Due: Wednesday, February 19th 11:59PM in gradescope.

Problem 1 *Suppose $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ for some unknown $\sigma^2 > 0$. Let $\widehat{\boldsymbol{y}} = \boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{y}$.*

a) *Determine the distribution of*

$$\begin{bmatrix} \widehat{\boldsymbol{y}} \\ \boldsymbol{y} - \widehat{\boldsymbol{y}} \end{bmatrix}.$$

b) *Determine the distribution of $\widehat{\boldsymbol{y}}^\top \widehat{\boldsymbol{y}}$.*

Problem 2 *An experiment was conducted to study the durability of coated fabric subjected to abrasive tests. Three factors were considered. One factor was filler type with two levels (F1 and F2). Another was surface treatment with two levels (S1 and S2). The third factor was proportion of filler with three levels (25%, 50%, and 75%). Using a completely randomized design with two fabric samples per treatment, the amount of fabric lost in milligrams for each fabric sample was recorded following testing. Data are available in a tab delimited text file at http://dnett.github.io/S510/FabricLoss.txt.*

(a) *Consider a cell means model for these data. Estimate the mean and standard error for the treatment corresponding to F2, S1, and 50% filler.*

(b) *The concept of LSMEANS has been explained carefully in lecture and course notes for the special case of a two-factor study. The concept generalizes easily to multi-factor studies. For example, in a three-factor study, the LSMEAN for level i of the first factor is the OLS estimator of $\bar{\mu}_{i\cdot\cdot}$, the average of the cell means for all treatments that involve level i of the first factor. Find LSMEANS for the levels of the factor filler type.*

(c) *We can also compute LSMEANS for estimable marginal means like $\bar{\mu}_{\cdot jk}$, the average of the cell means for all treatments involving level j of the second factor and level k of the third factor. Find the LSMEAN for surface treatment S2 and 25% filler.*

(d) *Provide a standard error for the estimate computed in part (c).*

(e) *In a three-factor study we would say there are no main effects for the first factor if $\bar{\mu}_{i\cdot\cdot} = \bar{\mu}_{i'\cdot\cdot}$ for all levels $i \neq i'$. Conduct a test for filler type main effects. Provide an F-statistic, a p-value, and a conclusion.*

(f) *In a three-factor study in which the third factor has K levels, we would say there are no three-way interactions if, for all $i \neq i'$ and $j \neq j'$,*

$$\mu_{ij1} - \mu_{ij'1} - \mu_{i'j1} + \mu_{i'j'1} = \mu_{ij2} - \mu_{ij'2} - \mu_{i'j2} + \mu_{i'j'2} = \cdots = \mu_{ijK} - \mu_{ij'K} - \mu_{i'jK} + \mu_{i'j'K}.$$

Note that each linear combination above can be viewed as a two-way interaction effect involving the first two factors while holding the level of the third factor fixed. If these interaction effects are *all the same regardless of which level of the third factor is selected, we say there are no three way interactions. Put another equivalent way, there are no three-factor interactions if*

$$\mu_{ijk} - \mu_{ij'k} - \mu_{i'jk} + \mu_{i'j'k} - \mu_{ijk'} + \mu_{ij'k'} + \mu_{i'jk'} - \mu_{i'j'k'} = 0$$

for all $i \neq i'$, $j \neq j'$, and $k \neq k'$. Conduct a test for three-way interactions among the factors filler type, surface treatment, and filler proportion. Provide an F-statistic, a p-value, and a conclusion.

(g) In a three-factor study, we would say there are no two-way interactions between the first and third factors if

$$\bar{\mu}_{i\cdot k} - \bar{\mu}_{i\cdot k'} - \bar{\mu}_{i'\cdot k} + \bar{\mu}_{i'\cdot k'} = 0$$

for all $i \neq i'$ and $k \neq k'$. Conduct a test for two-way interactions between the factors filler type and filler proportion. Provide an F-statistic, a p-value, and a conclusion.

Problem 3 *When $\boldsymbol{X}$ does not have full rank, let's see why $\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^- \boldsymbol{X}^\top$ is invariant to the choice of generalized inverse. Let $\boldsymbol{G}$ and $\boldsymbol{H}$ be two generalized inverses of $\boldsymbol{X}^\top \boldsymbol{X}$. For an arbitrary $\boldsymbol{v} \in \mathbb{R}^n$, let $\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{v}_2$ with $\boldsymbol{v}_1 = \boldsymbol{X}\boldsymbol{b} \in \boldsymbol{C}(\boldsymbol{X})$ for some $\boldsymbol{b}$.*

a) Show that $\boldsymbol{v}^\top \boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{v}^\top \boldsymbol{X}$, so that $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{X}$ for any generalized inverse.

b) Show that $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top \boldsymbol{v} = \boldsymbol{X}\boldsymbol{H}\boldsymbol{X}^\top \boldsymbol{v}$, and thus $\boldsymbol{X}\boldsymbol{G}\boldsymbol{X}^\top$ is invariant to the choice of generalized inverse.

Problem 4 *An experiment was conducted to study the effect of two diets (1 and 2) and two drugs (1 and 2) on blood pressure in rats. A total of 40 rats were randomly assigned to the 4 combinations of diet and drug, with 10 rats per combination. Let y_{ijk} be the decrease in blood pressure from the beginning of the study to the end of the study for diet i, drug j, and rat k ($i = 1, 2$; $j = 1, 2$; $k = 1, \ldots, 10$). Suppose*

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}, \tag{1}$$

where the μ_{ij} terms are unknown parameters and the ϵ_{ijk} terms are independent and identically distributed as $\mathcal{N}(0, \sigma^2)$ for some unknown variance parameter $\sigma^2 > 0$.

A researcher suspects the mean reduction in blood pressure will be the same for all combinations of diet and drug except for the combination of diet 1 with drug 1. This leads to consideration of the null hypothesis

$$H_0 : \mu_{12} = \mu_{21} = \mu_{22}.$$

Assuming model (1) holds, determine the distribution of the F statistic you would use to test this null hypothesis. State the degrees of freedom of the statistic and provide a fully simplified expression for the noncentrality parameter of the statistic in terms of model (1) parameters.