

Home Assignment 1: Multiple Linear Regression

due ~~Wednesday, 11th November 2020~~ (24hours) by email

due **Sunday, 22nd November 2020** (24hours) by email

1. Objectives

The objective of this HA is to improve your capabilities and ease of use of multiple linear regression models. In this case study, you have to estimate a multiple linear regression model that predicts Average number of trips during the day (per person) (“atrip”) of the Lisbon Metropolitan Area.

2. Software

The reference software is Excel or SPSS. However, you can use any other regression software of your choice that has enough functionality to complete the assigned tasks (e.g. R, SAS, STATA, etc.).

3. Data

You are given a set of aggregate data for 273 traffic analysis zones in the Lisbon Metropolitan Area (40 zones within the council of Lisbon and the other 243 are boroughs of the remaining 17 councils of the south and north bounds of the LMA) collected in a survey conducted in 2010. For each of the 273 zones you have available the information corresponding to the variables that are explained in detail in the appendix.

The data files (TDM_LinearRegression_HA_data_LMA.xlsx or

TDM_LinearRegression_HA_data_LMA.sav) are available in the course’s website.

4. Your tasks

1. Give a brief description of the cause/effect relationships you think are relevant in predicting “atrip” and formulate a hypothesis (a priori beliefs of the relationship). You may find it useful to do some preliminary statistical analysis on the data before this.
2. Estimate the model to predict the “atrip” to reflect the hypotheses.
3. Test different specifications (with any independent variable transformation you might find useful - although not forcedly) and perform statistical tests comparing the specifications.
4. Select the best model specification.

5. Report content

Your final report should include:

1. Description and discussion of the cause/effect relationships you think are relevant for “atrip”;
2. Description of no more than three model specifications that you believe reflect your a priori (i.e., part 1) considerations;
3. Presentation of your “best” specification (it is up to you to define “best” and indicate the statistical analysis that support your criteria for deciding) and a discussion of your selection criteria; and
4. A discussion of the similarities and differences between the causal inferences from your “best” specification and respective *a priori* considerations.

6. Some Comments

1. The following criteria will be applied for grading:
 - a. Your understanding of the problem (e.g. causal relationships);
 - b. Your understanding of linear regression analysis;
 - c. Your utilization of the regression software (evidence you estimated a regression model, etc.);
 - d. Your understanding of regression statistics and hypothesis testing (explain what the statistics mean).
2. Remember that you must always examine and comment your results. Computer outputs without explanations are not acceptable.
3. There are no formatting rules except that you should write a concise report that isn't longer than 15 pages without annexes (where you should include tables you might find useful to complete your report).

Appendix

The variables available for your specification are:

Variable	Description
DCBD	Distance to CBD (specified by the three main employment locations: Saldanha, Av. Columbano Bordalo Pinheiro and Parque das Nações) (km)
atrip	Average number of trips during the day (per person)
AvCar	Car or car and/or motorcycle daily availability
ATTD	Average Total distance travelled during the day per person (km)
ATTT	Average Total time spent travelling during the day (minutes)
Work	Percentage of interviews that referred Work as the main trip purpose within the trip chain
Shop	Percentage of interviews that referred Shopping as the main trip purpose within the trip chain
Family	Percentage of interviews that referred Family related purpose as the main trip purpose within the trip chain
Personal	Percentage of interviews that referred Personal reasons as the main trip purpose within the trip chain
AActT	Total number of hours spend in activity until the final trip of the day (hours)
aChildren	Average number of children in the households
ElderRetired	Number of elder or retired persons in the households
TCPass	Percentage of interviews holding a Public Transport Pass
AcessMetroResid	Distance to closest subway or suburban rail stations near the residence (meters)
ParkingOrigin	Indicator (between 0 and 1) that indicates the parking pressure close to the residential area. Higher values indicate that the availability of parking places near home is scarce.