Applied Statistics **Preliminary Examination** August 2012

1. Let $U = \{x \in \mathbb{R}^4 : x_1 + x_2 - x_3 + 2x_4 = 0\}$, and $V = \{x \in \mathbb{R}^4 : 2x_3 + x_4 = 0\}$. Consider the coordinate-free linear model $Y = \eta + E$, $\eta \in \mathbb{L} = (U \cap V)^\perp$, and $E \stackrel{d}{=} N_4(0, \sigma^2 I_4), 0 < \sigma^2 < \infty$. The null hypothesis is given by $H_0 : \eta \in \mathbb{L}_0 = U^\perp$. Define $\mathbb{L}_1 = \mathbb{L}_0^\perp \cap \mathbb{L}$.
   (a) Give an explicit expression for the projections P, P_0, and P_1 on the subspaces $\mathbb{L}, \mathbb{L}_0$, and $\mathbb{L}_1$, respectively, and find the dimensions p, p_0, and p_1 of these subspaces.
   (b) Let $\widehat{\eta}$ and $\widehat{\eta_0}$ be the least squares estimators of η under the full model ($\eta \in \mathbb{L}$), and the null hypothesis model ($\eta \in \mathbb{L}_0$) respectively. What is the joint distribution of $\widehat{\eta_0}$ and $\widehat{\eta} - \widehat{\eta_0}$?
   (c) Compute the test statistic for testing H_0, and specify its distribution under the null hypothesis.

2. Let $Y_i = \alpha + \beta x_i + E_i$, for $i = 1, \dots, n$, where $\alpha, \beta \in \mathbb{R}$, $x = (x_1, \dots, x_n)^* \notin [\![1_n]\!]$ with $1_n = (1, \dots, 1)^* \in \mathbb{R}^n$, and the E_i are independent and identically distributed with common distribution $N(0, \sigma^2)$, for some $0 < \sigma^2 < \infty$.
   (a) Write this model in vector notation, specify the design matrix, and compute the least squares estimators $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\sigma}^2$ of α, β, σ^2, respectively.
   (b) Compute the test statistic for testing $H_0 : \beta = 0$, and give its distribution under the null hypothesis.
   (c) Consider now the canonical model $Y_i = \alpha + \beta(x_i - \bar{x}) + E_i$, $i = 1, \dots, n$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$, and write $\widehat{\beta}_C$ for the least squares estimator in this special case. Show that $\widehat{\beta} = \widehat{\beta}_C$.

3. Consider the model $Y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + E_i$, for $i = 1, \dots, n$, where $\alpha, \beta_1, \beta_2 \in \mathbb{R}$, $x_1 = (x_{11}, \dots, x_{1n})^*, x_2 = (x_{21}, \dots, x_{2n})^* \in \mathbb{R}^n$, and the E_i are independent and identically distributed with $N(0, \sigma^2)$ distribution ($0 < \sigma^2 < \infty$). Suppose $x_1 \neq 0, x_2 \neq 0$.
   (a) Write this model in vector notation and specify the design matrix X.
   (b) Under what condition on X are all linear parameter functions estimable?

   Henceforth let us assume that $x_1^* 1_n = x_2^* 1_n = x_1^* x_2 = 0$, where $1_n = (1, \dots, 1)^* \in \mathbb{R}^n$.

   (c) Compute explicitly the least squares estimator of $\theta = (\alpha, \beta_1, \beta_2)^*$.
   (d) Determine the test statistic for testing $H_0 : \beta_1 + 2\beta_2 = 0$, and give its distribution under the null hypothesis.
   (e) Give a $(1 - \alpha)$-confidence interval for $\beta_1 + 2\beta_2$, for some $0 < \alpha < 1$.

4. Consider the model $Y_{jk} = \mu + \alpha_j + \beta x_j x_k + E_{jk}$, $1 \le j \le m$, $1 \le k \le m$. The parameters μ, α_j, and β are all real. Writing $a = (\alpha_1, \dots, \alpha_m)^*$ and $1_m = (1, \dots, 1)^* \in \mathbb{R}^m$ the parameter vector a satisfies the restriction $a^* 1_m = 0$. The x_j and x_k are coordinates of a vector $x = (x_1, \dots, x_m)^*$ that satisfies $||x|| = 1$ and $x^* 1_m = 0$. The E_{jk} are independent with the same $N(0, \sigma^2)$ distribution for some $0 < \sigma^2 < \infty$. Using tensor products, the model says that Y (the vector of all Y_{jk}) lies in a subspace $\mathbb{L} \subset \mathbb{R}^m \otimes \mathbb{R}^m$.

(a) Give a concise description of the model using tensor products.

(b) The subspace $\mathbb{L}$ corresponding to the model can be written as $\mathbb{L} = \mathbb{L}_1 \oplus \mathbb{L}_2 \oplus \mathbb{L}_3$. Specify the $\mathbb{L}_j$.

(c) Compute the projection of Y onto the subspaces $[\![1_m]\!] \otimes [\![1_m]\!]$ and $[\![x]\!] \otimes [\![x]\!]$.

(d) Are the projections in (c) stochastically independent? Why?

(e) Give a detailed computation of the statistic for testing $H_0 : a = 0$, and specify its distribution under the null hypothesis.

Design of Experiments: Preliminary Examination
Summer 2012

Please Do All the Problems

1. A plant contains a large number of coil winding machines. A production analyst studied a certain characteristic of the wound coils produced by these machines by selecting four machines at random and then choosing 10 coils at random from the day's output of each selected machine. The following summary statistics are available:

   SS(Machine) = 602.50, SSE = 357.40, $\bar{Y}_{..} = 205.05$

   a) Write the linear model for this experiment.

   b) Write the ANOVA table including the E(MS) column and test whether or not the mean coil characteristic is the same for all machines in the plant at the 0.05 level of significance.

   c) Obtain a point estimate of Var(machine) and of the intraclass correlation.

   d) Find a 95% confidence interval for σ^2, the error variance.

   e) Derive the expression for the $100(1-\alpha)\%$ confidence interval for the intraclass correlation.

   f) Find a 95% confidence interval for the intraclass correlation.

2. A production engineer studied the effects of machine model and operator on the output in a bottling plant. Three bottling machines were used, each a different model. Twelve operators were employed. Four operators were assigned to a machine and worked six-hour shifts each. Data on the number of cases produced by each machine and operator per hour were collected for five days. The following summary data were obtained:

Mean(*Cases of Bottles per Hour*)				
	Machine			
Operator	1	2	3	Overall
1	61.80	75.80	76.80	71.47
2	67.80	75.20	69.60	70.87
3	62.60	55.80	74.40	64.27
4	52.60	77.00	73.40	67.67
Overall	61.20	70.95	73.55	68.57

   SS(Machine) = 1695.633 SSE = 1132.800 SSTotal = 5100.733

   a) Write the appropriate linear model for the data.

   b) Write the ANOVA table including the $E(MS)$ column.

   c) Test the relevant hypotheses at the 0.05 level of significance each.

   d) Calculate the main effects of machines and operators.

e) Conduct pairwise comparisons of the means for machines using Tukey's simultaneous confidence intervals.

f) Conduct pairwise comparisons of the means of the operators assigned to machine 1 using Bonferroni's procedure.

g) Operator 4 assigned to machine 1 has relatively little experience compared to the other three operators. Estimate the contrast:

$$\Gamma = \frac{\mu_{11} + \mu_{12} + \mu_{13}}{3} - \mu_{14}$$

using a 95% confidence interval.

3. In a two-factor random effects factorial design, let Y_{ijk} denote the k-th replicate from the i-th level of factor A and j-th level of factor B. Let σ_τ^2 denote the variance component of factor A, σ_β^2 the variance component of factor B, $\sigma_{\tau\beta}^2$ the variance component of AB, and σ^2 the error variance.

a) Show that:

$\text{Var}(Y_{ijk}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2$

$\text{Cov}(Y_{ijk}, Y_{ij'k'}) = \sigma_\tau^2, \quad j \neq j'$

$\text{Cov}(Y_{ijk}, Y_{i'jk'}) = \sigma_\beta^2, \quad i \neq i'$

$\text{Cov}(Y_{ijk}, Y_{ijk'}) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2, \quad k \neq k'$

$\text{Cov}(Y_{ijk}, Y_{i'j'k'}) = 0, \quad i \neq i',\ j \neq j'$

b) Find the expression for the $100(1-\alpha)\%$ confidence interval for the ratio:

$$\frac{\text{E(MS(AB))}}{\text{E(MSE)}}$$