Ekaterina Kazak
University of Manchester

The purpose of this session is to introduce you to the OLS estimator and its theoretical properties.

Statistical properties of OLS

a) Simulate a bivariate linear regression model:

$$y_i = 1 - 3x_i + u_i, \tag{1}$$

where $u \sim \mathcal{N}(0, 0.25)$, $x \sim \mathcal{N}(5, 16)$ and $n = 500$. Plot a scatter plot of y against x and the histogram for y.

b) Create a function which computes the *sum of squared residuals* as a function of intercept and slope parameter α_0 and α_1. Create a 3D visualization with the help of `plotly` package, which plots the sum of squared residuals surface for a range of parameters $\alpha_0, \alpha_1 \in [-30, 30]$. Zoom in the plot and find the minimum of RSS.

c) Plot a partial derivative $\dfrac{\partial RSS}{\partial \alpha_0}$ fixing $\alpha_1 = -3$ next to a plot of partial derivative $\dfrac{\partial RSS}{\partial \alpha_1}$ fixing $\alpha_0 = 1$. For each plot find the intersection point with a horizontal zero line. Interpret the plot.

d) Write a function which solves a bivariate OLS problem for a given (y, x) sample. Simulate the distribution of OLS parameters over a 1000 datasets, identical to the one defined in a). Draw the histograms of simulated β_0 and β_1 next to each other.

e) Examine how does the variance of OLS estimates changes with the change in σ_u^2.

Exercises to solve at home

Consider Proposition 2.1 from the lecture notes. Define a simple linear regression

$$y_i = \beta x_i + u_i, \tag{2}$$

where $i = 1, ..., 1000$, $u \sim \mathcal{N}(0, 100)$, $x \sim \mathcal{N}(5, 100)$. Consider two estimators for the slope parameter:

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \qquad \tilde{\beta} = \frac{\bar{y}}{\bar{x}}$$

f) Simulate a 1000 draws from a bivariate model defined in (2).

g) Plot the histograms of $\hat{\beta}$ and $\tilde{\beta}$ over these draws next to each other.

h) On each histogram plot the vertical red line at the true parameter value β.

i) Which estimator, $\hat{\beta}$ or $\tilde{\beta}$ is more efficient and why?

j) Compare your results with the homework answers which will be available on Blackboard on Friday.