

1. 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 $1 \leq i < j \leq 5$,

$$y_{ij} = \beta_i - \beta_j + \epsilon_{ij}, \quad (1)$$

where β_1, \dots, β_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} .

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

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix} = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_3 - \beta_4 \\ \beta_2 - \beta_5 \\ \beta_1 - \beta_5 \end{bmatrix} + \begin{bmatrix} \epsilon_{12} \\ \epsilon_{34} \\ \epsilon_{25} \\ \epsilon_{15} \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix}}_{\mathbf{X}} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} + \begin{bmatrix} \epsilon_{12} \\ \epsilon_{34} \\ \epsilon_{25} \\ \epsilon_{15} \end{bmatrix} \end{aligned}$$

- (b) Is $\beta_1 - \beta_2$ estimable?

Let $\mathbf{c}' = (1, -1, 0, 0, 0)$, then $\beta_1 - \beta_2$ can be written as $\mathbf{c}'\boldsymbol{\beta}$. The linear combination $\mathbf{c}'\boldsymbol{\beta}$ is estimable, if there exists \mathbf{a}' so that $\mathbf{c}' = \mathbf{a}'\mathbf{X}$ by slide 7 of set 2. We can find such $\mathbf{a}' = (1, 0, 0, 0)$, so $\beta_1 - \beta_2$ is estimable. Alternatively, note that $\beta_1 - \beta_2$ is an element of $\mathbf{X}\boldsymbol{\beta}$, so it is estimable.

- (c) Is $\beta_1 - \beta_3$ estimable?

Let $\mathbf{c}'_2 = (1, 0, -1, 0, 0)$, then $\beta_1 - \beta_3$ can be written as $\mathbf{c}'_2\boldsymbol{\beta}$. If $\mathbf{c}'_2\boldsymbol{\beta}$ is estimable, then there must be an $\mathbf{a}'_2 = (a_1, a_2, a_3, a_4)$ so that

$$\begin{aligned} \mathbf{c}'_2 &= \mathbf{a}'_2\mathbf{X} \\ [1 \quad 0 \quad -1 \quad 0 \quad 0] &= [a_1 \quad a_2 \quad a_3 \quad a_4] \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix} \end{aligned}$$

which implies

$$1 = a_1 + a_4; \quad 0 = -a_1 + a_3; \quad \underbrace{-1 = a_2; \quad 0 = -a_2; \quad 0 = -a_3 - a_4}_{\text{contradiction!}}$$

So $\beta_1 - \beta_3$ is not estimable.

Comments: Some students claimed that they can not find an $\mathbf{a}' = \mathbf{c}'\mathbf{X}$ but without any further proof, you need show work to support your statement.

(d) Find a generalized inverse of $\mathbf{X}'\mathbf{X}$.

Use the R function `ginv` in the MASS package.

```
> X=matrix(c(1,0,0,1,-1,0,1,0,0,1,0,0,0,-1,0,0,0,0,-1,-1),nrow=4)
> MASS::ginv( t(X) %*% X)
```

a generalized inverse matrix $(\mathbf{X}'\mathbf{X})^-$ (not unique) is

$$(\mathbf{X}'\mathbf{X})^- = \begin{bmatrix} 0.2222 & -0.1111 & 0 & 0 & -0.1111 \\ -0.1111 & 0.2222 & 0 & 0 & -0.1111 \\ 0 & 0 & 0.25 & -0.25 & 0 \\ 0 & 0 & -0.25 & 0.25 & 0 \\ -0.1111 & -0.1111 & 0 & 0 & 0.2222 \end{bmatrix}$$

(e) Write down a general expression for the normal equations.

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$$

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

$$\begin{aligned} \mathbf{b} &= (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{y} \\ &= \begin{bmatrix} 0.2222 & -0.1111 & 0 & 0 & -0.1111 \\ -0.1111 & 0.2222 & 0 & 0 & -0.1111 \\ 0 & 0 & 0.25 & -0.25 & 0 \\ 0 & 0 & -0.25 & 0.25 & 0 \\ -0.1111 & -0.1111 & 0 & 0 & 0.2222 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix} \\ &= \begin{bmatrix} 1/3 & 0 & 0 & 1/3 \\ -1/3 & 0 & 1/3 & 0 \\ 0 & 1/2 & 0 & 0 \\ 0 & -1/2 & 0 & 0 \\ 0 & 0 & -1/3 & -1/3 \end{bmatrix} \begin{bmatrix} y_{12} \\ y_{34} \\ y_{25} \\ y_{15} \end{bmatrix} = \begin{bmatrix} y_{12}/3 + y_{15}/3 \\ -y_{12}/3 + y_{25}/3 \\ y_{34}/2 \\ -y_{34}/2 \\ -y_{25}/3 - y_{15}/3 \end{bmatrix} \end{aligned}$$

The solution is not unique since $(\mathbf{X}'\mathbf{X})^-$ is not unique.

Comments: In this problem, you need to write out \mathbf{b} as a 5×1 matrix involving y_{ij} instead of dot product of two matrices.

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

Let $\mathbf{c}' = (1, 0, 0, 0, -1)$ then $\beta_1 - \beta_5 = \mathbf{c}'\boldsymbol{\beta}$. The OLSE of $\mathbf{c}'\boldsymbol{\beta}$ is $\mathbf{c}'\mathbf{b}$ by slide 7 of set 2, based on the results in part (f),

$$\mathbf{c}'\mathbf{b} = (y_{12}/3 + y_{15}/3) - (-y_{25}/3 - y_{15}/3) = y_{12}/3 + 2y_{15}/3 + y_{25}/3$$

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

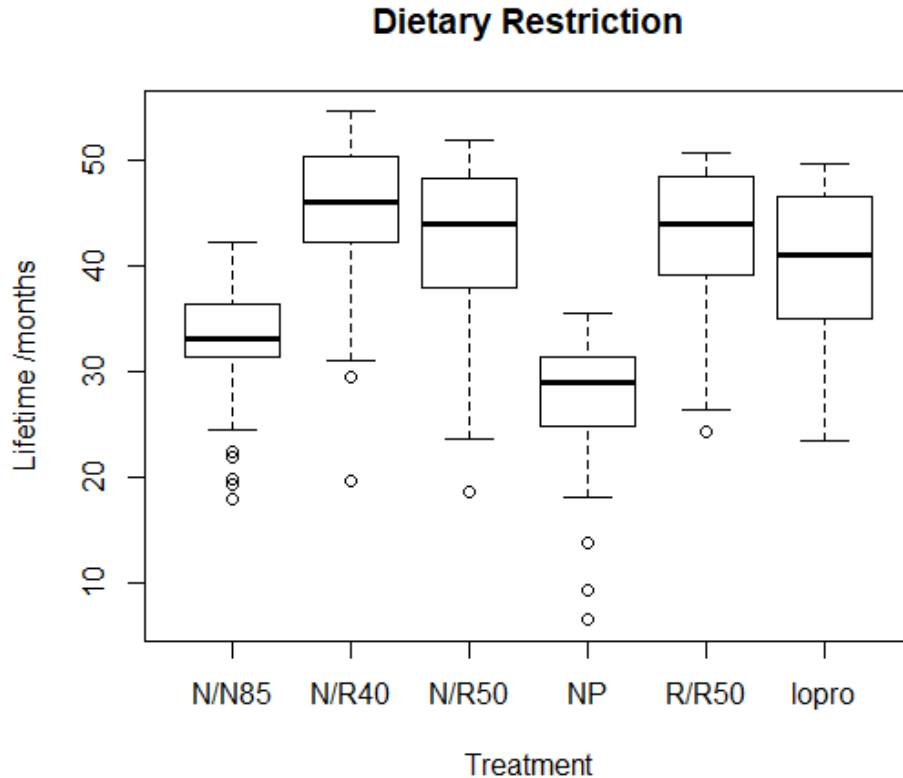
We need to find \mathbf{a}' such that $E(\mathbf{a}'\mathbf{y}) = \beta_1 - \beta_5$. The simplest one is y_{15} when $\mathbf{a}' = (0, 0, 0, 1)$. y_{15} is a linear unbiased estimator of $\beta_1 - \beta_5$, but it is not the OLSE of $\beta_1 - \beta_5$.

2. Case Study 5.1.1 from *The Statistical Sleuth*

```
> install.packages("Sleuth3")  
> library(Sleuth3)
```

(a) Create side-by-side boxplots of the response for this dataset, with one boxplot for each treatment group.

```
> boxplot(Lifetime~Diet,case0501, main = "Dietary Restriction",  
          xlab="Treatment", ylab = "Lifetime /months")
```



(b) Find the SSE (sum of squared errors) for the full model with one unrestricted mean for each of the six treatment groups.

From the code and output below,

$$SSE_{Full} = 15297.42.$$

```
> fit <- lm(Lifetime~Diet,case0501)  
> deviance(fit)  
[1] 15297.42
```

(c) Compute $\hat{\sigma}^2$ for the full model with 6 means.

$$\hat{\sigma}^2 = \frac{SSE_{Full}}{df_{Full}} = \frac{15297.42}{343} = 44.6.$$

```
> deviance(fit)/fit$df
[1] 44.59888
```

(d) Find the SSE for a reduced model that has one common mean for the N/N85, N/R50, N/R50lopro and N/R40 treatment groups and an unrestricted mean for each of the other two treatment groups.

From the code and output below,

$$SSE_{Reduced} = 20287.99$$

```
> ## merge levels for N/N85, N/R50, N/R50lopro and N/R40
> diet1 <- case0501$Diet
> levels(diet1)[c(1,3,6,2)] <- rep("N/N85+N/R50+N/R50lopro+N/R40", 4)
> newcase <- data.frame(Lifetime = case0501$Lifetime,diet1)

> fit1 <- lm(Lifetime~diet1,newcase)
> deviance(fit1)
[1] 20287.99
```

(e) Use the answers from parts (b) through (d) to compute an F statistic.

$$H_0 : E(\mathbf{y}) \in \mathcal{C}(\mathbf{X}_0)$$

$$H_a : E(\mathbf{y}) \in \mathcal{C}(\mathbf{X}) \setminus \mathcal{C}(\mathbf{X}_0)$$

From the code and output below,

$$F = \frac{(SSE_{Reduced} - SSE_{Full}) / (df_{Reduced} - df_{Full})}{SSE_{Full} / df_{Full}} = 37.3$$

```
> anova(fit1, fit)
```

Analysis of Variance Table

Model 1: Lifetime ~ diet1

Model 2: Lifetime ~ Diet

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	346	20288				
2	343	15297	3	4990.6	37.3	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(f) Explain to the scientists conducting this study what the F statistic in part (e) can be used to test. Consider the context of the study and use terms non-statistician scientists

will understand.

F statistic in part (e) indicates the variation contributed by the variable which is eliminated from the full model. It can be used to test whether the full model is significantly better than the reduced model, that is whether there is significant difference among the N/N85, N/R50, N/R50*lopro* and N/R40 treatment groups. In the context of this problem, F statistic can be used to test whether the expected lifetime is affected by different diets among mice who are treated with normal diet before weaning.

In this study, F statistic is 37.3 with p -value 0. We will reject the reduced model in favor of the full model at 0.05 level and conclude that there is significant difference among the N/N85, N/R50, N/R50*lopro* and N/R40 treatment group means.

- (g) Consider an F statistic of the form given on slide 20 of the notes entitled *A Review of Some Key Linear Model Results*. Provide the \mathbf{C} matrix and \mathbf{d} vector and compute the F statistic corresponding to the test of the hypotheses in part (e).

$$H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$$

$$H_a : \mathbf{C}\boldsymbol{\beta} \neq \mathbf{d}$$

where, $\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}$ according to the order of the treatments in dataset, $q = 3$ and $\mathbf{d} = \mathbf{0}$. (Other \mathbf{C} is possible)

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \mathbf{d})/q}{\hat{\sigma}^2}$$

and

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \hat{\sigma}^2 = \frac{\mathbf{y}'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{y}}{n - r}$$

From the code and output below, $F = 37.3$ the same result as in part (e).

```
> y=case0501$Lifetime
> I=diag(1, length(y))
> r=length(levels(case0501$Diet)) ## rank(X)

> xmat=model.matrix(~0+case0501$Diet)
> proj=function(x){x %*% MASS::ginv(t(x)%*%x) %*% t(x)}
> hat.sig2 = t(y) %*% (I-proj(xmat)) %*% y / (length(y)-r) ##hat.sigma^2
> hat.b=solve(t(xmat)%*%xmat) %*% t(xmat) %*% y ##hat.beta

> levels(case0501$Diet) # order of treatments
[1] "N/N85" "N/R40" "N/R50" "NP" "R/R50" "lopro"
> C <- matrix(c(1, -1, 0, 0, 0, 0,
               1, 0, -1, 0, 0, 0,
               1, 0, 0, 0, 0, -1), byrow = TRUE, nrow = 3)
```

```

> Fstat=t(C %*% hat.b) %*% solve(C %*% solve(t(xmat)%*%xmat) %*% t(C))
      %*% (C %*% hat.b)/3/hat.sig2
      [,1]
[1,] 37.29968

```

3. There are infinitely many possible examples. One example can be given using

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}.$$

We can verify that \mathbf{G} meets the definition of a generalized inverse of \mathbf{A} :

$$\mathbf{AGA} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{A}.$$

Clearly \mathbf{A} is symmetric (i.e., $\mathbf{A} = \mathbf{A}'$), yet \mathbf{G} is *not* symmetric:

$$\mathbf{G} = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & 3 \\ 2 & 0 \end{bmatrix} = \mathbf{G}'.$$

Hence, a generalized inverse of a symmetric matrix need not be symmetric.

4. We are given that for any two matrices \mathbf{U} and \mathbf{V} that allow for the product matrix \mathbf{UV} ,

$$\text{rank}(\mathbf{UV}) \leq \min \{ \text{rank}(\mathbf{U}), \text{rank}(\mathbf{V}) \} \quad (2)$$

This says $\text{rank}(\mathbf{UV})$ is no larger than the smaller of the two quantities $\text{rank}(\mathbf{U})$ and $\text{rank}(\mathbf{V})$, which implies

$$\text{rank}(\mathbf{UV}) \leq \text{rank}(\mathbf{U}) \quad \text{and} \quad \text{rank}(\mathbf{UV}) \leq \text{rank}(\mathbf{V}).$$

(a) Let \mathbf{X} be any matrix. Then,

$$\begin{aligned}
\text{rank}(\mathbf{X}'\mathbf{X}) &\leq \min \{ \text{rank}(\mathbf{X}'), \text{rank}(\mathbf{X}) \} && \text{by (??)} \\
&\leq \text{rank}(\mathbf{X}) \\
&= \text{rank}(\mathbf{P}_X \mathbf{X}) && \text{since } \mathbf{P}_X \mathbf{X} = \mathbf{X} \\
&= \text{rank}(\mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{X}) && \text{since } \mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}' \\
&\leq \min \{ \text{rank}(\mathbf{X}(\mathbf{X}'\mathbf{X})^-), \text{rank}(\mathbf{X}'\mathbf{X}) \} && \text{by (??)} \\
&\leq \text{rank}(\mathbf{X}'\mathbf{X}).
\end{aligned}$$

The above says that $\text{rank}(\mathbf{X}'\mathbf{X})$ is simultaneously no larger and no smaller than $\text{rank}(\mathbf{X})$, that is,

$$\text{rank}(\mathbf{X}) \leq \text{rank}(\mathbf{X}'\mathbf{X}) \quad \text{and} \quad \text{rank}(\mathbf{X}) \geq \text{rank}(\mathbf{X}'\mathbf{X}),$$

which implies that

$$\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}'\mathbf{X}).$$

Comments: An alternative solution can be given by appealing to some linear algebra facts. Some students approached this problem by stating that \mathbf{X} and $\mathbf{X}'\mathbf{X}$ must have the same null space because

$$\mathbf{X}\mathbf{z} = \mathbf{0} \iff \mathbf{X}'\mathbf{X}\mathbf{z} = \mathbf{0}. \quad (3)$$

However, a few additional steps are needed for this to establish the desired result. First, why does (??) imply that \mathbf{X} and $\mathbf{X}'\mathbf{X}$ have the same null space? For an $m \times n$ real-valued matrix \mathbf{A} , its null space, $\text{Null}(\mathbf{A})$, is defined as

$$\text{Null}(\mathbf{A}) = \{ \mathbf{z} \in \mathbb{R}^n : \mathbf{A}\mathbf{z} = \mathbf{0} \}.$$

From the definition, clearly (??) implies $\text{Null}(\mathbf{X}) = \text{Null}(\mathbf{X}'\mathbf{X})$. Now, how does this imply equality of ranks? The rank-nullity theorem says that for any $m \times n$ matrix \mathbf{A} ,

$$\text{rank}(\mathbf{A}) + \dim(\text{Null}(\mathbf{A})) = n.$$

Noting that if \mathbf{X} is $m \times n$, then $\mathbf{X}'\mathbf{X}$ is $n \times n$, so the rank-nullity theorem says

$$\text{rank}(\mathbf{X}) + \dim(\text{Null}(\mathbf{X})) = n = \text{rank}(\mathbf{X}'\mathbf{X}) + \dim(\text{Null}(\mathbf{X}'\mathbf{X})).$$

Since $\text{Null}(\mathbf{X}) = \text{Null}(\mathbf{X}'\mathbf{X})$, the two sets must have the same dimension, implying that $\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}'\mathbf{X})$.

(b) Let \mathbf{X} be any matrix. Then,

$$\begin{aligned} \text{rank}(\mathbf{X}) &= \text{rank}(\mathbf{P}_\mathbf{X}\mathbf{X}) && \text{since } \mathbf{P}_\mathbf{X}\mathbf{X} = \mathbf{X} \\ &\leq \min \{ \text{rank}(\mathbf{P}_\mathbf{X}), \text{rank}(\mathbf{X}) \} && \text{by (??)} \\ &\leq \text{rank}(\mathbf{P}_\mathbf{X}) \\ &= \text{rank}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}') && \text{since } \mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' \\ &\leq \min \{ \text{rank}(\mathbf{X}), \text{rank}((\mathbf{X}'\mathbf{X})^{-}\mathbf{X}') \} && \text{by (??)} \\ &\leq \text{rank}(\mathbf{X}). \end{aligned}$$

Inequality in both directions implies equality; therefore,

$$\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{P}_\mathbf{X}).$$

(c) Given $\mathbf{W}\mathbf{P}_\mathbf{X} = \mathbf{W}$, show $\text{rank}(\mathbf{W}\mathbf{X}) = \text{rank}(\mathbf{W})$.

$$\begin{aligned} \text{(i)} \quad \text{rank}(\mathbf{W}\mathbf{X}) &\leq \text{rank}(\mathbf{W}) && \text{by } \text{rank}(\mathbf{UV}) \leq \min \{ \text{rank}(\mathbf{U}), \text{rank}(\mathbf{V}) \} \\ \text{(ii)} \quad \text{rank}(\mathbf{W}) &= \text{rank}(\mathbf{W}\mathbf{P}_\mathbf{X}) && \text{by } \mathbf{W}\mathbf{P}_\mathbf{X} = \mathbf{W} \\ &= \text{rank}(\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}') && \text{by def. of } \mathbf{P}_\mathbf{X} \\ &\leq \text{rank}(\mathbf{W}\mathbf{X}) && \text{by } \text{rank}(\mathbf{UV}) \leq \min \{ \text{rank}(\mathbf{U}), \text{rank}(\mathbf{V}) \} \end{aligned}$$

$\therefore \text{rank}(\mathbf{W}\mathbf{X}) = \text{rank}(\mathbf{W})$ by (i) and (ii).

(d) Let \mathbf{X} be an $n \times p$ matrix. Suppose \mathbf{C} is a $q \times p$ matrix of rank q and that there exists a matrix \mathbf{A} such that $\mathbf{C} = \mathbf{A}\mathbf{X}$.

First, let us verify that $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-}\mathbf{C}'$ is $q \times q$. Because \mathbf{X} is a $n \times p$ matrix, $\mathbf{X}'\mathbf{X}$ is $p \times p$ matrix. Consequently, any generalized inverse $(\mathbf{X}'\mathbf{X})^{-}$ of $\mathbf{X}'\mathbf{X}$ is also $p \times p$. Then,

$$\mathbf{C} \underbrace{(\mathbf{X}'\mathbf{X})^{-}}_{p \times p} \mathbf{C}'$$

is clearly a $q \times q$ matrix.

Next, we will show that $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$ has rank q . In order to do so, I will use following fact: for any matrix \mathbf{X} ,

$$\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}'), \quad (4)$$

because the rank of a matrix can be defined as the number of linearly independent rows or columns.

Now, consider

$$\begin{aligned} \text{rank}(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}') &= \text{rank}(\mathbf{A}\mathbf{P}_X\mathbf{A}') && \text{since } \mathbf{C} = \mathbf{A}\mathbf{X} \\ &= \text{rank}(\mathbf{A}\mathbf{P}_X\mathbf{P}_X'\mathbf{A}') \\ &= \text{rank}(\mathbf{P}_X'\mathbf{A}') && \text{by part (a)} \\ &= \text{rank}(\mathbf{A}\mathbf{P}_X) \\ &= \text{rank}(\mathbf{A}\mathbf{P}_X\mathbf{X}) && \text{by part(c) with } \mathbf{W} = \mathbf{A}\mathbf{P}_X \\ &= \text{rank}(\mathbf{A}\mathbf{X}) \\ &= \text{rank}(\mathbf{C}) = q. \end{aligned}$$

Therefore, $\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'$ is a $q \times q$ matrix of rank q .

5. We have

$$\begin{aligned} \mathbf{P}_X = \mathbf{P}_W &\iff \mathbf{P}_X - \mathbf{P}_W = \mathbf{0} \\ &\iff (\mathbf{P}_X - \mathbf{P}_W)'(\mathbf{P}_X - \mathbf{P}_W) = \mathbf{0} \quad (\text{Homework 1.6}) \\ &\iff \mathbf{P}_X'\mathbf{P}_X - \mathbf{P}_X'\mathbf{P}_W - \mathbf{P}_W'\mathbf{P}_X + \mathbf{P}_W'\mathbf{P}_W = \mathbf{0} \\ &\iff \mathbf{P}_X - 2\mathbf{P}_X\mathbf{P}_W + \mathbf{P}_W = \mathbf{0} \quad (\text{Symmetric and idempotent}) \end{aligned}$$

Show $\mathbf{P}_X\mathbf{P}_W = \mathbf{P}_W$ next. Since each column of $\mathbf{P}_W = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ is in the column space $\mathcal{C}(\mathbf{W})$, it's also in the column space $\mathcal{C}(\mathbf{X})$. So $\mathbf{P}_W = \mathbf{X}\mathbf{C}_w$ for some \mathbf{C}_w . Then we have:

$$\begin{aligned} \mathbf{P}_X\mathbf{P}_W &= \mathbf{P}_X\mathbf{X}\mathbf{C}_w \\ &= \mathbf{X}\mathbf{C}_w \\ &= \mathbf{P}_W \end{aligned}$$

Similarly, we have $\mathbf{P}_X\mathbf{P}_W = \mathbf{P}_X$. So,

$$\mathbf{P}_X - 2\mathbf{P}_X\mathbf{P}_W + \mathbf{P}_W = \mathbf{0}$$

Then,

$$\mathbf{P}_X = \mathbf{P}_W$$