Statistics 101B - Spring 2020
Homework 3 (Due Saturday April 25 at 5:00pm)

** Please upload your homework on CCLE. Please write your full name, student ID number, and section number on your homework.
Must be PDF.
NO late work will be accepted!**
You do not need to write the questions in your homework. Please show your work!

1. Problem 4.8
2. Problem 4.24
3. Problem 4.37
4. Problem 4.44

5. Consider the following two block designs:

Design 1	Design 2
$\{1, 2, 3, 4\}$	$\{1, 2, 3, 4\}$
$\{2, 3, 4, 5\}$	$\{5, 1, 2, 3\}$
$\{3, 4, 5, 1\}$	$\{2, 3, 4, 5\}$
$\{4, 5, 1, 2\}$	$\{3, 4, 5, 1\}$
$\{5, 1, 3, 4\}$	$\{1, 2, 4, 5\}$

(a) Obtain the incidence matrices for both designs.

(b) Is any of the two a BIBD? Show your work.

Chapter 4 Problems

4.1 Suppose that a single-factor experiment with four levels of the factor has been conducted. There are six replicates and the experiment has been conducted in blocks. The error sum of squares is 500 and the block sum of squares is 250. If the experiment had been conducted as a completely randomized design the estimate of the error variance σ^2 would be.

(a) 25.0 **(b)** 25.5 **(c)** 35.0

(d) 37.5 **(e)** None of the above

SS **4.2** Suppose that a single-factor experiment with five levels of the factor has been conducted. There are three replicates and the experiment has been conducted as a complete randomized design. If the experiment had been conducted in blocks, the pure error degrees of freedom would be reduced by

(a) 3 **(b)** 5 **(c)** 2

(d) 4 **(e)** None of the above

4.3 Blocking is a technique that can be used to control the variability transmitted by uncontrolled nuisance factors in an experiment.

(a) True

(b) False

4.4 The number of blocks in the RCBD must always equal the number of treatments or factor levels.

(a) True

(b) False

4.5 The key concept of the phrase "Block if you can, randomize if you can't" is that:

(a) It is usually better to not randomize within blocks.

(b) Blocking violates the assumption of constant variance.

(c) Create blocks by using each level of the nuisance factor as a block and randomize within blocks.

(d) Randomizing the runs is preferable to randomizing blocks.

4.6 Consider the single-factor completely randomized experiment shown in Problem 3.7. Suppose that this experiment had been conducted in a randomized complete block design and that the sum of squares for blocks was 80.00. Modify the ANOVA for this experiment to show the correct analysis for the randomized complete block experiment.

4.7 A chemist wishes to test the effect of four chemical agents on the strength of a particular type of cloth. Because there might be variability from one bolt to another, the chemist decides to use a randomized block design, with the bolts of cloth considered as blocks. She selects five bolts and applies all four chemicals in random order to each bolt. The resulting tensile strengths follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions. SS

	Bolt				
Chemical	**1**	**2**	**3**	**4**	**5**
1	73	68	74	71	67
2	73	67	75	72	70
3	75	68	78	73	68
4	73	71	75	75	69

4.8 Three different washing solutions are being compared to study their effectiveness in retarding bacteria growth in 5-gallon milk containers. The analysis is done in a laboratory, and only three trials can be run on any day. Because days could represent a potential source of variability, the experimenter decides to use a randomized block design. Observations are taken for four days, and the data are shown here. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Days			
Solution	**1**	**2**	**3**	**4**
1	13	22	18	39
2	16	24	17	44
3	5	4	1	22

4.21 Suppose that the observation for chemical type 2 and bolt 3 is missing in Problem 4.7. Analyze the problem by estimating the missing value. Perform the exact analysis and compare the results.

4.22 Consider the hardness testing experiment in Problem 4.11. Suppose that the observation for tip 2 in coupon 3 is missing. Analyze the problem by estimating the missing value.

4.23 An industrial engineer is conducting an experiment on eye focus time. He is interested in the effect of the distance of the object from the eye on the focus time. Four different distances are of interest. He has five subjects available for the experiment. Because there may be differences among individuals, he decides to conduct the experiment in a randomized block design. The data obtained follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions.

	Subject				
Distance (ft)	**1**	**2**	**3**	**4**	**5**
4	10	6	6	6	6
6	7	6	6	1	6
8	5	3	3	2	5
10	6	4	4	2	3

4.24 The effect of five different ingredients (A, B, C, D, E) on the reaction time of a chemical process is being studied. Each batch of new material is only large enough to permit five runs to be made. Furthermore, each run requires approximately $1\frac{1}{2}$ hours, so only five runs can be made in one day. The experimenter decides to run the experiment as a Latin square so that day and batch effects may be systematically controlled. She obtains the data that follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Day				
Batch	**1**	**2**	**3**	**4**	**5**
1	$A = 8$	$B = 7$	$D = 1$	$C = 7$	$E = 3$
2	$C = 11$	$E = 2$	$A = 7$	$D = 3$	$B = 8$
3	$B = 4$	$A = 9$	$C = 10$	$E = 1$	$D = 5$
4	$D = 6$	$C = 8$	$E = 6$	$B = 6$	$A = 10$
5	$E = 4$	$D = 2$	$B = 3$	$A = 8$	$C = 8$

4.25 An industrial engineer is investigating the effect of four assembly methods (A, B, C, D) on the assembly time for a color television component. Four operators are selected for the study. Furthermore, the engineer knows that each assembly method produces such fatigue that the time required for the last assembly may be greater than the time required for the first, regardless of the method. That is, a trend develops in the required assembly time. To account for this source of variability, the engineer uses the Latin square design that follows. Analyze the data from this experiment ($\alpha = 0.05$) and draw appropriate conclusions.

	Operator			
Order of Assembly	**1**	**2**	**3**	**4**
1	$C = 10$	$D = 14$	$A = 7$	$B = 8$
2	$B = 7$	$C = 18$	$D = 11$	$A = 8$
3	$A = 5$	$B = 10$	$C = 11$	$D = 9$
4	$D = 10$	$A = 10$	$B = 12$	$C = 14$

4.26 Consider the randomized complete block design in Problem 4.8. Assume that the days are random. Estimate the block variance component.

4.27 Consider the randomized complete block design in Problem 4.11. Assume that the coupons are random. Estimate the block variance component.

4.28 Consider the randomized complete block design in Problem 4.13. Assume that the trucks are random. Estimate the block variance component.

4.29 Consider the randomized complete block design in Problem 4.14. Assume that the software projects that were used as blocks are random. Estimate the block variance component.

4.30 Consider the gene expression experiment in Problem 4.15. Assume that the subjects used in this experiment are random. Estimate the block variance component.

4.31 Suppose that in Problem 4.24 the observation from batch 3 on day 4 is missing. Estimate the missing value and perform the analysis using the value.

4.32 Consider a $p \times p$ Latin square with rows (α_i), columns (β_k), and treatments (τ_j) fixed. Obtain least squares estimates of the model parameters α_i, β_k, and τ_j.

4.33 Derive the missing value formula (Equation 4.28) for the Latin square design.

4.34 ***Designs involving several Latin squares.*** [See Cochran and Cox (1957), John (1971).] The $p \times p$ Latin square contains only p observations for each treatment. To obtain more replications, the experimenter may use several squares, say n. It is immaterial whether the squares used are the same or different. The appropriate model is

$$y_{ijkh} = \begin{matrix} \mu + \rho_h + \alpha_{i(h)} \\ +\tau_j + \beta_{k(h)} \\ +(\tau\rho)_{jh} + \varepsilon_{ijkh} \end{matrix} \quad \begin{cases} i = 1, 2, \ldots, p \\ j = 1, 2, \ldots, p \\ k = 1, 2, \ldots, p \\ h = 1, 2, \ldots, n \end{cases}$$

where y_{ijkh} is the observation on treatment j in row i and column k of the hth square. Note that $\alpha_{i(h)}$ and $\beta_{k(h)}$ are the row and column effects in the hth square, ρ_h is the effect of the hth square, and $(\tau\rho)_{jh}$ is the interaction between treatments and squares.

(a) Set up the normal equations for this model and solve for estimates of the model parameters. Assume that appropriate side conditions on the parameters are $\sum_h \hat{\rho}_h = 0$, $\sum_i \hat{\alpha}_{i(h)} = 0$, and $\sum_k \hat{\beta}_{k(h)} = 0$ for each h, $\sum_j \hat{\tau}_j = 0$, $\sum_j (\hat{\tau\rho})_{jh} = 0$ for each h, and $\sum_h (\hat{\tau\rho})_{jh} = 0$ for each j.

(b) Write down the analysis of variance table for this design.

4.35 Discuss how you would determine the sample size for use with the Latin square design.

4.36 Suppose that in Problem 4.24 the data taken on day 5 were incorrectly analyzed and had to be discarded. Develop an appropriate analysis for the remaining data.

4.37 The yield of a chemical process was measured using five batches of raw material, five acid concentrations, five standing times (A, B, C, D, E), and five catalyst concentrations $(\alpha, \beta, \gamma, \delta, \varepsilon)$. The Graeco-Latin square that follows was used. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Acid Concentration				
Batch	**1**	**2**	**3**	**4**	**5**
1	$A\alpha = 26$	$B\beta = 16$	$C\gamma = 19$	$D\delta = 16$	$E\varepsilon = 13$
2	$B\gamma = 18$	$C\delta = 21$	$D\varepsilon = 18$	$E\alpha = 11$	$A\beta = 21$
3	$C\varepsilon = 20$	$D\alpha = 12$	$E\beta = 16$	$A\gamma = 25$	$B\delta = 13$
4	$D\beta = 15$	$E\gamma = 15$	$A\delta = 22$	$B\varepsilon = 14$	$C\alpha = 17$
5	$E\delta = 10$	$A\varepsilon = 24$	$B\alpha = 17$	$C\beta = 17$	$D\gamma = 14$

4.38 Suppose that in Problem 4.25 the engineer suspects that the workplaces used by the four operators may represent an additional source of variation. A fourth factor, workplace $(\alpha, \beta, \gamma, \delta)$ may be introduced and another experiment conducted, yielding the Graeco-Latin square that follows. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Operator			
Order of Assembly	**1**	**2**	**3**	**4**
1	$C\beta = 11$	$B\gamma = 10$	$D\delta = 14$	$A\alpha = 8$
2	$B\alpha = 8$	$C\delta = 12$	$A\gamma = 10$	$D\beta = 12$
3	$A\delta = 9$	$D\alpha = 11$	$B\beta = 7$	$C\gamma = 15$
4	$D\gamma = 9$	$A\beta = 8$	$C\alpha = 18$	$B\delta = 6$

4.39 Construct a 5×5 hypersquare for studying the effects of five factors. Exhibit the analysis of variance table for this design.

4.40 Consider the data in Problems 4.25 and 4.38. Suppressing the Greek letters in problem 4.38, analyze the data using the method developed in Problem 4.34.

4.41 Consider the randomized block design with one missing value in Problem 4.22. Analyze this data by using the exact analysis of the missing value problem discussed in Section 4.1.4. Compare your results to the approximate analysis of these data given from Problem 4.22.

4.42 An engineer is studying the mileage performance characteristics of five types of gasoline additives. In the road test, he wishes to use cars as blocks; however, because of a time constraint, he must use an incomplete block design. He runs the balanced design with the five blocks that follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions. SS

	Car				
Additive	**1**	**2**	**3**	**4**	**5**
1		17	14	13	12
2	14	14		13	10
3	12		13	12	9
4	13	11	11	12	
5	11	12	10		8

4.43 Construct a set of orthogonal contrasts for the data in Problem 4.42. Compute the sum of squares for each contrast. SS

4.44 Seven different hardwood concentrations are being studied to determine their effect on the strength of the paper produced. However, the pilot plant can only produce three runs each day. As days may differ, the analyst uses the BIBD that SS follows. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

Hardwood Concentration (%)	Days 1	2	3	4
2	114			
4	126	120		
6		137	117	
8	141		129	149
10		145		150
12			120	
14				136

Hardwood Concentration (%)	Days 5	6	7
2	120		117
4		119	
6			134
8			
10	143		
12	118	123	
14		130	127

SS **4.45** Analyze the data in Example 4.4 using the general regression significance test.

SS **4.46** Prove that $k\sum_{i=1}^{a} Q_i^2/(\lambda a)$ is the adjusted sum of squares for treatments in a BIBD.

SS **4.47** An experimenter wishes to compare four treatments in blocks of two runs. Find a BIBD for this experiment with six blocks.

4.48 An experimenter wishes to compare eight treatments in blocks of four runs. Find a BIBD with 14 blocks and $\lambda = 3$.

4.49 Perform the interblock analysis for the design in Problem 4.42.

4.50 Perform the interblock analysis for the design in Problem 4.44.

SS **4.51** Verify that a BIBD with the parameters $a = 8$, $r = 8$, $k = 4$, and $b = 16$ does not exist.

SS **4.52** Show that the variance of the intrablock estimators $\{\hat{\tau}_i\}$ is $k(a-1)\sigma^2/(\lambda a^2)$.

4.53 Suppose that a single-factor experiment with five levels of the factor has been conducted. There are three replicates and the experiment has been conducted as a complete randomized design. If the experiment had been conducted in blocks, the pure error degrees of freedom would be reduced by (choose the correct answer):

(a) 3 **(b)** 5 **(c)** 2

(d) 4 **(e)** none of the above

4.54 Physics graduate student Laura Van Ertia has conducted a complete randomized design with a single factor, hoping to solve the mystery of the unified theory and complete her dissertation. The results of this experiment are summarized in the following ANOVA display:

Source	DF	SS	MS	F
Factor	?	?	14.18	?
Error	?	37.75	?	
Total	23	108.63		

Answer the following questions about this experiment.

(a) The sum of squares for the factor is ____.

(b) The number of degrees of freedom for the single factor in the experiment is ____.

(c) The number of degrees of freedom for error is ____.

(d) The mean square for error is ____.

(e) The value of the test statistic is ____.

(f) If the significance level is 0.05, your conclusions are not to reject the null hypothesis.

Yes

No

(g) An upper bound on the P-value for the test statistic is ____.

(h) A lower bound on the P-value for the test statistic is ____.

(i) Laura used ____ levels of the factor in this experiment.

(j) Laura replicated this experiment ____ times.

(k) Suppose that Laura had actually conducted this experiment as a randomized complete block design and the sum of squares for blocks was 12. Reconstruct the ANOVA display above to reflect this new situation. How much has blocking reduced the estimate of experimental error?