Graphical user interface, text

Description automatically generated

# Coversheet for submission of MSc coursework

* DO NOT WRITE YOUR NAME ON YOUR WORK. Instead, please write your student number on this coversheet.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Student No. | 2 | 4 | 7 | 8 | 6 | 8 | 8 |

|  |  |
| --- | --- |
| Date Submitted: | 16/01/2024 |
| Word Count: | 694 |

PLAGIARSIM DECLARATION:

|  |
| --- |
| This assignment is entirely my own work, and it adheres to the University of Bristol’s policy on plagiarism and academic integrity.  Quotations from secondary literature are indicated by the use of inverted commas around ALL such quotations AND by citations in the text or notes to the author concerned. ALL primary and secondary literature used in this piece of work is indicated in the bibliography placed at the end, and dependence upon ANY source used is indicated at the appropriate point in the text.   * I confirm that no sources have been used other than those stated. * I confirm that I have not used artificial intelligence or chatbot software to create any writing or content in this assessment * I confirm that I have not written this work in another language and translated it into English using translation tools * I confirm that I have not used grammar checkers that suggest rewrites   **I understand that plagiarism, collusion, and cheating constitute misconduct and may result in disciplinary action being taken.** |

**Research Compendium Report**

The association between fruit consumption and BMI

According to the WHO, around 1.9 billion adults are estimated to be overweight, over 650 of whom are obese [1]. Obesity is known to be associated with high blood pressure, type 2 diabetes, and many other health conditions. Therefore, finding new associations with weight is of great clinical importance. In this research compendium, we looked into whether there is an association between consumption of fruit, and Body Mass Index (BMI), confounded by income and gender.

To undertake such a task, we merged data from the NHANES, namely 2005-2006 Examination, Demographic and Dietary Data [2]. We removed missing observations on consumption of different fruits, as well as missing BMI and income values and subset the data to just these columns of interest.

Next, we created our ‘fruit’ variable by taking the levels of fruit consumption and converting them in to an approximate numeric value. For example, a value of “3” in the FFQ0016 column indicated consumption of 7-11 apples in a year, and so this would be converted to a mean value of 9, and combined with an individuals estimated consumption of all other fruits in the dataset. We also created an binary income variable, which took the Poverty Income ratio of each individual and dichotomised it, to track those of relative wealth and poverty. These 2 variables were combined into a new dataset, alongside BMI and gender, and saved within the “clean” subdirectory.

Our first ggplot involved first identifying the most suitable regression model between BMI and fruit consumption via a fractional polynomial process, which suggested a log transformation for the explanatory variable, fruit. Upon inspection of its distribution, such a transformation seemed intuitive given the cluster of small values followed by some very large observations. Additionally, a log transformation is still interpretable, as a % change in our covariate rather than a unit change. Then, we used the geom\_hex function in ggplot to make our plot of the regression more visually clear, as some of the clusters of data were so dense it cluttered the diagram.

For the second ggplot, we wanted to see if fruit consumption varied significantly between different levels of BMI. Thus, we categorised BMI into the 4 levels, based on medical standard, and used the geom\_boxplot() function in ggplot to create box plots of weekly fruit consumption for each level.

For the shiny app, we added in the additional complication of adjusting the regression model by some third confounder, as we wanted to observe how the relationship we are studying is present across different incomes and genders.

In order to view the shiny visualisation, simply locate app.R within the “/visualisation” subdirectory of the compendium and open within Rstudio, and then click “Run App”. All referenced directories in the code should be relative to the compendium and so run fine. This will produce a shiny app with some level of intractability. Users can select between “Income” and “Gender” and see the resulting regression lines of BMI against fruit consumption for both levels of the confounder.

A diagram of a body mass index

Description automatically generated

Our first ggplot shows how the regression line estimates a slight negative association between fruit consumption and BMI. For example, an individual whom consumes around 55 fruit a year is estimated to have a BMI of 26, whereas someone who consumes 400 is estimated to have BMI around 25. Clearly there is somewhat of a trend present, even if the size of association isn’t very substantial.

A graph with a number of squares

Description automatically generated with medium confidence

Our second ggplot supports the association present in the first plot. Individuals in higher BMI categories have slightly lower fruit consumption on average. For example, the mean fruit consumed in a week within obese individuals in the dataset is around 4, whereas it is closer to 5 in the ‘normal’ BMI category.

The shiny app shows a slight difference between predicted BMI by fruit consumption for both income and gender brackets. The same downwards relationship between BMI and fruit consumption is retained in this model, even after controlling for the confounders. Females are predicted to have slightly higher BMI for each level of fruit consumption versus males. Those in higher incomes are predicted marginally greater BMI than lower incomes.

**References**

[1] *Obesity and overweight statistics,* *World Health Organization*. Available at: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> (January 2024).

[2] <https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Dietary&Cycle=2005-2006>, <https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Demographics&CycleBeginYear=2005> <https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Examination&CycleBeginYear=2005#>