

Tutorial 8 - Linear Probability Model

AUTHOR

Victoria Hünewaldt

Tutorial 8

This tutorial will cover the linear probability model (LPM) and the root mean squared error. You will learn:

- how to run & interpret a LPM
- the problems coming with a LPM
- how to calculate & interpret the in-sample root mean squared error

Exercises

Create a .qmd-file and solve the tasks there. Store it in the JupyterHub folder "Session 8".

Linear Probability Model [🔗](#)

Create a model explaining "high_rating" (= binary endogenous variable) and run a linear probability model.

You first have to generate the dummy variable "high_rating" which should be equal to 1 if `n_review_scores_rating > 94` and 0 otherwise.

Choose the exogenous variables that you think explain "high_rating" most in your model. What can be problematic about linear probability models?

Root mean squared error

Calculate the in-sample root mean squared error (RMSE). It measures the average difference between values predicted by a model and the observed values and provides an estimation of how well the model is able to predict the target value.