You have 40 minutes to complete the test. Please explain each step of your derivations and state all the assumptions employed. Note that different problems can give you different points. Maximum for the test is 10 points.

**Problem 1**

You are given the following data on 2,000 respondents:

* hourly earnings (Y);
* educational attainment (years) (EDUC);
* total expenditure in year (TE);
* value of the respondent’s house (H);
* mother’s and father’s educational attainment (MEDUC and FEDUC);
* Weight (W);
* Sex (S);
* whether the main job was in the government sector or the private (G).

As a policy analyst, you are asked to investigate whether there is a difference in estimated impact of education on earnings for different genders.

Write one equation you can estimate to test this hypothesis. Explain why you chose these variables, type of model (nonlinear or linear). [1.5 points]

Tell whether a Chow test can be used to test this hypothesis. If not, explain why. If yes, show how you would perform this test. [1 point]

**Problem 2**

An econometrician estimated two models with 100 observations (s.e. in brackets):

Assuming the disturbance term has a standard normal distribution, calculate the 95 per cent confidence intervals for coefficients estimates for and . Write an interpretation of one of the intervals. [1 point]

*where D is dummy variable*

Give interpretation of the coefficients estimates for **both** models. [2 points]

**Problem 3**

An econometrician gained some data from a university and estimated a model based on 300 observations:

where *GPA* is the average grade of a student, *CLASS* is the percentage of attended classes, is a disturbance term.

The econometrician thinks that talent also must influence the average grade a student gets, however, this factor is impossible to observe and measure. Thus, the researcher supposes that there is an endogeneity problem since the effect of talent goes to the disturbance term and talent might correlate with the percentage of attended classes.

What will happen with the coefficients estimates in case of the endogeneity problem? [0.5 points]

The researcher discovered that the university could provide him with the data on the distance from a student's house to the university. Explain why this distance factor can be used as an instrumental variable. [1 point]

**Problem 4**

An econometrician estimated the following model with 100 observations:

Suspecting that the regression was subject to heteroskedasticity where , the researcher prepared the following data:

|  |  |
| --- | --- |
| RSS from a regression based on the 50 observations with the smallest values of Z | 200 |
| RSS from a regression based on the 50 observations with the largest values of Z | 250 |
| R2 from auxiliary regression (squared errors from basic regression –dependent variable and X, X2, Z, Z2, X\*Z - independent variables) | 0,3 |

Perform the Goldfeld–Quandt test and White test and state your conclusions. [2 points]

How do you deal with heteroskedasticity? Which methods do you know? Write at least two methods [1 point]