**STAT 40001/STAT 50001 Statistical Computing Fall 2024**

**Lab -21**

The dataset attached with this assignment shows three different studying technique each student used (A, B, or C), their current grade in the class when they started using the studying technique, and their exam score they received after using the studying technique for one month to prepare for the exam for 90 students in a class.

1. Import the dataset in R and calculate the summary statistics.
2. Calculate the average scores by study techniques.
3. Display the Exam scores by study techniques using a parallel box plot.
4. Perform ANOVA to check whether the exam score varies by study technique?
5. Fit the ANCOVA model to determine whether the current grade influence your findings in (d)

Note that **A**nova() function in the car package with **A**nova(model,type=”III”) could be used

> # file.choose()

> data <- read.csv("C:\\Users\\PNW\_checkout\\Downloads\\vaishak\\PNW\_COURSE-WORK\\FALL24\\STATISTICAL COMPUTING\\Assignment\\Assignment 21\\ANCOVA data.csv")

> head(data)

X technique current\_grade exam

1 1 A 83 92

2 2 A 76 91

3 3 A 80 89

4 4 A 67 84

5 5 A 76 92

6 6 A 80 81

> dim(data)

[1] 90 4

> names(data)

[1] "X" "technique" "current\_grade" "exam"

> new = data[,-c(1)]

> summary(new)

technique current\_grade exam

Length:90 Min. :65.00 Min. :71.00

Class :character 1st Qu.:71.25 1st Qu.:75.00

Mode :character Median :80.00 Median :83.00

Mean :79.38 Mean :82.94

3rd Qu.:86.75 3rd Qu.:88.75

Max. :95.00 Max. :95.00

> table(new$technique)

A B C

30 30 30

> # A B C

> attach(new)

> aggregate(exam, by = list(technique), FUN = mean)

Group.1 x

1 A 87.03333

2 B 82.76667

3 C 79.03333

> boxplot(exam~technique) #VISUAL analysis

> summary(aov(exam~technique))

Df Sum Sq Mean Sq F value Pr(>F)

technique 2 961 480.7 11.45 3.84e-05 \*\*\*

Residuals 87 3651 42.0

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> cat("p value < 0.05, reject null hypothesis -> there's a significnat difference")

p value < 0.05, reject null hypothesis -> there's a significnat difference

> cat("To check which pair, has that difference")

To check which pair, has that difference

> TukeyHSD(aov(exam~technique))

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = exam ~ technique)

$technique

diff lwr upr p adj

B-A -4.266667 -8.255190 -0.2781436 0.0331316

C-A -8.000000 -11.988523 -4.0114769 0.0000207

C-B -3.733333 -7.721856 0.2551898 0.0714575

> cat("->-> There is no significant difference between B and C")

->-> There is no significant difference between B and C

> m1 = lm(exam~technique+current\_grade)

> anova(m1)

Analysis of Variance Table

Response: exam

Df Sum Sq Mean Sq F value Pr(>F)

technique 2 961.4 480.71 11.3706 4.154e-05 \*\*\*

current\_grade 1 15.5 15.50 0.3667 0.5464

Residuals 86 3635.8 42.28

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

