Factors Affecting Scot's Satisfaction with Public Transport

Group 28: Aishwin Tikku, Mengran Li, Steven Kwok, Shaoquan Li, Shuning Li

Introduction & Purpose

Public transport is necessary for most Scots. Therefore, its comfort and customer satisfaction are important for operators. To help improve public transport, we aim to discover factors having correlation with passengers' satisfaction in this project.

Data

- The research is based on a poll by the Scottish Household Survey (SHS).
- The whole dataset, combined from seven datasets obtained from website of Scottish government statistics, has 24 factors and 460 observations.
- The explanatory variables are Road Casualties, Road Transport Expenditure, Public Transport, Road Vehicles, Concessionary Travel Cards, Road Network and Traffic and Travel to work and other purposes.

Methodology

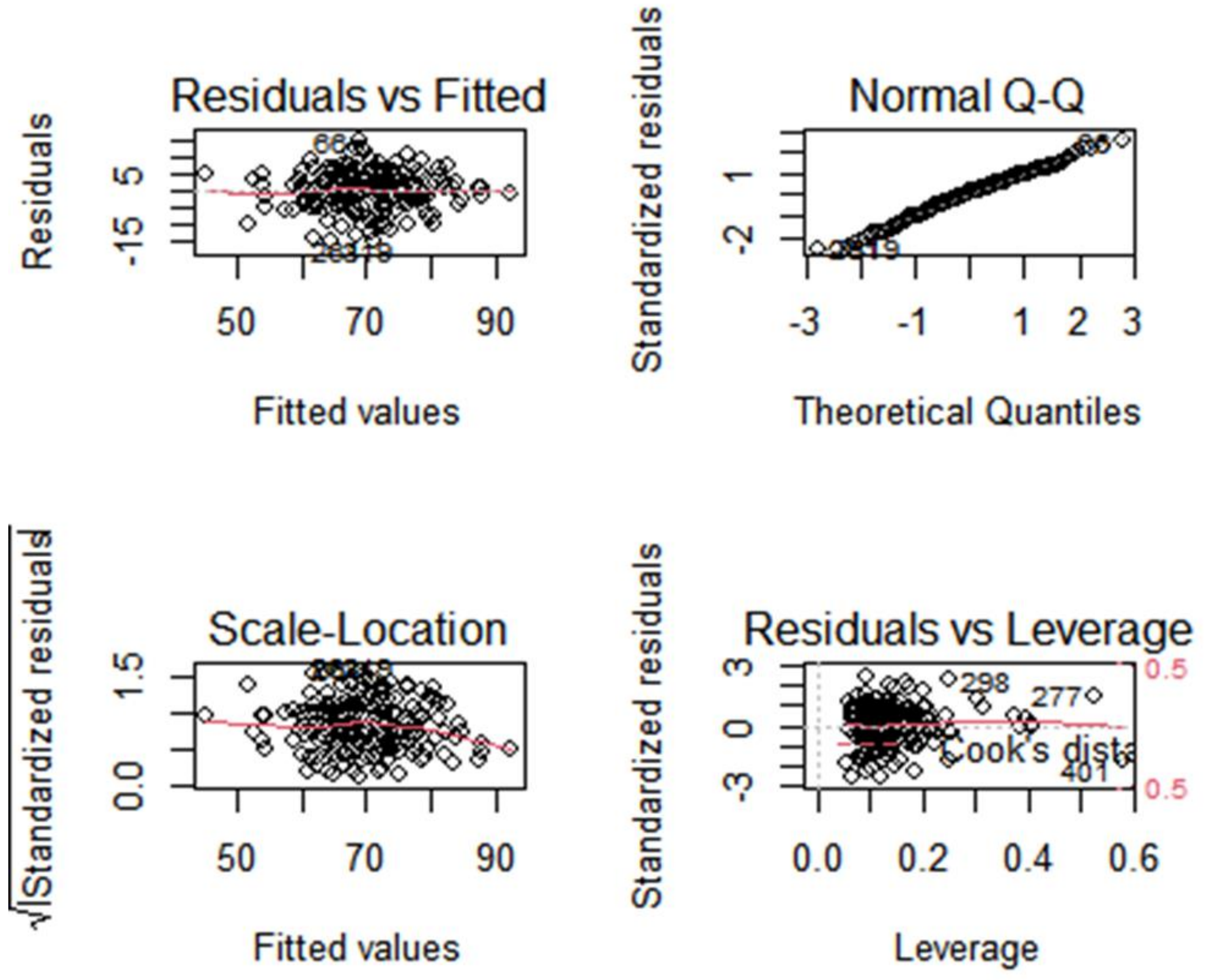
Linear regression model, a model for examining and discovering relations between response variable and explanatory variable(s), is applied in this project. The formula is presented below:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$$

where y is the response variable, x1 to Xp is the number of columns selected from 1 to p, β0 is the intercept of data, β1 to βp is the coefficients of corresponding columns of X from 1 to p. The ε is the error terms of the estimations.

Data Summary

Plot the Fitted values against Residuals and Q-Q plot to assess our assumptions.



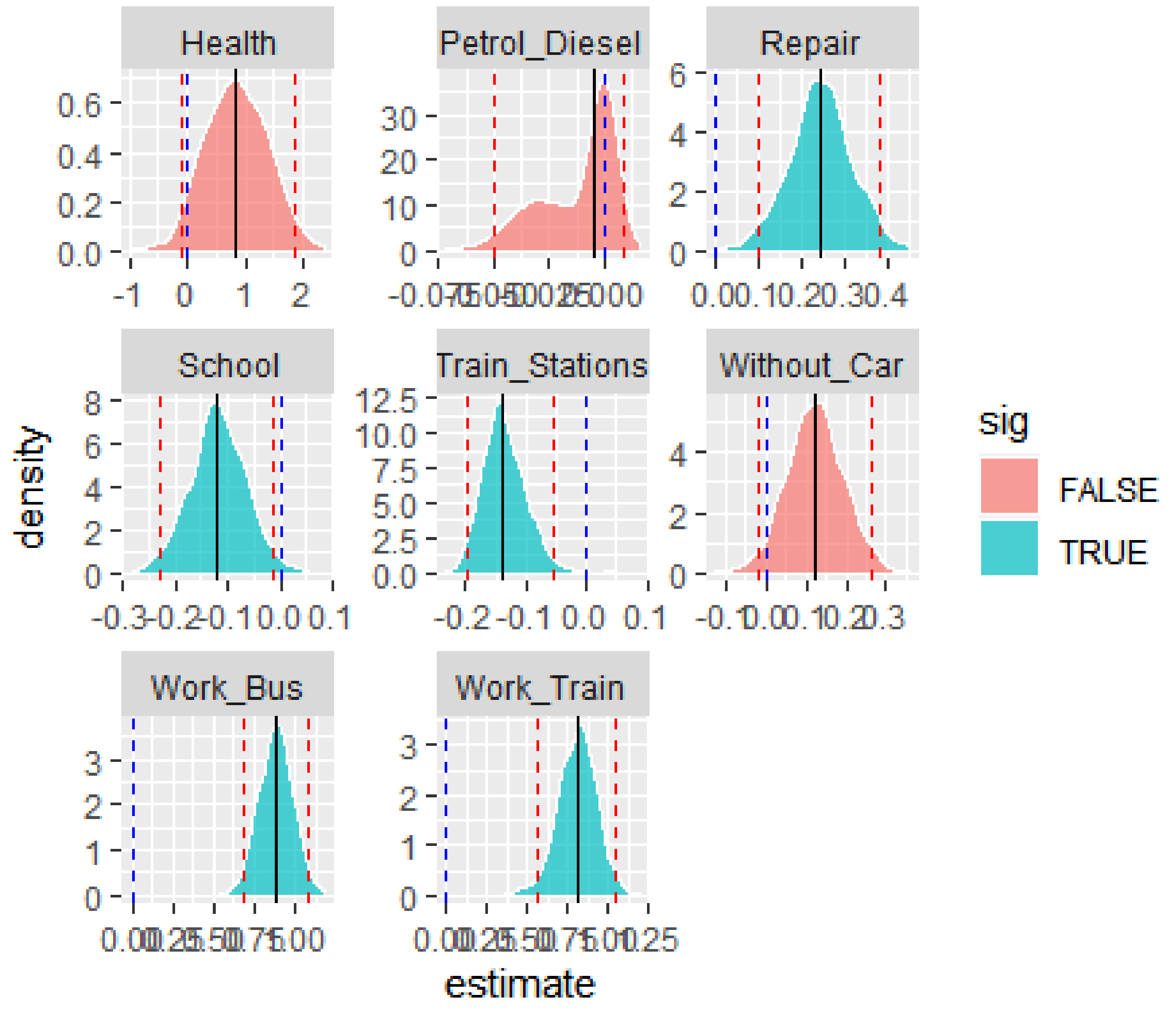
Model Selection

- A total of four regressions were performed. First, all the independent variables were regressed, and then the highly correlated factors were sequentially removed, and finally four models were obtained.
- AIC and adjusted R square method are applied as a certeria for model selection.
- Fit3 is the best choice.

	model 1	model 2	model 3	model 4
gfit.AIC	1271.288	1268.812	1266.846	1268.405
gfit.adjR2	0.547848	0.545727	0.552382	0.550709

Parameter estimation and distribution

Using Bootstrap to select significant variables at 95% level and obtain its confidence interval of parameter estimation and their observations.
This process should be repeated 1000 times.



Reference

- Jim Hester and Hadley Wickham, (2020). *fs: Cross-Platform File System Operations Based on 'libuv'*. R package version 1.5.0
- Kuhn et al., (2020). Tidymodels: a collection of packages for modeling and
- machine learning using tidyverse principles.
- Wickham et al., (2019). *Welcome to the tidyverse. Journal of Open Source Software,* e, 4(43), 1686