

Predicting Treatment and Rehabilitation Costs

Week Thirteen (Part One of Three)

This is part one of a three part assignment, part one will be assigned this week, part two assigned in week fourteen, and the final part assigned for the final project. Grading for this assignment is as follows:

Part 1 and 2 = the “Big Data Project” in the syllabus

Final Project (Part 3) = 33% of the “Big Data Project” grade + 64% of Part 3’s assignment.

Information and Background

Your hospital is running a clinical drug trial for stroke patients and one critical data point to the overall success of the trial is the total treatment and rehabilitation costs for each patient. Your Data Analytics group had provided you with data from 2,421 stroke patients who are either actively or have previously participated in the clinical trial. No differentiation is necessary for active and past participants in the trial, the indicator for the control and experimental groups is: `rx_clottx = [0,1]`, respectively. You will be responsible for building a prediction model for treatment and rehabilitation costs for patients participating in the clinical trials.

Codebook

PT_ID	Patient ID	RX_CLOTTX	Clot-dissolving drugs prescribed in Tx
HOSPTYPE	Hospital Type	TRANSP_TIME	Time to hospital
PT_AGE	Age in years	RANKIN_INIT	Initial Rankin Score
PT_FEMALE	Female	CATSCAN	CAT scan result
PT_ACTIVE	Physically active	PE_SURG	Post-event preventative surgery
PT_OBESITY	Obesity	POST_SURG	Post-event preventative surgery
PT_SMOKE	Smoker	POST_MI	Post-event Myocardial infarction
HX_DIAB	History of diabetes	POST_IS	Post-event Ischemic stroke
HX_HYPERT	History of hypertension	POST_HS	Post-event Hemorrhagic stroke
HX_AF	History of atrial fibrillation	LOS_REHAB	Length of stay for rehabilitation
HX_CHOLS	History of high cholesterol	COST	Total treatment and rehabilitation costs in thousands
HX_ANGINA	History of angina		
HX_MI	History of myocardial infarction		
HX_IS	History of ischemic stroke		
HX_TIA	History of transient ischemic attack		
RX_NITRO	Prescribed nitroglycerin		
RX_ANTICLOT	Taking anti-clotting drugs		

Questions

1. Build a prediction model using multiple regression to predict the total treatment and rehabilitation costs for patients who were in the clinical trial ($rx_clottx=1$).
2. Write a short interpretation of the variables and the overall model for patients participating in the clinical trial.
3. Is there a difference in significant contributing factors to a prediction model for patients who are in the control group ($rx_clottx=0$)? If so, what are the factors and their p-values in the final model?
4. If you are able to build a model for the control group, write a short interpretation of the variables and the overall model for the patients in the control group.

Hints and Information

You will need information from this week and previous weeks to properly build and interpret the results of the regression model. I will provide feedback on your answers to **only questions 1 and 2** if you wish to submit a draft before Sunday, April 18th 2021, by 11:59 PM EST.

A link will be available in the Week Thirteen folder to submit your draft.

Resubmissions

This assignment **is not eligible** for resubmission after grading.