

Applied Statistics Comprehensive Examination
Regression & Linear Models

1. (30 Points) 38 wines were graded on overall quality, aroma, body, and flavor. Researchers wanted to determine if the quality rating could be predicted using the aroma, body, and flavor scores. The full model with all three explanatory variables was fit to the data where X_{i1} is the aroma score, X_{i2} is the body score, and X_{i3} is the flavor score. Use the following information to answer the questions below:

$$(X'X)^{-1} = \begin{bmatrix} 0.97 & -0.04 & -0.14 & -0.02 \\ -0.04 & 0.05 & -0.01 & -0.04 \\ -0.14 & -0.01 & 0.07 & -0.03 \\ -0.02 & -0.04 & -0.03 & 0.07 \end{bmatrix} \quad X'Y = \begin{bmatrix} 472.6 \\ 2348.8 \\ 2248.0 \\ 2315.0 \end{bmatrix}$$

The mean squared error for this model is 1.55.

- (a) Write the mathematical model and state the assumptions. Be sure to define each term in the model.
 - (b) Obtain the parameter estimates.
 - (c) Test the following hypothesis at the 0.05 level of significance: $H_0: \beta_2 = \beta_3$.
 - (d) Discuss how hierarchical models can be used to test a hypothesis as in part c.
2. (20 Points) Consider the model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$.
- (a) Discuss the difference between interpolation and extrapolation for this model.
 - (b) Now consider the model $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$. Explain the difficulties in determining if a prediction is an example of interpolation or extrapolation for such a model.
 - (c) How might one resolve the conflict found in part b?

3. (30 points) An experiment was run to compare the mean lifetimes for four brands of flashlight batteries. The lifetimes (in hours) and an ANOVA table are given below.

Brand	Values				Average
1	61	54	54	59	57.0
2	92	79	83	90	86.0
3	45	49	38	40	43.0
4	47	58	49	46	50.0

Source	df	SS	MS	F
Brand	3	4280.0	1426.67	54.87
Error	12	312.0	26.00	

- (a) Using an effects model, write out the mathematical model for this design. Be sure to explain all terms and include all model assumptions.
- (b) State the normal equations for these data, assuming the model from (a).
- (c) Using a level 0.10 test, determine whether there is a difference between the average of the mean lifetimes for Brands 1 and 2 and the average of the mean lifetimes for Brands 3 and 4.
4. (20 points) An experiment was run to study the effects of two factors A and B on a response variable. Factor A has two levels, while factor B has three levels. The data are given below. Assume that the data will be analyzed using a model $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$, where α_i and β_j are fixed main effects, γ_{ij} is an interaction effect, and ϵ_{ijk} is an error term.

Level of A	Level of B		
	1	2	3
1	7, 7	8, 12	21, 17
2	13, 17	18, 18	21, 21

- (a) Write down two orthogonal contrasts in the interaction effects.
- (b) Determine if the following two contrasts are estimable and justify your answers.
- $\beta_1 - \beta_2 + \gamma_{11} - \gamma_{12}$
 - $\alpha_1 - \alpha_2 + \beta_1 - \beta_2$