Question 4. [30 marks] The researcher is studying the effect of professional trainings on the increase of sales in chain of kitchen stores in Moscow and Moscow region. The target is the number of contracts (variable S) sold by a particular seller during the year. The experience of seller, measured by the full number of years worked in that retail chain (variable E from 0 to 5), and the results of the IQ test , which each seller passes on hiring (variable IQ, the average value for all categories of sellers was equal to about 100), are considered as independent variables. Seller's participation in professional trainings is characterized by a dummy variable T (equal to one for trained sellers and zero for other sellers).

(a) [8 marks] Based on the data characterizing 2000 sellers, the researcher calculates the following equations

$$\hat{S}_i = -20.4 + 5.2E_i + 1.8IQ_i \quad RSS^{*} = 2735345 \qquad (1)$$

$$\hat{S}_i = -22.8 + 5.3E_i + 1.7IQ_i + 19.0T_i \quad RSS = 2579858 \qquad (2)$$

$$\hat{S}_i = -18.1 + 4.8E_i + 1.7IQ_i + 14.3T_i + 1.6E_i * T_i + \cancel{0.1IQ_i * T_i} \quad RSS = 2577011 \qquad (3)$$

$$H_0: \beta_T = 0$$

$$F = \frac{(RSS_R - RSS_U)/1}{RSS_{UR}/2000-4} = 120$$

(b) [7 marks] In addition, the researcher decided to estimate regressions in the specification of equation (1) separately for sellers who did not participate in the training (Equation 4, 1373 observations) and for sellers who received professional training (Equation 5, 627 observations):

$$\hat{S}_i = -18.1 + 4.8E_i + 1.7IQ_i \quad RSS_1 = 1622106 \qquad (4)$$

$$\hat{S}_i = -14.4 + 6.4E_i + 1.8IQ_i \quad RSS_2 = 954905 \qquad (5)$$

$$F = \frac{(RSS^{*} - (RSS_1 + RSS_2))/3}{(RSS_1 + RSS_2)/(2000-6)} = 40{,}8$$

(c) [8 marks] The researcher decided to additionally take into account the factor M of the store location (M equal to one for sellers working in Moscow stores and zero for the region), for which he estimated the following equations.

$$\hat{S}_i = -24.3 + 5.3E_i + 1.7IQ_i + 18.9T_i + 3.3M_i \quad RSS = 2574433 \qquad (6)$$

$$\hat{S}_i = -23.2 + 5.3E_i + 1.7IQ_i + 16.5T_i + 1.9M_i + 4.6T_i * M_{ii} \quad RSS = 2572170 \qquad (7)$$

Question 5. [30 marks] The researcher studies the factors that affect the volume of paid services per capita V_i in 82 regions of Russia (in rubles). He suggests that this indicator may depend primarily on average per capita monthly income in rubles I_i (from 14000 to 70000 rubles depending on the region), as well as on the level of unemployment in percent U_i for each region. In addition, the researcher suggests that the situation with paid services in the central (near Moscow) and northwestern regions of Russia (near the city of St. Petersburg) may differ from the rest of the country, so he introduces a dummy variable R_i equal to 1 for central and northwestern regions, and equal to 0 for other regions of Russia.

(a) [7 marks] To assess the impact of income on paid services, the researcher first runs a simple linear regression

$$\hat{V}_i = -2448.2 + 2.05 I_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2 I_i) \qquad R^2 = 0.78 \qquad (1)$$
$$(3546.8) \quad (0.12)$$

$$\frac{V}{I} = \beta_1 \cdot \frac{1}{I} + \beta_2 + \frac{\varepsilon_i}{I_i} \sim N(0, \sigma_\varepsilon^2)$$

□ The researcher then rank all regions in order of increasing per capita income, and then regresses first for the 20 regions with the lowest income (getting RSS_1 value equal $4.81 \cdot 10^8$), and then for the 30 regions with the highest income (getting RSS_2 value equal $5.87 \cdot 10^9$). How can this information be used to check the data for heteroscedasticity? Carry out the necessary calculations, explaining your actions, and make the conclusion.

$s.e(\varepsilon_i)$ prop I

$Var(\varepsilon_i) = \sigma_\varepsilon^2 \cdot I_i^2$

WLS with I

$$GQ = \frac{RSS_2 / 30-2}{RSS_1 / 20-2}$$

Question 1. (17 marks)

A researcher, using a sample of 626 individuals from the National Longitudinal Survey of Youth, is investigating how the probability of a respondent obtaining a bachelor's degree from a four-year college is related to the respondent's score on a test of cognitive ability, $SCORE_i$ ranged from 0 to 100 (22 to 65 in the sample, with mean value 50.2, and most scores lying in the range 40 to 60). $ETHBLACK_i = 1$ and $ETHHISP_i = 1$ for the ethnic black and ethnic hispanic persons correspondingly and zero for the others.

26.7 percent of the respondents earned bachelor's degrees. Defining a variable BA_i to be equal to one if the respondent has a bachelor's degree (or higher degree) and zero otherwise, the researcher fitted the OLS regression (1) (standard errors in parentheses): and logit regressions (2) и (3) (asymptotic standard errors in parentheses):

$$\hat{BA}_i = -0.54 + 0.015 \cdot SCORE_i \qquad R^2 = 0.10 \tag{1}$$
$$\quad (0.09) \quad (0.0018)$$

Z_i

$$\hat{BA}_i = -8.54 + 0.13 \cdot SCORE_i \quad McFadden \;\; R^2 = 0.13 \tag{2}$$
$$\quad (0.99) \quad (0.018) \qquad\qquad LR\ stat. = 72.49$$

Z_i

$$\hat{BA}_i = -9.05 + 0.14 \cdot SCORE_i + 1.26 \cdot ETHBLACK_i - 0.34 \cdot ETHHISP_i \quad McFadden \;\; R^2 = 0.14 \tag{3}$$
$$\quad (1.09) \quad (0.019) \qquad (0.45) \qquad\qquad (0.66) \qquad\qquad LR\ stat. = 80.24$$

(a) □ Evaluate marginal effect of $ETHBLACK_i$ for the ethnic black person with $SCORE_i = 70$? (Use the method of direct comparison of two probabilities, explaining why it is most sutable here and presenting and explaining all steps of your calculations).

$$\frac{\partial \Lambda(z_i)}{\partial\, score}$$

$$\Delta = p(B_i = 1 \mid score = 70, Black = 1, Hisp = 0)$$
$$- \; p(B_i = 1 \mid 70, 0, 0)$$