Course: Data Analysis

Student’s Name and Surname Pavel Drank

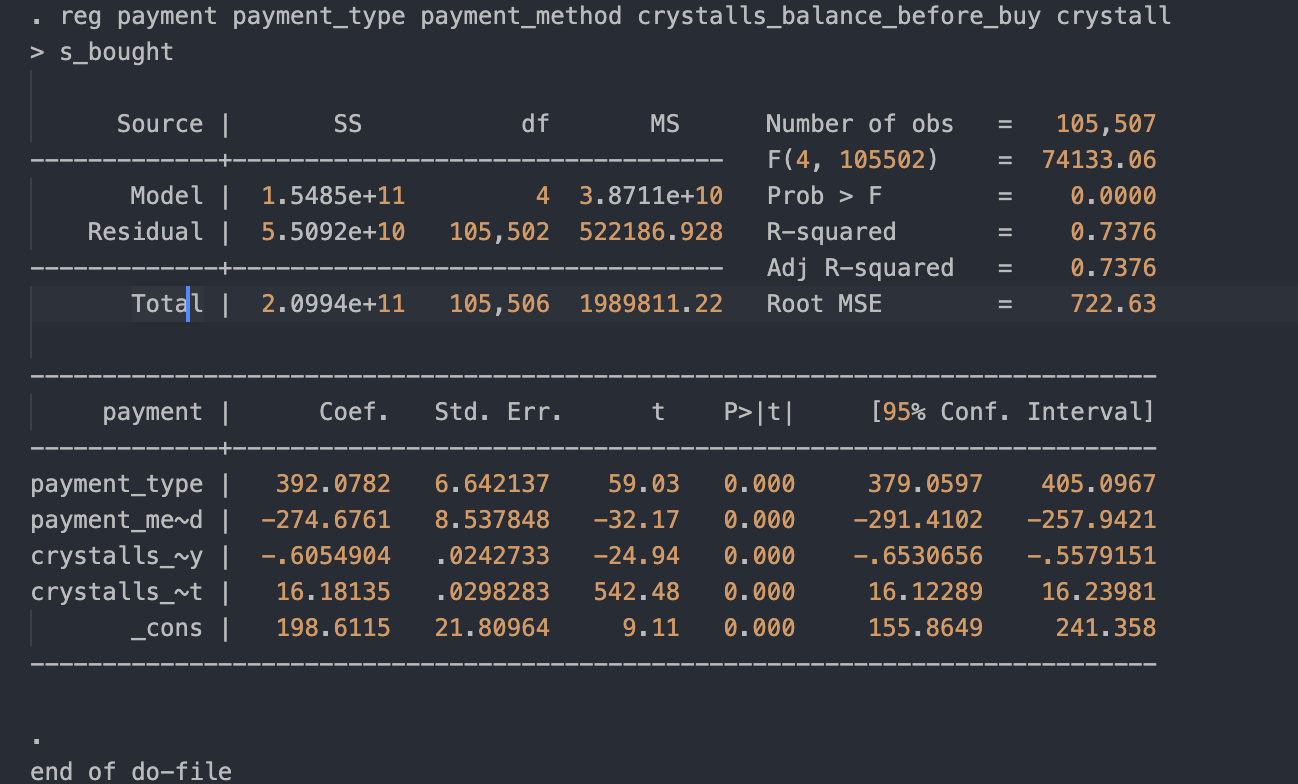
**Regression analysis**

Do the multiple regression analysis using the following variables from **data\_games.dta**:

Dependent variable: payment

Independent variables: payment\_type, payment\_method, crystalls\_balance\_before\_buy, crystalls\_bought

*Describe the model:*

1. Specify the regression equation.

payment = 392.0782 \* payment\_type - 274.6761 \* payment\_method - 0.6054904 \* crystalls\_balance\_before\_buy 16.18135 \* crystall + 198.6115

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

R-square is 0.73 which is a decent fit.

3. Which independent variables significantly influence the dependent variable (explain your answer)?

payment\_type’s and payment\_method’s influence is significant, because of high coef.

4 is. Describe the relationship between each independent and the dependent variable.

*Do the diagnostics of the model:*

*payment\_type: regular payments are used to be higher in sum then offer payments*

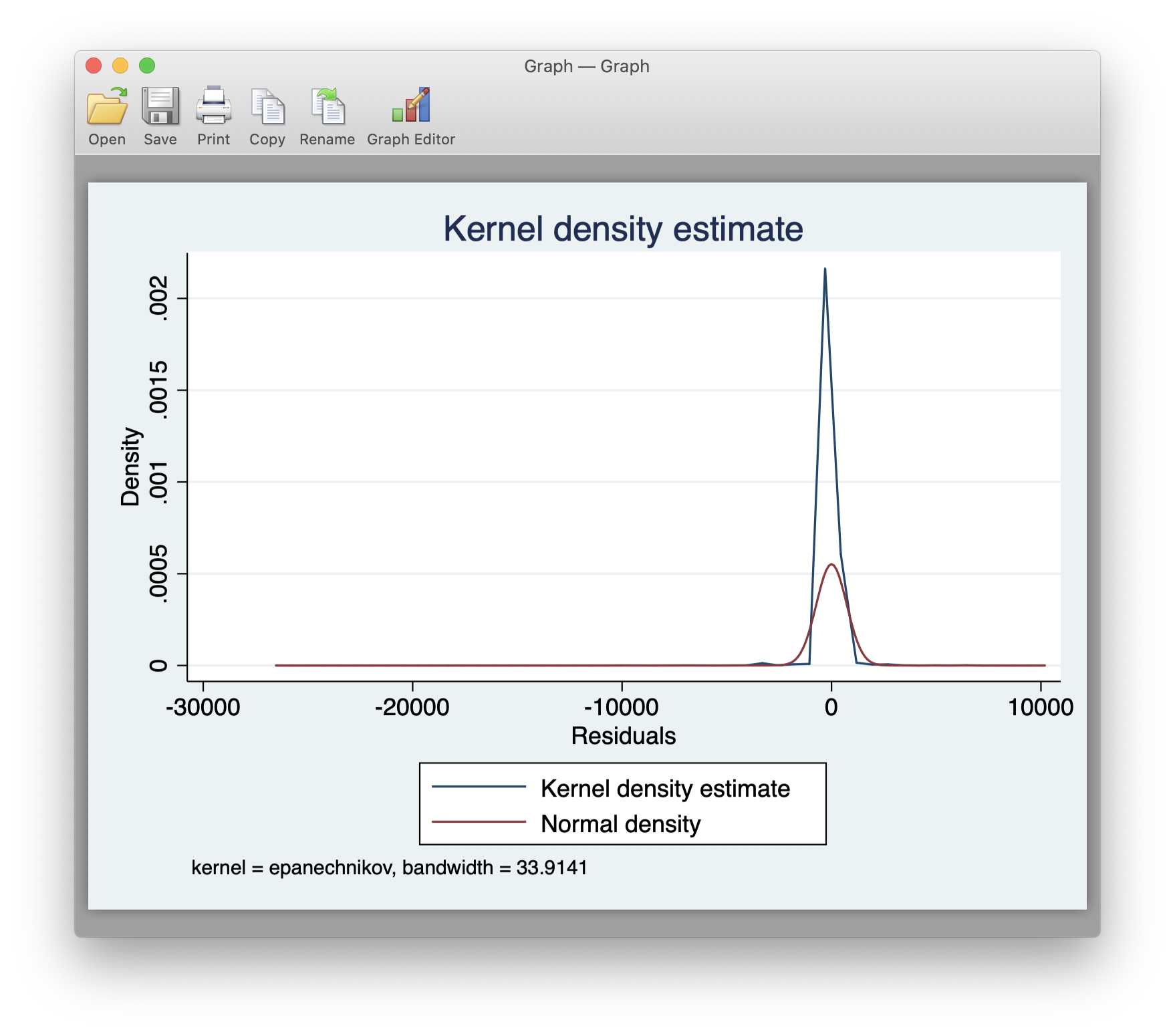
*payment\_method: mobile payment type a way more less then general and promotion methods*

*crystalls\_balance\_before\_buy: the more crystals a user have, the less resulted payment*

*Crystall: the more crystals, the higher payment*

5. Are the residuals normally distributed?

kdensity r, normal

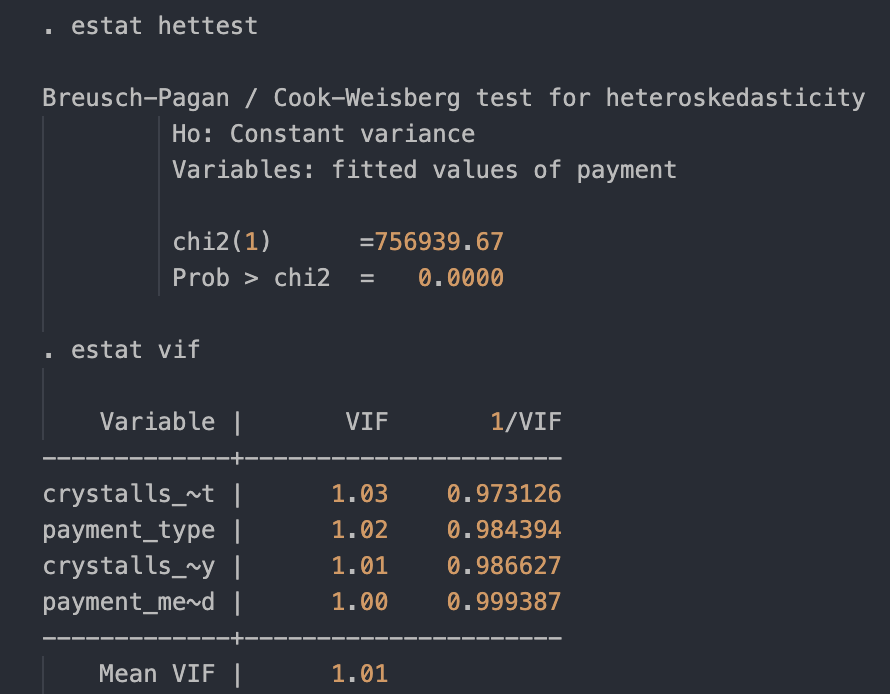


Yes, they are

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

Yes, look at picture in the prev. subtask.

7. Test the multicollinearity and heteroscedasticity.

There is no multicollinearity and heteroscedasticity:

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