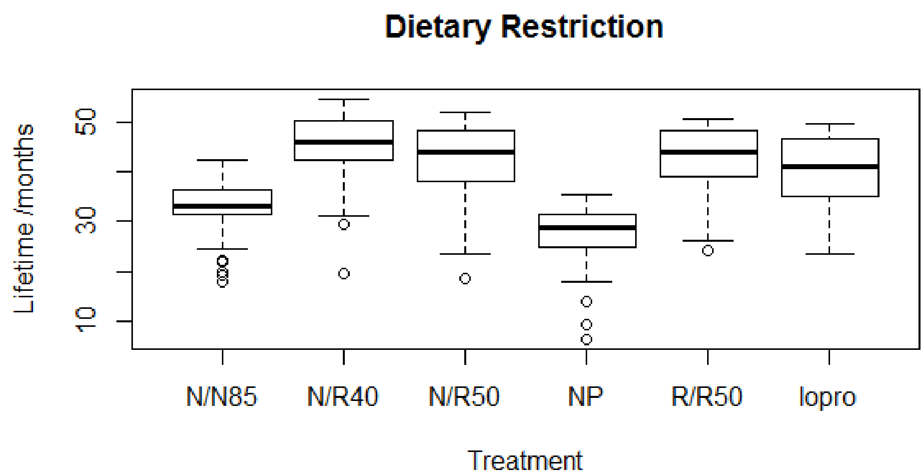
1. Case Study 5.1.1 from *The Statistical Sleuth*

```
> 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/R50 and N/R50 *lopro* treatment groups and an unrestricted mean for each of the other four treatment

groups.
From the code and output below,

$$SSE_{Reduced} = 15510.92.$$

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

> fit1 <- lm(Lifetime~diet1,newcase)
> deviance(fit1)
[1] 15510.92
```

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

$H_0 : E(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}_0)$
$H_a : E(\boldsymbol{y}) \in \mathcal{C}(\boldsymbol{X}) \backslash \mathcal{C}(\boldsymbol{X}_0)$

From the code and output below,

$$F = \frac{(SSE_{Reduced} - SSE_{Full})/(df_{Reduced} - df_{Full})}{SSE_{Full}/df_{Full}} = 4.7873$$

```
> 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    344 15511
2    343 15297  1    213.51 4.7873 0.02935 *
---
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 between the N/R50 and N/R50 *lopro* treatment groups. In the context of this problem, F statistic can be used to test whether the expected lifetime is affected by reduced protein among mice who are treated with normal diet before weaning and reduced calorie (50 kcal/week) after weaning.
In this study, F statistic is 4.7873 with p-value 0.0293. We will reject the reduced model

in favor of the full model at 0.05 level and conclude that there is significant difference between the N/R50 and N/R50 *lopro* treatment groups.

(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 $\boldsymbol{C}$ matrix and $\boldsymbol{d}$ vector and compute the F statistic corresponding to the test of the hypotheses in part (e).

$H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$
$H_a : \boldsymbol{C\beta} \neq \boldsymbol{d}$
where $\boldsymbol{C} = (0, 0, 1, 0, 0, -1)$ according to the order of the treatments in dataset, $q = 1$ and $\boldsymbol{d} = 0$.

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

and

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

.

From the code and output below, $F = 4.787275$ 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=t(c(0,0,1,0,0,-1))

> Fstat=t(C %*% hat.b) %*% solve(C %*% solve(t(xmat)%*%xmat) %*% t(C))
    %*% (C %*% hat.b)/1/hat.sig2
         [,1]
[1,] 4.787275
```

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

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

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

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

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

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

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

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

$$\text{rank}(\boldsymbol{UV}) \leq \min\{\, \text{rank}(\boldsymbol{U}),\ \text{rank}(\boldsymbol{V})\,\} \tag{1}$$

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

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

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

$$\begin{aligned}
\text{rank}(\boldsymbol{X}'\boldsymbol{X}) &\leq \min\{\, \text{rank}(\boldsymbol{X}'),\ \text{rank}(\boldsymbol{X})\,\} && \text{by (1)} \\
&\leq \text{rank}(\boldsymbol{X}) \\
&= \text{rank}(\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}) && \text{since } \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X} = \boldsymbol{X} \\
&= \text{rank}(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{X}) && \text{since } \boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}' \\
&\leq \min\left\{\, \text{rank}(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}),\ \text{rank}(\boldsymbol{X}'\boldsymbol{X})\,\right\} && \text{by (1)} \\
&\leq \text{rank}(\boldsymbol{X}'\boldsymbol{X}).
\end{aligned}$$

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

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

which implies that

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

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

$$\boldsymbol{Xz} = \boldsymbol{0} \iff \boldsymbol{X}'\boldsymbol{Xz} = \boldsymbol{0}. \tag{2}$$

However, a few additional steps are needed for this to establish the desired result. First, why does (2) imply that $\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{X}$ have the same null space? For an $m \times n$ real-valued matrix $\boldsymbol{A}$, its null space, $\text{Null}(\boldsymbol{A})$, is defined as

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

From the definition, clearly (2) implies $\text{Null}(\boldsymbol{X}) = \text{Null}(\boldsymbol{X}'\boldsymbol{X})$. Now, how does this imply equality of ranks? The rank-nullity theorem says that for any $m \times n$ matrix $\boldsymbol{A}$,

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

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

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

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

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

$$\begin{aligned}
\text{rank}(\boldsymbol{X}) &= \text{rank}(\boldsymbol{P_X}\boldsymbol{X}) && \text{since } \boldsymbol{P_X}\boldsymbol{X} = \boldsymbol{X} \\
&\leq \min\{\, \text{rank}(\boldsymbol{P_X}),\ \ \text{rank}(\boldsymbol{X}) \,\} && \text{by (1)} \\
&\leq \text{rank}(\boldsymbol{P_X}) \\
&= \text{rank}\left(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\right) && \text{since } \boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}' \\
&\leq \min\left\{\, \text{rank}(\boldsymbol{X}),\ \ \text{rank}\left((\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\right) \right\} && \text{by (1)} \\
&\leq \text{rank}(\boldsymbol{X}).
\end{aligned}$$

Inequality in both directions implies equality; therefore,

$$\text{rank}(\boldsymbol{X}) = \text{rank}(\boldsymbol{P_X}).$$

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

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

$$\underset{q\times p}{\boldsymbol{C}}\underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-}}_{p\times p}\underset{p\times q}{\boldsymbol{C}'}$$

is clearly a $q \times q$ matrix.

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

$$\text{rank}(\boldsymbol{X}) = \text{rank}(\boldsymbol{X}'), \tag{3}$$

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

Now, consider

$$\begin{aligned}
\text{rank}\big(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{C}'\big) &\le \min\big\{\,\text{rank}\big(\boldsymbol{C}\big),\ \ \text{rank}\big((\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{C}'\big)\,\big\} && \text{by (1)}\\
&\le \text{rank}\big(\boldsymbol{C}\big)\\
&= q && \text{by assumption}\\
&= \text{rank}(\boldsymbol{C})\\
&= \text{rank}(\boldsymbol{A}\boldsymbol{X})\\
&= \text{rank}(\boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X}) && \text{since } \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X} = \boldsymbol{X}\\
&\le \min\{\,\text{rank}(\boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}),\ \ \text{rank}(\boldsymbol{X})\,\} && \text{by (1)}\\
&\le \text{rank}(\boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}})\\
&= \text{rank}\big([\boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}]'\big) && \text{by (3)}\\
&= \text{rank}\big(\boldsymbol{P}_{\boldsymbol{X}}'\boldsymbol{A}'\big)\\
&= \text{rank}\big([\boldsymbol{P}_{\boldsymbol{X}}'\boldsymbol{A}']'\boldsymbol{P}_{\boldsymbol{X}}'\boldsymbol{A}'\big) && \text{by part (a) using } \boldsymbol{X} = \boldsymbol{P}_{\boldsymbol{X}}'\boldsymbol{A}'\\
&= \text{rank}\big(\boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}}'\boldsymbol{A}'\big)\\
&= \text{rank}\big(\boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{A}'\big) && \boldsymbol{P}_{\boldsymbol{X}} \text{ is symmetric}\\
&= \text{rank}\big(\boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{A}'\big) && \boldsymbol{P}_{\boldsymbol{X}} \text{ is idempotent}\\
&= \text{rank}\big(\boldsymbol{A}[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}']\boldsymbol{A}'\big) && \text{since } \boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\\
&= \text{rank}\big([\boldsymbol{A}\boldsymbol{X}](\boldsymbol{X}'\boldsymbol{X})^-[\boldsymbol{X}'\boldsymbol{A}']\big)\\
&= \text{rank}\big([\boldsymbol{A}\boldsymbol{X}](\boldsymbol{X}'\boldsymbol{X})^-[\boldsymbol{A}\boldsymbol{X}]'\big)\\
&= \text{rank}\big(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{C}'\big).
\end{aligned}$$

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

<u>Comments</u>: In this proof, you need to be careful to avoid claiming $(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^- = (\boldsymbol{X}'\boldsymbol{X})^-$ or $\boldsymbol{GAG} = \boldsymbol{G}$. This is not nessessary true, for example, $\boldsymbol{A} = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$ and one of its generalized inverse is $\boldsymbol{G} = \begin{bmatrix}1 & 2\\ 3 & 0\end{bmatrix}$, we can verify that $\boldsymbol{GAG} = \begin{bmatrix}1 & 2\\ 3 & 6\end{bmatrix} \ne \boldsymbol{G}$.
I will recommod to use the idempotent property of projection matrix $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^-\boldsymbol{X}' = \boldsymbol{P}_{\boldsymbol{X}}$ or $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{X} = \boldsymbol{X}$

(d) Suppose $\boldsymbol{X}$ is an $n \times p$ matrix and $\boldsymbol{A}$ is a matrix with n columns satisfying $\boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{A}$. Then,

$$\begin{aligned}
\operatorname{rank}\big(\boldsymbol{A}\big) &= \operatorname{rank}\big(\boldsymbol{A}\boldsymbol{P}_{\boldsymbol{X}}\big) & \\
&= \operatorname{rank}\big(\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\big) & \text{(def. of } \boldsymbol{P}_{\boldsymbol{X}}\text{)} \\
&\leq \min\big\{\operatorname{rank}\big(\boldsymbol{A}\boldsymbol{X}\big),\ \operatorname{rank}\big((\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\big)\big\} & \text{by (1)} \\
&\leq \operatorname{rank}\big(\boldsymbol{A}\boldsymbol{X}\big) & \\
&\leq \min\big\{\operatorname{rank}\big(\boldsymbol{A}\big),\ \operatorname{rank}\big(\boldsymbol{X}\big)\big\} & \text{by (1)} \\
&\leq \operatorname{rank}\big(\boldsymbol{A}\big).
\end{aligned}$$

Inequality in both directions implies equality; therefore,

$$\operatorname{rank}(\boldsymbol{A}\boldsymbol{X}) = \operatorname{rank}(\boldsymbol{A}).$$

4. Note that $\boldsymbol{A}$ is the matrix whose column space is the same as that of $\boldsymbol{X}$ because there exist $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ such that

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \boldsymbol{X}\boldsymbol{C}_1$$

and

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} = \boldsymbol{A}\boldsymbol{C}_2.$$

Then, $\mathcal{C}(\boldsymbol{A}) = \mathcal{C}(\boldsymbol{X})$. It implies that $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{X}}$. Also, note that $\mathcal{C}(\boldsymbol{A}) \subset \mathcal{C}(\boldsymbol{B})$ because there exists $\boldsymbol{C}$ such that $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{C}$ where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 1 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \boldsymbol{B}\boldsymbol{C}.$$

Thus, $\boldsymbol{P}_{\boldsymbol{A}}\boldsymbol{P}_{\boldsymbol{B}} = \boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{P}_{\boldsymbol{X}}$ where

$$\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}' = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$