Experimental Design: Comprehensive Exam (2020)

Name: Student id#: Department:

Please show your work in your answer sheet.

1. (20 points) An experiment was performed to compare the effectiveness of two diets (Diet A and Diet B) as a strategy for losing weight. Sixteen subjects were recruited for the study. Eight subjects were assigned to Diet A and the other eight were assigned for Diet B in a completely randomized fashion. The weight loss in pounds is reported by the subjects after three weeks shown below.

	Weight loss (pounds)	mean	variance
Diet A	6, 6, 5, 2, 2, 3, 2, 0	$\bar{y}_1 = 3.25$	$S_1^2 = 33.5/7 = 4.79$
Diet B	4, 2, 2, 2, 1, -1, -2, -2	$\bar{y}_2 = 0.75$	$S_2^2 = 33.5/7 = 4.79$

Place your answers by hand.

(a) Write out the ANOVA table for the analysis of this completely randomized experiment, and fill in the correct numbers of SS, df, MS and F_0. Does there appear to be an effect of diet?

(b) Suppose it is now discovered that the groups were highly unbalanced with respect to sex. Group Diet A had six males (6, 6, 5, 2, 2, 3) and two females (2, 0), whereas group Diet B had two males (4, 2) and six females (2, 2, 1, -1, -2, -2). Ignoring treatment, the results for males and females are shown below:

	Weight loss (pounds)	mean	variance
Male	6, 6, 5, 2, 2, 3, 4, 2	$\bar{y}_1 = 3.75$	$S_1^2 = 21.5/7 = 3.07$
Female	2, 0, 2, 2, 1, -1, -2, -2	$\bar{y}_2 = 0.25$	$S_2^2 = 21.5/7 = 3.07$

Write out the ANOVA table to test for an effect of sex, ignoring the Diets. Does there appear to be a sex effect?

(c) Two students were asked to investigate the combined effects of treatment and sex using two-way ANOVA. The first student reported that the Diet effect is significant:

	df	SS	MS	F0	p-value
Diet	1	25.000	25.000	7.5	0.01798
Sex	1	27.000	27.000	8.1	0.01473
Diet*Sex	1	0.000	0.000	0.0	1.00000
Residuals	12	40.000	3.333		

The second student reported that the Diet effect is not significant:

	df	SS	MS	F0	p-value
Sex	1	49.000	49.000	14.7	0.002378
Diet	1	3.000	3.000	0.9	1.361497
Diet*Sex	1	0.000	0.000	0.0	1.000000
Residuals	12	40.000	3.333		

Explain why the two answers are different. Which answer is better?

2. (20 points) An experiment is conducted to compare 4 navigation techniques (Factor A, Fixed), 2 input methods (Factor B, Fixed), in 36 subjects (Factor C, Random). Each subject is measured in each combination of levels of A and B once. Consider the following **unrestricted model**:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk},$$

where $\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0$, $\gamma_k \sim^{iid} N(0, \sigma_\gamma^2)$, $(\alpha\gamma)_{ik} \sim^{iid} N(0, \sigma_{\alpha\gamma}^2)$, $(\beta\gamma)_{jk} \sim^{iid} N(0, \sigma_{\beta\gamma}^2)$, $\epsilon_{ijk} \sim^{iid} N(0, \sigma^2)$, and all random effects and errors are pairwise independent.

   (a) Using the EMS rules, find the following $E(MS_A)$, $E(MS_B)$, $E(MS_C)$, $E(MS_{AB})$, $E(MS_{AC})$, $E(MS_{BC})$, and $E(MS_E)$.

   (b) Complete the following ANOVA table, testing all main effects and 2-factor interactions.

Source	df	SS	MS	F_0	df_{num}	df_{den}	$F_{df_{num},df_{den}}$
A		66996					
B		30636					
C		84008					
AB		18710					
AC		148797					
BC		68605					
Error		97282					
Total		515034					

3. (20 points) Consider the following SAS output from analysis of a balanced incomplete block design (BIBD). The statistical model is

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}.$$

We assume that all factors are fixed. τ_i and β_j are respectively treatment and block effects.

(a) Write out the remaining conditions for the above model.

(b) What are the hypotheses of interest? Should the hypothesis be rejected? Why, or why not?

(c) If the grand mean is 72.50, compute $\hat{\tau}_1, \cdots, \hat{\tau}_4$ (treatment effect).

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	6	77.7500000	12.95833333	19.94	0.0024
Error	5	3.2500000	0.65000000		
Total	11	81.0000000			

Source	DF	Type III SS	Mean Square	F Value	Pr>F
Block	3	66.08333333	22.02777778	33.89	.0010
Trt	3	22.75000000	7.58333333	11.67	.0107

Trt	y LSMEAN	Standard Error	LSMEAN Number
1	71.37500000	0.4868051	1
2	71.62500000	0.4868051	2
3	72.00000000	0.4868051	3
4	75.00000000	0.4868051	4

4. (14 points) Among a population of lakes, the mean adult fish lengths are normally distributed, with approximately 95% of the lake means lying between 50.2 and 69.8 centimeters. Within lakes, approximately 95% of the fish have lengths within 11.76 centimeters of the lake mean. Consider a one-way random effects model, where a sample of g lakes is selected and n fish are sampled from each lake.

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad \tau_i \sim^{iid} N(0, \sigma_\tau^2), \quad \epsilon_{ij} \sim^{iid} N(0, \sigma^2),$$

for $i = 1, \ldots, g$; $j = 1, \ldots, n$, τ_i's and ϵ_{ij} are independent.

(a) Obtain μ, σ_τ^2, and σ^2.

(b) Give the expectation and variance of $\bar{Y}_{\cdot\cdot} = \frac{1}{gn} \sum_{i=1}^{g} \sum_{j=1}^{n} Y_{ij}$.