Course: Data Analysis (task № 7)

Student’s Name and Surname Pavel Drank

Please do the tasks below.

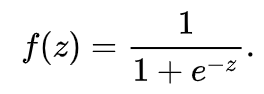
Open **Warranty.dta** file.

The aim of analysis is to find out what circumstances encourage customers to purchase extended warranties after a major appliance purchase. The response variable is an indicator of whether or not a warranty is purchased (Bought). The predictor variables are:

* Customer’s gender (Gender)
* Customer’s age (Age)
* Whether a gift is offered with the warranty or not (Gift)
* Price of the appliance (Price100)
* Customer’s race (Race)

Use the binary logistic regression to analyze the data.

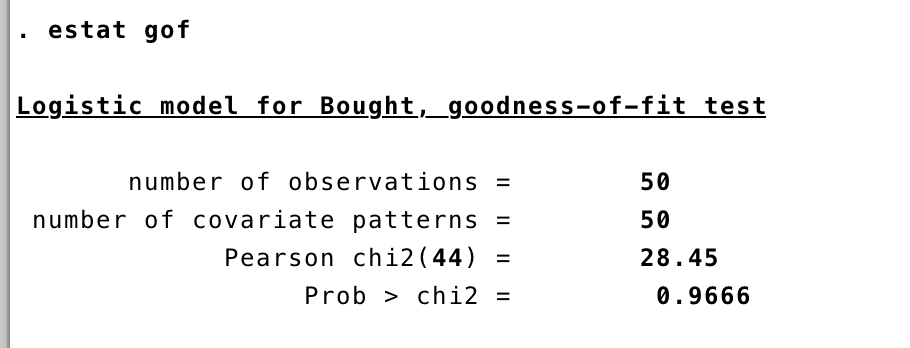
1. Specify the regression equation (the linear part of the formula).



Z = -2.20201 \* Gender 2.848165 \* Gift 0.0926411 \* Age 0.5782489 \* Race 0.0550196 \* Price100 -8.442717

2) Assess the goodness-of-fit of the model and interpret the results of the analysis.

Pseudo R^2=0.6637, which is higher than 0.5. It means that the model fit’s well.

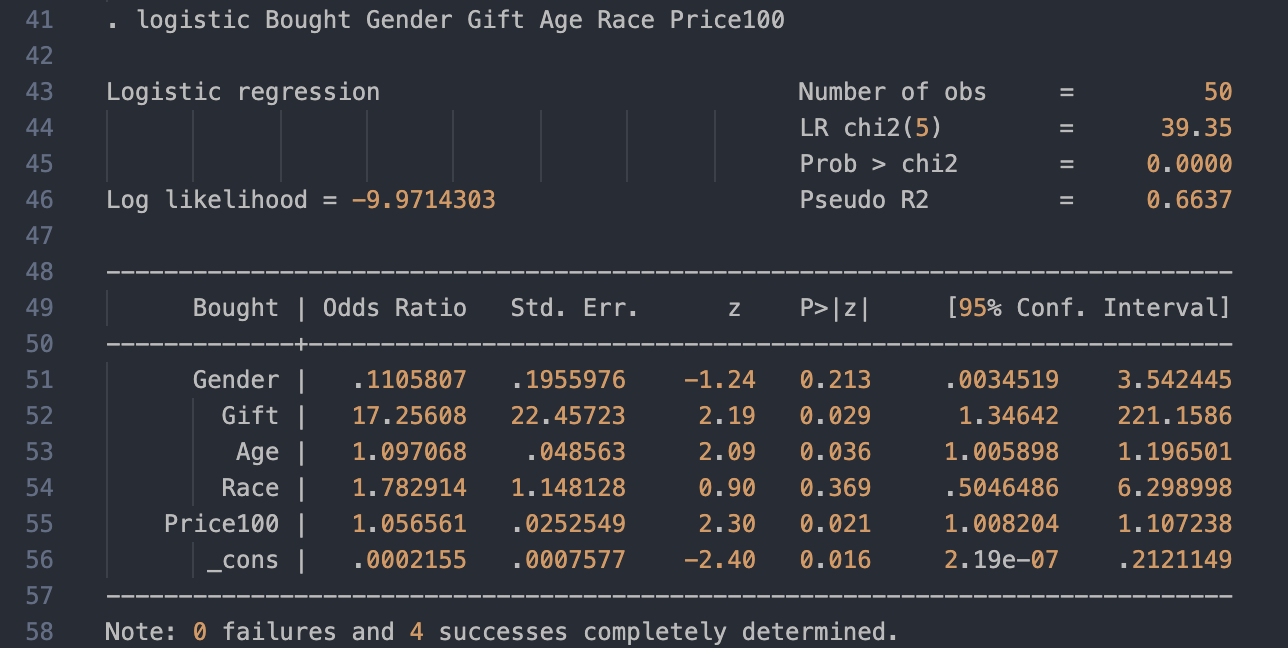


The goodness-of-fit test is positive. There is no significant difference between real and predicted values pf dependent variable.

H0: there is no difference between real and predicted values

H1: there is deference

Prob = 0.9666 > 0.05 => reject H1

3) Interpret the influence of any predictor variable on the dependent variable using Exp(b).

Gender: males purchase warranty 89% times less often

Gift: a gift increases a chance to purchase a warranty by 17 times

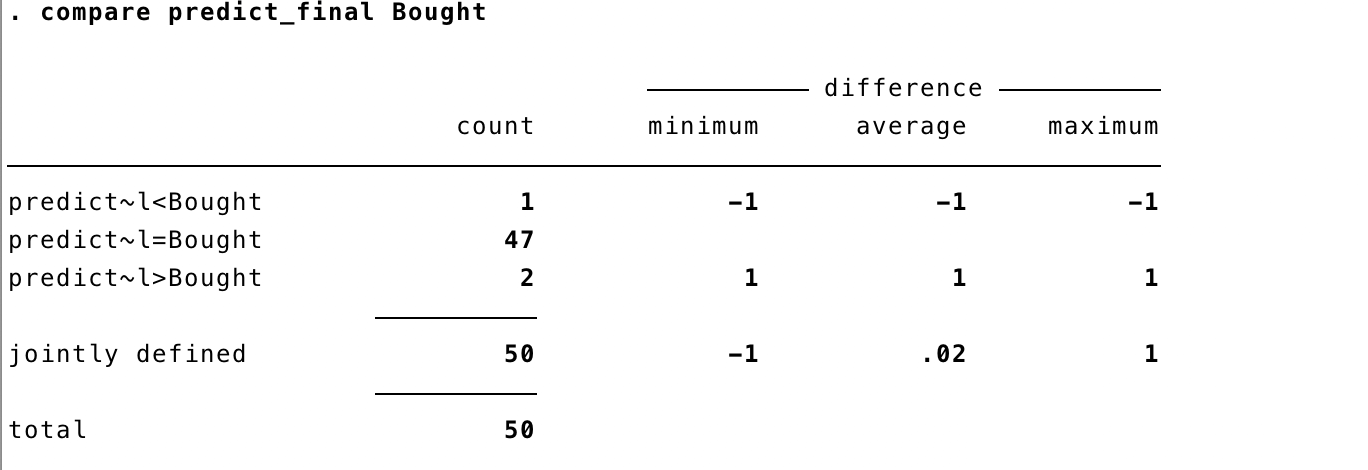
Age: any additional year of age increases warranty purchase probability by 9%

Race: African American people purchase warranty 78% more often, Hispanic 216%(1,78 \* 1.78 = 3.16) more often, Other 463%(1,78 \* 1.78 \* 1.78= 5,63) more often.

Price: Each unit of price increase warranty purchase change by 5.6%

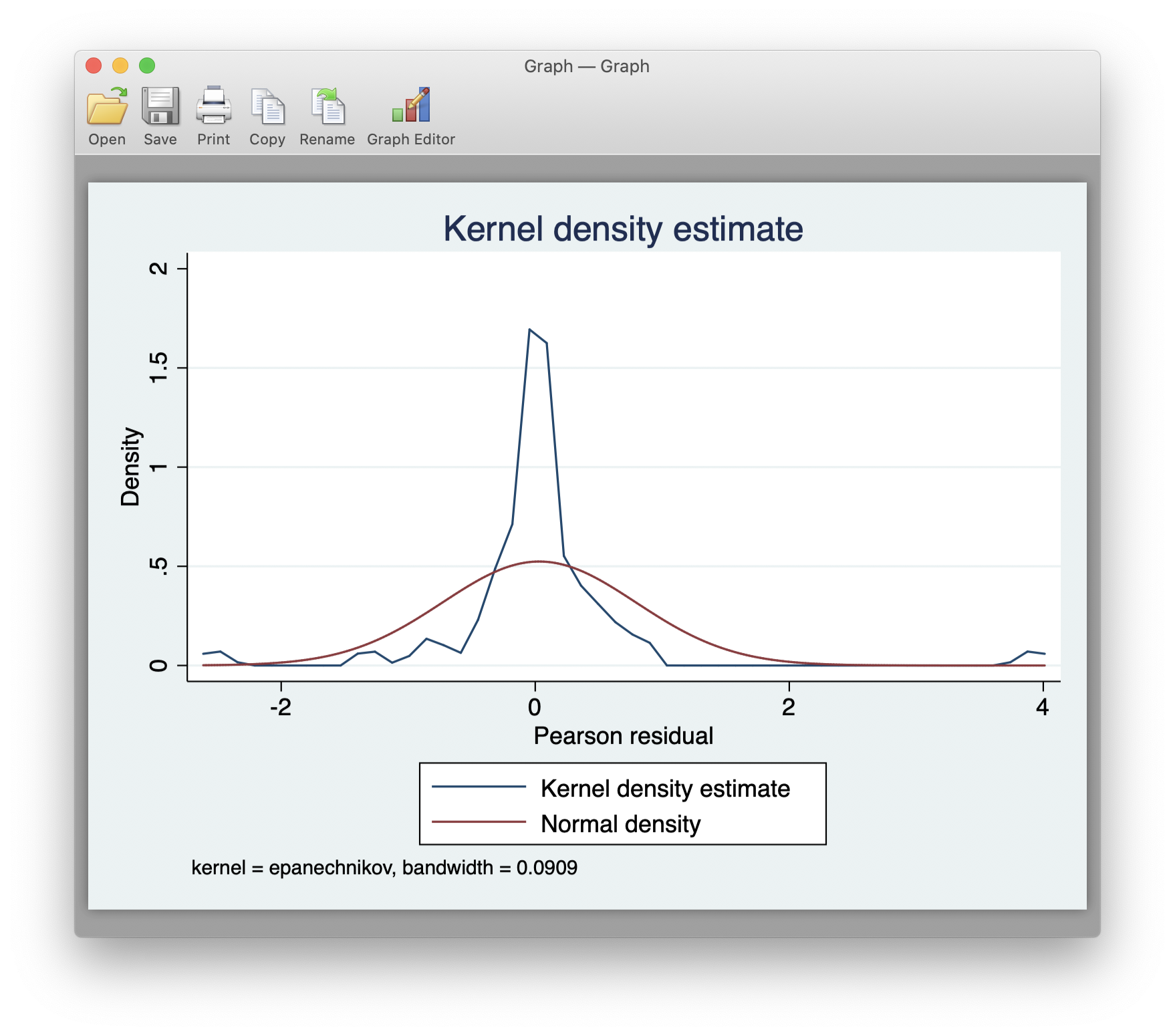
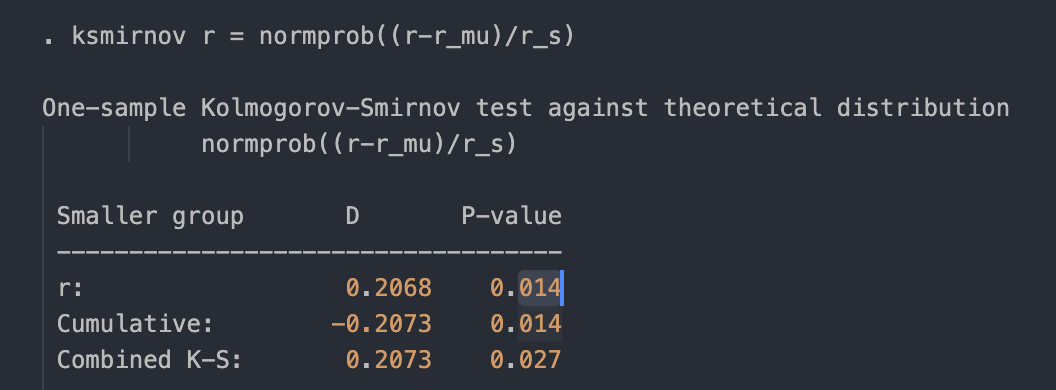
4) Which gradients are statistically significant?

Price100(0.05>0.021), Age(0.05>0.036), Gift(0.05>0.029)

5) What is the percentage of correctly predicted cases by the model?

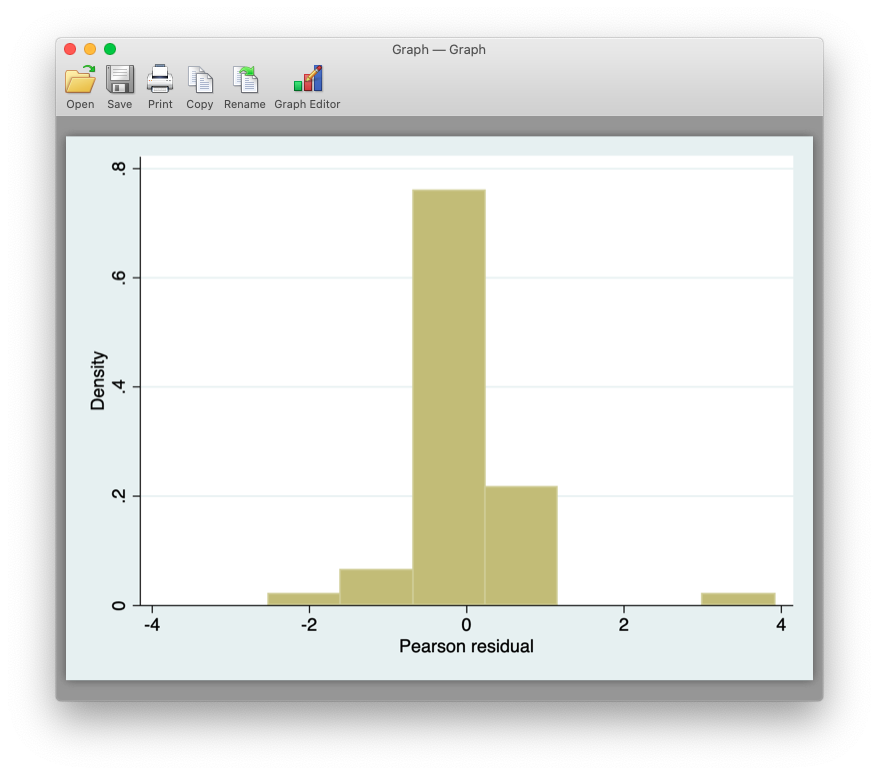
In 47 of 50 cases(94%), model was right.

6) Do the diagnostics of the model.

* Are the residuals normally distributed?

According to Kolmogorov-Smirnov test it is 98,6% that the residuals are not normally distributed.

* Are there any outliers? If yes, how many? Yes, only one:



* + Test the multicollinearity.