HSE and University of London Double Degree Programme in Data Science and Business Analytics

Statistical Methods for Market Researh, 2023-2024

Classteacher: Ksenia Kasianova.

Class 5: Diff-in-Diff

Task 1

Prove

$$\hat{\beta}_2 = \bar{Y}_1 - \bar{Y}_0$$

by deriving formulas for least squares estimates of both coefficients in the equation $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 \cdot D_i$, where D_i- binary variable.

Task 2

Let $\alpha-$ in the conditions of the previous task be the proportion of observations belonging to the test group, and $(1-\alpha)-$ is, respectively, the proportion of observations belonging to the control group. Assuming that the variance of the random error is the same for all observations and is equal to σ^2, calculate the conditional variance of the OLS estimate of the coefficient for the variable var $\left(\hat{\beta}_2 \mid D_1, D_2, \ldots, D_n\right)$ (express it in terms of σ^2, α, n). Which should there be a proportion of observations belonging to the test group in the total number of observations for the OLS estimator to be the most accurate?

Task 3

Prove that in model

$$Y_{it} = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot z_t + \delta \cdot x_i \cdot z_t + \varepsilon_{it},$$

the OLS estimate of the coefficient for the product $x_i \cdot z_t$ is equal to:

$$\hat{\delta} = \left[\bar{Y}_{\text{treatment, after}} - \bar{Y}_{\text{treatment, before}}\right] - \left[\bar{Y}_{\text{control, after}} - \bar{Y}_{\text{control, before}}\right].$$

Task 4

Prove that in model

$$\Delta Y_i = \beta_2 + \delta \cdot x_i + u_i,$$

the OLS estimate of the coefficient for the regressor is equal to:

$$\hat{\delta} = \left[\bar{Y}_{\text{treatment, after}} - \bar{Y}_{\text{treatment, before}}\right] - \left[\bar{Y}_{\text{control, after}} - \bar{Y}_{\text{control, before}}\right].$$

Task 5

In 2013, no minimum wage was established in any of the regions of the Kingdom of Westeros. In 2014, four regions of the kingdom (Winterfell, Highgarden, Riverrun and Pyke) passed minimum wage laws, while the remaining regions still had no restrictions on salary levels.

Economic theory suggests that minimum wages can increase unemployment, but researchers note that the average unemployment rate for regions that implemented minimum wage laws fell in 2014 compared to 2013. For simplicity of calculations, we assume that the size of the economically active population in all regions under consideration is the same. The dynamics of unemployment in the regions of Westeros is shown in the table.

Region	Unemployment rate, %	
	2013	**2014**
Winterfell	10	8
Highgarden	10	8
Riverrun	9	9
Pike	11	7
King's Landing	11	7
Dorn	12	7
Eagle's Nest	13	7
Staromest	13	8
Lannisport	11	9

a. Is the introduction of a minimum wage really causing unemployment to rise? To answer this question, estimate the effect of the minimum wage using available data and the difference-in-differences method. Interpret the result. Provide a graphical illustration of the solution.

b. Find OLS estimates of the parameters of the regression equation:

$$Y_{it} = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot z_t + \delta \cdot x_i \cdot z_t + \varepsilon_{it},$$

where $\boldsymbol{Y}_{it}-$ the unemployment rate in region i in year t; x_i- is a binary variable that is equal to one for regions that have introduced a minimum wage; z_t- is a binary variable that is equal to one for all observations relating to the second period (2014). Compare your results with the results of part (a).

c. Let the robust standard errors of coefficient estimates also be known:

$$\text{se}\left(\hat{\beta}_0\right) = 0.40; \text{se}\left(\hat{\beta}_1\right) = 0.53; \text{se}\left(\hat{\beta}_2\right) = 0.61; \text{se}(\hat{\delta}) = 0.93.$$

Using the 5% significance level, check the significance of the coefficients for the variables and give a meaningful interpretation of the results obtained.

Task 6

A researcher analyzes the impact of a law prohibiting the sale of alcohol after 11 p.m. on alcohol consumption. The researcher has information on per capita alcohol consumption in eight regions in 2014 and 2015. In 2014, alcohol was sold without restrictions in all regions. In 2015, this law was introduced in regions A, B, C, D, but in other regions it was not applied. Data on alcohol consumption (liters per person per year) are given in the table.

Region	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***
2014	6	6	8	4	4	3	3	2
2015	6	8	9	5	6	5	5	4

a. The researcher uses a fixed effects model to estimate the impact of interest: $y_{it} = \beta \cdot x_{it} + \alpha_i + \varepsilon_{it}$, where α_i- is the fixed effect of the i-th region ; x_{it} is a dummy variable equal to one if there was a law restricting the sale of alcohol in the i-th region in year t, and equal to zero otherwise; y_{it} - alcohol consumption per capita in the i-th region in the year t. Using the within-group transformation, find the estimate of the parameter β and interpret the result.

b. Now estimate the effect of the alcohol restriction law using the difference-in-differences method. Interpret the result. Provide a graphic illustration of the solution (do not forget to indicate in the figure the coordinates of all key points, as well as the magnitude of the impact effect).

c. What could cause such a discrepancy in estimates?