

COMPUTER PRACTICE 3

REGRESSION ANALYSIS

OBJECTIVES

- Obtain practical skills of regression analysis using Python libraries and self-developed scripts

GUIDELINES (CONSISTS FROM 14 POINTS)

1. Recall theoretical concepts of correlations and regression analysis.
2. Create Python Notebook for scripting and analysis running. You can use Jupyter Notebook or Google Colab as a work environment.
3. Load the data set provided for the Computer Practice 3 in the course and read through its description. Identify the dependent variable that is primary point of interest in this work and should be predicted by created model.
4. Run the exploratory analysis and data visualisation on the base of code used for the Computer Practice 1.
5. Calculate data sample correlation:
 - a. Report variables, significantly correlated with dependent variable, and specify the correlation direction
 - b. Report several most significant correlations between other variables.
6. Create scatter plots between dependent variable and list of continues independent variables (predictors). Make preliminary conclusions on the relationships.
7. Create the dummy variables for categorical variables
8. Estimate the linear regression model. Report:
 - a. Model goodness-of-fit
 - b. Significance of model coefficients
 - c. Direction of significant variables
 - d. Effect of variables with significant relationships
9. Construct histogram of residuals. Provide your explanations for the histogram form.
10. Create scatter plot of residuals and make a decision regarding the outliers and need in their exclusion from data sample. If you consider the need to exclude the outliers, please do it and create the new regression model. Compare the new model with previous one.
11. Conduct stepwise elimination of insignificant independent variables. Describe the resulting model.
12. Apply the resulting model for prediction and save the predicted values. Create the confidence interval of prediction. Review the results of prediction and report your conclusion.
13. Summarise all your findings and compose the report.

14. Upload your report (.pdf) and Python Notebook (.ipynb) with all your scripts used for analysis and generated results to the learning environment.