Preliminary Examination: LINEAR MODELS

Answer all questions and show all work.
Q1, Q3, and Q4 are 20 points each, and Q2 is 10 points.

1. Consider the model

$$Y_i = \beta + \epsilon_i, \; i = 1, \ldots, n,$$

where β is a scalar and

$$\epsilon_i = \epsilon_1^* + \cdots + \epsilon_i^*,$$

for $\epsilon_1^*, \ldots, \epsilon_n^*$ being a sequence of uncorrelated random variables with mean zero and unit variance.

To answer the following questions, you may wish to use the fact that the inverse of an $m \times m$ matrix of the form

$$\mathbf{M} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 2 & 3 & \cdots & 2 & 2 \\ 1 & 2 & 3 & \cdots & 3 & 3 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & 2 & 3 & \cdots & m-1 & m-1 \\ 1 & 2 & 3 & \cdots & m-1 & m \end{pmatrix}$$

can be expressed as

$$\mathbf{M}^{-1} = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}.$$

a. Provide a *simplified* expression for the ordinary least squares estimator $\hat{\beta}_{OLS}$ of β.

b. Similarly, provide a *simplified* expression for the generalized least squares estimator $\hat{\beta}_{GLS}$ of β.

c. Calculate the variances Var$(\hat{\beta}_{OLS})$ and Var$(\hat{\beta}_{GLS})$ of the two estimators in parts (a) and (b) and compare them.

d. A friend of yours, observing the structure of the model assumed hear, suggests that it would be natural to instead consider estimating β using the differences $D_i = Y_i - Y_{i-1}$, $i = 1, \ldots, n$, where $Y_0 = 0$. Comment on this idea and compare what you will get in this case using OLS and GLS.

2. Consider the cell means version of the one-way ANOVA model with three groups, two replicates per group:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \; i = 1, 2, 3; \; j = 1, 2;$$

where $\{\epsilon_{ij}\} \overset{iid}{\sim} N(0, \sigma^2)$ and $\mathbf{Y} \equiv (Y_{11}, Y_{12}, Y_{21}, Y_{22}, Y_{31}, Y_{32})' = (1, 3, 2, 6, 1, -1)$. Consider the hypothesis $H_0 : \mu_2 = \frac{\mu_1 + \mu_3}{2}$.

a. Show that under H_0, the fitted response, $\hat{\mathbf{Y}}$, will be in $\mathcal{L}(\mathbf{1}_6, \mathbf{x})$, the vector space spanned by $\mathbf{1}_6$ and $\mathbf{x} = (-1, -1, 0, 0, 1, 1)'$.

b. Conduce the test H_0 using F *test*. You need only compute the F statistic and give its distribution under H_0. *You don't need to compute the* p*-value, critical value or give the conclusion.*

3. In longitudinal data analysis, we usually have repeated and irregularly spaces measurements per subject. We assume that the observed data are realizations of a smooth random function. Let Y_{ij} be the jth observation of the random function $X_i(\cdot)$, made at time T_{ij}, where $i = 1, \ldots, n$, and $j = 1, \ldots, N_i$. We assume that $\{X_i(\cdot)\}$ are independent across n subjects, and further assume that

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t),$$

where ϕ_k is called the kth eigenfunction; corresponding eigenvalues are nonincreasing $\lambda_1 \geq \lambda_2 \geq \cdots$; $\{\xi_{ik}\}$ are uncorrelated random variables with mean 0 and variance $E(\xi_{ik}^2) = \lambda_k$. Let ϵ_{ij} be the additional measurement errors that are assumed to be iid and independent of the random coefficients $\{\xi_{ik}\}$. Then we have the following model:

$$Y_{ij} = X_i(T_{ij}) + \epsilon_{ij} = \mu(T_{ij}) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(T_{ij}) + \epsilon_{ij},$$

where $E(\epsilon_{ij}) = 0$, and $\text{Var}(\epsilon_{ij}) = \sigma^2$.

a. Show that $E(X_i(t)) = \mu(t)$, and $G(s,t) \equiv \text{Cov}(X_i(s), X_i(t)) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t)$. Here, $G(\cdot, \cdot)$ is called the covariance function.

b. The random coefficient ξ_{ik} is called the functional principal component (FPC) score of the kth principal component for the ith subject. Now we assume that the FPC scores $\{\xi_{ik}\}$ and measurement errors $\{\epsilon_{ij}\}$ are *jointly Gaussian*. Define

$$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})',$$

$$\mathbf{X}_i = (X_i(T_{i1}), \ldots, X_i(T_{iN_i}))',$$
$$\boldsymbol{\mu}_i = (\mu(T_{i1}), \ldots, \mu(T_{iN_i}))',$$
$$\boldsymbol{\phi}_{ik} = (\phi_k(T_{i1}), \ldots, \phi_k(T_{iN_i}))'.$$

Define $\boldsymbol{\Sigma}_{\mathbf{Y}_i} = \text{Cov}(\mathbf{Y}_i, \mathbf{Y}_i)$. Give its expression in terms of $G(\cdot,\cdot)$ and σ^2.

c. Define $\tilde{\xi}_{ij} = E(\xi_{ik}|\mathbf{Y}_i)$. *Consider the K leading eigenfunctions only.* Define $\tilde{\boldsymbol{\xi}}_{K,i} = (\tilde{\xi}_{i1}, \ldots, \tilde{\xi}_{iK})'$ and $\boldsymbol{\xi}_{K,i} = (\xi_{i1}, \ldots, \xi_{iK})'$. Show that

$$\tilde{\xi}_{ik} = \lambda_k \boldsymbol{\phi}'_{ik} \boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_i} (\mathbf{Y}_i - \boldsymbol{\mu}_i)$$

and derive the distribution of $\tilde{\boldsymbol{\xi}}_{K,i} - \boldsymbol{\xi}_{K,i}$.

d. Recall the following result from STAT 7024: Let $\mathbf{V}$ be a positive definite matrix. Then for any vector $\mathbf{b}$,

$$sup_{\mathbf{h}\neq\mathbf{0}} \frac{(\mathbf{h}'\mathbf{b})^2}{\mathbf{h}'\mathbf{V}\mathbf{h}} = \mathbf{b}'\mathbf{V}^{-1}\mathbf{b}.$$

Prove that for a fixed non-zero p-vector $\mathbf{x}$ and a constant $c > 0$, $\mathbf{x}'\mathbf{x} \leq c^2$ if and only if $|\mathbf{a}'\mathbf{x}| \leq c\sqrt{\mathbf{a}'\mathbf{a}}$, for all $\mathbf{a} \in \mathbb{R}^p$.

e. Let $\boldsymbol{\Omega}_K \equiv \text{Cov}(\tilde{\boldsymbol{\xi}}_{K,i} - \boldsymbol{\xi}_{K,i}, \tilde{\boldsymbol{\xi}}_{K,i} - \boldsymbol{\xi}_{K,i})$. Let $A \subset \mathbb{R}^K$ be a vector space with dimension $d \leq K$. Prove that for all linear combinations $\mathbf{l}'\boldsymbol{\xi}_{K,i}$ simultaneously, where $\mathbf{l} \in A$,

$$\mathbf{l}'\boldsymbol{\xi}_{K,i} \in \mathbf{l}'\tilde{\boldsymbol{\xi}}_{K,i} \pm \sqrt{\chi^2_{d,1-\alpha} \mathbf{l}'\boldsymbol{\Omega}_K \mathbf{l}},$$

with probability $(1-\alpha)$. *Hint: You may use the result in part (d), no matter whether you derive the proof there.*

4. An industrial engineer is studying the hand insertion of electronic components on printed circuit boards to improve the speed of the assembly operation. He has designed a assembly fixtures ($i = 1, \ldots, a$) and b workplace layouts ($j = 1, \ldots, b$) that seem promising. Operators are required to perform the assembly, and it is decided to randomly select c operators ($k = 1, \ldots, c$) for each fixture-layout combination. However, because the workplaces are in different locations within the plants, it is difficult to use the same c operators for each layout. The treatment combinations in this design are run in random order, and n replicates ($l = 1, \ldots, n$) are obtained. Assume the errors, $\epsilon_{ijkl} \sim N(0, \sigma^2)$, identically and independently distributed.

   a. Write down an appropriate ANOVA model with all possible interactions including model assumptions for this design.

   b. Find the expected mean squares (EMS) for all of main effects and interaction effects.

   c. Describe how to test each of main effects as well as interaction effects.