

**Cork Institute of Technology**  
**Institiúid Teicneolaíochta Chorcaí**  
**Semester 2 Examinations 2019/20**

<b>Module Title:</b>	Data Analysis with R
<b>Module Code:</b>	STAT9006
<b>Internal Examiner:</b>	Dr Seán Lacey
<b>Instructions:</b>	Write a Report
<b>Duration:</b>	2 hours 30 minutes
<b>Sitting:</b>	Assessment 03, April 2020

Total cholesterol levels less than 200 milligrams per deciliter (mg/dL) are considered desirable for adults. A reading between 200 and 239 mg/dL is considered borderline high and a reading of 240 mg/dL and above is considered high. A physician randomly assigned 16 subjects, with borderline high (or high) cholesterol readings, to one of two groups (8 subjects per group) for a two-month period. Subjects in both groups were placed on a particular diet for two months. While 8 subjects were also put on a strict exercise regime for the same two-month period. Their cholesterol levels were measured before, during and after the study, and the physician wants to know if subject's cholesterol levels have changed over the two months. Test the hypotheses that there is no difference in subjects cholesterol levels:

1. Over time;
2. With respect to group;
3. Over time with respect to group.

Open *Assessment 03: Dataset* on Canvas. Report and interpret your findings using 0.05 level of significance.

**Requirements:** In your assessment, you are expected to use *R* to compile the appropriate tables/graphs/tests etc. You are expected to **explain all concepts and procedures** used in the analysis of the data - i.e., all the **descriptive statistics and statistical inference should be fully explained!**

**Breakdown of marks (200 marks)**

- 15 marks for a clear introduction to the assessment with reference to:
  - **Type of data** gathered, **type of study**, **sample size**, **related** or **independent** samples and the main questions to be answered in the assessment.
- 30 marks for the correct use and explanation of **boxplots** and **simple main effects plot** describing the difference [**use `grid.arrange()`** to pair the appropriate plots]:
  - 20 marks for presentation of plots;
  - 10 marks for a correct comment/observation on each graph.
- 30 marks to test the assumptions of **normality**, homogeneity of **variance** and **sphericity**. Clearly state the hypotheses when testing the three assumptions.

- 50 marks for the calculation and explanation of the **mean, standard deviation, min, max** and **confidence interval** for each group:
  - 40 marks for correct calculations presented. Output needs to be presented in correct table format for full marks;
  - 10 marks for comments comparing **only** the **means** of the subgroups.
- 30 marks for the test of **differences** between the subgroups:
  - 10 marks for clearly stating the null and alternative hypotheses (note there are three null hypotheses);
  - 10 marks for the correct **conclusion**. The conclusion should state whether or not a difference exists.
  - 10 marks for **strength** of the difference found. Marks for **presentation**, full **description** of what effect size measures (in general) and comment on the actual **number**;
- 45 marks for **pairwise comparison** tests between the subgroups (using Tukey HSD):
  - 30 marks for **presentation** of the results;
  - 15 marks for the **correct conclusions**.

Upon completion of your report, upload your report **AND** *R* script as an assignment submission on Canvas