Quantitative Methods - Coursework 1
Investigating Data

Deadline: 17:00 GMT, Mon 15 November 2021
Submission: PDF via Moodle

Late submission of the written investigation will be subject to the standard penalties, according to UCL regulations.

Weighting:

This assessment makes up 20% of the marks for the Quantitative Methods module (CASA0007).

Context:

'Condition X' denotes the childhood obesity, which is known to be associated with a variety of environmental factors. Over the period 2008 to 2018, local authorities across England were allocated funding to tackle this condition in their communities. They could choose how to distribute this funding across six different areas: improving air quality; cleaning public spaces; training health professionals; raising awareness in schools; raising awareness through the media; and subsidising counselling services. The data source is https://data.london.gov.uk/dataset/prevalence-childhood-obesity-borough, and we have manipulated the dataset so that there are some interesting things to find.
Please use the dataset provided in this assessment.

Data:

Download the file *coursework_1_data_2019.csv* from Moodle.

The file contains data on 152 local authority areas across the country. The columns of the table are as follows:

- *local_authority_area*: Listed by name;
- *2008_cases_total*: Counts of those who developed 'Condition X' in 2008;
- *2013_cases_total*: Counts of those who developed 'Condition X' in 2013;
- *2018_cases_total*: Counts of those who developed 'Condition X' in 2018;
- *2008_pop_total*: Populations of areas in 2008;
- *2013_pop_total*: Populations of areas in 2013;
- *2018_pop_total*: Populations of areas in 2018;
- *2008_cases_male*; *2013_cases_male*; *2018_cases_male*;
  2008_cases_female; *2013_cases_female*; *2018_cases_female*;
  2008_ pop _male; *2013_ pop _male*; *2018_ pop _male;*
  2008_ pop _female; *2013_ pop _female*; *2018_ pop _female*:
  As above, but separated by gender;
- *total_budget*: Annual budget allocated to each local authority 2008-18 (£ / yr);
- *clean_air*: Part of budget allocated to improving air quality (£ / yr);
- *clean_environ*: Part of budget allocated to cleaning public spaces (£ / yr);
- *health_training*: Part of budget allocated to training health professionals (£ / yr);
- *school_awareness*: Part of budget allocated to raising awareness in schools (£ / yr);
- *media_awareness*: Part of budget allocated to raising awareness through the media (£ / yr);
- *sub_counselling*: Part of budget allocated to subsidising counselling (£ / yr);
- *region*: Nine regions of England;

- *local_authority_type*: Five categories of local authority.

[Note that this is a mixture of real data and artificial data created for this assignment.]

Task:

- Create a simple research question;
- Investigate the data using methods seen in the Quantitative Methods course (or by extending those methods using techniques you have discovered through your own reading);
- Write a written report of your investigation, accompanied by appropriate diagrams and following the structure for academic writing set out in the lectures.

Length:

The maximum length of the text is 1000 words, excluding graphs, tables, other figures and code extracts. Footnotes are included in the word limit. In addition, the work as a whole must not exceed 4 pages, with a minimum font size of 12. The word count must be stated at the end of the text.

Please don't add your student number or real name to the assessment, as submissions are anonymised during marking.

Any references should appear on their own page at the end of the document, and are not counted in the 1000-word limit or the 4-page limit.

Penalties for exceeding these limits will be in line with UCL regulations.

Assessment

The investigation will be assessed using the mark scheme for written work that is available on Moodle and that was distributed in the first lecture. However, it will be marked **only** on the streams “Communication”, “Accuracy” and “Technical Difficulty”. Additionally, the weighting of the “Technical Difficulty” strand will be halved, producing a total mark out of 50, as follows:

- Communication: 20 marks available, as in mark scheme
- Accuracy: 20 marks available, as in mark scheme
- Technical Difficulty: 10 marks available (mark scheme divided by 2)
- Total: 50 marks available

Purpose

There are two main purposes for this piece of work. The first is for you to have experience of writing a quantitative essay following the mark scheme. The second is to identify points for improvement in your work. Both of these objectives will be useful in preparing you to write your final assessments.

The top priority for the task is clarity and communication. Make sure you write a clear, well-structured account of your investigation of the data, following the guidelines on writing quantitative essays that were set out in the lectures, with the appropriate headings.

Note that the technical difficulty section is weighted less heavily than the other two sections. This means that, while credit will be given for using more sophisticated techniques, the level of difficulty is of secondary importance to producing a clear, well-argued piece of writing with accurate quantitative analysis.

Suggestions

- An interesting research question might relate to how things changed between 2008 and 2018 (if at all); which areas have the best and worst performance; which interventions appear to be most effective, the relevance of the region or local authority type, etc.

- Please be aware of the difference between the number of cases and the proportion of Condition X (i.e. the ratio of the number of cases to the population). Which one should be used in the research question and the analysis?
- If performing a linear regression – simple or multiple – think carefully about what to choose as your response variable (y).
- It is likely that you will want to add additional columns to the data table: e.g. columns of deaths and injuries *per hundred thousand people* in each area; columns for the *change* in these quantities; spending *per person* in each area (both in terms of the total budget and in terms of the subcategories); etc.
- Consider whether the data contains any outliers that may affect your analysis.

Frequently Asked Questions

Do I have to investigate all the data?

No. It is fine to focus on certain columns of the data, but make sure that your research question relates to the investigation you perform. Clarity of presentation is key, so do not try to put more material into the report than you can comfortably fit. You may wish to consider several possible approaches before deciding which ones to write about.

Do I have to use all the methods seen in the course?

No. Look at the Technical Difficulty stream of the mark scheme. *It would be possible to get a pass in this section simply by applying a single method seen in the course. Combining this with a good write up and accurate application would result in a good mark.* Considering more than one approach or introducing additional methods or complexity, would then result in a higher mark (though remember that the Technical Difficulty section is weighted lower than Communication and Accuracy).

Which methods should I use?

Possible approaches include: single or multiple regression (and appropriate visualisation); considering coefficients of determination, the Pearson correlation coefficient and Spearman's rank correlation coefficient; identification of how data is distributed; comparison of data from different years through hypothesis testing, etc. You can remove outliers or consider specific subsets of the data if you feel that this would be a valid approach. Just be sure to explain your reasoning clearly.

What must be included in the report?

Follow the advice on structuring quantitative essays given in the lectures and use the appropriate headings (though this work does not require a Literature Review). As a minimum, you must state a research question, explain your methodology, present your results, interpret your results, and present your conclusions.

Can I use methods not covered in the course? Can I use Python?

Yes and yes. Be sure to explain the purpose of your methods clearly in your report, such that other students taking the course could understand them. Either of these approaches would increase the technical difficulty of your work. However, there is no obligation to use Python or to include any methods not covered in the course, although using Python is highly recommended.

What about references?

If you use any information from other sources, it should be clearly referenced, using the Harvard system (see Moodle for details), with sources included in your list of references on a page at the end of your project (not included in the word limit). If you use Mendeley or Zetero to manage references, you can use the template of Harvard style from this link (http://www.zotero.org/styles/ucl-institute-of-education-harvard).

However, referencing is not specifically assessed for this work, so you are not expected to use a wide range of sources. Sticking to material provided in the lectures will be sufficient to get a good mark.

How should diagrams be presented?

As clearly as possible. Clarity of graphs and figures will count towards your "Communication" mark. Make sure axes are clearly labelled and that figures are large enough to read easily.

Help! I can't find any relationships in the data!

While there *are* interesting relationships to be uncovered in this data, if you don't find anything significant, there is no need to panic. If you write up your investigation clearly and you have applied your chosen methods accurately, you can still get a good mark. Remember: negative results are still results!