ECOM20001 Econometrics 1

Tutorial 7 (Week 7)

Zheng Fan

The University of Melbourne

Linear regression

This is a simple regression we have seen before

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

Linear regression

This is a simple regression we have seen before

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

Indeed, the intercept is actually based on a variable only contains 1

$$y_i = \beta_0 1 + \beta_1 x_i + \epsilon_i$$

Multicollinearity

If we have a multiple linear regression

$$y_i = \beta_0 1 + \beta_1 x_i + \beta_2 z_i + \epsilon_i$$

Multicollinearity

If we have a multiple linear regression

$$y_i = \beta_0 1 + \beta_1 x_i + \beta_2 z_i + \epsilon_i$$

and unfortunately $x = z$ for all observation i

$$y_i = \beta_0 1 + (\beta_1 + \beta_2) x_i + \epsilon_i$$

Multicollinearity

If we have a multiple linear regression

$$y_i = \beta_0 1 + \beta_1 x_i + \beta_2 z_i + \epsilon_i$$

and unfortunately $x = z$ for all observation i

$$y_i = \beta_0 1 + (\beta_1 + \beta_2) x_i + \epsilon_i$$

we can only estimate $\beta_1 + \beta_2$ the combined coefficient, but not each of them.

$\Rightarrow$ **perfect multicollinearity**

Dummy Variable Trap

An example of perfect multicollinearity

- Occurs when we include all dummy variables for a categorical variable in a regression model along with the intercept.

Dummy Variable Trap: Transmission Type Example

Suppose we model fuel efficiency based on car transmission type.

Dummy Variable Trap: Transmission Type Example

Suppose we model fuel efficiency based on car transmission type.

Two dummy variables:

$$D_{\text{Auto},i} = \begin{cases} 1 & \text{if automatic} \\ 0 & \text{otherwise} \end{cases}, \quad D_{\text{Manual},i} = \begin{cases} 1 & \text{if manual} \\ 0 & \text{otherwise} \end{cases}$$

Dummy Variable Trap: Transmission Type Example

Suppose we model fuel efficiency based on car transmission type.

Two dummy variables:

$$D_{\text{Auto},i} = \begin{cases} 1 & \text{if automatic} \\ 0 & \text{otherwise} \end{cases}, \quad D_{\text{Manual},i} = \begin{cases} 1 & \text{if manual} \\ 0 & \text{otherwise} \end{cases}$$

Full model with intercept:

$$y_i = \beta_0 1 + \beta_1 D_{\text{Auto},i} + \beta_2 D_{\text{Manual},i} + \varepsilon_i$$

Dummy Variable Trap: Transmission Type Example

Suppose we model fuel efficiency based on car transmission type.

Two dummy variables:

$$D_{\text{Auto},i} = \begin{cases} 1 & \text{if automatic} \\ 0 & \text{otherwise} \end{cases}, \quad D_{\text{Manual},i} = \begin{cases} 1 & \text{if manual} \\ 0 & \text{otherwise} \end{cases}$$

Full model with intercept:

$$y_i = \beta_0 1 + \beta_1 D_{\text{Auto},i} + \beta_2 D_{\text{Manual},i} + \varepsilon_i$$

But since every car is either automatic or manual:

$$D_{\text{Auto},i} + D_{\text{Manual},i} = 1$$

Dummy Variable Trap: Transmission Type Example

Therefore

$$
\begin{aligned}
y_i &= \beta_0 1 + \beta_1 D_{\text{Auto},i} + \beta_2 D_{\text{Manual},i} + \varepsilon_i \\
&= \beta_0 1 + \beta_1 D_{\text{Auto},i} + \beta_2 (1 - D_{\text{Auto},i}) + \varepsilon_i \\
&= (\beta_0 + \beta_2) 1 + (\beta_1 - \beta_2) D_{\text{Auto},i} + \varepsilon_i
\end{aligned}
$$

Again, we could only estimate the combined effect of $(\beta_0 + \beta_2)$ and $(\beta_1 - \beta_2)$, but not each of them.

$\Rightarrow$ **perfect multicollinearity**

Perfect Multicollinearity

- Note that $D_{\text{Auto},i} + D_{\text{Manual},i} = 1$ for all observations i.
- So, if we also include an intercept term, we have:

$$\text{Intercept} = 1 = D_{\text{Auto},i} + D_{\text{Manual},i}$$

- $\Rightarrow$ Perfect linear dependence among regressors.
- $\Rightarrow$ OLS fails.

How to Avoid the Trap

- Drop one dummy variable — use it as the **reference category**.
- Example: Include only $D_{\text{Auto},i}$
- The omitted category ($D_{\text{Manual},i}$) is captured in the intercept.

Correct Model:

$$y_i = \beta_0 1 + \beta_1 D_{\text{Auto},i} + \varepsilon_i$$

Interpretation:

- β_0: The mean value of y_i for manual cars.
- $\beta_0 + \beta_1$: The mean value of y_i for automatic cars.
- β_1: The effect (or difference) of having an automatic transmission compared to a manual transmission.

How to Avoid the Trap: Alternatively

Less common: Drop the intercept

Recall before

$$y_i = \beta_0 1 + \beta_1 D_{\text{Auto},i} + \beta_2 D_{\text{Manual},i} + \varepsilon_i$$

Removing intercept

$$y_i = \beta_1 D_{\text{Auto},i} + \beta_2 D_{\text{Manual},i} + \varepsilon_i$$

No problem, but no reference group.

$$\begin{aligned} y_i &= \beta_1 D_{\text{Auto},i} + \beta_2 (1 - D_{\text{Auto},i}) + \varepsilon_i \\ &= \beta_2 1 + (\beta_1 - \beta_2) D_{\text{Auto},i} + \varepsilon_i \end{aligned}$$