1. Write the following models using matrix notation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$:

(a) Multiple regression

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i;\ i = 1, 2, ..., 6.$$

(b) Multiple polynomial regression

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1}^2 + \beta_4 x_{i2}^2 + \beta_5 x_{i1} x_{i2} + \epsilon_i;\ i = 1, 2, ..., 6.$$

(c) Two-way ANCOVA with no interaction and common slope

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma x_{ijk} + \epsilon_{ijk};\ i = 1, 2, 3,\ j = 1, 2,\ k = 1, 2.$$

2. Consider an experiment to compare six oil refineries. Three refineries are in Texas and three are in Oklahoma. The observed variable is the amount of gasoline produced in a day. For each refinery, there are observations on two days. We write a model for these data as

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk},$$

for $i = 1, 2$, $j = 1, 2, 3$, and $k = 1, 2$. Here α refers to the state, and β refers to the refinery within state. This model can be written as a linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, \beta_{23})'$.

(a) Write out $\mathbf{Y}$, $\mathbf{X}$, and $\boldsymbol{\epsilon}$ for this model.
(b) Compute $\mathbf{X}'\mathbf{X}$.
(c) Show that $\mathbf{X}$ is not of full column rank. Specifically, show that there are three columns of $\mathbf{X}$ that can be written as linear combinations of the other six columns.
(d) Show that the remaining six columns of $\mathbf{X}$ are linearly independent.

3. For the linear model in Problem 2, place the following restrictions on the parameters:

- $\alpha_1 + \alpha_2 = 0$
- $\beta_{11} + \beta_{12} + \beta_{13} = 0$
- $\beta_{21} + \beta_{22} + \beta_{23} = 0$.

Under these restrictions, the model can be written as a linear model $\mathbf{Y} = \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, where $\boldsymbol{\gamma} = (\mu, \alpha_1, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})'$.

(a) Write out $\mathbf{Y}$, $\mathbf{W}$, and $\boldsymbol{\epsilon}$ for the model under these restrictions.
(b) Show that $\mathbf{W}$ is of full column rank.
(c) Compute $\mathbf{W}'\mathbf{W}$ and $(\mathbf{W}'\mathbf{W})^{-1}$.
(d) Show that $\mathcal{C}(\mathbf{W}) = \mathcal{C}(\mathbf{X})$, where $\mathbf{X}$ is the design matrix from Problem 2.

4. Consider the linear model

$$Y_{ij} = \mu + \alpha_i + \alpha_j + \epsilon_{ij},$$

for $i = 1, 2, 3$ and $j = 1, 2, 3$, where μ, α_1, α_2, and α_3 are fixed unknown parameters and $\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \alpha_3)'$.

(a) Write this model using the matrix notation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.
(b) Are the columns of $\mathbf{X}$ linearly independent? If they are, prove it. If they are not, demonstrate where the linear dependencies are.

REMARK: Note that this is not a two-way ANOVA model. As a frame of reference, imagine that we have three fertilizers and that we apply a standard dose of two of them to plots, laid out in a 3×3 square (so that there are 9 plots). If $i = j$, we are applying twice the standard dose of fertilizer i. If $i \neq j$, we are applying a standard dose of fertilizer i and fertilizer j. The response Y_{ij} denotes the yield for the ijth plot.

(c) For this part only, suppose that the model is misspecified in that if the standard dose of a fertilizer is doubled, the effect on yield is less than or more than doubled by a fixed amount, say, q, which doesn't depend on the fertilizer, whereas the model is correct if two different fertilizers are applied. If this misspecification occurs, compute $E(Y_{ij})$ for each cell, under the assumption that $E(\epsilon_{ij}) = 0$. How would you "test" $H_0 : q = 0$? Provide a strategy to do this (it can be conceptual).

5. A researcher studies the effects of 3 different diets on weight gains in rats. She chooses 12 rats at random (from a large population) and assigns the rats to exactly one of the 3 diets. She also records the covariate x (initial weight), because she knows the response Y (weight gain) will depend on x. Assume that 4 rats are assigned to each diet.

(a) Write out a linear model, in non-matrix notation, that relates the response Y to the covariate x and the diets. Clearly define all of your notation. Use appropriate subscripts to denote the different diets and rats.
(b) Take your model in (a) and write it in the form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Define all vectors and matrices.
(c) What assumptions must be true for your model to be a Guass-Markov model?