Good Luck!   Name **Sabrina Morse - Student A**
                                                                    Partner : Sam Olson

Please answer each question in the space provided, and **show all your work**. Credit cannot be given if work is not shown. Ask for extra paper if you need it. Good luck.

1. If I asked your best friend about one of your character strengths, what would they say?

    _____

2. **Technical and Conceptual.** A study was conducted to see if there are differences in the quality of steel produced by $m = 3$ machines. It is also felt that there may be differences in the feedstock obtained from $l = 3$ different suppliers. Nine samples of feedstock were selected from each supplier, and $n = 3$ samples were randomly assigned to each machine.

    In class we talked extensively about two different types of models to analyze data from this type of experiment. Most recently you encountered both models on Homework 5, Problem 1.

    (a) Name each statistical model. The order in which you name the models that does not matter!

        Model 1: **Cell Means Model**   Model 2: **Additive Model**

    (b) For each, write out the statistical model to describe the quality y_{ijk} of the kth piece of steel produced by the ith machine using feedstock from supplier j.

        - Do **not** write the model in matrix or vector form!
        - Be sure to properly define all parameters, use clear and complete notation and state the **necessary assumptions** such that the model is also a **Gauss-Markov Model** with **Normal Errors**.
        - Your answer should look like one of the slides in the notes when we introduce a new statistical model.

        <u>Model 1:</u> $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$

μ = combined mean from both factors

α_i = machine main effect

β_j = supplier main effect

γ_{ij} = interaction effect between machine & supplier

ϵ_{ijk} = random error

Assumptions:
  - $\epsilon_{ijk} \sim N(0, \sigma^2)$
  - independence of random errors

Model 2: $y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}$

μ = overall mean

α_i = machine main effect

β_j = supplier main effect

ϵ_{ijk} = random error

Assumptions:

- $\epsilon_{ijk} \sim N(0, \sigma^2)$

- independence of random error

- no interaction between treatments

(c) **In the context of the data**, what is the main difference between both types of statistical models in terms of model complexity?

Cell means is more complex than additive model, since it includes a term for potential interaction between machines and suppliers

3. Suppose that $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2, \boldsymbol{I})$, and let $q_i = \boldsymbol{y}^\top \boldsymbol{A}_i\, \boldsymbol{y}, i = 1, 2$ where

$$\boldsymbol{A}_1 = \frac{1}{3}\boldsymbol{1}\boldsymbol{1}^\top \qquad \text{and} \qquad \boldsymbol{A}_2 = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Note that $\boldsymbol{1}$ is a vector of three ones, i.e., $\boldsymbol{1} = (1\ 1\ 1)^\top$.

(a) Determine the distribution of each quadratic form and explain how you know that this indeed is the distribution of $q_i, i = 1, 2$.

quadratic form eq: $y_T A y \sim \chi^2_m\left(\frac{\mu^T A \mu}{2}\right)$

↳ rank of A

$A_1 = \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}[1\ 1\ 1] = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \Rightarrow$ rank = 1

$A_2 = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow$ rank = 1

$\therefore\ y^T A_i y \sim \chi^2_1\left(\frac{\mu^T A \mu}{2}\right)$ → now find ncp for each A_i

$\mu^T = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix}$

$A_1: A_1\mu = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 + \mu_2 + \mu_3 \\ \mu_1 + \mu_2 + \mu_3 \\ \mu_1 + \mu_2 + \mu_3 \end{bmatrix}$

$\mu^T A_1 \mu = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix}\begin{bmatrix} \mu_1 + \mu_2 + \mu_3 \\ \mu_1 + \mu_2 + \mu_3 \\ \mu_1 + \mu_2 + \mu_3 \end{bmatrix}$

$= (\mu_1 + \mu_2 + \mu_3)^2$

so $y^T A_1 y \sim \chi^2_1\left(\frac{(\mu_1 + \mu_2 + \mu_3)^2}{6}\right)$

$A_2: A_2\mu = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_1 \\ 0 \end{bmatrix}$

$\mu^T A_2 \mu = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix}\begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_1 \\ 0 \end{bmatrix} = \mu_1(\mu_1 - \mu_2) + \mu_1(\mu_2 - \mu_1) + \mu_2(\mu_1 - \mu_2) + \mu_2(\mu_2 - \mu_1)$

$= \mu_1^2 - \mu_1\mu_2 + \mu_1\mu_2 - \mu_1^2 + \mu_2\mu_1 - \mu_2^2 + \mu_2^2 - \mu_2\mu_1$

$= 0$ so $y^T A_2 y \sim \chi^2_1(0)$

we know these are distributions of q_i for $i = 1, 2$ since the normality of y ensures the quadratic form will follow χ^2 distribution of rank m & ncp of $\frac{\mu^T A_i \mu}{2}$

(b) Show that q_1 and q_2 are independent.

Show $A_1 \perp A_2$ = independence of q_1 & $q_2 \Rightarrow A_1 \cdot A_2 = 0$ means orthogonal

$\begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 1/3 - 1/3 + 0 - 1/3 + 1/3 + 0 + 0 + 0 = 0$

since $A_1 \cdot A_2 = 0 \Rightarrow q_1$ & q_2 are independent ✓

3

4.

Consider a one-way ANOVA model with two levels and two observations at each level,

$$E(y_{ij}) = \mu + \alpha_i, \quad i, j = 1, 2$$

Using the given information, we may represent it in matrix form to aid in soliving the following questions:

$$\mathbf{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

(a) Is α_1 estimable? Show work to justify why it is or is not estimable.

- If C is any $q \times p$ matrix, we say that the linear function of β given by $C\beta$ is estimable if and only if $C = AX$ for some matrix $q \times n$ matrix A.

Figure 1: Ch. 2: Some Key Linear Models Results

The definition of estimability is given above as first defined in Chapter 2 of lecture ntoes.

For α_1: Can we find a matrix A to satisfy the above?

We have:

$$c^T = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

There in fact does not exist a matrix A such that:

$$C = A\mathbf{X}$$

This means that α_1 is not estimable (this is also descriptively found by noting that the matrix C is not in row space of $\mathbf{X}$).

Generally, parameters (elements of the β vector) are only estimable if they can be expressed as a linear combination of the rows of the design matrix, $\mathbf{X}$. For our purposes, the model matrix does not have full rank, so we know there are elements of the parameter vector that are not estimable, though we still have linear combinations that are (more on that in the following).

(b) Provide a quantity that is estimable. $\alpha_1 - \alpha_2$ or $\mu + \alpha_1$.

We again utilize the above definition to determine estimability. We find there are more than one quantity (linear combination of parameters) that are estimable, despite our design matrix not being full rank. Two examples:

For $\alpha_1 - \alpha_2$, an estimable quantity:

$$c^T = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} = 1 \cdot \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} + (-1) \cdot \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$$

The corresponding matrix A (a row vector) satisfying $C = A\mathbf{X}$ is:

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

Another estimable quantity, $\mu + \alpha_1$ has C matrix:

$$c^T = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}$$

But this is just the first row of $\mathbf{X}$! So the corresponding A matrix is:

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

The remaining questions refer to any general linear model as discussed in class. Thus, provide answers for a general **X** instead of referring to the particular **X** defined above.

Additional Note: The following are related to the "Prerequisite Knowledge" Slides and an extension of the OLS Slides (Chapters 2 and 3) from lectures. Of note:

- The set of all possible linear combinations of the columns of $\boldsymbol{A}$ is called the <u>column space</u> of $\boldsymbol{A}$ and is written as

$$\mathcal{C}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^m : \boldsymbol{x} = \boldsymbol{A}\boldsymbol{c} \text{ for some } \boldsymbol{c} \in \mathbb{R}^n\}.$$

Any member of the set $\mathcal{C}(\boldsymbol{A})$ is an $m \times 1$ vector, so $\mathcal{C}(\boldsymbol{A})$ is a subset of $\mathbb{R}^m$.

Figure 2: Linear Algebra Overview

Suppose $\quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \mathrm{E}(\boldsymbol{\epsilon}) = 0, \qquad \mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$

- $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta} \in \mathcal{C}(\boldsymbol{X})$ with $\boldsymbol{\beta}$ unknown, $\boldsymbol{X}$ is full-rank

Figure 3: Column Rank

(c) True or False Circle the appropriate choice. The expected value of any observation is only estimable when **X** has full column rank.

False.

The expected value of any observation is estimable as long as it is a linear combination of the columns of **X**. Having full column rank just ensures that every element of the parameter vector, β, is estimable. There are examples of observations able to be estimated without having full column rank, such as the second estimable quantity in the above part b).

(d) The set of vectors **c** for which $\mathbf{c}^T\boldsymbol{\beta}$ is estimable forms a vector space. Specify the vector space.

Answer: **The column space of X**.

The above answer is based on the Linear Algebra overview, as well as review of "Alternative Parametrization" Lecture Slides, where we discuss how/what estimates are able to be produced when working with a design matrix that is not full rank, e.g. for the Additive Model.

Span, Basis, Dimension of a Set of Vectors.

- A vector space $\mathcal{S}$ is said to be generated by a set of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ if for

$$\boldsymbol{x} \in \mathcal{S} \Rightarrow \boldsymbol{x} = \sum_{i=1}^{n} c_i \boldsymbol{x}_i \ \text{ for some } \ c_1, \ldots, c_n \in \mathbb{R}.$$

- $\text{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} = \mathcal{S}$ is the vector space generated by $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$.

- If a vector space $\mathcal{S}$ is generated by LI vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, then $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ form a basis for $\mathcal{S}$.

- A basis for a vector space is not unique, but the number of vectors in the basis, known as dimension of the vector space, is unique.

Column and Row spaces. Let $\boldsymbol{A}$ denote a $m \times n$ matrix (m – number of rows & n – number of columns). Then,

- The *row space* is the subspace of $\mathbb{R}^n$ spanned by the m row vectors of $\boldsymbol{A}$.

- The *column space* is the subspace of $\mathbb{R}^m$ spanned by the n column vectors of $\boldsymbol{A}$.

Note, that following the definition of vector spaces the following is true

- The set of all possible linear combinations of the columns of $\boldsymbol{A}$ is called the column space of $\boldsymbol{A}$ and is written as

$$\mathcal{C}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^m : \boldsymbol{x} = \boldsymbol{A}\boldsymbol{c} \text{ for some } \boldsymbol{c} \in \mathbb{R}^n\}.$$

Any member of the set $\mathcal{C}(\boldsymbol{A})$ is an $m \times 1$ vector, so $\mathcal{C}(\boldsymbol{A})$ is a subset of $\mathbb{R}^m$.

- The set of all possible linear combinations of the rows of $\boldsymbol{A}$ is called the row space of $\boldsymbol{A}$ and is written as

$$\mathcal{R}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{d} : \boldsymbol{d} \in \mathbb{R}^m\}.$$

Any member of the set $\mathcal{R}(\boldsymbol{A})$ is an $n \times 1$ vector, so $\mathcal{R}(\boldsymbol{A})$ is a subset of $\mathbb{R}^n$. Additionally, observe that $\mathcal{R}(\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A}^\top)$.

- Note that $\mathcal{C}(\boldsymbol{A}) \subseteq \mathbb{R}^m$ and $\mathcal{R}(\boldsymbol{A}) \subseteq \mathbb{R}^n$.

- If $\boldsymbol{A}$ denote a $m \times n$ matrix, the row space and column space of $\boldsymbol{A}$ have the same dimension.

- The dimension of the row (or column) space of $\boldsymbol{A}$ is called *rank* of $\boldsymbol{A}$ denoted as $\text{rank}(\boldsymbol{A})$. (see more below)

Figure 4: Vector Space

(e) Fill in the blank.

The column rank of a model matrix **X** is always **equal to** the number of linearly independent vectors that span the vector space in part (d).

The above answer is directly tied back to the definition of estimability, in addition to the Linear Algebra definition of span.

(f) What is the relationship/connection between the column rank of **X** and the estimability of $\boldsymbol{\beta}$? Answer using a short sentence.

The column rank of **X** determines the number of estimable parameters, i.e., the higher/larger/greater the column rank (Integer) of **X**, the more parameters in $\boldsymbol{\beta}$ that are estimable.

5.

Consider the following linear model with $n = 5$ observations:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}$$

Note that the columns of $\mathbf{X}$ are perpendicular so that $\mathbf{X}^T\mathbf{X}$ is diagonal.

(a) In a Gauss-Markov version of this model, which of the parameters, $\beta_1, \beta_2, \beta_3$, can be estimated with greatest precision? Explain carefully.

In a Gauss-Markov version of this model, β_1 can be estimated with the greatest precision.

β_1 corresponds to the first column of the design matrix, $\mathbf{X}$, which has the largest diagonal entry in $\mathbf{X}^T\mathbf{X}$, corresponding to the 5 rows we can use in estimating β_1. This means we have more information to estimate β_1 which in turn corresponds to a smaller variance in estimation.

Calculating explicitly:

$$\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

By contrast, the other two parameter vectors, β_2 and β_3 can only be estimated using 4 of the rows, so they use comparatively less information and would have greater variance, less precision.

(b) Suppose $\mathbf{y}$ is such that $\text{SSE} = 3$ and $\hat{\boldsymbol{\beta}} = (5\ 6\ 2)^T$. Consider an analysis under the Gauss-Markov model with Normal errors and the following two null hypotheses:

$$H_{0,1} : E(y_1) = E(y_2) \quad \text{and} \quad H_{0,2} : E(y_1) = E(y_5)$$

  i. Write $H_{0,1}$ and $H_{0,2}$ in testable form $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$ by identifying an appropriate matrix $\mathbf{C}$. (**Hint:** Start out by expressing each expected value as a function of $\boldsymbol{\beta}$ given $\mathbf{X}$ and $\boldsymbol{\beta}$ as defined above.)

From the model, the expected value for each observation is given by:

$$E[y_i] = \mathbf{x}_i^\top \boldsymbol{\beta}$$

Row 1: y_1, $\mathbf{x}_1^\top = (1,0,0) \Rightarrow \mathbb{E}[y_1] = \beta_1$

Row 2: y_2, $\mathbf{x}_2^\top = (1,-1,-1) \Rightarrow \mathbb{E}[y_2] = \beta_1 - \beta_2 - \beta_3$

Row 5: y_5, $\mathbf{x}_5^\top = (1,1,1) \Rightarrow \mathbb{E}[y_5] = \beta_1 + \beta_2 + \beta_3$

For $H_{0,1} : \mathbb{E}[y_1] = \mathbb{E}[y_2]$

$$\beta_1 = \beta_1 - \beta_2 - \beta_3 \rightarrow 0 = -\beta_2 - \beta_3 \rightarrow [0\ 1\ 1]\boldsymbol{\beta} = 0$$

For $H_{0,2} : \mathbb{E}[y_1] = \mathbb{E}[y_5]$

$$\beta_1 = \beta_1 + \beta_2 + \beta_3 \rightarrow 0 = \beta_2 + \beta_3 \rightarrow [0\ 1\ 1]\boldsymbol{\beta} = 0$$

Taken together, we have:

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$$

Noting we can have a matrix $\mathbf{C}$ with just one row for this problem uniquely determined by the above linear combinations simplifying to the same expressions.

ii. Based on $\mathbf{C}$ compute an F statistic for testing H_0 (you need not do the arithmetic, but plug correct numbers into a correct formula).

$$
\begin{aligned}
F &= (\textbf{estimate} - d)^\top \left[\widehat{\mathrm{Var}}(\textbf{estimator}) \right]^{-1} (\textbf{estimate} - d)/q \\[2mm]
&= (C\widehat{\beta} - d)^\top [\widehat{\mathrm{Var}}(C\widehat{\beta})]^{-1}(C\widehat{\beta} - d)/q \\[2mm]
&= (C\widehat{\beta} - d)^\top [\widehat{\sigma}^2 C(X^\top X)^- C^\top]^{-1}(C\widehat{\beta} - d)/q \\[2mm]
&= \frac{(C\widehat{\beta} - d)^\top [C(X^\top X)^- C^\top]^{-1}(C\widehat{\beta} - d)/q}{\widehat{\sigma}^2}
\end{aligned}
$$

Figure 5: F-Statistic Definition

To test the hypothesis $H_0 : \mathbf{C}\beta = 0$, we the F-statistic defined by:

$$F = \frac{(\mathbf{C}\widehat{\beta})^T \left[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T \right]^{-1} (\mathbf{C}\widehat{\beta})/q}{\mathrm{MSE}}$$

With the following known values already calculated:

- $\mathbf{C} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$ From part ii).
- $\widehat{\beta} = \begin{pmatrix} 5 \\ 6 \\ 2 \end{pmatrix}$ As given in the beginning of the problem
- $\mathbf{X}^T\mathbf{X} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$ As calculated in part a).
- $q = 1$ (the number of constraints, as given from the number of rows in our matrix $\mathbf{C}$
- MSE $=$ SSE$/(n - p) = 3/(5 - 3) = 3/2 = 1.5$ where SSE was given and the denominator calculated using known information.

Computing this value directly then, and noting known method for finding the inverse of a 3x3 matrix, we have:

8

$$\mathbf{C}\hat{\boldsymbol{\beta}} = 0 \cdot 5 + 1 \cdot 6 + 1 \cdot 2 = 8$$

And:

$$\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1/5 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1/4 + 1/4 = 1/2$$

Combining the computations, the F-statistic is then given by:

$$F = \frac{8^2/1}{(1/2) \cdot 1.5} = \frac{64}{0.75} = 85.33$$

iii. Specify the reference distribution of F under the null hypothesis.

The F–Test ($H_0 : C\beta = d$)

To test $H_0 : C\beta = d$, we can use the following statistic

$$
\begin{aligned}
F &\equiv (C\hat{\beta} - d)^\top [\widehat{\mathrm{Var}}(C\hat{\beta})]^{-1}(C\hat{\beta} - d)/q \\
&= (C\hat{\beta} - d)^\top [\hat{\sigma}^2 C(X^\top X)^- C^\top]^{-1}(C\hat{\beta} - d)/q \\
&= \frac{(C\hat{\beta} - d)^\top [C(X^\top X)^- C^\top]^{-1}(C\hat{\beta} - d)/q}{\hat{\sigma}^2}.
\end{aligned}
$$

F has a non-central F–distribution with non-centrality parameter

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

and df q and $n - r$.

Figure 6: F-Statistic - Central

The F–Test continued ($H_0 : C\beta = d$)

The non-negative non-centrality parameter

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

is equal to zero if and only if $H_0 : C\beta = d$ is true.

If $H_0 : C\beta = d$ is true, the statistic F has a central F–distribution with q and $n - r$ degrees of freedom ($F_{q,n-r}$).

Figure 7: F-Statistic - Non-Central

Under the null hypothesis $H_0 : \mathbf{C}\boldsymbol{\beta} = 0$, the F-statistic:

$$F = \frac{(\mathbf{C}\hat{\boldsymbol{\beta}})^T \left[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\right]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}})/q}{\text{MSE}}$$

follows a central F-distribution with Numerator degrees of freedom: $q = 1$ and Denominator degrees of freedom: $n - p = 5 - 3 = 2$ (central meaning a non-central parameter of 0), i.e.:

$$H_0 : F \sim F_{1,2}(0)$$

Under the Alternative Hypothesis, the F-statistic follows a non-central F-distribution with the same numerator and denominator degrees of freedom, i.e.:

$$H_a : F \sim F_{1,2}(\lambda)$$

Where the non-centrality parameter λ is defined as:

$$\lambda = \frac{(\mathbf{C}\boldsymbol{\beta})^T \left[\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{C}^T\right]^{-1} (\mathbf{C}\boldsymbol{\beta})}{\sigma^2}$$

6.

Consider a completely randomized experiment in which a total of 10 freshly cut Gerber daisies were placed into 10 vases (one daisy per vase). The Gerber daisies were randomly assigned to five treatment groups with two Gerber daisies in each treatment group. The treatment corresponds to the amount of a chemical compound added to the water in each vase. Of interest is the longevity of the Gerber daisies measured in days.

Treatment	1	2	3	4	5
Amount of compound (g)	0	2	4	10	16

Suppose for $i = 1,$
$dots, 5$ and $j = 1, 2$, y_{ij} denotes the longevity in days of the study of the jth Gerber daisy from treatment group i. Furthermore, suppose

$$y_{ij} = \mu_i + \varepsilon_{ij},$$

where the μ_i are unknown parameters and the ε_{ij} terms are $\mathcal{N}(0, \sigma^2)$ for some unknown $\sigma^2 > 0$.

Use the R code and partial output provided with this exam to answer the following questions.

(a) For the first model fit in R, called M1, specify the model matrix $\mathbf{X}$ used by R.

The model matrix $\mathbf{X}$ used by R is a design matrix with an intercept and indicator columns for each level of `amt`. By default R produces a design matrix of full rank using the baseline constraint where the first level of the factor `amt` is used as the baseline. This corresponds to:

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

(b) Consider the following information from the output associated with model M1:

```
F-statistic: 30.84 on 4 and 5 DF,    p-value: 0.001019
```

Specify the null and alternative hypothesis associated with this test:

H_0: All treatment means (average response per level of `amt`) are equal ($\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$).

H_a: At least one treatment mean differs from the treatment mean of the other groups (levels of the factor `amt`).

(c) Provide the BLUE of μ_2:

By default the estimate of $\beta_2 = \mu_2 - \mu_1$ due to the R baseline constraint, as noted more in part a). Additionally, R by default and as specified provides the OLS estimated effects, which are BLUE as they correspond to the normal equations under GMM. That is all to say:

The BLUE of μ_2 is the estimated expression using "intercept + coefficient", which corresponds to $\hat{\mu}_2 = 2.32$ + 2.41 = 4.73.

(d) What is the standard error of the BLUE of μ_2?

The standard error of the BLUE of μ_2 is 0.8747, as given in the R output.

(e) Provide the BLUE of $\mu_1 - \mu_2$:

Under R's default treatment contrast coding, the model fitted with `lm(y ~ factor(amt))` uses the first level of the factor as the baseline, which is for `amt == 0`.

The model is of the form:

$$y_{ij} = \beta_0 + \beta_2 \cdot \mathbb{K}_{\text{amt}=2} + \beta_3 \cdot \mathbb{K}_{\text{amt}=4} + \cdots + \varepsilon_{ij},$$

Mean of treatment one is given by $\beta_0 = \mu_1$

However, mean of treatment two is not β_2, as $\beta_2 = \mu_2 - \mu_1$.

So we must consider a linear combination of rows in the β vector to get at $\mu_1 - \mu_2$. To that end:

$$\mu_1 - \mu_2 = \mu_1 - (\mu_1 + \beta_2) = -\beta_2$$

So, the BLUE of $\mu_1 - \mu_2$ is:

$$-\hat{\beta}_2 = -2.41$$

Using the above output

(f) What is the standard error of the BLUE of $\mu_1 - \mu_2$?

Since $\mu_1 - \mu_2 = -\beta_2$, this is a direct transformation of a single coefficient estimate.

Thus, the standard error of the BLUE is also, similar to the prior SE question, given from a single standard error estimate, i.e.:

$$\text{SE}(\mu_1 - \mu_2) = \text{SE}(-\hat{\beta}_2) = \text{SE}(\hat{\beta}_2) = 0.8747$$

(g) What is the value of $\mathbf{y}^T(\mathbf{I} - \mathbf{P_1})\mathbf{y}$, where $\mathbf{y}$ denotes the vector containing the values of longevity?

The quantity $\mathbf{y}^T(\mathbf{I} - \mathbf{P}_1)\mathbf{y}$ is the residual Sums of Squares from the fitted model, M1.

From the provided R output, the residual sum of squares is:

$$\mathbf{y}^T(\mathbf{I} - \mathbf{P}_1)\mathbf{y} = \text{RSS} = 3.826$$

(h) Provide the value of the F-statistic, numerator and denominator df, and the p-value associated with the following ANOVA table:

The F-Statistic is of the form:

$$F = \frac{\frac{SS_{amt}}{df_{amt}}}{\frac{SS_{Res}}{df_{Res}}} = \frac{94.398/4}{3.826/5} \approx 30.841$$

With p-value:

$$P(F_{4,5} > 30.84)$$

Taken together, we have:

F-statistic: 30.84
amt Df (Numerator Df): Number of Groups − 1 = 5 − 1 = 4 Residual Df (Denominator Df): Total Observations − Number of Groups = 10 − 5 = 5 p-value: 0.001019

In R, the p-value can be directly calculated. However, on the exam, I believe keeping it closed form would have been the only viable solution since the p-value was not explicitly provided in the output and could not have been backward-inducted using known values.

```
pf(30.84, df1 = 4, df2 = 5, lower.tail = FALSE)
```

```
## [1] 0.001018976
```

```
OLD
> anova(M1)
Analysis of Variance Table

Response: longevity
          Df   Sum Sq  Mean Sq  F value  Pr(>F)
amt       ---  94.398  ---      ---      ---
Residuals ---   3.826

NEW
> anova(M1)
Analysis of Variance Table

Response: longevity
          Df   Sum Sq  Mean Sq   F value   Pr(>F)
amt        4   94.398  23.5995   30.84     0.001019
Residuals  5    3.826   0.7652
```

For the general case, consider the test statistic

$$F = \frac{y^\top (P_X - P_{X_0})y / [\mathrm{rank}(X) - \mathrm{rank}(X_0)]}{y^\top (I - P_X)y / [n - \mathrm{rank}(X)]}.$$

Figure 8: Matrix Notation: Form of Partial F-Test

Thus, the F statistic has the familiar form

$$\frac{(SSE_{\mathrm{REDUCED}} - SSE_{\mathrm{FULL}})/(DFE_{\mathrm{REDUCED}} - DFE_{\mathrm{FULL}})}{SSE_{\mathrm{FULL}}/DFE_{\mathrm{FULL}}}.$$

Figure 9: Familiar Form of Partial F-Test

(i) Look at the output associated with Model 2, M2.

    i. Fill in the missing entries in the ANOVA table produced by the R command `anova(M2, M1)`.

```
OLD
> anova(M2, M1)
Analysis of Variance Table

Model 1: longevity ~ amount
Model 2: longevity ~ amt

  Res.Df   RSS     Df    Sum of Sq   F      Pr(>F)
1   ---    ---
2   ---    ---    ---    ---(*)      ---    0.006412 **


NEW
> anova(M2, M1)
Analysis of Variance Table

Model 1: longevity ~ amount
Model 2: longevity ~ amt

  Res.Df   RSS     Df    Sum of Sq   F      Pr(>F)
1   9      37.757
2   5       3.826   4     33.931     2.02   0.006412 **
```

The above values were found using the known relation between Sums of Squares and RSS, and the corresponding F-statistic, similar to the calculations done previously, but with the p-value provided explicitly.

The other key feature of this problem is recognizing the Residual Df. of 9 for the simplified model (only taking one degree of freedom when allowing `amt` to be treated as a continuous variable instead of multiple group, cell, means).

    ii. Provide an interpretation of Sum of Squares in part (i). This is the value denoted by (*).

The value (*) = 33.931 represents the additional variability, the sum of squared errors, explained by using a more complex model that treats `amt` as a factor, compared to a simpler model that treats amount as a continuous numeric variable.

This sum of squares quantifies the improvement in model fit when we allow each treatment level to have its own group (cell) mean, rather than assuming a single linear relationship across all levels of the compound amount.

    iii. Provide a conclusion in the context of the data about the null hypothesis that is tested in part (i).

The small p-value (0.006412) from the partial F-test comparing `M2` to `M1` provides strong evidence against the null hypothesis.

The null hypothesis for this test is that the simpler model (with a linear relationship between compound amount and longevity) is sufficient, i.e., there is no significant improvement in model fit by allowing separate means for each treatment level.

In the context of this experiment, we say there is evidence that the relationship between the amount of chemical compound added to the water and the longevity of Gerber daisies is not well explained by a simple linear trend, and instead (contra-positively) have evidence that treatment effects vary non-linearly across dose levels, and the more flexible model—where each treatment level has its own mean (i.e. the cell means model) better captures the pattern in the data and explains more of the observed variability in flower longevity.

```
> amount<-c(0, 0, 2, 2, 4, 4, 10, 10,  16, 16)
> plot(amount, longevity, ylim = c(0,14))
> amt=as.factor(amount)
> M1<-lm(longevity~amt)
> summary(M1)

Call:
lm(formula = longevity ~ amt)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.3200     0.6185   3.751 0.013281 *
amt2          2.4100     0.8747   2.755 0.040063 *
amt4          6.0850     0.8747   6.957 0.000943 ***
amt10         8.2400     0.8747   9.420 0.000227 ***
amt16         7.0300     0.8747   8.037 0.000482 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.8747 on 5 degrees of freedom
Multiple R-squared:  0.9611,Adjusted R-squared:  0.9299
F-statistic: 30.84 on 4 and 5 DF,  p-value: 0.001019

> anova(M1)
Analysis of Variance Table

Response: longevity
          Df Sum Sq Mean Sq F value   Pr(>F)
amt          94.398
Residuals     3.826
---

> is.numeric(amount)
[1] TRUE

> M2<-lm(longevity~amount)

> anova(M2)
Analysis of Variance Table

Response: longevity
          Df Sum Sq Mean Sq F value   Pr(>F)
amount       60.467
Residuals    37.757
```

Figure 10: R Output, Last Page