a_4 = 1$ (for β_1),
2. $-a_1 + a_3 = 0$ (for β_2),
3. $a_2 = -1$ (for β_3),
4. $-a_2 = 0$ (for β_4),
5. $-a_3 - a_4 = 0$ (for β_5).

From equation 3, $a_2 = -1$. From equation 4, $-a_2 = 0$, which implies $a_2 = 0$. This is a contradiction, meaning there is no solution for $\mathbf{a}$ that satisfies all the equations, meaning that the linear combination $\beta_1 - \beta_3$ is not estimable based on the observed data $y_{12}, y_{34}, y_{25}, y_{15}$.

d)

Find a generalized inverse of $\mathbf{X}^\top \mathbf{X}$.

Start by noting again the design matrix $\mathbf{X}$ defined previously:

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

Note, the transpose of $\mathbf{X}$ is:

$$\mathbf{X}^\top = \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 \end{bmatrix}$$

Computing $\mathbf{X}^\top \mathbf{X}$, we have:

```
X <- matrix(c(1, -1, 0, 0, 0,
              0, 0, 1, -1, 0,
              0, 1, 0, 0, -1,
              1, 0, 0, 0, -1),
              nrow = 4,
              ncol = 5,
              byrow = TRUE)
XT <- t(X)
XTX <- XT %*% X

X
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1   -1    0    0    0
## [2,]    0    0    1   -1    0
## [3,]    0    1    0    0   -1
## [4,]    1    0    0    0   -1
```

```
XT
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    0    0    1
## [2,]   -1    0    1    0
## [3,]    0    1    0    0
## [4,]    0   -1    0    0
## [5,]    0    0   -1   -1
```

```
XTX
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    2   -1    0    0   -1
## [2,]   -1    2    0    0   -1
## [3,]    0    0    1   -1    0
## [4,]    0    0   -1    1    0
## [5,]   -1   -1    0    0    2
```

$$\mathbf{X}^\top\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 0 & 2 \end{bmatrix}$$

Using the above, then note, by definition, a generalized inverse $\mathbf{G}$ satisfies the relation:

$$\mathbf{X}^\top\mathbf{X}\mathbf{G}\mathbf{X}^\top\mathbf{X} = \mathbf{X}^\top\mathbf{X}$$

And

$$\mathbf{G} = (\mathbf{X}^\top\mathbf{X})^-$$

Note the method used in the previous problemset for calculating a generalized inverse (might be above, maybe below, knitr can be weird):

Using the above method gives us:

```
X <- matrix(c(1, -1, 0, 0, 0,
              0, 0, 1, -1, 0,
              0, 1, 0, 0, -1,
              1, 0, 0, 0, -1),
              nrow = 4,
              ncol = 5,
              byrow = TRUE)
XT <- t(X)
```

Finding a Generalized Inverse of a Matrix $\boldsymbol{A}$.

- Find any $n \times n$ nonsingular submatrix of $\boldsymbol{A}$ where $n = \text{rank}(\boldsymbol{A})$. Call this matrix $\boldsymbol{W}$.
- Invert and transpose $\boldsymbol{W}$, i.e., compute $(\boldsymbol{W}^{-1})^{\top}$.
- Replace each element of $\boldsymbol{W}$ in $\boldsymbol{A}$ with the corresponding element of $(\boldsymbol{W}^{-1})^{\top}$.
- Replace all other elements in $\boldsymbol{A}$ with zeros.
- Transpose the resulting matrix to obtain $\boldsymbol{G}$, a generalized inverse for $\boldsymbol{A}$.

Figure 1: CocoMelon

```
XTX <- XT %*% X
# Did a whole roundabout calculation, but this proved easiest
library(MASS)
qr(X)$rank
```

```
## [1] 3
```

```
G <- ginv(XTX)
round(G, digits = 2)
```

```
##       [,1]  [,2]  [,3]  [,4]  [,5]
## [1,]  0.22 -0.11  0.00  0.00 -0.11
## [2,] -0.11  0.22  0.00  0.00 -0.11
## [3,]  0.00  0.00  0.25 -0.25  0.00
## [4,]  0.00  0.00 -0.25  0.25  0.00
## [5,] -0.11 -0.11  0.00  0.00  0.22
```

```
# Verify generalized inverse property
XTX
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    2   -1    0    0   -1
## [2,]   -1    2    0    0   -1
## [3,]    0    0    1   -1    0
## [4,]    0    0   -1    1    0
## [5,]   -1   -1    0    0    2
```

```
mult <- XTX %*% G %*% XTX
round(mult, digits = 2)
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    2   -1    0    0   -1
## [2,]   -1    2    0    0   -1
## [3,]    0    0    1   -1    0
## [4,]    0    0   -1    1    0
## [5,]   -1   -1    0    0    2
```

```r
all.equal(mult, XTX)
```

```
## [1] TRUE
```

As a result of the above, one (of many) possible generalized inverse(s) of $\mathbf{X}^\top\mathbf{X}$ is:

$$\mathbf{G} = \begin{bmatrix} \frac{2}{9} & -\frac{1}{9} & 0 & 0 & -\frac{1}{9} \\ -\frac{1}{9} & \frac{2}{9} & 0 & 0 & -\frac{1}{9} \\ 0 & 0 & \frac{1}{4} & -\frac{1}{4} & 0 \\ 0 & 0 & -\frac{1}{4} & \frac{1}{4} & 0 \\ -\frac{1}{9} & -\frac{1}{9} & 0 & 0 & \frac{2}{9} \end{bmatrix}$$

Note: I did use the manual algorithm method to derive a generalized inverse. I just included the R for ease of reading and to easily validate that it is in fact a generalized inverse.

Another possible generalized inverse is:

$$\mathbf{G} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

For parts e), f), and g), I will use the latter matrix, primarily because it's much easier to use when multiplying/doing other matrix operations on it.

e)

Find a solution to the normal equations in this particular problem involving table tennis players.

The normal equations are given by:

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

From part d), we have:

$$\mathbf{X}^\top\mathbf{X} = \begin{bmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 1 \\ 0 & 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 & 2 \end{bmatrix}$$

Now, we compute:

$$\mathbf{X}^\top\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix} = \begin{bmatrix} y_{12} + y_{15} \\ -y_{12} + y_{25} \\ y_{34} \\ -y_{34} \\ -y_{25} - y_{15} \end{bmatrix}$$

Using the generalized inverse from part d), we have::

$$
\mathbf{G} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}
$$

Computing $\boldsymbol{\beta}$ then:

$$
\boldsymbol{\beta} = \mathbf{G}\mathbf{X}^\top \mathbf{y}
$$

Computing:

```
G <- matrix(c(2/3, 1/3, 0, 0, 0,
              1/3, 2/3, 0, 0, 0,
              0, 0, 1, 0, 0,
              0, 0, 0, 0, 0,
              0, 0, 0, 0, 0),
            nrow = 5, byrow = TRUE)

XT <- matrix(c(1,  0,  0,  0,  1,
               -1, 0,  1,  0,  0,
                0,  1,  0,  0,  0,
                0, -1,  0,  0,  0,
                0,  0, -1, -1,  0),
             nrow = 5, byrow = TRUE)

G %*% XT
```

```
##            [,1] [,2]      [,3] [,4]      [,5]
## [1,]  0.3333333    0 0.3333333    0 0.6666667
## [2,] -0.3333333    0 0.6666667    0 0.3333333
## [3,]  0.0000000    1 0.0000000    0 0.0000000
## [4,]  0.0000000    0 0.0000000    0 0.0000000
## [5,]  0.0000000    0 0.0000000    0 0.0000000
```

$$
\boldsymbol{\beta} = \mathbf{G}\mathbf{X}^\top \mathbf{y} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ -\frac{1}{3} & 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix} = \begin{bmatrix} \frac{1}{3}y_{12} + \frac{1}{3}y_{25} + \frac{2}{3}y_{55} \\ -\frac{1}{3}y_{12} + \frac{2}{3}y_{25} + \frac{1}{3}y_{55} \\ y_{34} \\ 0 \\ 0 \end{bmatrix}
$$

Thus, the solution to the normal equations is:

$$
\boldsymbol{\beta} = \mathbf{G}\mathbf{X}^\top \mathbf{y} = \begin{bmatrix} \frac{1}{3}y_{12} + \frac{1}{3}y_{25} + \frac{2}{3}y_{55} \\ -\frac{1}{3}y_{12} + \frac{2}{3}y_{25} + \frac{1}{3}y_{55} \\ y_{34} \\ 0 \\ 0 \end{bmatrix}
$$

As defined above.

f)

Find the Ordinary Least Squares (OLS) estimator of $\beta_1 - \beta_5$.

From the results of part e), we note:

$$\beta_1 = \frac{1}{3}y_{12} + \frac{1}{3}y_{25} + \frac{2}{3}y_{55}, \quad \beta_5 = 0$$

Thus, we compute:

$$\beta_1 - \beta_5 = \left(\frac{1}{3}y_{12} + \frac{1}{3}y_{25} + \frac{2}{3}y_{55}\right) - (0) = \frac{1}{3}y_{12} + \frac{1}{3}y_{25} + \frac{2}{3}y_{55}$$

So the OLS estimator for $\beta_1 - \beta_5$ as defined is:

$$\theta_{OLS} = \hat{\beta}_1 - \hat{\beta}_5 = \frac{1}{3}y_{12} + \frac{1}{3}y_{25} + \frac{2}{3}y_{55}$$

g)

Give a linear unbiased estimator of $\beta_1 - \beta_5$ that is not the OLS estimator.

Note our results from part e):

$$\beta_1 = \frac{1}{3}y_{12} + \frac{1}{3}y_{25} + \frac{2}{3}y_{55}, \quad \beta_5 = 0 \to \theta_{OLS} = \hat{\beta}_1 - \hat{\beta}_5 = \frac{1}{3}y_{12} + \frac{1}{3}y_{25} + \frac{2}{3}y_{55}$$

To construct/give an alternative and unbiased estimator that is not OLS, consider our initial matrices:

The design matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

And our other givens:

$$\mathbf{y} = \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{12} \\ \epsilon_{34} \\ \epsilon_{25} \\ \epsilon_{15} \end{bmatrix}$$

Where our model is given by:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Note then, using the model as defined, that we may construct a new estimator:

$$\hat{\theta} = y_{15}$$

Where:

Via the model as defined:

$$y_{15} = \beta_1 - \beta_5 + \epsilon_{15}$$

And it follows that:

$$E[\hat{\theta}] = E[y_{15}] = \beta_1 - \beta_5$$

Which is an alternative non-OLS estimator of $\beta_1 - \beta_5$.

A Quick Proof of Unbiasedness

Let's make sure that alternative estimator is unbiased.

Given:

$$E[y_{15}] = E[\beta_1 - \beta_5 + \epsilon_{15}] = E[\beta_1] - E[\beta_5] + E[\epsilon_{15}] = \beta_1 - \beta_5$$

Assuming $E[\epsilon_{15}] = 0$, and noting linearity of expectations.

So the new non-OLS estimator we created is an unbiased estimator of $\beta_1 - \beta_5$, as:

$$E[\hat{\theta}] = E[y_{15}] = \beta_1 - \beta_5$$

Problem 4

Consider a linear model for which

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}, \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{bmatrix}$$

a)

Obtain the normal equations for this model and solve them.

By the definition of normal equations:

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

And given the design matrix as specified above,

$\mathbf{X}^\top$ is:

$$\mathbf{X}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

We then compute $\mathbf{X}^T\mathbf{X}$:

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$$

This is a good matrix for us! Good in the sense that the diagonal elements are all 8 and 0 elsewhere (on the off diagonal).

We then note the given response vector and compute $\mathbf{X}^T\mathbf{y}$:

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix}$$

This results in:

$$\mathbf{X}^T\mathbf{y} = \begin{bmatrix} y_1 + y_2 + y_3 + y_4 + y_5 + y_6 - y_7 - y_8 \\ y_1 + y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8 \\ y_1 + y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8 \\ -y_1 - y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 \end{bmatrix}$$

Returning then to the normal equations:

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

We use the above calculations to derive:

$$\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} y_1 + y_2 + y_3 + y_4 + y_5 + y_6 - y_7 - y_8 \\ y_1 + y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8 \\ y_1 + y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8 \\ -y_1 - y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 \end{bmatrix}$$

Taking advantage of only diagonal elements being non-zero, we thus have:

$$\beta_1 = \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6 - y_7 - y_8}{8}$$

$$\beta_2 = \frac{y_1 + y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8}{8}$$

$$\beta_3 = \frac{y_1 + y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8}{8}$$

$$\beta_4 = \frac{-y_1 - y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8}{8}$$

The least squares estimates of $\boldsymbol{\beta}$ are:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} \frac{y_1+y_2+y_3+y_4+y_5+y_6-y_7-y_8}{8} \\ \frac{y_1+y_2+y_3+y_4-y_5-y_6+y_7+y_8}{8} \\ \frac{y_1+y_2-y_3-y_4+y_5+y_6+y_7+y_8}{8} \\ \frac{-y_1-y_2+y_3+y_4+y_5+y_6+y_7+y_8}{8} \end{bmatrix}$$

b)

Are all functions $\mathbf{c}^\top\boldsymbol{\beta}$ estimable? Justify your answer.

By definition, a linear function $\mathbf{c}^\top\boldsymbol{\beta}$ is estimable if and only if $\mathbf{c}$ lies in the row space of the design matrix $\mathbf{X}$.

Hand-in-hand (equally true) is: a linear function $\mathbf{c}^\top\boldsymbol{\beta}$ is estimable implies that $\mathbf{c}$ can be expressed as a linear combination of the rows of $\mathbf{X}$.

We are given the design matrix, $\mathbf{X}$:

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

By definition, the rank of $\mathbf{X}$ is the number of linearly independent rows (or columns). In the above design matrix, we have 4 unique rows and 4 unique columns, making the rank of $\mathbf{X}$ is 4. This means that $\mathbf{X}$ has full column rank.

This is a desired property to have (helpful for our purposes)! The implications of $\mathbf{X}$ having full column rank, is that: The normal equations $\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$ have a unique solution for $\boldsymbol{\beta}$, the row space of $\mathbf{X}$ spans the entire $\mathbb{R}^4$ space (since $\mathbf{X}$ has 4 linearly independent columns), and finally that any vector $\mathbf{c} \in \mathbb{R}^4$ can be expressed as a linear combination of the rows of $\mathbf{X}$.

Since $\mathbf{X}$ has full column rank, the row space of $\mathbf{X}$ spans $\mathbb{R}^4$. It then follows that any vector $\mathbf{c} \in \mathbb{R}^4$ can be expressed as a linear combination of the rows of $\mathbf{X}$. Therefore, all linear functions $\mathbf{c}^\top\boldsymbol{\beta}$ are estimable, i.e. under our assumption we ensure the definition of estimability.

c)

Obtain the least squares estimator of $\beta_1 + \beta_2 + \beta_3 + \beta_4$.

From part a), the least squares estimates of $\boldsymbol{\beta}$ are:

$$\beta_1 = \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6 - y_7 - y_8}{8}$$

$$\beta_2 = \frac{y_1 + y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8}{8}$$

$$\beta_3 = \frac{y_1 + y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8}{8}$$

$$\beta_4 = \frac{-y_1 - y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8}{8}$$

Then we note that the least squares estimator of a linear combination of the parameters is the same linear combination of the least squares estimates of the individual parameters. So for our purposes, we evaluate $\beta_1 + \beta_2 + \beta_3 + \beta_4$ using these estimates.

Adding the four estimates together:

$$\beta_1+\beta_2+\beta_3+\beta_4 = \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6 - y_7 - y_8}{8} + \frac{y_1 + y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8}{8} + \frac{y_1 + y_2 - y_3 - y_4 + y_5 + y_6 \cdots}{8}$$

Combining the terms:

$$\beta_1+\beta_2+\beta_3+\beta_4 = \frac{(y_1 + y_2 + y_3 + y_4 + y_5 + y_6 - y_7 - y_8) + (y_1 + y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8) + (y_1 + y_2 - y_3 - y_4 + y_5 \cdots}{8}$$

Sorry, I think this runs off the page, and I couldn't manage text wrapping in an R Markdown pdf file.

Simplifying terms:

$$\beta_1 + \beta_2 + \beta_3 + \beta_4 = \frac{2(y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8)}{8} = \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8}{4}$$

So the least squares estimator of $\beta_1 + \beta_2 + \beta_3 + \beta_4$ is:

$$\beta_1 + \widehat{\beta_2 + \beta}_3 + \beta_4 = \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8}{4}$$

Problem 5

Suppose the Gauss-Markov model with normal errors (GMMNE) holds.

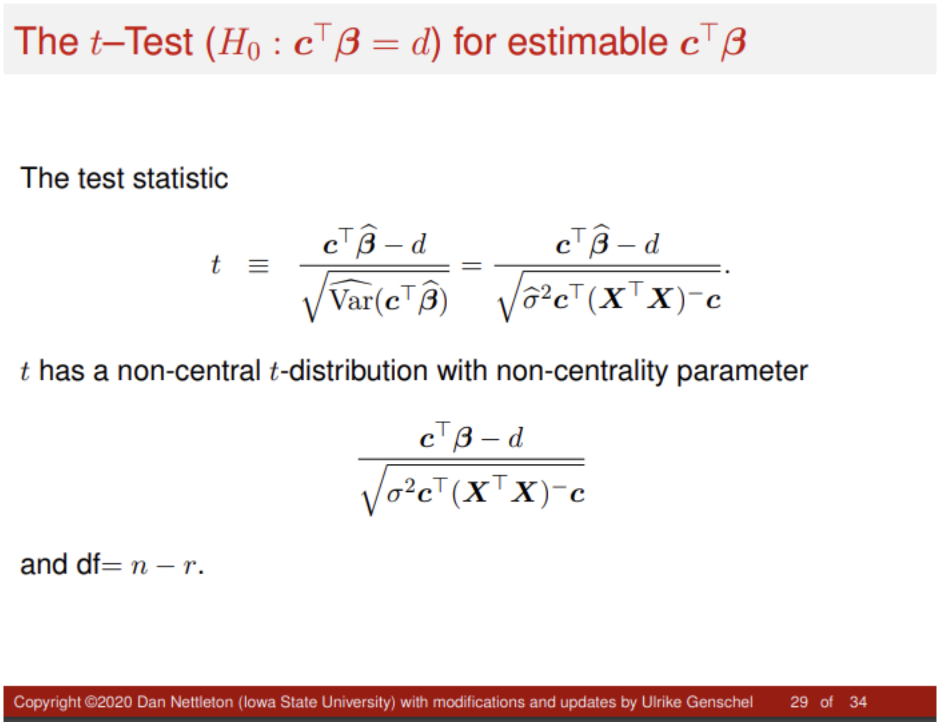

Figure 2: CocoMelon

a)

Suppose $\mathbf{C}\boldsymbol{\beta}$ is estimable. Derive the distribution of $\mathbf{C}\hat{\boldsymbol{\beta}}$, the OLSE of $\mathbf{C}\beta$.

Given the Gauss-Markov model with normal errors, i.e., assuming:

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}).$$

It follows that $\mathbf{C}\boldsymbol{\beta}$ is estimable, which by the definition of estimability means $\mathbf{C} = \mathbf{AX}$ for some matrix $\mathbf{A}$.

The OLS equation of $\boldsymbol{\beta}$ is given by the expression:

$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^- \mathbf{X}^\top \mathbf{y}$

where $(\mathbf{X}^\top \mathbf{X})^-$ is a generalized inverse.

Since $\hat{\boldsymbol{\beta}}$ is a linear transformation of $\mathbf{y}$, by the normality assumption:

$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \sigma^2 \mathbf{I})$

We then know:

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^-\right)$$

Because $\mathbf{C}\boldsymbol{\beta}$ is estimable, $\mathbf{C}\hat{\boldsymbol{\beta}}$ is also a linear transformation of $\hat{\boldsymbol{\beta}}$.

Consider then a linear combination with $\mathbf{C}$, where $\mathbf{C} = \mathbf{AX}$:

For the mean calculation, we have:

$$\mathbf{AP_X X}\beta = \boldsymbol{AX}\beta = \boldsymbol{C}\beta$$

For the variance calculation, we have:

$$\boldsymbol{AP_X}\sigma^2\boldsymbol{I}(\boldsymbol{AP_X})^\top = \sigma^2\boldsymbol{AP_X P_X^\top A^\top} = \sigma^2\boldsymbol{AP_X A^\top} = \sigma^2\boldsymbol{AX(X^\top X)^- X^\top A^\top} = \sigma^2\boldsymbol{C(X^\top X)^- C^\top}$$

Furthermore, knowing the distribution of $\hat{\boldsymbol{\beta}}$ is normal, we not only know that $\mathbf{C}\hat{\boldsymbol{\beta}}$ is also normally distribution, but has parameters given from the above calculations, namely:

$$\mathbf{C}\hat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\mathbf{C}\boldsymbol{\beta}, \sigma^2\mathbf{C}(\mathbf{X}^\top\mathbf{X})^-\mathbf{C}^\top\right)$$

b)

Now suppose $\mathbf{C}\boldsymbol{\beta}$ is NOT estimable. Provide a fully simplified expression for $\text{Var}\left(\mathbf{C}(\mathbf{X}^\top\mathbf{X})^\top\mathbf{X}^\top\mathbf{y}\right)$.

To determine the variance of $\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$, via our model assumptions, we still assume that:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

However, since my assumption $\mathbf{C}\boldsymbol{\beta}$ is not estimable, there does not exist a matrix $\mathbf{A}$ such that $\mathbf{C} = \mathbf{AX}$.

However, let us consider the variance of the linear transformation.For any linear transformation, regardless of our assumption of estimability, we may write:

$$\text{Var}(\mathbf{Ay}) = \mathbf{A}\text{Var}(\mathbf{y})\mathbf{A}^\top$$

Let:

$$\mathbf{A} = \mathbf{C}(\mathbf{X}^\top\mathbf{X})^-\mathbf{X}^\top$$

As $\text{Var}(\mathbf{y}) = \sigma^2\mathbf{I}$, it then follows:

$$\text{Var}\left(\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}\right) = (\mathbf{C}(\mathbf{X}^\top\mathbf{X})^-\mathbf{X}^\top)\boldsymbol{\sigma^2}\mathbf{I}(\mathbf{C}(\mathbf{X}^\top\mathbf{X})^-\mathbf{X}^\top)^\top = \sigma^2\mathbf{C}(\mathbf{X}^\top\mathbf{X})^-\mathbf{X}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^-\mathbf{C}^\top$$

And we cannot simplify any further and hence we cannot the distribution of $\mathbf{C}\hat{\boldsymbol{\beta}}$, the OLSE of $\mathbf{C}\beta$.

This is all assuming generalized inverse, not inverse in particular.

c)

Now suppose $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$ is testable and that $\mathbf{C}$ has only one row and $\mathbf{d}$ has only one element so that they may be written as $\mathbf{c}^\top$ and $\mathbf{d}$, respectively. Prove the result on slide 29 of slide set 2 of Key Linear Model Results.

Given the hypothesis $H_0 : \mathbf{c}^\top\boldsymbol{\beta} = d$ is testable, this implies that $\mathbf{c}^\top\boldsymbol{\beta}$ is estimable (linear transformation combination of estimable functions is itself estimable).

Under the assumption of GMMNE:

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

The test statistic for testing $H_0 : \mathbf{c}^\top\boldsymbol{\beta} = \mathbf{d}$ is given by:

$$t = \frac{\mathbf{c}^\top\hat{\boldsymbol{\beta}} - d}{\sqrt{\widehat{\mathrm{Var}}(\mathbf{c}^\top\hat{\boldsymbol{\beta}})}}$$

We'll return to the above shortly, but before then, note that under H_0, the test statistic t defined above follows a t-distribution with $n - r$ degrees of freedom, where r is the rank of $\mathbf{X}$.

Furthermore, from part a), we know:

$$\mathbf{c}^\top\hat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\mathbf{c}^\top\boldsymbol{\beta}, \sigma^2\mathbf{c}^\top(\mathbf{X}^\top\mathbf{X})^-\mathbf{c}\right)$$

Thus,

$$\frac{\boldsymbol{c}^\top\hat{\boldsymbol{\beta}} - d}{\sqrt{\sigma^2\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{c}}} \sim \mathcal{N}\left(\frac{\boldsymbol{c}^\top\boldsymbol{\beta} - d}{\sqrt{\sigma^2\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{c}}}, 1\right)$$

The estimated variance is given by:

$$t = \frac{\mathbf{c}^\top\hat{\boldsymbol{\beta}} - d}{\sqrt{\widehat{\mathrm{Var}}(\mathbf{c}^\top\hat{\boldsymbol{\beta}})}} \to \widehat{\mathrm{Var}}(\mathbf{c}^\top\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2\mathbf{c}^\top(\mathbf{X}^\top\mathbf{X})^-\mathbf{c}$$

where $\hat{\sigma}^2$ is the unbiased estimator of σ^2.

The above assumed the type of distribution for the test statistic t. To explicitly show it is in fact t-distributed,

Define:

$$u = \frac{\boldsymbol{c}^\top\hat{\boldsymbol{\beta}} - d}{\sqrt{\sigma^2\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{c}}}$$

$$\delta = \frac{\boldsymbol{c}^\top\boldsymbol{\beta} - d}{\sqrt{\sigma^2\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^-\boldsymbol{c}}}$$

Noting prior results of the proof thus far, $u \sim \mathcal{N}(\delta, 1)$ and δ is our non-centrality parameter (Euclidean distance).

Additionally, note the difference between u and δ is $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ respectively.

Then:

$$\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi^2_{n-r}}{n-r} \to w = \frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r}$$

With note that $\boldsymbol{c}^\top\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent, meaning u and w are independent (interpretation being the mean and variance, parameters of interest, are independent of one another).

Referring back to our initial test statistic:

$$t = \frac{u}{\sqrt{w/(n-r)}} = \frac{\boldsymbol{c}^\top \hat{\boldsymbol{\beta}} - d}{\sqrt{\hat{\sigma}^2 \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{c}}} \sim t_{n-r}(\delta)$$

Under H_0, $\delta = 0$, we can simplify the expression as:

$$t = \frac{\mathbf{c}^\top \hat{\boldsymbol{\beta}} - d}{\sqrt{\hat{\sigma}^2 \mathbf{c}^\top (\mathbf{X}^\top \mathbf{X})^- \mathbf{c}}}$$

I am assuming from here on that the proof is complete. Otherwise I would be replicating information from the Slides regarding a proof of Independence between $\boldsymbol{c}^\top \hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, which is a necessary but not guarenteed condition; for the purposes of this proof, and from our discussion, I will consider this point a ncessary condition of the above.

Problem 6

Provide an example that shows that a generalized inverse of a symmetric matrix need not be symmetric. (Comment: For this reason, we cannot assume that $(\mathbf{X}^\top\mathbf{X})^- = [(\mathbf{X}^\top\mathbf{X})^-]^\top$.)

A generalized inverse $\mathbf{A}^-$ of a matrix $\mathbf{A}$ satisfies the condition:

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$$

However, $\mathbf{A}^-$ need not be symmetric even if $\mathbf{A}$ is symmetric.

We start with a Symmetric Matrix $\mathbf{A}$:

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

Then, we have a Generalized Inverse $\mathbf{A}^-$ (that is not $\mathbf{A}$!). We need to ensure the generalized inverse property holds, $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$.

One possible generalized inverse we may have is:

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

First, we check that the necessary property of a generalized inverse holds:

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

Followed by $\mathbf{A}\mathbf{A}^-\mathbf{A}$:

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

So our chosen $\mathbf{A}^-$ satisfies the generalized inverse condition.

Let us then consider whether $\mathbf{A}^-$ is symmetric

The transpose of $\mathbf{A}^-$ is:

$$(\mathbf{A}^-)^\top = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

Notably,

$$\mathbf{A}^- = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = (\mathbf{A}^-)^\top$$

So $\mathbf{A}^-$ is not symmetric, even though $\mathbf{A}$ is symmetric!

This counterexample shows that a generalized inverse of a symmetric matrix need not be symmetric. Ding dang!