STAA 551 - Case Study

Due Friday, March 11, 11:30pm

Students will be divided into groups of 3 (subject to the class size being divisible by 3). Data will be provided to each group, along with a brief description of the data. Each group will submit one final Case Study paper containing all of the components listed below. The written portion should be 7 pages (or less), written in the same style as a journal article, but with more detail and focus on the analysis. Each student will also independently provide a peer evaluation of their group members.

Required Elements/Grading (20 points total)

Introduction (4 points)

- Background/identification of the purpose of the study.
- State response and predictor variables.
- Identify each variable as quantitative or categorical (with levels specified for categorical variable).

Summary Statistics and Graphics (4 points)

- Do this before formal model fitting.
- Include information on data cleaning/restructuring.

Analysis (5 points)

- Description/discussion of analysis with enough detail that someone else could recreate your results.
- You are encouraged to try different approaches, but please restrict yourselves to the least squares methods we have covered in this class (no other methods, such as loess or splines, should be used).
- Justify any choices that you made as part of the analysis.
- Discuss model assumptions and include diagnostic checks (plots and quantitative checks).
- As you are working toward a final model, keep in mind all that we have covered (checking model assumptions, transformations, WLS, GLS, influential observations, collinearity, etc).

Results and Conclusions (5 points)

- Final model and fitted equation
- ANOVA tables and tables of estimated coefficients/standard errors, confidence intervals, etc.
- Other results as appropriate
- Interpretation and discussion
- Refer back to the purpose of the study (how will your fitted model be of benefit)

Overall Style (2 points)

- 7 pages or less (including graphs, but not R code or references)
- Use complete sentences and correct grammar.
- R code should **not** be included in main body of the report.
- While some tables and results can be taken directly from R output, the reader should **not** have to sift through superfluous output (reduce R tables down to only what is needed).

- Graphs should be clearly labeled.

R Code Appendix

- I will check congruence of R code vs written description.
- If R code is not included, there will be a 2 point deduction.

Data Description

About the Data

These data are from the Prevention of REnal and Vascular END-stage Disease (PREVEND) study, which took place in the Netherlands. The study collected various demographic and cardiovascular risk factors.

> https://www.maelstrom-research.org/mica/individual-study/prevend

This dataset is from the third survey, which participants completed in 2003-2006; data is provided for a subset of individuals who completed cognitive testing.

The Ruff Figure Fluency Test (RFFT) is one measure of cognitive function that provides information about cognitive abilities such as planning and the ability to switch between different tasks. The test consists of drawing as many unique designs as possible from a pattern of dots, under timed conditions; scores range from 0 to 175 points (worst and best score, respectively).

Statins are a class of drug widely used to lower cholesterol. Research suggests that adults with elevated low density lipoprotein (LDL) cholesterol may be at risk for adverse cardiovascular events. A set of guidelines released in 2013 recommended statin therapy in individuals who are at high risk of adverse cardiovascular events, including individuals with Type II diabetes and moderately high LDL and non-diabetic individuals with atherosclerotic cardiovascular disease and high LDL. If these guidelines were to be followed, almost half of Americans ages 40 to 75 and nearly all men over 60 would be prescribed a statin.

However, some physicians have raised the question of whether treatment with a statin might be associated with an increased risk of cognitive decline.

The goal of this analysis is to examine which factors are associated with cognitive decline, with particular interest in statin use.

There are three natural candidates for potential confounders: age, educational level, and presence of cardiovascular disease (CVD). Many studies have found a negative association between cognitive function and age. Individuals with more education tend to have higher incomes and consequently, better access to health care and medication; also, individuals with more education may be more comfortable with assessments like the RFFT. Individuals with cardiovascular disease are often prescribed statins to lower cholesterol; cardiovascular disease can lead to vascular dementia and cognitive decline.

This analysis uses data from the Prevention of REnal and Vascular END-stage Disease (PREVEND) study. Clinical and demographic data for 4,095 individuals are stored in the `prevend` dataset. Full details of the variables in this subset of data are provided below. *Be on the lookout for things like missing values* and deal with them appropriately (and document what you did).

Variable Descriptions

```
Casenr
case number, numberic

Age
age in years, recorded at time of enrollment.

Gender
Numeric vector: 0 = males; 1 = females.
```

Ethnicity
Numeric vector: 0 = Western European; 1 = African; 2 = Asian; 3 = Other.

Education
Highest level of education. Numeric: 0 primary school; 1 = lower secondary education;
3 = university.

RFFT
Performance on the Ruff Figural Fluency Test. Scores range from 0 (worst) to 175 (best).

VAT
Visual Association Test score. The VAT is a learning task based on image recognition. Scores
may range from 0 (worst) to 12 (best)

CVD
History of cardiovascular event. Numeric vector: 0 = No; 1 = Yes.

DM
Diabetes mellitus status at enrollment. Numeric vector: 0 = No; 1 = Yes.

Smoking
Smoking at enrollment. numeric vector: 0 = No; 1 = Yes.

Hypertension
status of hypertension at enrollment. Numeric vector: 0 = No; 1 = Yes.

BMI
Body mass index, in kg/m^2

SBP
Systolic blood pressure, in mmHg

DBP
Diastolic blood pressure, in mmHg

MAP
Mean arterial pressure, in mmHg

eGFR
Estimated glomerular filtration rate, a measure of kidney function. Low values indicate
possible kidney damage, in mL/min.

Albuminuria
Albuminuria (mg/24hr) in two categories. Numeric vector: 0 = (< 30); 1 = (>= 30)

Chol
Total cholesterol, in mmol/L.

HDL
HDL cholesterol, in mmol/L.

Statin

Statin use at enrollment. Numeric vector: 0 = No; 1 = Yes.

FRS
Framingham risk score. Numeric vector. The score, a measure of risk for a cardivascular event within 10 years. Higher values imply increased use. For details see D'Agostino RBS, Vasan RS, Pencina MJ, Wolf PA, Cobain M, et al. (2008) General cardiovascular risk profile for use in primary care: The framingham heart study. Circulation 117: 743-753.