INDIAN STATISTICAL INSTITUTE
Second Semester of A.Y. 2018-19
Assignment

B.Stat. - 3rd Year
Design of Experiments

Maximum Marks: 200
Submission Deadline: 21.04.2019
Note: Need to Submit the report along with all codes in replicable form.

1. Four different designs for a digital circuit are being studied to the amount of noise present. The following data have been obtained:

Circuit Design	Noise Observed				
1	19	20	19	30	8
2	80	61	73	56	80
3	47	26	25	35	50
4	95	46	83	78	97

(a) Is the amount of noise present the same for all four designs? Use $\alpha = 0.05$.

(b) Analyze the residuals from this experiment. Are the analysis of variance assumptions satisfied?

(c) Which circuit design would you select for use? Low noise is best.

[5+7+3]=15

2. 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 shown below.

	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

Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions. [10]

3. 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 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 as follows.

	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	75	71	75	75	69

Analyze the data from this experiment (use $\alpha = 0.05$) and draw appropriate conclusions.

[10]

4. Suppose that in Problem 3, the observations for chemical type 2 and bolt 3 and chemical type 4 and bolt 4 are missing.

(a) Analyze the design by iteratively estimating the missing values.

(b) Differentiate SSE with respect to the two missing values, equate the result to zero, and solve for estimates of the missing values. Analyze the design using these two estimates of the missing values.

(c) Derive general formulas for estimating two missing values when the observations are in different blocks.

(d) Derive general formulas for estimating two missing values when the observations are in same blocks.

[5+5+5+5]=20

5. Three brands of batteries are under study. It is suspected that the lives (in weeks) of the three brands are different. Five batteries of each brand are tested with the following results:

Weeks of Life		
Brand 1	Brand 2	Brand 3
100	76	108
96	80	100
92	75	96
96	84	98
92	82	100

(a) Are the lives of these brands of batteries different?

(b) Analyze the residuals from this experiment and comment on the model's adequacy.

(c) Construct a 95 % confidence interval estimate on the mean life of battery brand 2. Construct a 99 % confidence interval estimate on the mean difference between the lives of battery brands 2 and 3.

[3+5+(3+4)]=15

6. 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 experiment 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.

	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

Analyze the data from this experiment (use $\alpha = 0.05$) and draw the conclusions. [10]

7. Suppose a researcher comes to you with a table of data, obtained from a block design, that looks as follows, where each x represents one observation. She asks you to analyze the data.

	Treatment		
	1	2	3
Block 1	xxx	xxx	xxx
Block 2	xxx	xxx	xxx
Block 3	xxx	xxx	xxx
Block 4	xxx	xxx	xxx

(a) How would you determine how to analyze the data?

(b) Describe two experimental plans, with proper justifications, that could have given rise to this data set.

(c) For each of the plans(designs) you have identified in (b) give,

(i) an associated linear model,

(ii) the ANOVA table based on this model,

(iii) the variance of a simple treatment comparison,

(iv) the estimator for the variance in (iii). [3+(2+2)+2×(1+3+2+3)]=25

8. A horticultural experiment conducted in a green house was laid out as a Latin square design, where the blocking factors represent temperature and light intensity, respectively. The treatment have a 2^2 factorial structure, that is, 2 factors A and B each at 2 levels. The layout of the design and the results from the experiment (in parentheses) are given below:

	Light Intensity			
Temperature	1	2	3	4
1	a_0b_0 (5)	a_1b_1 (10)	a_0b_1 (8)	a_1b_0 (7)
2	a_1b_1 (12)	a_1b_0 (8)	a_0b_0 (6)	a_0b_1 (10)
3	a_1b_0 (10)	a_0b_1 (8)	a_1b_1 (15)	a_0b_0 (7)
4	a_0b_1 (9)	a_0b_0 (9)	a_1b_0 (11)	a_1b_1 (16)

(a) Give a linear model for analyzing the data from this experiment. Justify.

(b) Obtain the ANOVA table, giving sources of variation and d.f.

(c) Give a numerical expression for the estimate of the interaction $A \times B$, and its variance.

(d) The experiment is to be repeated at different times, so that in the end data from 3 different times, T_1, T_2, T_3, say, will be available (the experiment above represents T_1). Even though the temperature and light intensity trends remain, they may be assumed to differ from one time period to the next. It is expected that there is interaction between the time factor and the treatment factors.
Give a linear model for data from this experiment and sketch the ANOVA table, giving sources of variation and d.f.

(e) For the experiment described in (d), what are the variances of the estimated main effects and interaction?

[2+5+3+7+3]=20

9. [**Wildflower experiment (1986)**] An experiment was run to determine whether or not the germination rate of the endangered species of Ohio plant *Froelichia floridana* is affected by storage temperature or storage method. The two levels of the factor *temperature* were *spring temperature*, $14°C - 24°C$ and *summer temperature*, $18°C - 27°C$. The two levels of the factor *storage* were *stratified* and *unstratified*. Thus, there were four treatment combinations in total. Seeds were divided randomly into sets of 20 and the sets assigned at random to the treatments. Each stratified set of seeds was placed in a mesh bag, spread out to avoid overlapping, buried in two inches of moist sand, and placed in a refrigeration unit for two weeks at $50°F$. The unstratified sets of seeds were kept in a paper envelope at room temperature. After the stratification period, each set of seeds was placed on a dish with 5ml of distilled deionized water, and the dishes were put into one of the two growth chambers for two weeks according to their assigned level of temperature. At the end of this period, each dish was scored for the number of germinated seeds. The resulting data are given in the following table:

Data for The wildflower experiment

Treatment Combination	Number Germinating	$\overline{y_i}$	s_i
1:Spring/Stratified	12 13 2 7 19	8.4	6.995
	0 0 3 17 11		
2:Spring/Unstratified	6 2 0 2 4	2.5	3.308
	1 0 10 0 0		
3:Summer/Stratified	6 4 5 7 6	5.0	1.633
	5 7 5 2 3		
4:Summer/Unstratified	0 6 2 5 1	3.6	2.271
	5 2 3 6 6		

(a) For the original data, evaluate the constant variance assumption on the one way analysis of variance model both graphically and by comparing sample variances.

(b) It was noted by the experimenter that since the data were the numbers of germinated seeds out of a total of 20 seeds, the observations Y_{it} should have a binomial distribution. Does the corresponding transformation help to stabilize the variances.

(c) Plot $\ln(s_i^2)$ against $\ln(\overline{y_i})$ and discuss whether or not a suitable power transformation might equalize the variances.

(d) Use Scheffe's method of multiple comparisons, in conjunction with *Satterthwaite's* approximation, to construct 95% confidence intervals for all pairwise comparisons and for the two contrasts $\frac{1}{2}(1, 1, -1, -1)$ and $\frac{1}{2}(1, -1, 1, -1)$, which compare the effects of temperature and storage methods, respectively.

[(2+3)+5+5+(5+5)]=25

10. The percentage of hardwood concentration in raw pulp, the vat pressure, and the cooking time of the pulp are being investigated for their effects on the strength of paper. Three levels of hardwood concentration, three levels of pressure, and two cooking times are selected. A factorial experiment with two replicates is conducted, and the following data are obtained:

Percentage of hardwood Concentration	Cooking Time 3.0 Hours Pressure			Cooking Time 4.0 Hours Pressure		
	400	500	650	400	500	650
2	196.6	197.7	199.8	198.4	199.6	200.6
	196.0	196.0	199.4	198.6	200.4	200.9
4	198.5	196.0	198.4	197.5	198.7	199.6
	197.2	196.9	197.6	198.1	198.0	199.0
8	197.5	195.6	197.4	197.6	197.0	198.5
	196.6	196.2	198.1	198.4	197.8	199.8

(a) Analyze the data and draw conclusions. Use $\alpha = 0.05$.

(b) Prepare appropriate residual plots and comment on the model's adequacy.

(c) Under what set of conditions would you operate this process? Why?

[10+5+10]=25

11. Consider the Latin Square Design (LSD) with dimension v.

(a) Given v, how many LSD can be constructed with standard form, semi-standard form, and al together?

(b) Write a computer code to generate all possible LSD with a given positive integer v. Verify what the maximum value of v would be that your program can handle.

(c) Using your code in (b), or otherwise, show that there does not exists no mutually orthogonal Latin Squares (MOLS) of order $v = 6$.

(d) Write down another computer code to generate all possible MOLS of order $v = p^r$, for given prime p and positive integer r (as input).

(e) Can you extend your code in (d) to accommodate all positive integers v? Justify.

[5+5+5+5+5]=25