SCHOOL OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF DATA SCIENCE AND ANALYTICS

SUMMER 2024

END SEMESTER

COURSE CODE: STA 4020A

UNIT NAME: DESIGN AND ANALYSIS OF EXPERIMENTS

DATE: 5TH AUGUST 2024 **TOTAL MARKS**: 40 MARKS

INSTRUCTIONS:

For this exercise:

1. ANSWER ALL QUESTIONS
2. Do all your working in the Rmarkdown (.rmd).
3. Submissions should be in either a **`.ipynb`** or **`.rmd` file**
4. NO SUBMISSIONS SHOULD BE DONE VIA EMAIL

QUESTION 1: [20 MARKS]

Imagine you are conducting a factorial experiment to study the effects of diet, exercise, and supplementation on health outcomes. The factors and their levels are as follows:

Factor A: Diet (3 levels: Low, Medium, High) - Fixed Effect
Factor B: Exercise (2 levels: None, Daily) - Fixed Effect
Factor C: Supplement (2 levels: With, Without) - Fixed Effect

1. Calculate the total number of possible treatment combinations for the given levels of each factor.
2. Create a full factorial design matrix that lists all possible combinations of the levels of Factors A, B, and C. Each row should represent a unique treatment combination.
3. Assume you have collected data on a response variable (e.g., a health outcome) for each treatment combination. Perform a factorial ANOVA to assess the main effects and interaction effects of Factors A, B, and C on the response variable.
4. Provide the ANOVA table summarizing the results, including the sum of squares, degrees of freedom, mean squares, F-values, and p-values for each main effect and interaction.
5. Interpret the p-values to determine which factors and interactions have statistically significant effects on the response variable. Discuss the implications of these findings.

QUESTION 2: [20 MARKS]

Consider a split-plot experiment designed to investigate the effects of soil type and watering frequency on plant growth. The experiment involves the following factors:

Main Plot Factor: Soil Type (2 levels: Sandy, Clay) - Fixed Effect
Subplot Factor: Watering Frequency (3 levels: Low, Medium, High) - Random Effect

1. Explain how the split-plot design is structured, including the distinction between main plots and subplots. Describe how the treatments are assigned to whole plots (main plots) and split-plots (subplots) within each whole plot.
2. Create a design matrix that shows all treatment combinations, detailing how the levels of the main plot factor (Soil Type) and the subplot factor (Watering Frequency) are arranged.
3. Given data on a response variable (e.g., plant height) collected from this experiment, perform a split-plot ANOVA to evaluate the effects of Soil Type and Watering Frequency on the response variable.
4. Provide the ANOVA table, including the sources of variation (main plot factor, subplot factor, interaction), sum of squares, degrees of freedom, mean squares, F-values, and p-values.
5. Interpret the p-values to identify which factors and interactions have statistically significant effects on the response variable. Discuss any significant interactions and their practical implications.

QUESTION 3: [20 MARKS]

You have yield data (in bushels per acre) from an experiment designed to investigate the effects of different fertilization treatments on corn growth. The experiment follows a Latin square design with the following treatments:

A: High Nitrogen
B: Low Nitrogen
C: High Phosphorus
D: Low Phosphorus
E: Control

The yields for each treatment are recorded as follows:

Treatment	Row 1	Row 2	Row 3	Row 4	Row 5
A	10.2	4.5	7.0	5.3	10.0
B	5.3	6.8	6.5	8.2	8.2
C	12.4	10.0	16.0	13.8	16.5
D	6.8	5.9	8.2	9.2	12.4
E	3.9	5.7	5.0	6.9	6.9

1. Evaluate the effects of the different fertilization treatments on corn yield, taking into account potential random effects due to environmental variability. Perform an appropriate statistical analysis to determine the significance of the treatment effects.

2. If necessary, conduct post-hoc tests to determine which specific treatments significantly differ from each other in terms of yield. Include a visual representation (e.g., a bar plot or box plot) comparing the treatments.

3. Discuss the results in the context of agricultural practices for enhancing corn growth. Consider both the statistical significance and the practical implications of the findings, offering recommendations for optimal fertilization strategies based on the results.

QUESTION 4: [20 MARKS]

In a feeding trial involving 5 dairy cows, each cow receives one of 5 diets in 5 successive periods, following a Latin square design. The milk production (in liters) for each cow and period is as follows:

Cow	Period 1	Period 2	Period 3	Period 4	Period 5
1	24.1 (B)	26.3 (D)	25.6 (C)	24.9 (E)	22.8 (A)
2	22.9 (A)	23.1 (E)	25.0 (D)	24.0 (C)	26.7 (B)
3	26.5 (C)	24.2 (A)	23.1 (B)	26.2 (E)	24.0 (D)
4	26.7 (E)	27.4 (B)	25.1 (D)	25.1 (C)	29.6 (A)
5	27.8 (D)	24.7 (C)	24.6 (E)	22.2 (A)	24.7 (B)

The diets are:

A: Grass alone
B: Grass with supplement 1
C: Grass with supplement 2
D: Grass with supplement 3
E: Grass with supplement 4

1. Evaluate the effects of the different diets on milk production by performing an ANOVA to test the significance of diet effects, period effects, and cow effects. Provide the ANOVA table, including sources of variation, sum of squares, degrees of freedom, mean squares, F-values, and p-values.

2. If necessary, conduct post-hoc tests to identify which specific diets significantly differ in terms of milk production. Include a visual representation (e.g., a bar plot or box plot) comparing the diets.

3. Interpret the results in the context of nutritional strategies for dairy farming. Highlight any potential benefits or drawbacks of specific diets based on the milk production results, offering recommendations for optimal feeding practices.