

Elements of Econometrics, 2023-2024

Class teacher: Ksenia Kasianova.

Seminar 3. Pair linear regression

Problem 1.

Recall Gauss Markov theorem. What problems with can lead to violation of GMT?

Problem 2.

Let Y_t be the aggregated consumption and X_t be aggregated disposable income. The regression model is

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \quad t = 1, \dots, n$$

	t	Y_t	X_t		t	Y_t	X_t
1986	1	152	170	1992	7	177	200
1987	2	159	179	1993	8	179	207
1988	3	162	187	1994	9	184	215
1989	4	165	189	1995	10	186	216
1990	5	170	193	1996	11	190	220
1991	6	172	199	1997	12	191	225

$$n = 12, \quad \bar{X} = 200$$

$$\sum X = 2400, \quad \bar{Y} = 173.92$$

$$\sum Y = 2087, \quad \sum Y^2 = 364741$$

$$\sum X^2 = 483236, \quad \sum XY = 419782$$

- (a) (5 points) Calculate coefficient for pair linear regression, give an interpretation.
- (b) (10 points) Calculate $s^2, s_\alpha^2, s_\beta^2$.
- (c) (5 points) Perform t -tests for both coefficients ($\alpha = 0.05$). Explain the results.
- (d) (5 points) Calculate 95%-confidence intervals.
- (e) (5 points) Calculate R^2 .

Problem 3.

We are interested in investigating the factors governing the precision of regression coefficients. Consider the model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

with OLS parameter estimates $\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_3$. Under the Gauss Markov assumptions, we have

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2} \times \frac{1}{1 - r_{X_2 X_3}^2},$$

where σ_ε^2 is the variance of ε and $r_{X_2 X_3}$ is the sample correlation between X_2 and X_3 .

- (a) **(4 marks)** Provide four factors that help with obtaining more precise parameter estimates for, say, $\hat{\beta}_2$.

Problem 4.

Consider a model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- (a) Derive confidence interval for prediction \hat{y}_{n+1} given new data x_{n+1} .
Hint: Check if \hat{y}_{n+1} unbiased then consider the variance of $\hat{y}_{n+1} - y_{n+1}$.
(b) Say, we have data for 10 observations:

$$\sum x_i = 20, \sum x_i^2 = 50, \sum y_i = 8, \sum y_i^2 = 26, \sum x_i y_i = 10$$

Find best linear unbiased prediction for y_{11} given that $x_{11} = 5$ and estimate its precision by constructing confidence interval.