SMM634 - Group Assignment

(worth 50% of final grading)

Deadline 23 October 2024

For this assignment, you will use the data frame `expenditure`, which contains information on 10,638 individuals from the USA, sampled in 2012. The data frame includes the following variables:

- `general`: General health: 1 excellent, 2 very good, 3 good, 4 fair, 5 poor.

- `mental`: Mental health (as above).

- `bmi`: Body mass index.

- `income`: Income measured in thousand of dollars.

- `age`: Age in years.

- `gender`: Male 1, Female 0.

- `ethnicity`: 1 white, 2 black, 3 native american, 4 others.

- `education`: Education in years.

- `region`: 1 Northeast, 2 Midwest, 3 South, 4 West.

- `hypertension`: 1 if hypertension present and 0 otherwise.

- `hyperlipidemia`: 1 if hyperlipidemia present and 0 otherwise.

- `dvisit`: Number of doctor (physicians) visits.

- `ndvisit`: Number of non doctor visits (non-physician providers).

- `dvexpend`: Expenditure on doctor visits.

- `ndvexpend`: Expenditure on non doctor visits.

Using and analysing these data, write a report addressing the following points:

1. **Identify a Response Variable**: Select one variable as the response (dependent) variable and a subset of the remaining variables as regressors (independent variables). Justify your choice. [8]

2. **Fit a General Linear Regression Model**: Perform a linear regression analysis on the data. Justify your choice of the regression model and discuss the assumptions underlying the model. [8]

3. **Summarize and Interpret the Results**: Using the fitted model, provide a summary of the results, including tables and figures where appropriate. Interpret the empirical findings. [10]

4. Comment on the Limitations: Discuss the limitations of your analysis.   [8]

5. Suggestions for Improvement: Provide recommendations for how the analysis could be improved.   [8]

The report must be in PDF format and its structure must follow the five points outlined above. The report should not exceed 1,500 words, excluding tables, figures, references, and programming code.

An additional 8 marks will be awarded for the overall structure of the report, including neatness, presentation style, and any extra effort.

[Total marks: 50]