STAT 510 **Exam 2** **Spring 2020**

Instructions: This exam is open notes, open internet, and open anything except that you are not allowed to give help to or to receive help from another person. Write your name and answers on your own paper, scan your answers to a single PDF, and upload the PDF file to the STAT 510 Canvas Assignments page before 9:15 P.M.

1. An experiment was conducted to compare two ice cream recipes (labeled 1 and 2). A total of 14 people each prepared one batch of ice cream according to one of the two recipes. The two recipes were randomly assigned to the 14 people using a completely randomized design with 7 people per recipe. Four bowls of ice cream were taken from each batch. A total of 56 judges each ate one bowl of ice cream and provided a numerical score from 0 to 10 to indicate how good the ice cream tasted, with higher scores indicating better tasting ice cream. Let y_{ijk} be the score assigned to the kth bowl from the jth batch made with recipe i ($i = 1, 2$; $j = 1, \ldots, 7$; $k = 1, \ldots, 4$). Let

$$\boldsymbol{y} = (y_{111}, y_{112}, y_{113}, y_{114}, y_{121}, y_{122}, y_{123}, y_{124}, \ldots, y_{261}, y_{262}, y_{263}, y_{264}, y_{271}, y_{272}, y_{273}, y_{274})'.$$

   (a) A linear mixed-effects model of the form $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$ may be reasonable for this dataset (assuming that the usual normality assumptions apply to the scores). Using Kronecker product notation, give expressions for $\boldsymbol{X}$ and $\boldsymbol{Z}$ for the linear mixed-effects model you would propose for this dataset.

   (b) Complete the ANOVA table for this dataset with columns labeled "Source" and "Degrees of Freedom".

   (c) Give the formula for the sum of squares in the second line of your ANOVA table as a function of the y_{ijk} terms.

   (d) State the degrees of freedom for the t statistic used to test whether the mean score for recipe 1 is equal to the mean score for recipe 2.

2. Another experiment was conducted to compare two ice cream recipes (labeled 1 and 2). A total of 7 people each prepared two batches of ice cream. For each person, one of the batches was prepared using recipe 1, and the other was prepared using recipe 2. The preparation order was randomized separately for each person with a coin flip to determine which recipe was used for the first batch made by that person.

   Four bowls of ice cream were taken from each batch. Four different toppings (labeled 1, 2, 3, and 4) were randomly assigned to the four bowls from each batch. A total of 56 judges each ate one bowl of ice cream with its assigned topping and provided a numerical score from 0 to 10 to indicate how good the ice cream with its topping tasted. Higher scores indicate better taste.

   Let y_{prt} be the score assigned to the bowl associated with person p, recipe r, and topping t ($p = 1, \ldots, 7$; $r = 1, 2$; $t = 1, \ldots, 4$).

   (a) Assuming the usual normality assumptions hold for the scores, write down the model you would propose for this dataset. Rather than using vector and matrix form, write your answer like

$$y_{prt} = \ldots, \text{ where } \ldots$$

   (b) Provide an expression in terms of your model parameters for $\mathrm{Var}(\bar{y}_{\cdot r \cdot})$.

   (c) Complete the ANOVA table for this dataset with columns labeled "Source" and "Degrees of Freedom".

3. Consider an experiment with two factors A and B. Suppose each factor has two levels. Let n_{ij} be the number of experimental units for level i of factor A and level j of factor B ($i = 1, 2$; $j = 1, 2$). Let y_{ijk} be the value of the response variable for the kth experimental unit treated with level i of factor A and level j of factor B ($i = 1, 2$; $j = 1, 2$; $k = 1, \ldots, n_{ij}$). Suppose a completely randomized design was used to assign treatments to experimental units. Furthermore, suppose

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk},$$

where μ_{11}, μ_{12}, μ_{21}, and μ_{22} are unknown parameters, and the ε_{ijk} terms are independent and identically distributed as $N(0, \sigma^2)$ for some unknown variance component $\sigma^2 > 0$. Suppose estimates of the μ_{ij} parameters along with the number of experimental units per treatment are as follows:

$$\hat{\mu}_{11} = 4.50 \quad \hat{\mu}_{12} = 6.90 \qquad\qquad n_{11} = 4 \quad n_{12} = 2$$

$$\hat{\mu}_{21} = 9.00 \quad \hat{\mu}_{22} = 8.55 \qquad\qquad n_{21} = 2 \quad n_{22} = 4$$

An ANOVA table with sequential (Type I) sums of squares is as follows:

Source	Type I Sum of Squares
A	34.680
B	2.535
$A * B$	5.415
Error	1.250
C. Total	43.880

Note the table above was produced by the R command

```
anova(lm(y ~ A + B + A:B)).
```

Give the sequential (Type I) sums of squares column of the ANOVA table that would be produced by the R command

```
anova(lm(y ~ B + A + A:B)).
```

4. Suppose the same experiment was conducted independently at three research labs around the world. At each lab, two treatments were assigned to a total of 10 mice using a completely randomized design with 5 mice per treatment group. Let y_{ijk} be the response variable value for the kth mouse that received treatment j in lab i ($i = 1, 2, 3$; $j = 1, 2$; $k = 1, \ldots, 5$). Suppose

$$y_{ijk} = \mu_i + \tau_j + \varepsilon_{ijk},$$

where μ_1, μ_2, μ_3, τ_1, and τ_2 are unknown parameters, all ε_{ijk} terms are independent, and $\varepsilon_{ijk} \sim N(0, \sigma_i^2)$, where σ_1^2, σ_2^2, and σ_3^2 are unknown positive variance components. Consider the following results obtained from the data produced by the three labs.

Lab	$\widehat{\tau_1 - \tau_2}$	MSE
1	5.6	9.1
2	-3.2	15.4
3	1.3	6.2

(a) Provide an estimate of $\tau_1 - \tau_2$ that combines the results of all three experiments.

(b) Determine a standard error for your estimate in part (a).