MATH11188 Statistical Research Skills

Assignment 2: Scientific Report

Deadline: Friday 28th March 13.00 (1pm)

Submission of assessments is via Learn

This assignment is separated into two distinct but related tasks. Task 1 focuses on producing a *group* scientific report relating to a statistical analysis; Task 2 relates to an associated *individual* (non-technical) executive summary to describe the main findings. The assignment contributes 40% of the final grade.

The scientific report and executive summary should be submitted in the form of pdf files (only pdfs will be marked!). You need to submit two separate submissions - one group report (to "Assignment 2: Group Report") and an individual report (to "Assignment 2: Individual Exec Summary").

The marking rubric for the assignment is available in Learn in the folder: "Assignment 2" with associated marking categories: Content = 40%; Structure = 20%; Executive Summary = 40%. (Content and Structure relate to Task 1; Executive Summary to Task 2).

The data comes from Universities and Colleges Admissions Service (UCAS) and The Guardian.

In this assignment, you are asked to **choose** a question of interest, and **respond** it based on the dataset.

Dataset description: an overview of diversity in UK higher education

I merged two datasets, from The Guardian university guide ranking[1] and the UCAS equality report of 2022. Note that this dataset is of low quality as it is mostly viewed as a toy dataset. The questions you will ask are nevertheless very serious and I expect you take problems of diversity and inclusion with gravity.

Your task is to analyse the dataset and respond to an interesting question about UK higher education it using the dataset and statistical methods you know. You will have to find yourself a question of interest that can be related to the success rates of the universities, the way they attract students, the effect of diversity, etc. The question must be related to the dataset and the response rely on the available data. In case of doubt, you can contact me to check if your question is good. The goal is not to get significant result but to see if you are capable of finding questions of interest and appropriate methods.

Data description

The data is constituted of several measurements for each university. The first columns identify the universities

- `INSTITUTION_CODE` corresponds to a code designating each university (as the names might change with time);
- `INSTITUTION_NAME` corresponds to the usual name of the university.

The data from the Guardian correspond to the first columns:

- `satisfied_teaching` corresponds to the satisfaction of the students about teaching, from NSS;
- `satisfied_teaching` corresponds to the satisfaction of the students about assessment and feedback, from the national students survey (NSS);
- `students_staff_ratio` is the number of members of student per member of staff;
- `spent_per_student` is a rating from 0 to 10 estimating the spending of the university per student;

[1] `https://www.theguardian.com/education/ng-interactive/2024/sep/07/the-guardian-university-guide-2025-the-rankings`

- `avg_entry_tariff` is the average UCAS of young (under 21) entrants at the university;
- `added_value` measures how better the students succeed at their final year exam compared to their entry qualification;
- `career_after_15_month` is the proportion of students with a graduate-level job, or continuing studies at 15 month after their graduation;
- `continuation` is the percentage of students of first year that continue to second year.

The next columns come from UCAS and correspond to diversity measurements. Note that people belong to several categories (gender, socioeconomic background, origin, etc.).

- `White.ethnic.group`, `Black.ethnic.group`, `Asian.ethnic.group`, `Mixed.ethnic.group` and `Other.ethnic.group` correspond to the proportion of each ethnic groups in the new students.
- `POLAR4.Q1`, `POLAR4.Q2`, `POLAR4.Q3`, `POLAR4.Q4`, `POLAR4.Q5` measure the proportion of students arriving from each of the POLAR regions. "POLAR4 classifies local areas across the UK according to the young participation rate in higher education. The young participation rate is calculated by dividing the number of young people from each area who enter higher education aged 18 or 19 by the young population of that area. POLAR4 was calculated using data on students who begun their studies between 2009-10 and 2013-14. The areas are then ranked by participation rate and split into five quintiles, each of which represents about a fifth of the young population. The 20% of areas with the lowest participation rates are designated as "quintile 1", the top 20% are "quintile 5" and everywhere else is somewhere in between" [2];
- `SIMD.2016.Q1` to `SIMD.2016.Q5` correspond to geographical areas defined using the Scottish deprivation index [3] measuring a more general poverty. Unfortunately only Scotland uses this metric;
- `Men` and `Women` correspond to the gender;
- `Total` is the total number of new entrants.

Task 1: Group Scientific Report - max 2 pages

You have been assigned to a group of size three or four (see Learn for group membership). You are to write a scientific report for statististically informed people, responding to your question of interest. You should clearly describe any relevant data pre-processing, any statistical method used, explaining why you have chosen this method. The results should be presented, discussed and interpreted accordingly in relation to the goodness-of-fit measures (relative and absolute), and limitations of the data and model should be highlighted. Include any scientific and statistical information that you deem relevant to your analysis.

Note 1: You do not need to describe the data - the focus is on the application of the statistical analyses, and associated interpretation of the results.

Note 2: You may consider either a classical or Bayesian approach to the problem - as long as the statistical approaches are consistent with the framework applied.

Note 3: The group scientific report should be at most two A4 pages (minimum font size of 10). An additional appendix should be included providing the computer code used within the statistical analysis - the appendix is not included in the 2 page limit. You may include figures which will be included in the 2 page limit.

Note 4: All members of a group need to contribute to the submitted group report. However, it is expected that members of the group may contribute in different ways and different aspects of the work, including for example, investigating potentially suitable statistical models, identifying relevant relative/absolute goodness-of-fit approaches, undertaking the statistical analyses in `R`, creating illustrative figures, providing feedback on the analyses, interpreting the results, drafting parts of the report, proof-reading etc. Thus you may want to think about the different skills and interests of the members

[2] `https://educationblog.buckingham.ac.uk/2020/01/30/what-is-polar-and-how-useful-is-this-measurement-of-disadvantage/` for more details

[3] `https://simd.scot/#/simd2020/BTTTFTT/9/-4.0000/55.9000/`

of the group when working together and/or if assigning individual tasks. If one or more member(s) of the group fails to engage and does not reasonably contribute within the group assignment, please contact the course lecturer (gclarte@ed.ac.uk) detailing the issues **before** 12.00 noon on Thursday 27th March.

Task 2: Individual Executive Summary - max 400 words

Using the information from the group project, write an individual executive summary for the presidents of universities. The summary should outline the problem to be addressed, briefly describe the data and summarise the main results or findings of the statistical analyses undertaken in the context of the given problem. The summary should be concise including only the necessary information and be non-technical. Thus, the executive summary should not simply be text "cut-and-pasted" from the scientific report, the audience of the summary is very different to that of the scientific report.

Note 1: There is an upper limit of 400 words for the executive summary, but it is quite possible that the summary may be shorter than this (400 words are the limit not necessarily the aim). Also see the marking rubric (conciseness and relevant information are explicit criteria).

Note 2: The executive summary needs to be written independently by each member of the group. Thus, whilst the material on which the summary is based will be common to individuals within the same group, there are numerous choices to be made when writing the summary, such as the material to present, structure, style etc so that the summaries should be clearly distinct from each other.