THE UNIVERSITY OF AUCKLAND
BUSINESS SCHOOL

INFOSYS 750, Assignment 1

Before you start:

1. Your assignment should be in the form of a report that responds to the parts of this assignment.
2. Sections from your R output should be embedded in appropriate places in your report. PLEASE NOTE THAT INCLUDING NON-RELEVANT OUTPUT MIGHT BE PENALIZED.
3. Please mention the additional R packages you are using for producing the outputs of analysis.
4. You will need different datasets for completing this assignment provided in Canvas:
    a. DIET.csv
    b. STATPAK.csv

Question 1

You will be using the dataset **DIET.csv** for this question. This data contains information on 78 people using one of three diets.

Variable name	Variable
Person	Participant number
Gender	Gender, 1 = male, 0 = female
Age	Age (years)
Height	Height (cm)
Preweight	Weight before the diet (kg)
Diet	Diet 1, diet 2, diet 3
Weight6weeks	Weight after 6 weeks (kg)

Use an appropriate analysis step by step to investigate the effects of **diet** and **gender** on **weight loss.**

(a) Report the essential assumption checks.

(b) Interpret the main output and visualize the interaction plots.

(c) Is there any significant difference in the weight loss of males between diet1 and diet 2? If yes, how different is it compared to the female person?

(d) Which combination of diets and gender generates the highest efficiency for weight loss?

Question 2

For this question, you will be using the dataset stored in **'STATPAK.csv'** This dataset can be used to explore the relationships between different stat packages, computing platforms, and experience on the one hand, and time it takes to complete a task, satisfaction, and comprehension measures on the others when performing a stat package task.

Independent Variables

Stat Package	Which stat package was used?
Platform	Which type of computer was used?
Experience	Number of years that the subject has used stat packages

Dependent Variables

Comprehension	Objective quality of task output. The resulting stat package was given a score out of 120.
Time	Time needed to finish the task. The subjects were allowed as much time as they thought was necessary until they felt they could make no further progress.
Satisfaction	Self-reported level of satisfaction The subjects were asked for a single value up to 150.

Run MANOVA model examining the influences of various statistic packages on users' satisfaction and time needed to finish the task.

(a) Are there outliers? If yes, remove them and justify the dataset. Also, report the homogeneity of variance and covariance

(b) Is there a significant effect of the three different stat packages on the time and satisfaction as a group? And as each individual? (main effects)

(c) Test and report the significant differences between independent groups.