Analyze the data and apply a linear regression model

You are given a dataset which contains information about house prices in the suburbs of Chicago. Your task is first to analyze the data, and then to apply a regression model to it.

Data overview

The dataset consists of following variables:

Price price of the house

Bedroom. number of bedrooms

Space. size of the house (in square feet)

Room. number of rooms Lot width of a lot

Tax. amount of annual tax Bathroom. number of bathrooms

Garage. number of parking lots in the garage

Condition. condition of the house (1 if good, 0 otherwise)

The values in some of the columns may be missing, so you must handle this properly (e.g. by filtering out NA values from a given column before calculating any statistics or data frames that is dependent on it).

We want to describe the relationship between Price (which will be a dependent variable in the model) and all other variables (predictors) using a linear regression model.

To fit a model to the data, you can either use built-in functions or calculate the parameters of the model from scratch. If you choose the latter approach, here you will find all the equations you need to implement a least-squares method for calculating model parameters.

Task details

Write a function named analyse\_and\_fit\_1rm() which takes one argument (a path to a dataset) and returns a named list of the following objects (the order and names of the objects should be the same as below):

summary\_list- a named list of length 3 with the following elements:

• statistics - a numeric vector of length 5 specifying mean, standard deviation, median, minimum and maximum for a variable Tax for all houses with two bathrooms and four bedrooms (you don't need to name elements of the vector).

• data frame - a data frame with observations for which Space is bigger than 800, ordered by decreasing Price.

number\_of\_observations - a numeric value corresponding to the number of observations for which the value of a variable Lot is equal to or bigger than th 4th 5-quantile of this variable.

regression\_list-a named list of length 2 with the following elements:

o model\_parameters - a numeric vector of length 9 giving the model parameters. The first element of the vector should be named Intercept, and all other elements should have the same name as the respective variable.

o price\_prediction - a numeric value which corresponds to the prediction of the

price (using the applied model) for a house with the following specific

parameters: three bedrooms; 1500 square feet of space; eight rooms; width of lot is 40; $1000 tax; two bathrooms; one space in the garage; house is in bad condition.

Hints

Do not call analyse\_and\_fit\_lrm() function explicitly in your file. It will be automatically invoked with correct file\_path argument during the execution of unit tests.