Show your calculations and provide suitable motivation for full credit.

The asterisk indicates the transpose of a vector or matrix, $\oplus$ is the symbol for the direct sum of two subspaces of a vector space, and $\otimes$ is the symbol used for tensor products. For any $m \in N$ write $1_m = (1, \dots, 1)^* \in R^m$, $P_m = \frac{1}{m} 1_m 1_m^*$ for the orthogonal projection onto $[1_m]$, and $P_m^{\perp}$ for the orthogonal projection onto $[1_m]^{\perp} \cap \mathbb{R}^m$.

1. Let U, V denote the orthogonal complement of the subspace $\{u \in \mathbb{R}^3 : u_1 + 2u_2 - u_3 = 0\}$ respectively $\{v \in \mathbb{R}^3 : -3v_1 + v_2 - v_3 = 0\}$ of $\mathbb{R}^3$. Consider the coordinate-free linear model $Y = \eta + E$, $\eta \in \mathbb{L}$, $E \stackrel{d}{=} N_3(0, \sigma^2 I_3)$ for unknown $\sigma^2 \in (0, \infty)$, where $\mathbb{L} = U \oplus V$.
   (a) Compute the projection of Y onto $\mathbb{L}$ and the unbiased version of the ML estimator for σ^2.
   (b) Suppose $\mathbb{L}_0 = U$. Compute the LR statistic for testing $H_0 : \eta \in \mathbb{L}_0$ and determine its distribution under the null hypothesis.

2. Let be given the model $Y_{ij} = \mu + \alpha_i + E_{ij}$, $i = 1, 2$, and $j = 1, 2$, where μ and the α_i are real parameters, and the error variables E_{ij} are i.i.d. $N(0, \sigma^2)$ for unknown $\sigma^2 \in (0, \infty)$.
   (a) Express $Y^* = (Y_{11}, Y_{12}, Y_{21}, Y_{22})$ in terms of the parameter vector $\theta = (\mu, \alpha_1, \alpha_2)^*$ by means of a design matrix X.
   (b) Is this design matrix of full rank?
   (c) If not, find a matrix $\prod$ such that $(X^*X + \prod)^{-1}$ is a generalized inverse of X^*X.
   (d) Is $\alpha_1 - \alpha_2$ estimable? Why?
   (e) Find a 95% confidence interval for $\alpha_1 - \alpha_2$. (You may use the ML estimator $\widehat{\theta} = (\widehat{\mu}, \widehat{\alpha_1}, \widehat{\alpha_2})^*$ and the above generalized inverse $(X^*X)^-$ without explicit calculation.)

3. Let be given the regression model $Y = X\theta + E$, where Y is an $n \times 1$ vector of observations, X an $n \times m$ design matrix of full rank $m < n$, $\theta \in \mathbb{R}^m$ an unknown parameter, and $E \stackrel{d}{=} N_n(0, \sigma^2 I_n)$ the $n \times 1$ error vector with $\sigma^2 \in (0, \infty)$ unknown. Suppose X is of the form $X = (1_n | X')$, where X' is an $n \times (m-1)$ matrix.
   (a) Determine the ML estimator $\widehat{\theta}$ of θ and its exact distribution.
   (b) Define the vector of residuals $\widehat{E} = Y - X\widehat{\theta}$. Compute $\mathbb{E}\widehat{E}$ and prove that
   $$\sum_{i=1}^{n} \widehat{E_i} = 0.$$
   (c) Compute the covariance matrix of $\widehat{E}$.

4. Consider the model $Y_{jk} = \mu + \alpha_j + \gamma \cdot v_k + E_{jk}$, $j = 1, \dots, J$, and $k = 1, \dots, K$, where μ, γ, and the α_j are real parameters, $v^* = (v_1, \dots, v_K)$ a given vector in $\mathbb{R}^K$, and the E_{jk} are independent error variables with the same $N(0, \sigma^2)$ distribution, $\sigma^2 \in (0, \infty)$ unknown. We will write $a^* = (\alpha_1, \dots, \alpha_J)$, and assume that $a^* 1_J = v^* 1_K = 0$. Using tensor products the model says that Y (the vector of all Y_{jk}) lies in a subspace $\mathbb{L}$ of $\mathbb{R}^J \otimes \mathbb{R}^K$.

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

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

(c) Suppose we want to test the null hypothesis $H_0 : \gamma = 0$. Compute the numerator of the LR test statistic for this problem.

Design of Experiments: Prelim Problems
May 2014

- **Do All the Problems.**
- **For each test, state the null and alternative hypotheses in terms of the model parameters.**
- **All parts have equal weights.**

1. The yield of a chemical process is being studied. The two factors of interest are temperature and pressure. Three levels of each factor are selected; however, only 9 runs can be made in one day. The experimenter runs a complete replicate of the design on each day. The summary data for yield are shown in the following table.

Factor	Level	Mean
Day	1	87.944
	2	89.644
Temperature	Low	85.783
	Medium	89.067
	High	91.533
Pressure	250	88.517
	260	88.300
	270	89.567
Overall		88.794

In addition, we have

$$\text{SSE} = 4.250 \qquad \text{SST} = 127.069$$

(a) Write a linear model equation appropriate for analyzing the responses from this experiment. Clearly state the assumptions and conditions for this model.

(b) Calculate the sum of squares for Day, Temperature, and Pressure.

(c) Write the ANOVA table appropriate for your model.

(d) Test all relevant hypotheses. Make sure to write all the null and alternative hypotheses in terms of the parameters defined in part (a). Use $\alpha = 0.05$.

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 chosen at random. 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	
1	61.80	75.80	76.80	
2	67.80	75.20	69.60	
3	62.60	55.80	74.40	
4	52.60	77.00	73.40	
Overall	61.20	70.95	73.55	68.57

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

a) Write a linear model equation appropriate for analyzing the responses from this experiment. Clearly state the assumptions and conditions for this model.

b) Write the ANOVA table appropriate for your model and include the $E(MS)$ column.

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

d) Let $\mu_i,\ i = 1, 2, 3$ denote the mean number of cases produced by the three machines. Test $H_0 : \mu_1 = 0.5(\mu_2 + \mu_3)$ vs. $H_1 : \mu_1 < 0.5(\mu_2 + \mu_3)$ at the 0.05 level of significance. Clearly state your conclusion.

e) Estimate the intra-operator correlation coefficient.

f) Find a 95% confidence interval for the ratio σ^2_β/σ^2, where σ^2_β and σ^2 are the operator and error variances, respectively.