

R for Economics and Social Science Research

NE Milla, Jr.

2025-04-02

WORKSHOP 3

1. Filter the *Wage* data set to include only individuals with $\text{wage} < 200$ and $\text{age} < 60$. Then split it into training and testing sets.
2. Using the training set, fit linear regression models with *wage* (or *logwage*) as response variable and *age*, *marital status*, *education*, *job class*, *health condition*, and *health insurance* (*health_ins*) as possible predictors.
3. Check the model assumptions for each fitted model.
4. Use performance metrics such as Adj. R^2 , RMSE, AICc, BIC, etc) to identify top 3 models.
5. Evaluate the accuracy of the predictions obtained from the top three models using the test set.
6. Report and interpret the parameter estimates (and their significance) of the final model.