

# 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

<b>PT_ID</b>	Patient ID	<b>RX_CLOTTX</b>	Clot-dissolving drugs prescribed in Tx
<b>HOSPTYPE</b>	Hospital Type	<b>TRANSP_TIME</b>	Time to hospital
<b>PT_AGE</b>	Age in years	<b>RANKIN_INIT</b>	Initial Rankin Score
<b>PT_FEMALE</b>	Female	<b>CATSCAN</b>	CAT scan result
<b>PT_ACTIVE</b>	Physically active	<b>PE_SURG</b>	Post-event preventative surgery
<b>PT_OBESITY</b>	Obesity	<b>POST_SURG</b>	Post-event preventative surgery
<b>PT_SMOKE</b>	Smoker	<b>POST_MI</b>	Post-event Myocardial infarction
<b>HX_DIAB</b>	History of diabetes	<b>POST_IS</b>	Post-event Ischemic stroke
<b>HX_HYPERT</b>	History of hypertension	<b>POST_HS</b>	Post-event Hemorrhagic stroke
<b>HX_AF</b>	History of atrial fibrillation	<b>LOS_REHAB</b>	Length of stay for rehabilitation
<b>HX_CHOLS</b>	History of high cholesterol	<b>COST</b>	Total treatment and rehabilitation costs in thousands
<b>HX_ANGINA</b>	History of angina		
<b>HX_MI</b>	History of myocardial infarction		
<b>HX_IS</b>	History of ischemic stroke		
<b>HX_TIA</b>	History of transient ischemic attack		
<b>RX_NITRO</b>	Prescribed nitroglycerin		
<b>RX_ANTICLOT</b>	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 18<sup>th</sup> 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.