1. Suppose $\mathbf{y}$ is an $n \times 1$ response vector and X is an $n \times p$ design matrix.

   (a) State the Gauss-Markov linear model.

   (b) Provide a matrix formula for the best linear unbiased estimator (BLUE) of $E(\mathbf{y})$ in terms of X and $E(\mathbf{y})$.

   (c) State the normal equations.

(d) Suppose $\mathbf{b}$ is any solution to the normal equations. Is it necessarily true that the best linear unbiased estimator in part (b) is equal to $X\mathbf{b}$? Prove that your answer is correct.

2. Consider an experiment designed to study the effect of two dietary factors, protein source and protein amount, on weight gain in pigs. A total of 12 pigs were randomly assigned to treatment with one of six combinations of protein source (1 or 2) and protein amount (1, 2, or 3 units). A completely randomized design was used with two individually penned pigs per treatment group. Let y_{ijk} denote the amount of weight gained during the study period by the k^{th} pig fed j units of protein from source i ($i = 1; 2; j = 1, 2, 3; k = 1, 2$). Consider the model

$$y_{ijk} = \mu + \alpha_i + \beta x_j + \epsilon_{ijk}$$

where $\mu,\ \alpha_1,\ \alpha_2$ and β are unknown real-valued parameters, $x_j = j - 2$, the ϵ_{ijk}'s are $NID \sim (0, \sigma^2)$, and σ^2 is an unknown parameter. Suppose

$$\mathbf{y} = (y_{111}, y_{112}, y_{121}, y_{122}, \cdots, y_{231}, y_{232})^T$$

$$\boldsymbol{\epsilon} = (\epsilon_{111}, \epsilon_{112}, \epsilon_{121}, \epsilon_{122}, \cdots, \epsilon_{231}, \epsilon_{232})^T$$

and

$$\boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \beta)^T$$

(a) Provide the appropriate design matrix X so that the model may be written as $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

(b) For each of the quantities below, state whether the quantity is estimable and prove that your answer is correct.

i. $\mu + \alpha_1$

ii. $\mu + \alpha_1 + 10\beta$

iii. $\alpha_1 - \alpha_2$

iv. μ

(c) Write down a full-column-rank matrix that has the same column space as X in part (a).

(d) Use your answer to part (c) to find a simplified expression for the BLUE of $E(y_{111})$ in terms of the y_{ijk} values.

(e) Provide the least squares estimate of each estimable quantity in part (b).

3. Once again consider the experiment described in problem 2. Consider the model

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk} \; (i = 1, 2; j = 1, 2, 3; k = 1, 2)$$

where $\mu_{11}, \cdots, \mu_{23}$ are unconstrained, unknown, real-valued cell mean parameters; the ϵ_{ijk}'s are $NID(0, \sigma^2)$ and σ^2 is an unknown parameter. Use the R code and output provided on the last page of this exam to complete the following parts.

(a) Is there evidence of a difference among the six treatment means? Provide a test statistic, its degrees of freedom, a p-value, and a conclusion.

(b) Provide the BLUE of μ_{23}.

(c) If possible, provide the standard error for the estimate in part (b). If it is not possible to determine the standard error using the information provided, explain why.

(d) Provide a matrix C so that $C\% * \%coef(o)$ is an estimate of the main effect of protein source.

(e) If the model in problem 2 were fit to these data, what would the estimate of the error variance σ^2 be?

(f) Suppose the researchers would like to know if the model specified in the statement of problem 2 fits these data adequately relative to the cell means model specified in the statement of problem 3. Compute a test statistic that can be used to answer this question and state the degrees of freedom associated with this test statistic.

Your Score: ______/100

```
> y=#DATA NOT SHOWN#

> s=factor(rep(1:2,each=6))
> s
 [1] 1 1 1 1 1 1 2 2 2 2 2 2
Levels: 1 2

> x=rep(rep(c(-1,0,1),each=2),2)
> x
 [1] -1 -1  0  0  1  1 -1 -1  0  0  1  1

> o=lm(y~s+x+I(x^2)+s:x+s:I(x^2))

> anova(o)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq  F value    Pr(>F)
s          1  90.75   90.75  27.9231  0.001858 **
x          1 528.12  528.12 162.5000 1.430e-05 ***
I(x^2)     1   0.04    0.04   0.0128  0.913544
s:x        1  28.13   28.13   8.6538  0.025889 *
s:I(x^2)   1   0.37    0.37   0.1154  0.745670
Residuals  6  19.50    3.25
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1


> summary(o)

Call:
lm(formula = y ~ s + x + I(x^2) + s:x + s:I(x^2))

Residuals:
       Min         1Q     Median         3Q        Max
-2.000e+00 -1.125e+00 -1.712e-16  1.125e+00  2.000e+00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  17.5000     1.2748  13.728 9.29e-06 ***
s2            5.0000     1.8028   2.774 0.032273 *
x             6.2500     0.9014   6.934 0.000446 ***
I(x^2)       -0.2500     1.5612  -0.160 0.878035
s2:x          3.7500     1.2748   2.942 0.025889 *
s2:I(x^2)     0.7500     2.2079   0.340 0.745670
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.803 on 6 degrees of freedom
Multiple R-squared: 0.9708,     Adjusted R-squared: 0.9464
F-statistic: 39.84 on 5 and 6 DF,  p-value: 0.0001587
```