

HW4

Samuel Olson

Outline

- Q1: g2g
- Q2:
- Q3: g2g
- Q4: needs edits

Problem 1

Suppose $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ for some unknown $\sigma^2 > 0$. Let $\hat{\mathbf{y}} = \mathbf{P}_\mathbf{X} \mathbf{y}$.

a)

Determine the distribution of

$$\begin{bmatrix} \hat{\mathbf{y}} \\ \mathbf{y} - \hat{\mathbf{y}} \end{bmatrix}$$

Useful Property:

Linear transformation of a normal random variable is itself normal (distribution remains normal with known/calculable parameters):

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \rightarrow \mathbf{A}\mathbf{x} + \mathbf{b} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$$

To begin, note the [Useful Property] provides us reason to assert that the following follows an MVN distribution:

$$\begin{bmatrix} \hat{\mathbf{y}} \\ \mathbf{y} - \hat{\mathbf{y}} \end{bmatrix} \sim \mathcal{N}(?, ??)$$

Such that we need to identify the mean and covariance matrix of the above.

To that end, note that by definition:

$$\begin{bmatrix} \hat{\mathbf{y}} \\ \mathbf{y} - \hat{\mathbf{y}} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_\mathbf{X} \mathbf{y} \\ \mathbf{y} - \mathbf{P}_\mathbf{X} \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_\mathbf{X} \\ \mathbf{I} - \mathbf{P}_\mathbf{X} \end{bmatrix} \mathbf{y}$$

To calculate the mean, noting linearity of expectation:

$$E\left(\begin{bmatrix} \mathbf{P}_X \\ \mathbf{I} - \mathbf{P}_X \end{bmatrix} y\right) = \begin{bmatrix} \mathbf{P}_X \\ \mathbf{I} - \mathbf{P}_X \end{bmatrix} E(y) = \begin{bmatrix} \mathbf{P}_X \\ \mathbf{I} - \mathbf{P}_X \end{bmatrix} \mathbf{X}\beta = \begin{bmatrix} \mathbf{X}\beta \\ \mathbf{X}\beta - \mathbf{X}\beta \end{bmatrix} = \begin{bmatrix} \mathbf{X}\beta \\ \mathbf{0} \end{bmatrix}$$

To calculate the covariance, noting properties of the Projection Matrix (symmetric and idempotent):

$$\text{Var}\left(\begin{bmatrix} \mathbf{P}_X \\ \mathbf{I} - \mathbf{P}_X \end{bmatrix} y\right) = \begin{bmatrix} \mathbf{P}_X \\ \mathbf{I} - \mathbf{P}_X \end{bmatrix} \text{Var}(y) \begin{bmatrix} \mathbf{P}_X \\ \mathbf{I} - \mathbf{P}_X \end{bmatrix}^\top = \begin{bmatrix} \mathbf{P}_X \\ \mathbf{I} - \mathbf{P}_X \end{bmatrix} \sigma^2 \mathbf{I} [\mathbf{P}_X, (\mathbf{I} - \mathbf{P}_X)^\top] = \sigma^2 \begin{bmatrix} \mathbf{P}_X \mathbf{P}_X^\top & \mathbf{P}_X (\mathbf{I} - \mathbf{P}_X)^\top \\ (\mathbf{I} - \mathbf{P}_X) \mathbf{P}_X^\top & (\mathbf{I} - \mathbf{P}_X) (\mathbf{I} - \mathbf{P}_X)^\top \end{bmatrix}$$

$$\text{Var}\left(\begin{bmatrix} \mathbf{P}_X \\ \mathbf{I} - \mathbf{P}_X \end{bmatrix} y\right) = \sigma^2 \begin{bmatrix} \mathbf{P}_X & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \mathbf{P}_X \end{bmatrix}$$

Taken together, and again with note of the [Useful Property], we know:

$$\begin{bmatrix} \hat{\mathbf{y}} \\ \mathbf{y} - \hat{\mathbf{y}} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{X}\beta \\ \mathbf{0} \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{P}_X & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \mathbf{P}_X \end{bmatrix}\right)$$

b)

Determine the distribution of

$$\hat{y}^\top \hat{y}$$

Noting again the properties of the Projection Matrix (symmetric and idempotent):

$$\hat{y}^\top \hat{y} = [\mathbf{P}_X \mathbf{y}]^\top \mathbf{P}_X \mathbf{y} = \mathbf{y}^\top \mathbf{P}_X^\top \mathbf{P}_X \mathbf{y} = \mathbf{y}^\top \mathbf{P}_X \mathbf{P}_X \mathbf{y} = \mathbf{y}^\top \mathbf{P}_X \mathbf{y}$$

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

Important Distributional Result about Quadratic Forms. Suppose \mathbf{A} is an $n \times n$ symmetric positive definite matrix. Suppose further that \mathbf{A} is an $n \times n$ symmetric positive definite matrix such that $\mathbf{A}\boldsymbol{\Sigma}$ is idempotent (i.e., $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma} = \mathbf{A}\boldsymbol{\Sigma}$).

Then, $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ implies

$$\mathbf{y}^\top \mathbf{A} \mathbf{y} \sim \chi_m^2 ((\boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu})/2)$$

Note that this result holds even when $\boldsymbol{\Sigma}$ is nonnegative definite. Note that $\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}$ and $\text{trace}(\mathbf{A}\boldsymbol{\Sigma}) = m$. For a proof see Schott, 2001.

Figure 1: CocoMelon

Note the above “Distributional Result About Quadratic Forms”. This will be our guide here.

We want to find a matrix \mathbf{A} such that $\mathbf{A}\boldsymbol{\Sigma}$ is idempotent for $\boldsymbol{\Sigma} \equiv \text{Var}(\mathbf{y})$.

One candidate we consider: $\mathbf{A} = \frac{\mathbf{P}_X}{\sigma^2}$

We first know that this matrix will need to be symmetric. To that end note:

$$\mathbf{A}^\top = \left(\frac{\mathbf{P}_X}{\sigma^2} \right)^\top = \frac{1}{\sigma^2} (\mathbf{P}_X)^\top = \frac{1}{\sigma^2} \mathbf{P}_X = \mathbf{A}$$

So this matrix is symmetric (because we utilize the Projection Matrix and multiply it by $\frac{1}{\sigma^2}$, which preserves symmetry).

Then consider, using the symmetry property shown:

$$\mathbf{A}\boldsymbol{\Sigma} = \mathbf{A}\text{Var}(\mathbf{y}) = \frac{\mathbf{P}_X}{\sigma^2} \sigma^2 \mathbf{I} = \mathbf{P}_X$$

So, we then know that $\mathbf{A}\boldsymbol{\Sigma} = \mathbf{P}_X$ is idempotent, as the Projection Matrix is idempotent by definition.

We then note: $\Sigma = \sigma^2 \mathbf{I}$ is positive definite since $\sigma^2 > 0$.

We use this property to determine the rank of \mathbf{A} , with the goal of determining if it is full rank:

Using the expressions given above:

$$\text{rank}(\mathbf{A}) = \text{rank}\left(\frac{\mathbf{P}_X}{\sigma^2}\right) = \text{rank}(\mathbf{P}_X) = \text{rank}(\mathbf{X})$$

Noting that scalar multiplication (non-zero) does not affect rank calculations of a matrix.

Then, noting that both \hat{y}^\top and \hat{y} are normally distributed, we then know that the quadratic form $\hat{y}^\top \hat{y}$ is χ^2 distributed, and that $\frac{1}{\sigma^2} \hat{y}^\top \hat{y}$ is a scaled χ^2 distributed. The degrees of freedom parameter to this distribution is given by the rank, shown above, and we again make use of “Distributional Result About Quadratic Forms” regarding the centrality parameter.

Taken together, this gives us the expression:

$$\frac{1}{\sigma^2} \hat{y}^\top \hat{y} = y^\top \frac{\mathbf{P}_X}{\sigma^2} y \sim \chi_{\text{rank}(\mathbf{X})}^2 \left(\frac{\mathbf{X}^\top \mathbf{A} \mathbf{X} \beta}{2} \right) = \chi_{\text{rank}(\mathbf{X})}^2 \left(\frac{\mathbf{X}^\top \frac{\mathbf{P}_X}{\sigma^2} \mathbf{X} \beta}{2} \right) = \chi_{\text{rank}(\mathbf{X})}^2 \left(\frac{\mathbf{X}^\top \mathbf{P}_X \mathbf{X} \beta}{2\sigma^2} \right)$$

Noting the results of the quadratic form:

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma) \rightarrow \mathbf{y}^\top \mathbf{A} \mathbf{y} \sim \chi_m^2 \left(\frac{\boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu}}{2} \right)$$

The remaining point to detail then is the non-centrality parameter. Knowing the underlying distributions of \hat{y}^\top and \hat{y} , we then know:

$$\frac{\mathbf{X}^\top \mathbf{A} \mathbf{X} \beta}{2} = \frac{1}{2} \mathbf{X}^\top \frac{\mathbf{P}_X}{\sigma^2} \mathbf{X} \beta = \frac{1}{2\sigma^2} \beta^\top \mathbf{X}^\top \mathbf{P}_X \mathbf{X} \beta = \frac{1}{2\sigma^2} \beta^\top \mathbf{X}^\top \mathbf{X} \beta$$

Taken together, we then conclude with:

$$\hat{y}^\top \hat{y} \sim \sigma^2 \chi_{\text{rank}(\mathbf{X})}^2 \left(\frac{\beta^\top \mathbf{X}^\top \mathbf{X} \beta}{2\sigma^2} \right)$$

Problem 2

An experiment was conducted to study the durability of coated fabric subjected to abrasive tests. Three factors were considered: **Filler type** with two levels (F1 and F2), **Surface treatment** with two levels (S1 and S2), **Proportion of filler** with three levels (25%, 50%, and 75%).

Using a completely randomized design with two fabric samples per treatment, the amount of fabric lost (in mg) for each fabric sample was recorded. Data are available in a tab-delimited text file at: FabricLoss.txt.

```
FabricLoss<- read.table(file = "FabricLoss.txt",  
                        header = TRUE,  
                        sep = "\t")
```

a)

Consider a cell means model for these data. Estimate the mean and standard error for the treatment corresponding to F2, S1, and 50% filler.

b)

The concept of LSMEANS has been explained carefully in lecture and course notes for the special case of a two-factor study. The concept generalizes easily to multi-factor studies. For example, in a three- factor study, the LSMEAN for level i of the first factor is the OLS estimator of $\bar{\mu}_{i..}$, the average of the cell means for all treatments that involve level i of the first factor. Find LSMEANS for the levels of the factor filler type.

c)

We can also compute LSMEANS for estimable marginal means like $\bar{\mu}_{.jk}$, the average of the cell means for all treatments involving level j of the second factor and level k of the third factor. Find the LSMEAN for surface treatment S2 and 25% filler.

d)

Provide a standard error for the estimate computed in part (c).

e)

In a three-factor study we would say there are no main effects for the first factor if $\bar{\mu}_{i..} = \bar{\mu}_{i^{\top}..}$ for all levels $i \neq i^{\top}$. Conduct a test for filler type main effects. Provide an F-statistic, a p-value, and a conclusion.

f)

In a three-factor study in which the third factor has K levels, we would say there are no three-way interactions if, for all $i \neq i^\top$ and $j \neq j^\top$,

$$\mu_{ij1} - \mu_{ij^\top 1} - \mu_{i^\top j 1} + \mu_{i^\top j^\top 1} = \mu_{ij2} - \mu_{ij^\top 2} - \mu_{i^\top j 2} + \mu_{i^\top j^\top 2} = \cdots = \mu_{ijK} - \mu_{ij^\top K} - \mu_{i^\top j K} + \mu_{i^\top j^\top K}.$$

Note that each linear combination above can be viewed as a two-way interaction effect involving the first two factors while holding the level of the third factor fixed. If these interaction effects are all the same regardless of which level of the third factor is selected, we say there are no three way interactions. Put another equivalent way, there are no three-factor interactions if

$$\mu_{ijk} - \mu_{ij^\top k} - \mu_{i^\top j k} + \mu_{i^\top j^\top k} - \mu_{ijk^\top} + \mu_{ij^\top k^\top} + \mu_{i^\top j k^\top} - \mu_{i^\top j^\top k^\top} = 0$$

for all $i \neq i^\top$, $j \neq j^\top$, and $k \neq k^\top$. Conduct a test for three-way interactions among the factors filler type, surface treatment, and filler proportion. Provide an F-statistic, a p-value, and a conclusion.

g)

In a three-factor study, we would say there are no two-way interactions between the first and third factors if

$$\bar{\mu}_{i \cdot k} - \bar{\mu}_{i \cdot k^\top} - \bar{\mu}_{i^\top \cdot k} + \bar{\mu}_{i^\top \cdot k^\top} = 0$$

for all $i \neq i^\top$ and $k \neq k^\top$. Conduct a test for two-way interactions between the factors filler type and filler proportion. Provide an F-statistic, a p-value, and a conclusion.

Problem 3

When \mathbf{X} does not have full rank, let's see why $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^- \mathbf{X}^\top$ is invariant to the choice of generalized inverse. Let \mathbf{G} and \mathbf{H} be two generalized inverses of $\mathbf{X}^\top \mathbf{X}$. For an arbitrary $\mathbf{v} \in \mathbb{R}^n$, let $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ with $\mathbf{v}_1 = \mathbf{X}\mathbf{b} \in C(\mathbf{X})$ for some \mathbf{b} .

a)

Show that

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

so that $\mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X} = \mathbf{X}$ for any generalized inverse.

Answer

As given, we may write:

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$$

Then:

$$\mathbf{v}_1 \perp \mathbf{v}_2$$

And since:

$$\mathbf{v}_1 = \mathbf{X}\mathbf{b} \in C(\mathbf{X}) \rightarrow \mathbf{v}_2 \in (C(\mathbf{X}))^\perp$$

So may then write:

$$\mathbf{v}^\top \mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X} = (\mathbf{v}_1 + \mathbf{v}_2)^\top \mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X} = \mathbf{v}_1^\top \mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X} + \mathbf{v}_2^\top \mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X}$$

Each of these terms may be further evaluated. To (hopefully) make the proof more legible then, consider:

(1):

As $\mathbf{v}_1 = \mathbf{X}\mathbf{b}$, we may write:

$$\mathbf{v}_1^\top \mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X} = (\mathbf{X}\mathbf{b})^\top \mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X} = \mathbf{b}^\top \mathbf{X}^\top \mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X} = \mathbf{b}^\top \mathbf{X}^\top \mathbf{X} = \mathbf{b}^\top \mathbf{X}^\top \mathbf{X}$$

Noting \mathbf{G} is a generalized inverse.

However, as $\mathbf{X}\mathbf{b} = \mathbf{v}_1$, we may simplify:

$$\mathbf{v}_1^\top \mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X} = \mathbf{v}_1^\top \mathbf{X}$$

(2):

As indicated, \mathbf{v}_2 is orthogonal to $C(\mathbf{X})$, such that we may write:

$$\mathbf{v}_2^\top \mathbf{X} = 0 \rightarrow \mathbf{v}_2^\top \mathbf{X} \mathbf{G} \mathbf{X}^\top \mathbf{X} = 0$$

Taking (1) and (2) together:

$$\boldsymbol{v}^\top \boldsymbol{X} \boldsymbol{G} \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{v}_1^\top \boldsymbol{X} + 0 = \boldsymbol{v}^\top \boldsymbol{X}$$

Concluding:

$$\boldsymbol{v}^\top \boldsymbol{X} \boldsymbol{G} \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{v}^\top \boldsymbol{X}$$

So that $\boldsymbol{X} \boldsymbol{G} \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{X}$ for any generalized inverse.

b)

Show that

$$\mathbf{XGX}^\top \mathbf{v} = \mathbf{XHX}^\top \mathbf{v}$$

and thus \mathbf{XGX}^\top is invariant to the choice of generalized inverse.

Answer

As given: \mathbf{G} and \mathbf{H} are generalized inverses of $\mathbf{X}^\top \mathbf{X}$, so by definition the following holds:

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

And

$$(\mathbf{X}^\top \mathbf{X})\mathbf{H}(\mathbf{X}^\top \mathbf{X}) = \mathbf{X}^\top \mathbf{X}$$

For some vector $\mathbf{v} \in \mathbb{R}^n$, we may decompose \mathbf{v} like in part a), i.e.:

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$$

where $\mathbf{v}_1 \in C(\mathbf{X})$ and $\mathbf{v}_2 \in C(\mathbf{X})^\perp$:

(Hey, quick Q, Google wasn't helpful: $\mathbf{v}_2 \in C(\mathbf{X})^\perp \equiv \mathbf{v}_2 \perp C(\mathbf{X})$?)

At any rate:

$$\mathbf{v}_1 = \mathbf{Xb}$$

For some real-valued vector \mathbf{b} .

We may then write:

$$\mathbf{XGX}^\top \mathbf{v} = \mathbf{XGX}^\top (\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{XGX}^\top \mathbf{v}_1 + \mathbf{XGX}^\top \mathbf{v}_2$$

As defined, $\mathbf{v}_1 = \mathbf{Xb}$, so we may simplify:

$$\mathbf{XGX}^\top \mathbf{v} = \mathbf{XGX}^\top \mathbf{Xb} + \mathbf{XGX}^\top \mathbf{v}_2 = \mathbf{Xb} + \mathbf{XGX}^\top \mathbf{v}_2$$

As \mathbf{G} is a generalized inverse.

Also, as defined \mathbf{v}_2 is orthogonal to $C(\mathbf{X})$, so the second term simplifies as well, again, similar to part a):

$$\mathbf{XGX}^\top \mathbf{v} = \mathbf{Xb} + \mathbf{XGX}^\top \mathbf{v}_2 = \mathbf{Xb} + 0 = \mathbf{Xb}$$

Note: The choice of starting with \mathbf{G} was arbitrary. The same steps and derivations would follow for the generalized inverse \mathbf{H} , such that we'd also say:

$$\mathbf{XHX}^\top \mathbf{v} = \mathbf{Xb}$$

$\forall \mathbf{v}$.

This then shows:

$$\mathbf{XGX}^\top = \mathbf{XHX}^\top$$

And that \mathbf{XGX}^\top is invariant to the choice of the generalized inverse.

Problem 4

An experiment was conducted to study the effect of two diets (1 and 2) and two drugs (1 and 2) on blood pressure in rats. A total of 40 rats were randomly assigned to the 4 combinations of diet and drug, with 10 rats per combination. Let y_{ijk} be the decrease in blood pressure from the beginning to the end of the study for diet i , drug j , and rat k ($i = 1, 2; j = 1, 2; k = 1, \dots, 10$). Suppose:

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^2).$$

A researcher suspects that the mean reduction in blood pressure will be the same for all combinations of diet and drug except for the combination of diet 1 with drug 1. This leads to consideration of the null hypothesis:

$$H_0 : \mu_{12} = \mu_{21} = \mu_{22}.$$

Assuming model (1) holds, determine the distribution of the F-statistic you would use to test this null hypothesis: State the degrees of freedom of the statistic. Provide a fully simplified expression for the noncentrality parameter in terms of model (1) parameters.

Answer

We can conduct a lack of fit F-test for this problem. The full model is a cell mean model with 4 parameters, and the reduced model has 2 parameters, with one for the combination of diet 1 with drug 1 and one for all the other three combinations. Since we have 40 observations, the degree of freedom for the full model is

$$df_{\text{full}} = 40 - 4 = 36,$$

the degree of freedom for the reduced model is

$$df_{\text{reduced}} = 40 - 2 = 38,$$

and the difference is

$$df_{\text{reduced}} - df_{\text{full}} = 38 - 36 = 2.$$

Thus, the F-distribution we would use is

$$F_{2,36}.$$

Noncentrality Parameter

The noncentrality parameter is given by:

$$\frac{(C\beta - d)^\top [C(\mathbf{X}^\top \mathbf{X})^{-1} C^\top]^{-1} (C\beta - d)}{2\sigma^2}$$

where we define:

$$\boldsymbol{\beta} = \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{pmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad \mathbf{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{1}_{10 \times 1} & 0 & 0 & 0 \\ 0 & \mathbf{1}_{10 \times 1} & 0 & 0 \\ 0 & 0 & \mathbf{1}_{10 \times 1} & 0 \\ 0 & 0 & 0 & \mathbf{1}_{10 \times 1} \end{bmatrix}$$

Thus, we obtain:

$$\mathbf{C}\boldsymbol{\beta} - \mathbf{d} = \begin{pmatrix} \mu_{12} - \mu_{21} \\ \mu_{12} - \mu_{22} \end{pmatrix}$$

and:

$$[\mathbf{C}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{C}^\top]^{-1} = \frac{10}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Thus, the noncentrality parameter simplifies to:

$$\begin{aligned} & \frac{\begin{pmatrix} \mu_{12} - \mu_{21} \\ \mu_{12} - \mu_{22} \end{pmatrix}^\top \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{pmatrix} \mu_{12} - \mu_{21} \\ \mu_{12} - \mu_{22} \end{pmatrix}}{2\sigma^2} \times \frac{10}{3} \\ &= \frac{10}{3\sigma^2} [(\mu_{12} - \mu_{21})^2 + (\mu_{12} - \mu_{22})^2 - (\mu_{12} - \mu_{21})(\mu_{12} - \mu_{22})] \end{aligned}$$

Alternative Method

An alternative method can be used to obtain an equivalent expression. The null hypothesis can be tested with an F-test comparing a reduced model to a full model.

Let:

$$\mathbf{X}_0 = \begin{bmatrix} \mathbf{1}_{10 \times 1} & 0 \\ 0 & \mathbf{1}_{30 \times 1} \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} \mathbf{1}_{10 \times 1} & 0 & 0 & 0 \\ 0 & \mathbf{1}_{10 \times 1} & 0 & 0 \\ 0 & 0 & \mathbf{1}_{10 \times 1} & 0 \\ 0 & 0 & 0 & \mathbf{1}_{10 \times 1} \end{bmatrix}$$

Since $r_0 = 2$, $r = 4$, we use the noncentral F-distribution with degrees of freedom $r - r_0 = 2$ and $n - r = 36$.

Then:

$$\begin{aligned} \text{NCP} &= \frac{\boldsymbol{\beta}^\top \mathbf{X}^\top (\mathbf{P}_\mathbf{X} - \mathbf{P}_{\mathbf{X}_0}) \mathbf{X} \boldsymbol{\beta}}{2\sigma^2} \\ &= \frac{\|(\mathbf{P}_\mathbf{X} - \mathbf{P}_{\mathbf{X}_0}) \mathbf{X} \boldsymbol{\beta}\|^2}{2\sigma^2} \end{aligned}$$

$$= \frac{\|\mathbf{X}\boldsymbol{\beta} - \mathbf{P}_{\mathbf{X}_0}\mathbf{X}\boldsymbol{\beta}\|^2}{2\sigma^2}$$

$$\mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} \mu_{11}\mathbf{1}_{10 \times 1} \\ \mu_{12}\mathbf{1}_{10 \times 1} \\ \mu_{21}\mathbf{1}_{10 \times 1} \\ \mu_{22}\mathbf{1}_{10 \times 1} \end{bmatrix}$$

$$\mathbf{P}_{\mathbf{X}_0}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_0(\mathbf{X}_0^\top \mathbf{X}_0)^{-1} \mathbf{X}_0^\top \mathbf{X}\boldsymbol{\beta}$$

$$= \mathbf{X}_0 \begin{bmatrix} 1/10 & 0 \\ 0 & 1/30 \end{bmatrix} \mathbf{X}_0^\top \mathbf{X}\boldsymbol{\beta}$$

$$= \mathbf{X}_0 \begin{bmatrix} 10\mu_{11} \\ 10(\mu_{12} + \mu_{21} + \mu_{22})/3 \end{bmatrix}$$

$$= \mathbf{X}_0 \begin{bmatrix} \mu_{11} \\ \bar{\mu} \end{bmatrix}, \quad \text{where } \bar{\mu} = \frac{\mu_{12} + \mu_{21} + \mu_{22}}{3}$$

Thus:

$$\|\mathbf{X}\boldsymbol{\beta} - \mathbf{P}_{\mathbf{X}_0}\mathbf{X}\boldsymbol{\beta}\|^2$$

$$= 10[(\mu_{12} - \bar{\mu})^2 + (\mu_{21} - \bar{\mu})^2 + (\mu_{22} - \bar{\mu})^2]$$

So the noncentrality parameter is:

$$\frac{10[(\mu_{12} - \bar{\mu})^2 + (\mu_{21} - \bar{\mu})^2 + (\mu_{22} - \bar{\mu})^2]}{2\sigma^2} = \frac{5[(\mu_{12} - \bar{\mu})^2 + (\mu_{21} - \bar{\mu})^2 + (\mu_{22} - \bar{\mu})^2]}{\sigma^2}$$