Lab07: Regression Model Building

Handout date: Wednesday, October 14, 2020
Due date: Friday, October 23, 2020 in the link L̲ᴀʙ07S̲ᴜʙᴍɪᴛ in eLearning
                       This lab counts 4 % toward your total grade:
Objectives: You engage an empirical dataset to build an ***informed*** regression model to explain the variation of a dependent variable.

Format of answer: Your answers (graphs and verbal description) should be handed in as ***hard-copy*** in ***one*** document. Add a running title into the header of the document with the following information: ***your name***, ***Lab07*** and ***page numbers***. Format any code and computer output properly before inserting it into the document with your answer. Ⓡ-code and text output need to be in a ***monospaced*** font (i.e., fixed-pitch font) such as `Courier New` so proper spacing and alignments are preserved. Excessive, but irrelevant, output will lead to a deduction of your accumulated points.

This lab is based on the **`tractShp`** spatial data frame in the package **`TexMix`**. You can use the script **`Lab07StarterCode`** to set up the data.

Task 1: Model Specification (1point)

Select a set of <u>no more than 6 independent variables</u> which explain the variation in the median home values in each census tract, i.e., the dependent variable is **`MEDVALHOME`**.

<u>Caution:</u> Several census tracts have no reported median home value, i.e., **`NA`**s, and therefore need to be dropped from the analysis.

<u>Note:</u> Independent variables should be on relative scale rather than absolute values because the dependent variable is a regional average that is independent of the "size" of the census tract. (see the lecture notes **`Chapt04Relationships.pdf`** on page 6)

- Justify your choice of independent variables and formulate explicit hypotheses how and in which direction these independent variables are expected to influence the dependent variables.
- Explore the univariate and pairwise relationships of the variables. For variables with a strong positive skewness and all observations larger than zero consider a log-transformation.
- Include relevant analysis results, figures and maps as well as their interpretation

Task 2: Model Building (2 points)

Objective: Build a proper regression model and perform model diagnostics.

- Based on the information gained by the model diagnostics revise, if necessary, your model specification.
- Perform a Moran's scatterplot of the residuals of your final model.
- Interpret and document each relevant step of your analysis (skip broad trial analysis).
- Include important intermediate analysis steps into your report.
- Include relevant analysis results, figures and maps as well as their interpretation.

- In an appendix add the ℝ-code that you used for your report.

Task 3: Analysis Documentation (1 point)

Write a ***professional and properly formatted report*** which includes

- an objective statement in the introduction section,
- your hypothesis formulation and justification in task 1,
- your analysis steps with a justification how you proceeded and results in task 2, and
- a conclusions section summarizing the key findings of your study and an outlook into future studies.