STAT 510 Homework 5

Due Date: 11:00 A.M., Wednesday, February 15

1. Consider a completely randomized experiment in which a total of 10 rats were randomly assigned to 5 treatment groups with 2 rats in each treatment group. Suppose the different treatments correspond to different doses of a drug in milliliters per gram of body weight as indicated in the following table.

Treatment	1	2	3	4	5
Dose of Drug (mL/g)	0	2	4	8	16

Suppose for $i = 1, \ldots, 5$ and $j = 1, 2$, y_{ij} is the weight at the end of the study of the jth rat from the ith treatment group. Furthermore, suppose

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

where $\mu_1, \ldots, \mu_5$ are unknown parameters and the ϵ_{ij} terms are $iid\ N(0, \sigma^2)$ for some unknown $\sigma^2 > 0$. Use the R code and partial output provided below to answer the following questions.

```
> d=rep(c(0,2,4,8,16),each=2)
> #y is the data vector with entries ordered to appropriately
> #match the vector d.
> dose=factor(d)
> o1=lm(y~dose)
> summary(o1)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  351.000      6.576  53.372 4.37e-08 ***
dose2        -10.000      9.301  -1.075 0.331406
dose4         -6.000      9.301  -0.645 0.547277
dose8        -17.000      9.301  -1.828 0.127119
dose16       -70.500      9.301  -7.580 0.000634 ***

> anova(o1)

Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
dose         6505.6
Residuals     432.5

> is.numeric(d)
[1] TRUE

> o2=lm(y~d)
> anova(o2)
```

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
d            5899.6
Residuals    1038.5

> o3=lm(y~d+dose)
> anova(o3)

Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
d                                   0.0004245 ***
dose                                0.1907591
Residuals
```

(a) Provide the numerical value of the BLUE of μ_1.

(b) Provide the numerical value of the BLUE of μ_2.

(c) Determine the standard error of the BLUE of μ_2.

(d) Conduct a test of $H_0 : \mu_1 = \mu_2$. Provide a test statistic, the distribution of that test statistic (be very precise), a p-value, and a conclusion.

(e) Provide an F-statistic for testing $H_0 : \mu_3 = \mu_4$.

(f) Does a simple linear regression model with body weight as a response and dose as a quantitative explanatory variable fit these data adequately? Provide a test statistic, its degrees of freedom, a p-value, and a conclusion.

(g) Provide a matrix $\boldsymbol{C}$ and a vector $\boldsymbol{d}$ so that the null hypothesis of the test in part (f) may be written as $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$, where $\boldsymbol{\beta} = (\mu_1, \ldots, \mu_5)'$.

(h) Fill in the missing entries in the ANOVA table produced by the R command `anova(o3)`. (This is the last R command in the provided code.)

2. Suppose $\boldsymbol{X}$ is an $n \times p$ matrix and $\boldsymbol{B}$ is a $p \times p$ non-singular matrix. Prove that

$$\mathcal{C}(\boldsymbol{X}) = \mathcal{C}\left(\boldsymbol{X}\boldsymbol{B}^{-1}\right).$$

3. Suppose $\boldsymbol{X}$ and $\boldsymbol{W}$ are any two matrices with n rows for which $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{W})$. Show that $\boldsymbol{P_X} = \boldsymbol{P_W}$, i.e., show that $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}' = \boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-}\boldsymbol{W}'$.

4. Consider an experiment conducted at two research labs. Within each lab, four mice were assigned two treatments using a completely randomized design with two mice per treatment. Let y_{ijk} be the response variable measurement for the kth mouse that received treatment j in research lab i ($i = 1, 2;\ j = 1, 2;\ k = 1, 2$). Suppose

$$y_{ijk} = \mu + \lambda_i + \tau_j + \epsilon_{ijk}, \tag{1}$$

where μ, λ_1, λ_2, τ_1, and τ_2 are unknown parameters and the ϵ_{ijk} terms are independent normal random variables with mean 0 and some unknown variance σ^2. Let

$$\boldsymbol{y} = (y_{111}, y_{112}, y_{121}, y_{122}, y_{211}, y_{212}, y_{221}, y_{222})',$$

$$\boldsymbol{\epsilon} = (\epsilon_{111}, \epsilon_{112}, \epsilon_{121}, \epsilon_{122}, \epsilon_{211}, \epsilon_{212}, \epsilon_{221}, \epsilon_{222})',$$

and

$$\boldsymbol{\beta} = (\mu, \lambda_1, \lambda_2, \tau_1, \tau_2)'.$$

(a) Give the entries in a matrix $\boldsymbol{X}$ so that the model defined in equation (1) on page 2 can be written as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

(b) Prove that $\tau_1 - \tau_2$ is estimable.

(c) Give the entries in a matrix $\boldsymbol{X}^*$ that has all of the following properties:

- $\boldsymbol{X}^*$ has the same column space as $\boldsymbol{X}$.
- $\boldsymbol{X}^*$ has full-column rank.
- The columns of $\boldsymbol{X}^*$ are orthogonal; i.e., if $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ are any two columns of $\boldsymbol{X}^*$, then $\boldsymbol{x}_u'\boldsymbol{x}_v = 0$.

(d) Define the elements of a vector $\boldsymbol{\beta}^*$ in terms of μ, λ_1, λ_2, τ_1, and τ_2 so that $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}^*\boldsymbol{\beta}^*$.

(e) Without the help of a computer or calculator, derive the ordinary least squares estimator of $\tau_1 - \tau_2$. To get full credit, fully simplify your answer so that it contains no matrices or vectors. (*Hint: You may want to work with* $\boldsymbol{X}^*$ *rather than* $\boldsymbol{X}$ *to derive the ordinary least squares estimator of* $\tau_1 - \tau_2$.)