Course: Data Analysis

Student’s Name and Surname \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Regression analysis**

*Describe the model:*

1. Specify the regression equation.



1. Assess the goodness-of-fit of the model (R-square, significance of the model);



The regression model is statistically significant. The model explains 54,9% of price’s variation

1. Which independent variables significantly influences the dependent variable (explain your answer)?



3 out of 5 predictors significantly correlate with the predictor. Mpg and trunk are non-significant predictors.

1. Describe the relationship between each independent and the dependent variable.

There is a positive significant relationship between weight and price. One unit increase in weight will increase the price by 5.7.

There is no linear relationship between mpg and price.

The price of the foreign car is significantly higher than the price of the non-foreign car. The average difference is 3557.1

*Do the diagnostics of the model:*

1. Are the residuals normally distributed?



H1: the distribution of residuals is significantly different form the normal distribution

p-value < 0.05 => we accept H1 =>residuals are not normally distributed

1. Are there any outliers (standardized residuals greater than 3, or less than -3)?

We have no outliers. All the standardized residuals are < 3 or > -3.

We have no influential cases. All the dfbetas are <1 and >-1.

7. Test the multicollinearity and heteroscedasticity.

There is a multicollinearity. Correlation between length and weight is 0.94.

There is a heteroscedasticity – relationship between residuals and the fitted values of the price.

Please send this MS Word file with answers and the do-file to amelikyan@hse.ru from your personal e-mail.