DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
GRADUATE WRITTEN EXAMINATION
JANUARY 2011

Applied Statistics (M.A. Version)

Instructions to the Student

a. Answer any six questions Each will be graded from 0 to 10.

b. Use a different booklet for each question. Write the problem number and your code number (**NOT YOUR NAME**) on the cover

c. Keep scratch work on separate pages in the same booklet.

d. If you use a "well known" theorem in your solution to any problem, it is your responsibility to make clear which theorem you are using and to justify its use.

e. You may use calculators as needed.

---

1. Consider the linear model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \boldsymbol{\varepsilon}.$$

This can be thought of as a regression model with $Y_i = \beta_0 + \sum_j x_{ij}\beta_j + \varepsilon_i$

(a) Show that all parameters are estimable and write out the least squares equations.

(b) Suppose that one wanted to augment the model by adding a term $\beta_{12}x_1x_2$. How does this term affect estimability of the parameters?

2. A drug is administered to each of n subjects and a response is observed at times $t = 1, 2, \ldots, T$. It is believed that the effect of the drug occurs gradually over time, so the observation on subject i at time t is modeled as

$$Y_{it} = \mu + a_i + \beta(t - \bar{t}) + \varepsilon_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \quad \bar{t} = (T+1)/2.$$

The random subject effects a_i, $i = 1, \ldots, n$, are i.i.d. $N(0, \sigma_a^2)$ and the error terms ε_{it} are i.i.d. $N(0, \sigma_e^2)$.

(a) Suppose the data are reduced to the time averages $\bar{Y}_t$, $t = 1, \ldots, T$, and a least squares regression line $\hat{Y}_t = \alpha + \beta t$ is fitted to the reduced data. What is the joint distribution of the resulting estimates $\hat{\alpha}$ and $\hat{\beta}$? Verify that $\hat{\beta}$ is not a function of the a_i.

(b) Starting from the standard ANOVA table for a two way layout, derive an ANOVA table for this model with sums of squares for subject-to-subject variation, regression and residuals. Show that the mean square for residuals is an unbiased estimator for σ_e^2. What are the degrees of freedom for residuals?

(c) Use the results of (b) to derive a confidence interval for β

3. Let $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, where the parameters $\mu, \alpha_i, \beta_j, \gamma_{ij}$ are unrestricted. The error terms ε_{ijk} are i.i.d. $N(0, \sigma^2)$.

(a) Show that $\alpha_1 - \alpha_2$ is not estimable but that $\gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22}$ is estimable.

(b) If the additive model holds, that is, if $\gamma_{ij} = 0$ for all i, j, show that $\alpha_1 - \alpha_2$ is now estimable and give a confidence interval for this contrast.

(c) State the usual test for additivity and give its distribution under the general alternative.

4. Let $Y_{11}, \ldots, Y_{1m}$ be i.i.d. $N(\mu_1, \sigma^2)$ and let $Y_{21}, \ldots, Y_{2n}$ be i.i.d. $N(\mu_2, k\sigma^2)$, where μ_1, μ_2, σ^2 are unknown parameters and k is a known constant. Find the BLUE of $\mu_1 - \mu_2$ and provide a confidence interval for this quantity

5. For population values $y_1, \ldots, y_N$, consider the ith systematic sample of size n:

$$y_i, y_{i+k}, y_{i+2k}, \ldots, y_{i+(n-1)k}, \quad i = 1, \ldots, k, \quad N = kn.$$

Let y_{ij} denote the jth element, $j = 1, \ldots, n$, of systematic sample i.

(a) Obtain the basic ANOVA table for the above population.

(b) Show that the population variance S^2 can be expressed in terms of the between and within sample sums of squares.

(c) Define

$$S^2_{wsy} = \frac{1}{k(n-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2.$$

Obtain a condition in terms of S^2_{wsy} under which the mean of a systematic sample is more precise than the mean from a simple random sample. Interpret your result.

6. In most cases, positive geophysical data such as duration of snowstorms have skewed distributions. Suppose that for some positive geophysical data $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T$ and for some $\lambda \in (-3, 3)$ the transformation

$$Y_i^{(\lambda)} = g(Y_i, \lambda) = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

gives a linear model with independent errors $\varepsilon_i \sim N(0, \sigma^2)$, where

$$g(y, \lambda) = \begin{cases} (y^\lambda - 1)/\lambda, & \text{if } \lambda \neq 0, \\ \log y & \text{if } \lambda = 0. \end{cases}$$

(a) Obtain the likelihood for the original data $\boldsymbol{Y}$ in terms of $\boldsymbol{Y}^{(\lambda)} = (Y_1^{(\lambda)}, \ldots, Y_n^{(\lambda)})'$.

(b) For a fixed λ, write down the maximized log-likelihood.

(c) Suggest a way to estimate λ.

7. A sample $\mathcal{S}$ of size n is drawn sequentially from a population $\mathcal{U}$ of size N as follows:

(i) The first element drawn is element k with probability p_k, $k = 1, \ldots, N$.

(ii) A simple random sample of size $n - 1$ is drawn from the remaining $N - 1$ elements without replacement.

Here $p_1, \ldots, p_N$ are nonnegative numbers with $\sum_{k=1}^{N} p_k = 1$

(a) What is the probability that $\mathcal{S}$ contains element k?

(b) What is the probability that $\mathcal{S}$ contains both elements j and k, $j \neq k$?

(c) What is the probability that $\mathcal{S} = \{k_1, \ldots, k_n\}$?

8. An experimenter wishes to fit the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

using least squares, subject to the restriction that $\beta_1 = 1$. He asks specifically if he can just fit

$$W = Y - x_1 = \beta_0 + \beta_2 x_2 + \varepsilon$$

using least squares to get what he wants Prove that his proposal will work.