Lecture 1

Text: Dougherty, Christopher. *Introduction to Econometrics*.

Class web page: Click here

Linear Regression

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki} + u_i$$

Two (main) uses:

1. To test theories about how the world works;
2. Forecasting.

- To get the best answers to these questions: we need the best fitting model
    1. Specification;
    2. Estimation:
        1. Ordinary Least Squares (OLS) estimator:

           $$\underset{\beta_i}{min} \sum_{i=1}^{N} u_i^2$$

        2. Minimum Absolute Deviation (MAD) estimator:

           $$\underset{\beta_i}{min} \sum_{i=1}^{N} |u_i|$$

        - When the assumptions of Gauss-Markov Theorem are satisfied, OLS estimators are the best linear unbiased estimator (BLUE).
            - Best: lowest variance;
            - Linear: linear at y_i ;
            - Unbiased: $E(\hat{\beta}) = \beta$;
            - Estimator: rule for processing data.
        - What assumptions are needed to ensure estimator is BLUE?
            - Regression is linear in parameters (β_i) ;
            - The x-values are fixed in repeated sampling (not random);
            - The error term has zero mean $(E(u_i) = 0)$;
            - Homoskedasticity, *i. e.*, the variance of the errors is constant (not depending on i) : $u_i \sim (0, \sigma^2)$.
            - No autocorrelation in errors, *e. g.*, $u_i = \rho \cdot u_{i-1} + \varepsilon_i$, where ε_i is a random factor.
            - No correlation between observed values and errors, *i. e.*, $E(x_i \cdot u) = 0, \ \forall x_i$;
            - The number of observations is at least the same as that of variables $(N \geq k)$;
            - There must be (as much as possible) variability in x's (otherwise, variance would be infinite).
            - Regression specified correctly;
            - No <u>perfect</u> multicollinearity $(x_i = c \cdot x_j)$.