GROUND RULES:

- This exam contains 5 questions. Each question is worth 20 points. The maximum number of points on this exam is 100.
- This is a closed-book and closed-notes exam.
- Show all of your work. Explain all of your reasoning.
- Any discussion or otherwise inappropriate communication between examinees, as well as the appearance of any unnecessary material, will be dealt with severely.
- You have 3 hours to complete this exam. GOOD LUCK!

HONOR PLEDGE FOR THIS EXAM:

After you have finished the exam, please read the following statement and sign your name below it.

> *I promise that I did not discuss any aspect of this exam with anyone other than the instructor, that I neither gave nor received any unauthorized assistance on this exam, and that the work presented herein is entirely my own.*

1. Define the matrix

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

Note that $r(\mathbf{A}) = 3$.

(a) Find $\mathcal{C}(\mathbf{A})$.
(b) Find a 4×4 matrix $\mathbf{B}$ whose column space $\mathcal{C}(\mathbf{B}) \subset \mathcal{C}(\mathbf{A})$. Prove your assertion.
(c) For your choice of $\mathbf{B}$ in part (b), find a basis for $\mathcal{N}(\mathbf{B}')$.
(d) Let $\mathbf{M}$ denote the perpendicular projection matrix onto $\mathcal{C}(\mathbf{A})^{\perp}$. Find $tr(\mathbf{I} - \mathbf{M})$, where $\mathbf{I}$ is the identity matrix with the same dimensions as $\mathbf{M}$.
(e) Suppose that $\mathbf{Y} \sim \mathcal{N}_4(\mathbf{0}, \mathbf{I})$. Find the moment generating function of $\mathbf{X} = \mathbf{AY}$. Describe the distribution of $\mathbf{X}$.

2. Define

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & -1 \\ 1 & 1 & 0 & -1 & 0 \\ 1 & 1 & 0 & 0 & -1 \\ 1 & 1 & -1 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{pmatrix}.$$

Assume the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ holds, where $E(\boldsymbol{\epsilon}) = \mathbf{0}$. Let $\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, and $\mathbf{x}_4$ denote the columns of $\mathbf{X}$. Note that $\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 + \mathbf{x}_4 = \mathbf{0}$.

(a) Find the components of $E(\mathbf{Y})$ in terms of $\mu, \theta_1, \theta_2, \theta_3$, and θ_4.
(b) Let $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4)'$. Give conditions on $\boldsymbol{\lambda}$, of the form $\boldsymbol{\lambda}'\mathbf{c}_i = 0$, $i = 1, 2, ..., s$, that are necessary and sufficient for $\boldsymbol{\lambda}'\boldsymbol{\beta}$ to be estimable. What is the value of s?
(c) Show that $\mu - \theta_1 + \theta_2 + \theta_3 - \theta_4$ is estimable.
(d) Give a nonestimable function of the form $\boldsymbol{\lambda}'\boldsymbol{\beta}$. Explain your answer.
(e) Explain briefly how you could use your answer in part (d) to force a particular solution to the normal equations $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$. If you did this, would your nonestimable $\boldsymbol{\lambda}'\boldsymbol{\beta}$ in part (d) "become" estimable? Explain.

3. Consider the cell means ANOVA model

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

for $i = 1, 2, 3$ and $j = 1, 2, ..., n$, where ϵ_{ij} are iid $\mathcal{N}(0, \sigma^2)$. The restriction

$$\mu_3 = \mu_1 - \mu_2$$

is placed on the parameters. Set $\boldsymbol{\beta} = (\mu_1, \mu_2, \mu_3)'$.

(a) Write this as a general linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. How do you express the restriction in the form $\mathbf{P}'\boldsymbol{\beta} = \boldsymbol{\delta}$?
(b) Find the restricted least squares estimator, $\widehat{\boldsymbol{\beta}}_H$. Express this estimator in terms of the treatment means $\overline{Y}_{i+}$, for $i = 1, 2, 3$.
(c) Define

$$Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

and let $\widehat{\boldsymbol{\beta}}$ denote the unrestricted least squares estimator. How do $Q(\widehat{\boldsymbol{\beta}})$ and $Q(\widehat{\boldsymbol{\beta}}_H)$ compare?
(d) In part (c), find $E[Q(\widehat{\boldsymbol{\beta}})]$ and $\text{var}[Q(\widehat{\boldsymbol{\beta}})]$.
(e) Consider testing $H_0 : \mu_3 = \mu_1 - \mu_2$ using the statistic

$$\frac{(\mathbf{P}'\widehat{\boldsymbol{\beta}} - \boldsymbol{\delta})'\mathbf{H}^{-1}(\mathbf{P}'\widehat{\boldsymbol{\beta}} - \boldsymbol{\delta})/s}{\text{MSE}},$$

where $\mathbf{H} = \mathbf{P}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{P}$ and $s = r(\mathbf{P})$. What is the distribution of this statistic when H_0 is true?

4. Consider an experiment with $I = 3$ treatments and a covariate x. Corresponding to each data value Y_{ij}, there is a covariate value x_{ij}. The covariate values are known, fixed numbers. Consider the following two ANCOVA models:

$$\begin{aligned} &\text{Model A:} \quad Y_{ij} = \mu + \alpha_i + \beta_1 \overline{x}_{i+} + \beta_2(x_{ij} - \overline{x}_{i+}) + \epsilon_{ij} \\ &\text{Model B:} \quad Y_{ij} = \mu_i + \beta x_{ij} + \epsilon_{ij}. \end{aligned}$$

Both models are defined for $i = 1, 2, 3$ and $j = 1, 2, 3, 4$. The ϵ_{ij} are assumed to be uncorrelated random variables with zero mean and constant variance $\sigma^2 > 0$. Assume that $\sum_{j=1}^{4}(x_{ij} - \overline{x}_{i+})^2 > 0$ for each $i = 1, 2, 3$.

(a) Write Model A as a linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Identify $\mathbf{X}$ and $\boldsymbol{\beta}$ explicitly. Repeat for Model B. Identify $\mathbf{W}$ and $\boldsymbol{\gamma}$ explicitly.
(b) Show that $\mathcal{C}(\mathbf{X}) = \mathcal{C}(\mathbf{W})$.
(c) Let SSE_A denote the residual sum of squares from fitting Model A. Let SSE_B denote the residual sum of squares from fitting Model B. How do SSE_A and SSE_B compare?
(d) In Model B, consider testing

$$\begin{gathered} H_0 : \mu_1 = \mu_2 = \mu_3 = 0 \\ \text{versus} \\ H_1 : \text{not } H_0. \end{gathered}$$

Suppose that the F statistic for this test was $F = 0.12$. What would you conclude?
(e) In Model A, consider the reduced model that arises when $\beta_1 = \beta_2 = 0$. Write this reduced model as $\mathbf{Y} = \mathbf{U}\boldsymbol{\theta} + \boldsymbol{\epsilon}$ and compute $\mathbf{P}_\mathbf{U}$, the perpendicular projection matrix onto $\mathcal{C}(\mathbf{U})$.

5. Consider the general linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is $n \times p$ with rank $r \leq p$, $\boldsymbol{\beta}$ is $p \times 1$, and $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{V})$, where $\mathbf{V}$ is known and nonsingular. Let $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$ denote an ordinary least squares estimator. Define

$$\widehat{\sigma}^2 = (n-r)^{-1}\mathbf{Y}'(\mathbf{I} - \mathbf{P_X})\mathbf{Y},$$

where $\mathbf{P_X}$ is the perpendicular projection matrix onto $\mathcal{C}(\mathbf{X})$. Suppose that $\boldsymbol{\lambda}'\boldsymbol{\beta}$ is a scalar (linearly) estimable function.

(a) Define what it means for $\boldsymbol{\lambda}'\boldsymbol{\beta}$ to be (linearly) estimable.
(b) Suppose that $\mathbf{V} = \sigma^2\mathbf{I}$. Derive the sampling distribution of $\boldsymbol{\lambda}'\widehat{\boldsymbol{\beta}}$.
(c) Suppose that $\mathbf{VX} = \mathbf{XQ}$ for some matrix $\mathbf{Q}$. Prove that $\boldsymbol{\lambda}'\widehat{\boldsymbol{\beta}}$ and $(\mathbf{I} - \mathbf{P_X})\mathbf{Y}$ are independent.
(d) Suppose that $\mathbf{V} = \sigma^2(\mathbf{I} + \mathbf{P_X})$, for some $\sigma^2 > 0$. Define

$$T = \frac{\boldsymbol{\lambda}'\widehat{\boldsymbol{\beta}} - \boldsymbol{\lambda}'\boldsymbol{\beta}}{\sqrt{\widehat{\sigma}^2\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\boldsymbol{\lambda}}}.$$

Find the constant k so that kT follows a central t distribution with $n-r$ degrees of freedom.
(e) As in part (d), suppose that $\mathbf{V} = \sigma^2(\mathbf{I} + \mathbf{P_X})$. Use the result in part (d) to derive a $100(1-\alpha)$ percent confidence interval for $\boldsymbol{\lambda}'\boldsymbol{\beta}$.