1. John estimates a regression of a car's price in USD $PriceUSD_i$ on its maximum speed in mph $SpeedMPH_i$ and dummy variable $TransmissionAUTO_i$ that is 1 for cars with automatic transmission and 0 for cars with manual transmission. John's regression is:

$$\widehat{PriceUSD}_i = 10000 + 90SpeedMPH_i + 2700TransmissionAUTO_i$$

Ivan uses the same data to estimate the price of a car in roubles (for simplicity you can use the exchange rate 1 USD = 90 RUB) on maximum speed in kph (1 mile = 1.6 km) and dummy that is 1 for cars with manual transmission.

   (a) (1 point each) Find the coefficients in Ivan's model

   (b) (2 points) Compare the R^2 in John's and Ivan's regressions.

2. Vasya estimates a regression of ice cream price $Price_i$ on temperature outside $Temp_i$ and dummy variable $Weekends_i$ that is 1 for weekends and 0 for weekdays (Vasya uses daily data). From previous research, he knows that variance of errors in constant for weekdays $Var(\varepsilon_i) = \sigma^2$ and depends on temperature for weekends $Var(\varepsilon_i) = \sigma^2 \cdot Temp_i$.

   (a) (2 points) What will be the consequences of such error behaviour for $\hat{\beta}$ and $\hat{V}(\hat{\beta})$ estimators?

   (b) (3 points) What transformations with data Vasya needs to do to get BLUE estimators of $\hat{\beta}$?

   (c) (2 points) Vasya is lazy and decided to delete weekend observations from his sample instead of transforming his data. What problems with the estimation will he encounter?

3. Vasya studies how pets influence the probability of depression. Using data on 100 different cities all over the world, he estimates a regression of the share of citizens with symptoms of depression $Depr_i$ on the number of pets per capita $Pets_i$.

$$Depr_i = \beta_0 + \beta_1 Pets_i + \varepsilon_i$$

Unfortunately, the statistics about pets is far from perfect and the $Pets_i$ variable is measured with error: $Pets_i = Pets_i * + \nu_i$, where $Pets_i*$ is the real number of pets per capita, ν_i is measurement error. It is known, that ε_i and ν_i are independent, $Var(\varepsilon) = 10$, $Var(\nu) = 30$.

   (a) (2 points) Compare the $plim\hat{\beta}_1$ in Vasya's regression with the real β_1 value

   (b) (1 points) Explain how you can find a consistent estimator of β_1

   (c) (3 points) Describe what additional data you will need to obtain a consistent estimator of β_1. Describe the properties of the new variables(-s) you need and provide an example of such variable(-s).

4. During the 3024 year election campaign $\sqrt{2}$-Television Network actively supported Brain Slug Party. Dr. Zoidberg analyzes the influence of television on election results and considers the following simultaneous equations model:

$$\begin{gathered} y_i = \alpha_1 + \alpha_2 \cdot x_i + \varepsilon_i \quad (1); \\ x_i = \beta_1 + \beta_2 \cdot y_i + u_i \quad (2) \end{gathered}$$

where X_i is the number of people who watch this channel in the i-th region, Y_i is the number of votes received in this region by the Brain Slug Party, ε_i and u_i are identically and independently distributed disturbance terms with zero means. The sample consists of n observations (X_i, Y_i).

(a) (1 point) Explain why OLS estimates of the coefficients in a regression of Y on X would be invalid?

(b) (2 points) Show that OLS estimator $\hat{\alpha}_2^{\text{OLS}}$ of α_2 is inconsistent.

Now an instrument is introduced into the equation system - Z_i variable, which is equal to one if the $\sqrt{2}$-Television Network was broadcast in the i-th region in 3004 (long before the formation of the Brain Slug Party) and is equal to zero otherwise.

$$\begin{gathered} y_i = \alpha_1 + \alpha_2 \cdot x_i + \varepsilon_i \quad (1); \\ x_i = \beta_1 + \beta_2 \cdot y_i + \beta_3 \cdot z_i + u_i \quad (2) \end{gathered}$$

(c) (1 point) Why the chosen instrument is valid?

(d) (2 points) Show that the instrumental variable estimator $\hat{\alpha}_2^{IV}$ based on the instrument Z is consistent.

(e) (1 point) The researcher decides to use a two-stage least squares (TSLS) approach. First he fits OLS regression

$$\hat{x}_i = h_1 + h_2 z_i$$

saves the fitted values, and uses them as an instrument for X in equation (2).

Demonstrate that obtained TSLS estimator $\hat{\alpha}_2^{TSLS}$ is the same as $\hat{\alpha}_2^{N}$.

5. Consider the regression equation $y_i = \beta + \varepsilon_i, i = 1, \ldots, n$. Let the regression errors satisfy the following conditions:

$$E(\varepsilon_i) = 0, E(\varepsilon_i \varepsilon_j) = 0 \text{ for } i \neq j, E(\varepsilon_i^2) = \sigma^2 \cdot x_i, x_i > 0.$$

(a) Recall from ordinary least squares estimation that previously you calculated this variance under homoscedasticity conditions, but now you need to do it under heteroscedasticity conditions).

(b) Find the weighted least squares estimate. Show that it is unbiased. Calculate its variance.

(c) Compare the variances of the estimates obtained in parts (a) and (b). Interpret the result.