United States International University-Africa

Education to take you places

SCHOOL OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF DATA SCIENCE AND ANALYTICS

SUMMER 2024 – ASSIGNMENT 3

COURSE CODE: STA 4020A

UNIT NAME: DESIGN AND ANALYSIS OF EXPERIMENTS

DATE: 10TH JUNE 2024 **TOTAL MARKS**: 30 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

TASK: Optimizing Corn Yield with Simulated RCBD Data

The Agricultural Research Institute (ARI) is planning an experiment to investigate the impact of corn variety selection on yield, considering the influence of soil fertility. They need your help to simulate data and analyze it beforehand.

The experiment involves three promising corn varieties (V_1, V_2, V_3) planted within six distinct blocks representing different soil fertility levels. Each block will contain a randomized plot for each variety.

QUESTIONS:

a) Simulate the yield data with some random variation between varieties and blocks.

b) State the null and alternative hypotheses for the effect of variety (or blocking) on yield.

c) Perform an ANOVA to assess the effect of variety on yield while accounting for the block effect as a random factor.

d) Based on the p-value, can you conclude a significant difference in yield between the varieties at a 5% significance level? Explain your reasoning.

e) Create boxplots to visually compare the distribution of yield across the different corn varieties.

f) Based on your analysis of the simulated data, summarize your key findings regarding the effect of corn variety on yield.

g) How can the results from the actual experiment, if they confirm the simulated findings, be used to inform future corn breeding programs and selection strategies for maximizing yield?