STAT 510 Homework 13
No Due Date: Ungraded

1. Consider a generic repeated measures experiment like the experiment on strength training programs that we considered in class. Suppose there are three treatments indexed by $i = 1, 2, 3$ with n_i subjects indexed by $j = 1, \ldots, n_i$ for the ith treatment group. Suppose the response of interest is measured at t time points for each subject. Let y_{ijk} be the response for treatment i, subject j, and time point k ($i = 1, 2, 3; j = 1, \ldots, n_i; k = 1, \ldots, t$). For all i and j, let

$$\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijt})'.$$

Suppose all $\boldsymbol{y}_{ij}$ are mutually independent of one another and that, for all i and j,

$$\boldsymbol{y}_{ij} \sim N(\boldsymbol{\mu}_i, \boldsymbol{W}),$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{it})'$ and $\boldsymbol{W}$ is some unknown $t \times t$ positive definite and symmetric matrix. Let

$$\boldsymbol{y} = (\boldsymbol{y}'_{11}, \ldots, \boldsymbol{y}'_{1n_1}, \boldsymbol{y}'_{21}, \ldots, \boldsymbol{y}'_{2n_2}, \boldsymbol{y}'_{31}, \ldots, \boldsymbol{y}'_{3n_3})', \text{ and let } \boldsymbol{\beta} = (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2, \boldsymbol{\mu}'_3)'.$$

   (a) Use Kronecker product notation to specify a matrix $\boldsymbol{X}$ so that $E(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$.
   (b) Use Kronecker product notation to specify $\text{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma}$ in terms of $\boldsymbol{W}$.
   (c) Find a simplified expression for $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$.
   (d) Find a simplified expression for $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}$.
   (e) Find a simplified expression for $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$.
   (f) Give simplified expressions for the BLUEs of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\mu}_3$.

2. Do the following counts seem like they might be an independent and identically distributed sample from one Poisson distribution? Explain why or why not.

$$15, 9, 15, 23, 14, 18, 5, 7, 12, 11$$

3. Consider an experiment designed to compare the resistance of three plant genotypes (A, B, and C) to a fungal pathogen. Eight plants of each genotype were infected with the pathogen. After 24 hours, a leaf from each plant was sampled and examined under a microscope. The number of infected plant cells was recorded for each leaf. The smaller the number of infected cells the more resistant a plant tends to be to the fungal pathogen. Data are provided below. Is there evidence of a difference in resistance among the genotypes? Analyze these data and explain your conclusions to the researchers.

Genotype	Number of Infected Cells for Each Plant
A	39 31 43 31 34 36 34 24
B	23 28 24 19 16 20 25 12
C	36 38 33 22 23 17 29 16

4. For $i = 1, 2$ and $j = 1, \ldots, n_i$, suppose $\lambda_{ij} = \exp(\mu_i + e_{ij})$, where $e_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, and suppose $y_{ij}|\lambda_{ij} \overset{ind}{\sim} \text{Poisson}(\lambda_{ij})$. Consider three different tests of $H_0 : E(y_{1j}) = E(y_{2j})$. Test 1 is the Wald test conducted by assuming the data are Poisson distributed with no overdispersion (an incorrect assumption). Test 2 is like test 1 except that overdispersion is adjusted for using the quasilikelihood approach for Poisson data discussed in Slide Set 28. Test 3 is the Wald test conducted by fitting the generalized linear mixed effects model specified in this problem. Conduct a simulation study to estimate the type I error rate that will be incurred if the null hypothesis is rejected for p-values ≤ 0.05 using Test k ($k = 1, 2, 3$) for the case of $n_1 = n_2 = 5$, $\mu_1 = \mu_2 = 3$, and $\sigma = 0.25$.

5. This is essentially Computational Exercise 16 from Chapter 22 of *The Statistical Sleuth* by Ramsey and Schafer. Some sociologists suspect that highly publicized suicides may trigger additional suicides. In one investigation of this hypothesis, a researcher collected information about 17 airplane crashes that were known (because of notes left behind) to be murder-suicides. (That means that the pilot intentionally crashed the plane to kill him or herself and the passenger(s).) For each of these crashes, the researcher reported an index of the news coverage and the number of multiple-fatality plane crashes during the week following the publicized crash. The data are available at

   http://dnett.github.io/S510/PlaneCrashes.txt

   Is there evidence of an association between the news coverage index and the number of crashes in the following week? Conduct an analysis to address this question.

6. Suppose $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r]$ is an $n \times r$ matrix of rank r. A technique known as *Gram-Schmidt Orthogonalization* can be used to obtain an $n \times r$ matrix $\boldsymbol{W} = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_r]$ that has orthogonal columns and the same column space as $\boldsymbol{X}$. This can be useful because $\boldsymbol{P_W}$ is relatively easy to compute when the columns of $\boldsymbol{W}$ are orthogonal, and $\mathcal{C}(\boldsymbol{X}) = \mathcal{C}(\boldsymbol{W})$ implies $\boldsymbol{P_X} = \boldsymbol{P_W}$. An algorithm for carrying out Gram-Schmidt Orthogonalization is as follows:

   - Let $\boldsymbol{w}_1 = \boldsymbol{x}_1$.
   - Let $\boldsymbol{w}_2 = (\boldsymbol{I} - \boldsymbol{P}_{[\boldsymbol{w}_1]})\boldsymbol{x}_2$.
   - Let $\boldsymbol{w}_3 = (\boldsymbol{I} - \boldsymbol{P}_{[\boldsymbol{w}_1, \boldsymbol{w}_2]})\boldsymbol{x}_3$.

     $\vdots$
   - Let $\boldsymbol{w}_r = (\boldsymbol{I} - \boldsymbol{P}_{[\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{r-1}]})\boldsymbol{x}_r$.

   Now consider an experiment with two factors: A and B. Suppose that the levels of factor A are indexed by $i = 1, 2$. Suppose the levels of factor B are indexed by $j = 1, 2$. For $i = 1, 2$ and $j = 1, 2$, let n_{ij} be the number of observations for the treatment combination of level i of factor A and level j of factor B. For $i = 1, 2$ and $j = 1, 2$ and $k = 1, \ldots, n_{ij}$, suppose

   $$y_{ijk} = \mu_{ij} + e_{ijk},$$

   where the μ_{ij} terms are unknown parameters and the e_{ijk} terms are independent and identically distributed as $N(0, \sigma^2)$. The following table contains response averages and the number of observations for each treatment group.

Level of Factor A	Level of Factor B	Average Response ($\bar{y}_{ij\cdot}$)	Number of Observations (n_{ij})
1	1	3.0	2
1	2	5.0	8
2	1	7.0	6
2	2	3.0	4

   (a) Provide a model matrix $\boldsymbol{W}$ with orthogonal columns that corresponds to the additive model where $\mu_{ij} = \mu + \alpha_i + \beta_j$ for all $i = 1, 2$, $j = 1, 2$, and some unknown parameters μ, α_1, α_2, β_1, and β_2.

(b) Without the aid of a computer, find the value of $\boldsymbol{P_W y}$. Note that it is usually easiest to compute this by finding $(\boldsymbol{W}'\boldsymbol{W})^{-1}$, $\boldsymbol{W}'\boldsymbol{y}$, $(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{y}$, and then finally $\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{y}$. Even though you don't have all the elements of $\boldsymbol{y}$ it is still possible to compute $\boldsymbol{P_W y}$ from the averages in the table above.

(c) Without the aid of a computer, find the Type II sum of squares for factor B. This question can be answered by making use of part (b).