Urban Transport Masters Programme

Transport Planning Lab Assignment 2

Complete both questions and then submit your answers in a Word or PDF document through Turnitin. You should also submit your R or RMarkdown file or files showing your workings for both questions.

The deadline for submission is noon on February 27th 2023.

Question 1: Multiple linear regression

You should use multiple linear regression to examine the relationship between a dependent variable and at least two independent variables of your choice. You may use data from either the Puget Sound Household Travel Survey (https://household-travel-survey-psregcncl.hub.arcgis.com/) or the National Travel Survey for England (available from the UK Data Service).

Your R code must demonstrate how you obtained a random sample of 200 people using the datasets available online. In other words, you must not use any sample of data provided on Moodle. You may choose to limit your analysis to people with particular characteristics, for example employed people, but your remaining sample for analysis must contain 200 people.

You should provide responses for each of the following in the Word/PDF file:

a) State what the population is for your analysis and explain your choice.
b) Specify the dependent variable and independent variables you have chosen to examine and briefly justify your choices.
c) Write out the equation connecting your dependent variable with your independent variables and include the estimates of the coefficients from your model.
d) Using the relevant values from the model summary (including the appropriate value for R^2 and the relevant p values), explain how well the model fits the data and discuss whether all of the independent variables should be retained in the model.

e) Present suitable graphs to demonstrate whether the model assumptions hold in this case (for example using the check_model function). Discuss each graph and conclude whether or not all assumptions hold.
f) Based on your results from parts c) and d) only, explain the findings from your model to transport planners using non-statistical language. [**For part f) you should ignore your findings relating to the model fit and assume that these were satisfactory**]

Question 2: Logistic regression

You should run a binary logit model using variables of your choice. You may use data from either the Puget Sound Household Travel Survey (https://household-travel-survey-psregcncl.hub.arcgis.com/) or the National Travel Survey for England (available from the UK Data Service).

Your R code must demonstrate how you obtained a random sample of 200 people using the datasets available online. In other words, you must not use any sample of data provided on Moodle. You may choose to limit your analysis to people with particular characteristics, for example employed people, but your remaining sample for analysis must contain 200 people.

You should provide responses for each of the following in the Word/PDF file:

a) State what the population is for your analysis and explain your choice.
b) Specify the dependent variable and independent variables you have chosen to examine and briefly justify your choices.
c) Write out the equation connecting your dependent variable with your independent variables using the coefficients estimated by your model.
d) State and explain the odds-ratio for each of the independent variables.
e) Using the relevant values from the model summary, discuss the significance of each of the independent variables in the model.
f) Calculate a suitable pseudo-R^2 statistic for your model and explain the result.
g) Explain the findings from your logit model to transport planners using non-statistical language.