BIOS0052: Human and Ecosystem Health in a Changing World

Assessment 2: Data analysis report

Building on the data analysis and critical thinking skills you have developed during the workshop activities, for this assessment you will **conduct a statistical analysis of an environment and health dataset** (see below), and **present your findings in the form of a short scientific paper**.

This will involve developing a research question and hypotheses, data exploration and visualization, statistical analysis, and presenting your results for a scientific audience. This assignment aims to give you experience in developing and answering a research question on the socio-ecological factors shaping health, and critically reflecting and discussing your findings.

You have been provided with a dataset on **Lyme disease incidence in the northeastern USA** between 2000 and 2019, which contains data on annual notified case counts from New York and Pennsylvania states, as well as a number of ecological, social and climatic covariates (including landscape variables, biodiversity, human population, economic and climate indicators). Lyme disease is a significant tick-borne zoonotic infection, which has a complex multi-host transmission cycle, is climate-sensitive, and is a rising health threat in the US and Europe. These data originally come from the US Centers for Disease Control (CDC), and contain Lyme disease surveillance data aggregated annually at the county (admin-2) level.

The disease dataset is provided as a .csv file on the Moodle GitHub – click here to access it – along with a shapefile of the county borders (for any mapping and GIS analysis) and a data dictionary describing the full dataset in detail.

For the assignment, you will develop a research question (or set of questions) and address them using this dataset. You are free to choose the specific question(s), and you are not obliged to use all the data. For example, your analyses might choose to focus on a single State or subset of years, or might focus on a particular set of covariates rather than all of them (e.g. climate impacts).

Please submit your assessment report in the form of a short scientific paper, with a maximum length of 2500 words (not including figure legends or references) and up to 3 figures or tables.

The style should follow a typical scientific paper structure, including an Abstract, a concise Introduction (outlining the background, rationale, research questions and hypotheses), Methods, Results, and Discussion. A good example for style is a "Research Article" in the journal *Biology Letters*.

At the end of the Introduction, you should include a short paragraph clearly stating the objectives and hypotheses you are addressing with your analysis. The Abstract should be maximum 200 words, and briefly outline the study rationale, questions, and your key findings You are advised to make good use of Figures and Tables to communicate and visualize your analysis design and findings.

Marking criteria

Reports will be marked on the following criteria:

- **Structure and presentation** – e.g. does the report follow a logical scientific paper structure; are sections clearly labelled with headings/subheadings; does the report follow a clear narrative?
- **Study rationale** – e.g. are the research questions and hypotheses clearly explained and justified, including with references to appropriate literature? Is it clear what question is being tested and why?
- **Scientific content** – e.g. appropriate and justified choice of analysis methods; statements backed up with evidence from credible scientific literature sources; clear and consistent citation/referencing style.
- **Communication of results** – e.g. clear and accurate interpretation of results; good use of figures, tables and visualization to describe the dataset and communicate findings
- **Critical synthesis and discussion** – e.g. critical and balanced interpretation of results and their potential generality; comparison to other evidence; discussion of limitations, knowledge gaps and future research priorities.

General guidance and hints

- Consider what type of question and analysis you are doing – descriptive, explanatory or predictive – and communicate this clearly. Think about how this should influence your analytical choices and the way you interpret your findings.
- If you are testing hypotheses in an explanatory study, make sure you develop and clearly explain your hypotheses, and keep in mind the potential for confounding relationships among variables and the challenges of causal interpretation.
- Exploring and mapping the data is an important step to understanding your dataset and the questions it's possible to ask, as well as for visual communication in your report.
- You can use any of the methods we have used in the module so far, but you are also welcome to use other analytical approaches you are familiar with, provided they are appropriate and justified.
- The field "ADMcode" is a unique identifier for each County, and is shared between the disease data and the shapefile, which will allow you to combine them for mapping and GIS analysis.
- The dataset has a hierarchical structure, with annual case counts nested within counties, so you should consider carefully how to account for this in your analysis.
- Different counties have different human population sizes so you may wish to account for this in your analysis, if you are analysing disease incidence. The dataset includes columns for population and log-transformed population.
- The dataset includes a column for both the number of Lyme disease cases (NumCases) and a binary (1/0) indicator of whether any cases were detected in a given year (AnyCases) – either of these could be interesting as outcome variables, but would require using different model families.
- Certain covariates might be highly correlated with one other, so not suitable for inclusion in the same statistical model (collinearity) – it's always worth checking this!
- You are welcome to supplement the dataset with additional data if you wish, although this is not expected or necessary to get a high grade.

Submission details

Please submit your assessment via TurnItIn on the Moodle page by the deadline of **28th April at 23:59**. Assignments will be independently marked by Rory Gibb and a second marker. The marking process will be anonymous, ensuring your identity remains confidential throughout the assessment process, so **please use your candidate number on your work and as the title of your upload (not your name)**.