**CCT College Dublin Continuous Assessment**

| | | | |
|---|---|---|---|
| **Programme Title:** | *HDIP in Science in Data Analytics for Business* | | |
| **Cohort:** | *FT Sept. 2025* | | |
| **Module Title(s)**: | *Statistical Techniques for DA* | | |
| **Assignment Type:** | *Individual* | **Weighting(s)**: | *60%* |
| **Assignment Title:** | | | |
| **Lecturer(s)**: | *Dr. Shree Krishna Acharya* | | |
| **Issue Date:** | *01st Nov. 2025* | | |
| **Submission Deadline Date:** | *21st Dec. 2025* | | |
| **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.<br>Submissions received more than 5 calendar days after the deadline above **will not** be accepted and a mark of 0% will be awarded. | | |
| **Method of Submission:** | **Moodle** | | |
| **Instructions for Submission:** | *You must submit a zip file containing a Word with the report and a Jupyter Notebook file with the code and the outcome* | | |
| **Feedback Method:** | **Results posted in Moodle gradebook** | | |
| **Feedback Date:** | *After the approval from the exams board.* | | |

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.

## Learning Outcomes:

**Statistical Techniques for Data Analytics**

- Select several descriptive and inferential statistics and evaluate based on case study? Linked to PLO 1 and PLO 5
- Use software tools such as Excel and Python to produce results. Linked to PLO5
- Develop data analysis and create documentation with critical analysis. Use probability and random process for optimizing feature scaling the statistical dataset. Linked to PLO 3 and PLO 4
- Apply regression analysis to appropriate datasets and demonstrate an awareness of limitation of regression models. Linked to PLO 2
- Formulate and test hypotheses using appropriate statistical techniques. Linked to PLO 4.

**Classroom GitHub**

Students must use Classroom GitHub for version control in this assignment. The Python code should be developed and documented in a Jupyter Notebook, while the findings, analysis, and discussion should be compiled into an MS Word report. The GitHub repository link is provided below.

**Link:** https://github.com/Statistical-For-DA-Sept-2025

## Background :

CCT college is studying several factors that affect student success and health. The collected dataset has following features from 100 random students.

- Hours studied per week
- Hours sleep per night
- Daily Caffeine intake ( in terms of Cups)
- Exam scores between 0 to 100.
- Whether the student reports high stress. Binary output (Y/N)

CCT is making track on :

- Probability of catching a cold during exam season. Cold is assumed to enhance stress level.
- The cost for average student if they get sick.

## Assessment Details:

You can find the dataset in the module.

**Statistical Techniques for Data Analytics**

1. Compute parametric and non-parametric test based on exam scores between stressed and non-stressed students. Check the distribution and test methodologies (e.g. t-test, U-test) Explain in 250 words. [0-25%].
2. CCT introduced the App to identify stressed student. The lunched app flags 30% of student stressed but 25% students are stressed. The app correctly identifies stressed student 80% of time but also wrongly flags 10% of non-stressed students. If the app flags a student as stressed, what is the probability they are truly stressed? Solve and explain it in 250 words) [0-25%].
3. Build a model to predict exam scores using hours studied, hours of sleep and caffeine intake. Explore the strongest predictor factor of exam performance? Explain in 250 words) [0-25%].
4. Create an ipython file to compute expected value, flags probability, and regression model. Explain the model in 250 words [0-25%].

**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:

- Maximum number of words should be in the range of 1000±5%.
- Add required word count if applicable.
- Submission the report in Word format and. IPYNB file containing the code. The name of the file should be Hdip_YOUR_NAME_Year.
- Be submitted by the deadline date specified or be subject to late submission penalties.
- Use Harvard Referencing when citing third party material.
- Be the student's own work.

- Include the CCT assessment cover page.
- Students must use the classroom GitHub link (**https://classroom.github.com/a/o4rONclq**) and you should have more than 7 commits on GitHub on different dates and times.
- Use of Gen AI (chatGPT or Copilot or others) must be clearly stated for any part of your assignment.

# Statement of Acceptable Use of Artificial Intelligence

| **Acceptable and Unacceptable Use of AI** |
|---|
| ● The use of generative AI tools (e.g. ChatGPT, Dall-e, etc.) is permitted in this assignment for the following activities:<br>o Brainstorming and refining your ideas;<br>o Fine tuning your research questions;<br>o Finding information on your topic;<br>o Drafting an outline to organise your thoughts; and<br>o Checking grammar and style.<br>● The use of generative AI tools is not permitted in this course for the following activities:<br>o Impersonating you in classroom context<br>o Completing group work that your group has assigned to you<br>o Generating code for your assignment<br>o Writing a draft of a writing assignment<br>o Writing entire sentences, paragraphs or papers to complete class assignments.<br>● You are responsible for the information you submit based on an AI query. Your use of AI tools must be properly documented and cited.<br>● Any assignment that is found to have used generative AI tools in an unauthorised way will be subject to college disciplinary procedures as outlined in the QA Manual.<br>● When in doubt about permitted usage, please ask for clarification. |

| **Marking Criteria** | Weighting | Excellent (+70%) | Very Good (60-69%) | Good (50-59%) | Acceptable (40-49%) | Fail (<39%) |
|---|---|---|---|---|---|---|
| Parametric and non-parametric test | 0-20 % | Thoroughly compute the stressed and non-stressed students using tests (e.g. t-test, ANOVA, U-test), skewness and outliers. | Accurately compute the stressed and non-stressed students using tests (e.g. t-test, ANOVA, U-test), skewness and outliers. | Provides a good overview of the stressed and non-stressed students using tests (e.g. t-test, ANOVA, U-test), skewness and outliers. | Basic description of the stressed and non-stressed students using tests (e.g. t-test, ANOVA, U-test), skewness and outliers. | Insufficient or incorrect description of the stressed and non-stressed students using tests, skewness and outliers. |
| Expected Value and Random process | 0-20% | Thoroughly explore the covariance, variance and expected value of in terms of medical cost per student. | Accurately identifies the covariance, variance and expected value of in terms of medical cost per student. | Provides a good overview of the covariance, variance and expected value of in terms of medical cost per student. | Basic dataset description provided, the covariance, variance and expected value of in terms of medical cost per student. | Insufficient or incorrect description of the covariance, variance and expected value of in terms of medical cost per student. |
| Probability and | 0-20% | Thoroughly explore | Accurately identifies | Provides a good overview | Basic dataset description | Insufficient or incorrect |

| | | | | | | |
|---|---|---|---|---|---|---|
| Applications (Baye's Theory | | probability, conditional probability and baye's theory. | probability, conditional probability and baye's theory. | of the probability, conditional probability and baye's theory. | provided, lacking key details in one or more area of probability, conditional probability and baye's theory. | description of probability, conditional probability and baye's theory. |
| Regression Model | 0-20% | Thoroughly explore the feature and target variables, bias, slope, optimization and prediction for provided dataset. | Accurately identifies the feature and target variables, bias, slope, optimization and prediction for provided dataset. | Provides a good overview the feature and target variables, bias, slope, optimization and prediction for provided dataset. | Basic dataset description of the feature and target variables, bias, slope, optimization and prediction for provided dataset. | Insufficient or incorrect description of the feature and target variables, bias, slope, optimization and prediction for provided dataset. |
| Coding | 0-20 % | Thoroughly presenting coding for Expected value, baye's application and regression model. | Accurately implementing coding for Expected value, baye's application and regression model. | Provides a good coding for Expected value, baye's application and regression model. | Basic coding for Expected value, baye's application and regression model. | Insufficient or incorrect coding result for Expected value, baye's application and regression model. |

**The Irish Grading System**

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 experienced 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.

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, and review the marking criteria outlined in this assignment brief for a breakdown of the marking criteria for this specific assignment.

**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.
- 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.