

CCT College Dublin Continuous Assessment

Programme Title:	HDip in Science in Data Analytics for Business (FT)		
Cohort:	Feb 23 Start (Fulltime)		
Module Title(s):	Statistical Techniques for Data Analysis		
Assignment Type:	Individual	Weighting(s):	60%
Assignment Title:	CA2		
Lecturer(s):	John O'Sullivan		
Issue Date:	28th of April 2023		
Submission Deadline Date:	23:59, 19th of May 2023		
Late Submission Penalty:	Late submissions will be accepted up to 5 calendar days after the deadline. All late submissions are subject to a penalty of 10% of the mark awarded . Submissions received more than 5 calendar days after the deadline above <u>will not</u> be accepted and a mark of 0% will be awarded.		
Method of Submission:	Moodle		
Instructions for Submission:	Submissions should consist of: a .ipynb Notebook and a corresponding pdf		
Feedback Method:	Results posted in Moodle gradebook		
Feedback Date:			

Learning Outcomes:

Please note this is not the assessment task. The task to be completed is detailed on the next page.

This CA will assess student attainment of the following minimum intended learning outcomes:

1. Formulate and test hypotheses within a business context using appropriate statistical techniques and both evaluate and communicate the results effectively to peers and team members. (Linked to PLO 3, PLO 6)
2. Apply regression analysis to appropriate datasets and demonstrate an awareness of the limitations of regression models. (Linked to PLO 2)
3. Use and understand current software tools and languages to produce result sets from existing data(e.g. Excel, R, Python). (Linked to PLO 4)

Attainment of the learning outcomes is the minimum requirement to achieve a Pass mark (40%). Higher marks are awarded where there is evidence of achievement beyond this, in accordance with QQI

Assessment and Standards, Revised 2013, and summarised in the following table:

Percentage Range	CCT Performance Description	QQI Description of Attainment	
		Level 6, 7 & 8 awards	Level 9 awards
90% +	Exceptional	Achievement includes that required for a Pass and in most respects is significantly and consistently beyond this	Achievement includes that required for a Pass and in most respects is significantly and consistently beyond this
80 – 89%	Outstanding		
70 – 79%	Excellent		
60 – 69%	Very Good	Achievement includes that required for a Pass and in many respects is significantly beyond this	Achievement includes that required for a Pass and in many respects is significantly beyond this
50 – 59%	Good	Achievement includes that required for a Pass and in some respects is significantly beyond this	Attains all the minimum intended programme learning outcomes
40 – 49%	Acceptable	Attains all the minimum intended programme learning outcomes	
35 – 39%	Fail	Nearly (but not quite) attains the relevant minimum intended learning outcomes	Nearly (but not quite) attains the relevant minimum intended learning outcomes
0 – 34%	Fail	Does not attain some or all of the minimum intended learning outcomes	Does not attain some or all of the minimum intended learning outcomes

Please review the CCT Grade Descriptor available on the module Moodle page for a detailed description of the standard of work required for each grade band.

The grading system in CCT is the QQI percentage grading system and is in common use in higher education institutions in Ireland. The pass mark and thresholds for different grade bands may be different from what you have experience of in the higher education system in other countries. CCT grades must be considered in the context of the grading system in Irish higher education and not assumed to represent the same standard the percentage grade reflects when awarded in an international context.

Assessment Task

Students are advised to review and adhere to the submission requirements documented after the assessment task.

You must create a single notebook in Jupyter Notebook to complete the following tasks.

N.B.: For all hypothesis tests, all steps and conclusions must be clearly stated.

Tasks

:

1. Load the dataset *Q1.csv*. It contains the exam scores (in percentages) of a sample of 50 students from a Dublin secondary school.
 - a. Find and comment on important summary statistics and produce an appropriate plot to summarise the dataset.
 - b. One of the teachers is concerned about the performance of the students in the school. She suspects that their performance may be below the reported national average of 70%. Does the data show that her concerns are justified? Use a significance level of $\alpha = 0.05$.
 - c. Produce and comment on an appropriate plot to illustrate your findings.

(25 points)
2. Load the *diamonds* dataset, and print the first 5 rows. The *color* variable refers to the colour of the diamond, with categories from "D" to "J". Colourless diamonds are considered better than diamonds with a yellow tint. Diamonds from "D" to "F" are considered colourless, and diamonds from colour "G" to "J" are not considered colourless (that is, they have a very faint colour).
 - a. Create a new binary variable in the dataframe called "colourless" which records 1 in rows with colourless diamonds and 0 otherwise.
 - b. Perform an appropriate hypothesis test to determine whether there is any association between the clarity of a diamond and whether it is colourless or not. Use a significance level of $\alpha = 0.01$.
 - c. Produce and comment on an appropriate plot to illustrate your findings.
 - d. Find and interpret 90% confidence intervals for both the mean price of colourless diamonds and the mean price of non-colourless diamonds.

(25 points)
3. Load the *PlantGrowth* dataset from the *pydataset* library. It contains the results of a small study comparing the yields of plants obtained under a control and under two different treatment conditions.
 - a. Find and comment on important summary statistics by treatment and produce an appropriate plot to summarise the dataset.
 - b. Conduct an appropriate hypothesis test to see if there is evidence of a difference between the three means (that is, the control and the two treatments). Use a significance level of $\alpha = 0.05$.
 - c. If there is evidence of a difference between the three means, find and comment on where this difference may be.

(25 points)

4. Load the trees dataset from the *pydataset* library. It contains measurements of the diameter, height and volume of timber in 31 felled black cherry trees. Note that the diameter (in inches) is labelled girth in the dataset. It is measured at 4 foot 6 inches above the ground.
 - a. Perform a correlation analysis between all numerical variables. Include and comment on the results of hypothesis tests for the population correlation coefficients between all three pairs of variables (you can use the *pearsonr* function from the *scipy.stats* library).
 - b. There is interest in estimating the volume of timber from trees using either the girth or the height of the trees, or both. Perform a regression analysis to decide which of the three possible models you would recommend using. Interpret your results and provide a short conclusion of your findings.

(25 points)

Total: 100 points

Submission Requirements

All assessment submissions must meet the minimum requirements listed below. Failure to do so may have implications for the mark awarded.

All assessment submissions must adhere to or contain the following:

- Submissions should be in the form of:
 - A .ipynb Notebook file containing all relevant text, code, output, comments, plots, graphs, tables, etc.
 - A single corresponding pdf document created from the above .ipynb Notebook file.
 - Any source datasets used (ideally in csv format)
- A list of references should be included. These should be in alphabetical order, and use the Harvard referencing system.
- Be neat and presentable (well laid out, consistent font sizes and headings etc.), and be well structured
- Be submitted by the deadline date specified or be subject to late submission penalties
- Be submitted via Moodle upload
- Use Harvard Referencing when citing third party material
- Be the student's own work.
- Include the CCT assessment cover page.

Additional Information

- Lecturers are not required to review draft assessment submissions. This may be offered at the lecturer's discretion.
- In accordance with CCT policy, feedback to learners may be provided in written, audio or video format and can be provided as individual learner feedback, small group feedback or whole class feedback.
- Results and feedback will only be issued when assessments have been marked and moderated / reviewed by a second examiner.
- Additional feedback may be provided as individual, small group or whole class feedback. Lecturers are not obliged to respond to email requests for additional feedback where this is not the specified process or to respond to further requests for feedback following the additional feedback.
- Following receipt of feedback, where a student believes there has been an error in the marks or feedback received, they should avail of the recheck and review process and should not attempt to get a revised mark / feedback by directly approaching the lecturer. Lecturers are not authorised to amend published marks outside of the recheck and review process or the Board of Examiners process.
- Students are advised that disagreement with an academic judgement is not grounds for review.
- For additional support with academic writing and referencing students are advised to contact the CCT Library Service or access the [CCT Learning Space](#).
- For additional support with subject matter content students are advised to contact the [CCT Student Mentoring Academy](#)
- For additional support with IT subject content, students are advised to access the [CCT Support Hub](#).