Reminder TD_2

Marion Brouard, Pauline Leveneur

October 2, 2025

- **Recap Objective:**
  1. Understand the theoretical model notations and its matrix formulation.
  2. Review the FOCs and properties of the OLS estimator.
  3. Review R^2:
     - Its mathematical definition,
     - Its interpretation,
     - Its properties.

0 Notation Recap

1. Theoretical Model: (Often called the population model or true model in English)

$$Y_{(1,1)} = X'_{(1,p+1)}\beta_{0(p+1,1)} + \varepsilon_{(1,1)} \tag{1}$$

Y = dependent variable
X = regressors, independent variables, explanatory variables
ε = error: everything unobserved that impacts Y
$p+1$ = number of regressors + a constant
This expression is purely theoretical: it does not take the multiple observations into account.

2. Matrix Formulation: When we want to estimate the theoretical model, we use a matrix formulation that includes each observation.

$$\mathbf{Y}_{(n,1)} = \mathbf{X}_{(n,p+1)}\beta_{0(p,1)} + \varepsilon_{(n,1)}$$

$$\underbrace{\begin{pmatrix} Y_{1,1} \\ Y_{2,1} \\ \vdots \\ Y_{n,1} \end{pmatrix}}_{\mathbf{Y}_{(n,1)}} = \underbrace{\begin{pmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,p} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,p} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,p} \end{pmatrix}}_{\mathbf{X}_{(n,p+1)}} \underbrace{\begin{pmatrix} \beta_{0,1} \\ \beta_{1,1} \\ \vdots \\ \beta_{p,1} \end{pmatrix}}_{\beta_{0(p+1,1)}} + \underbrace{\begin{pmatrix} \varepsilon_{1,1} \\ \varepsilon_{2,1} \\ \vdots \\ \varepsilon_{p,1} \end{pmatrix}}_{\varepsilon_{(n,1)}}$$

Thus for individual i, we have:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_p X_{i,p} + \varepsilon_i$$

If we set $p = 2$, for example, we can rewrite this regression in vector form:

$$\mathbf{Y}_{(n,1)} = \beta_{0(1,1)} + \mathbf{X}_{1(n,1)}\beta_{1(1,1)} + \mathbf{X}_{2(n,1)}\beta_{1(1,1)} + \varepsilon_{(n,1)}$$

$$\underbrace{\begin{pmatrix} Y_{1,1} \\ Y_{2,1} \\ \vdots \\ Y_{n,1} \end{pmatrix}}_{\mathbf{Y}_{(n,1)}} = \beta_0 + \underbrace{\begin{pmatrix} X_{1,1} \\ X_{2,1} \\ \vdots \\ X_{n,1} \end{pmatrix}}_{\mathbf{X}_{1(n,1)}} \beta_1 + \underbrace{\begin{pmatrix} X_{1,2} \\ X_{2,2} \\ \vdots \\ X_{n,2} \end{pmatrix}}_{\mathbf{X}_{2(n,1)}} \beta_2 + \underbrace{\begin{pmatrix} \varepsilon_{1,1} \\ \varepsilon_{2,1} \\ \vdots \\ \varepsilon_{p,1} \end{pmatrix}}_{\varepsilon_{(n,1)}}$$

FOC:

Let $\hat{Y}$ be the predicted value of Y from the linear regression.
NB: Here X_i' is dimension $(1, p+1)$, Y_i is dimension $(1, 1)$, b is dimension $(p+1, 1)$
$\rightarrow$ *Carefully check the statements to verify how the matrices are defined.*

$$\hat{Y}_i = X_i'\hat{\beta}$$
$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$$

By following the ordinary least squares (OLS) minimization program:

$$\hat{\beta} = \arg\min_{b \in \mathbb{R}^p} \sum_{i=1}^{n} (Y_i - X_i'b)^2$$

We find the following two first-order conditions (FOCs):

$$\sum_{i=1}^{n} (Y_i - X_i'\hat{\beta}) = 0$$
$$\sum_{i=1}^{n} X_i(Y_i - X_i'\hat{\beta}) = 0$$

Which gives us the following two properties: $E(\hat{\varepsilon}) = 0$ and $E(X\hat{\varepsilon}) = 0$

1 Key Points to Remember:

1.1 Estimators

If we consider a random sample of n observations following the theoretical model (1), then:

- If **X** is full rank, then **X'X** is invertible, and the estimator $\hat{\beta}$ exists. In other words, there are no linear relationships between the independent variables. *(In English, this is often referred to as the "no perfect collinearity" assumption.)* Note: **X** can be correlated, but they simply cannot be perfectly correlated.

1.2 R^2

Definition: The R^2 is a measure between 0 and 1 that tells us how well X predicts Y. It is the proportion of the variance in Y explained by X. The higher the R^2, the better the predictive quality. If $R^2 = 1$, the prediction is perfect.

$$R^2 = \frac{\hat{V}(\hat{Y})}{\hat{V}(Y)}$$

Interpretation: $R^2 = 0.75 \rightarrow 75\%$ of the variance in Y is explained by X.

Key Properties to Remember:

- R^2 is invariant to all affine transformations of X or Y *(TD Question 1)*.
- R^2 converges in probability to $R^2_\infty = \frac{V(X'\beta)}{V(Y)}$, which is the proportion of variance in Y explained by the best linear prediction from X. *(TD Question 2)*
- **R^2 mechanically increases with the addition of new X variables.**
  An alternative option is the adjusted R^2 (which adds a penalty for the number of X variables). *(TD Question 4)*