OLS is the best procedure for estimating a linear regression model only under certain assumptions.

The word *classical* refers to these assumptions that are required to hold.

Assumptions of the Classical Linear Regression Model:

1. The regression model is linear, correctly specified, and has an additive error term.
2. The error term has a zero population mean.
3. All explanatory variables are uncorrelated with the error term
4. Observations of the error term are uncorrelated with each other (no serial correlation).
5. The error term has a constant variance (no heteroskedasticity).
6. No explanatory variable is a perfect linear function of any other explanatory variables (no perfect multicollinearity).
7. The error term is normally distributed (not required).

Essentially, the rest of the course deals with what happens when one or more of these assumptions do not hold and what we can do to remedy the situation.

Assumption 1: Linear Model, Correctly Specified, Additive Error

The model with *k* explanatory variables:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i.$$

The regression model is linear in the coefficients.

The random error is additive.

The model also contains the "right" independent variables.

Assumption 2: Error term has a population mean of zero

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i$$

In terms of notation: $E(\varepsilon_i) = 0$.

The error term itself cannot be observed. But let's suppose the mean of ε_i is 10.

So we could just add and subtract 10 to the model to force the error term to have a mean of zero:

$$Y_i = (\beta_0 + 10) + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + (\varepsilon_i - 10).$$

Assumption 3: Explanatory variables uncorrelated with error term

This assumption requires that the observed values of the independent variables be determined independently of the error term.

In terms of notation: $Cov(X_i, \varepsilon_i) = 0$.

Let's say that X_{1i} is correlated with the error term. OLS will not be able to estimate the slope of X_{1i} separately from movements in the error term. OLS will give credit to X_{1i} for movement in Y that is due to the error term.

Assumption 4: No serial correlation

Error term observations are drawn independently (and therefore not correlated) from each other. When observed errors follow a pattern, they are said to be **serially correlated** or **autocorrelated**.

In terms of notation: $Cov(\varepsilon_i, \varepsilon_j) = 0$.

Assumption 5: No heteroskedasticity (homoskedasticity)

The error term observations come from the same probability distribution with a constant variance. If the variance of the distribution of the error term changes for each observation or range of observations then we have what's called **heteroskedasticity**.

In terms of notation: $Var(\varepsilon_i) = \sigma^2$.

Assumption 6: No perfect multicollinearity

None of the independent variables have a perfect linear relationship (**perfect collinearity** or **multicollinearity**) with any of the other independent variables. If they do, OLS cannot separately estimate the coefficients for them.

Even is the variables are not perfectly collinear, there can still be a problem. If two or more variables are highly correlated, multicollinearity exists and could be problematic.

Assumption 7: Error term is normally distributed

Have already assumed that the observations of the error terms are drawn independently (Assumption 4) from a distribution that a mean of zero (Assumption 2) has a constant variance (Assumption 5). Now, we specify the shape of the distribution. The normal distribution is assumed.

In terms of notation: $\varepsilon_i \sim N(0, \sigma^2)$.