Statistical Modelling with Data
DATA 603 (Fall 2022)

Group:_____

Final Project: Multiple linear regression (written)

SPEC SHEET

The instructional team will place a score to the left of the box reflecting the degree to which we judge that the spec has been fulfilled. This assessment will be based on the strength of evidence you provide. We will use the submitted project materials as sources of evidence.

The large font text in each spec will serve as the primary guide for assessment, with bullet points offering suggestions for how you might achieve the words in **BOLD ALL-CAPS.** (Note that these suggestions are not an exhaustive list; i.e., there may be other ways to achieve these specs that are not described, here).

Scores: 0 = little or no evidence; 1 = some evidence; 2 = strong evidence

SPEC

☐ Administrative tasks required for this assignment were **ALL** completed on time.

These include:

- A group of 3 to 5 members was formed and their names submitted at the required link
- The project checkpoint document was submitted to D2L on the required date
- The group evaluation survey is completed and electronically submitted with the final document
- The final document, the data set and group survey were submitted to D2L on the required date

THIS SPEC ASSESSES GROUP EFFECTIVENESS AND COMPLIANCE WITH REQUIREMENTS

SPEC

☐ Formatting requirements of the final document are **ALL** met.

These include:

- The final document is provided as a PDF and as R Markdown.
- There is a cover page, and it includes an informative project title, and the names of all group members
- There are section headings
- R code and mathematical expressions, if included, are offset from discussion text in a consistent manner (e.g., by using a different font; and/or by placing these in their own paragraph)
- Figures and tables, if included, are numbered consecutively (i.e., Figure 1, Table 1, Figure 2, Table 2, etc.)
- Figures and tables, if included, each have an informative caption of one to three sentences concisely explaining their content and demonstrating their relevance to the document
- Excluding in the appendix, where raw R output may be included, the number of significant digits reported for statistics, P-values, estimated parameters, model selection metrics, etc. is appropriate for the research objective, meaningful in terms of the variables used, and consistently applied (i.e., may involve some rounding from raw R output; tiny P-values could be given as less than an alpha value).

THIS SPEC ASSESSES COMPLIANCE WITH REQUIREMENTS

SPEC

☐ An "Introduction" section **CLEARLY EXPLAINS** and **JUSTIFIES** the research question and/or research objectives.

Suggestions:

- We have a clear understanding about what you are researching.
- The importance of the work is justified (e.g., its importance to society, political discourse, business, or a discipline of science, medicine, engineering, social science or humanities is clearly explained).
- There is a brief overview of what others have found (if available or relevant) on this question or objective

THIS SPEC ASSESSES HIGHER-ORDER THINKING (i.e., the creation of a question and evaluation of its relevance)

SPEC

☐ The data set you will analyse is **CONCISELY INTRODUCED**.

Suggestions:

- The source of the data set is given (i.e., URL, website, citation)
- The agency that generated the data is stated
- The conditions under which it was collected or generated are explained (if available)
- The sampling of the data set is discussed (the sample size; was the sample random; was it systematically collected; are there biases in its collection that are likely to exist, or that you are aware of)
- What the rows (i.e., cases or observations) in the data represent is explained
- Relevant additional details about the data collection (e.g., metadata, geography, time of collection, etc.) are given (if available).

THIS SPEC ASSESSES KNOWLEDGE, UNDERSTANDING AND HIGHER-ORDER THINKING (i.e., synthesis of data details)

SPEC

☐ A "Methodology" section **CLEARLY INTRODUCES** and **JUSTIFIES** the specific variables you will analyse

Suggestions:

- All variables used are described (e.g., how they were measured; units; qualitative or quantitative).
- The focal response variable is quantitative and continuous (i.e., it could theoretically take on decimal values, even though it may not actually do so in your data set).
- The focal response variable is *not* a proportion, percentage, category, or binary variable (as these require advanced modelling techniques when used as response variables)
- The choice of predictor variables and their importance to the analysis is justified (i.e., *they can be any type of variable*, including quantitative, qualitative, proportions, percentages, categories or binary variables)

THIS SPEC ASSESSES KNOWLEDGE, UNDERSTANDING AND HIGHER-ORDER THINKING (i.e., justification of importance)

SPEC

☐ A "Methodology" section **CLEARLY DESCRIBES** the statistical analyses you will use and **CORRECTLY JUSTIFIES** the use of these techniques.

Suggestions:

- You state that you are using multiple linear regression modelling.
- You describe the order in which you will conduct your analyses, and if a particular step is dependent on the result of a previous step, this is explained or presented in a figure (i.e., workflow).
- Any techniques that you use are introduced and their use is justified (e.g., interactions, polynomial terms, individual t-tests, global and partial F-tests, automated model selection, statistical tests of residuals, etc.)
- An alpha value that will be used for statistical hypothesis tests is stated.
- The description of these techniques and the justification for their use is consistent with how they have been taught in DATA 603.

THIS SPEC ASSESSES KNOWLEDGE AND UNDERSTANDING

SPEC

☐ A "Methodology" section states how workload was distributed among group members (i.e., co-authors), and shows that this distribution **IS EQUITABLE**.

Suggestions:

- The roles of each group member with respect to the tasks outlined in the methodology is stated in a paragraph (tasks associated with the oral presentation may also be mentioned)
- These roles may *not be equal* in the sense that they split the entire task into pieces of equal size, but rather the roles are *equitable* in that they play to individual strengths among group members and consequently require individuals to exert roughly equivalent efforts.
- The equity of the division is justified.

THIS SPEC ASSESSES GROUP EFFECTIVENESS AND COMPLIANCE WITH REQUIREMENTS

SPEC

☐ A "Results" section **CONCISELY PRESENTS** and **INTERPRETS** the results of statistical modelling and statistical inference.

Suggestions:

- The structure of this section follows the one you outlined in the "Methodology" section (e.g., it demonstrates its alignment by using sub-headings).
- The outcomes of statistical modelling, inference, and hypothesis tests are presented either in the text, in tables or in figures.
- Results selected for communication are the most important ones to support the investigation
- Results have been simplified from their raw form to present only information relevant for interpretation (i.e., it is not a dump of R output).
- More detailed results or outputs are placed in an Appendix, and this is cited (if applicable).

THIS SPEC ASSESSES HIGHER-ORDER THINKING (i.e., synthesis of results to those most salient)

SPEC

☐ A "Results" section **CORRECTLY INTERPRETS** the results of statistical modelling and statistical inference.

Suggestions:

- Results that have been presented are correctly interpreted
- The equation of a final linear model is given, containing the selected variables and parameter estimates.
- The effects of focal variables in this final linear model are interpreted in terms of their units (if applicable)
- If statistical hypothesis tests are interpreted: null and alternative hypotheses are concisely presented; test statistics are provided; P-values are evaluated with respect to these hypotheses

THIS SPEC ASSESSES KNOWLEDGE, UNDERSTANDING AND ANALYSIS (i.e., of statistical results)

SPEC

☐ A "Discussion" section **SUMMARIZES** the most important findings and **EVALUATES** their implications with respect to the research question and/or objectives.

Suggestions:

- There is a statement of key results in a more concise manner while connecting these findings to questions or objectives.
- There is an evaluation of the success of the modelling exercise in terms of these questions or objectives.
- Future statistical approaches, data sets or follow-up analyses that could be considered to better achieve objectives are suggested.
- A concluding paragraph serves as brief abstract to the work, reframing key findings and their significance.
- The concluding paragraph explains your findings in a way that could be understood by someone without any education in statistical analysis.

THIS SPEC ASSESSES HIGHER-ORDER THINKING (i.e., synthesis and evaluation of results in terms of objectives)

SPEC

☐ Requirements for written submissions in a university course are **ALL** followed.

These include:

- All text was written by the authors with the exception of quotations from other written sources
- Quotations (if used at all) are: short, infrequent, serve only to illustrate a point and not as a replacement for a written sentence, are surrounded in quotation marks, are cited and referenced
- All ideas that are not unique to the authors are cited and referenced; they come from reputable sources such as journals and reports published by publicly-funded organizations; media and website citations are infrequent; Wikipedia is not used as a primary source!
- Citation and referencing use any conventional system (see academic journals for some examples)

THIS SPEC ASSESSES COMPLIANCE WITH REQUIREMENTS

COMMENTS

WRITTEN MARK

/22