Quantitative Methods 2

Tutorial 11

Nhan La

MLR assumptions

	Assumption	Violation
MLR1	$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i$	Nonlinearity
MLR2	$E(\varepsilon_i \| x_i) = 0$	Omitted variable
MLR3	$\mathrm{Var}(\varepsilon_i \| x_i) = \sigma^2$	Heteroskedasticity
MLR4	$E(\varepsilon_i \varepsilon_j \| x_i, x_j) = 0$	Autocorrelation
MLR5	No exact linear relationship among x_i	(Perfect) multicollinearity
MLR6	$\varepsilon_i \| x_i \sim N(0, \sigma^2)$	Unreliable hypothesis testing

Last week

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

- Multicollinearity
  - Effect of X_i on Y, holding all other X's constant?
- Heteroskedasticity : $\mathrm{Var}(\varepsilon_i|x_i) = \sigma_i^2$
- Consequence:
  - $\hat{\beta}$ are still linear and unbiased
  - $\mathrm{SE}(\hat{\beta})$ are incorrect, thus hypothesis tests are incorrect

Last week

Violation of MLR5: Multicollinearity

- High R^2 but only a few significant t-ratios with the logical signs
  - $t_k = \frac{\widehat{\beta}_k - 0}{SE(\widehat{\beta}_k)}$
  - $R^2 \geq 0.8$ (strict); $R^2 \geq 0.5$ (less strict)
- Strong correlation between X's. Stronger correlation between some X's than between these variables and Y
  - $|r| > 0.8$
- $VIF_j = \frac{1}{1 - R_j^2} \geq 5$

Last week

	Assumption	Violation
MLR1	$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i$	Nonlinearity
MLR2	$E(\varepsilon_i \| x_i) = 0$	Omitted variable
MLR3	$\mathrm{Var}(\varepsilon_i \| x_i) = \sigma^2$	Heteroskedasticity
MLR4	$E\left(\varepsilon_i \varepsilon_j \middle\| x_i, x_j\right) = 0$	Autocorrelation
MLR5	No exact linear relationship among x_i	(Perfect) multicollinearity
MLR6	$\varepsilon_i \| x_i \sim N(0, \sigma^2)$	Unreliable hypothesis testing

Last week

- Multiple linear regression

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

- General F test
  - General expression: $F = \frac{n-k-1}{m} \frac{SSE_r - SSE}{SSE}$
  - Same dependent variable: $F = \frac{n-k-1}{m} \frac{R^2 - R_r^2}{1-R^2} = \frac{(R^2 - R_r^2)/m}{(1-R^2)/(n-k-1)}$
- Conduct the linear restriction test
  - Use EViews' coefficient restriction test, or
  - Estimate separately restricted and unrestricted models

Tutorial 11

- Point prediction
$$\hat{y}_0 = \hat{E}(Y|x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{0,1} + \cdots + \hat{\beta}_k x_{0,k}$$

- Individual confidence interval prediction
  - $k = 1$: $s_{\hat{y}_0} = s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$
  - $k > 1$: $s_{\hat{y}_0} \approx s_\varepsilon$
  - $\hat{y}_0 \pm t_{\alpha/2, n-2} s_{\hat{y}_0}$
- Sub-population (group) confidence interval prediction
  - $k = 1$: $s_{\hat{E}(y_0)} = s_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{s_{\hat{y}_0}^2 - s_\varepsilon^2} < s_{\hat{y}_0}$
  - $\hat{y}_0 \pm t_{\alpha/2, n-2} s_{\hat{E}(y_0)}$

Tutorial 11

- Dummy independent variable models
  - Intercept dummy variable: $y = \beta_0 + \beta_1 X + \beta_2 D + \varepsilon$
  - Slope dummy variable: $y = \beta_0 + \beta_1 X + \beta_2 DX + \varepsilon$
  - Combined dummy variable: $y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 DX + \varepsilon$

- Interpretation
  - Group change instead of incremental change

- Dummy (index) variable trap
  - If a categorical variable consists of k categories, only need to create and include (at most) $k - 1$ dummy variables
  - If include k dummy variables: perfect multicollinearity

Exercise 3d

$$MBAGPA = \beta_0 + \beta_1 UGPA + \beta_2 MBAA + \beta_3 D_1 + \beta_4 D_2 + \beta_5 D_3 + \varepsilon$$

- Base group: BA
- Coefficients β_3, β_4 and β_5 of D_1, D_2 and D_3 show the effect on $MBAGPA$ of BCom, BEng, and BSc compared with BA, respectively
- For instance, when BA $= 1 \Longrightarrow D_1 = 0, D_2 = 0, D_3 = 0$
$$MBAGPA = \beta_0 + \beta_1 UGPA + \beta_2 MBAA + \varepsilon$$
- When BCom=1 $\Longrightarrow D_1 = 1, D_2 = 0, D_3 = 0$. Also BA $= 0$
$$MBAGPA = \beta_0 + \beta_1 UGPA + \beta_2 MBAA + \beta_3 \times 1 + \varepsilon$$
- Hence β_3 represents the difference of effect on $MBAGPA$ between BCom and BA
- What if we want to find effect of BCom (vs. Not BCom)?
$$D_1 = \begin{cases} 1 \text{ if BCom} \\ 0 \text{ if Not BCom} \end{cases}$$
$$MBAGPA = \beta_0 + \beta_1 UGPA + \beta_2 MBAA + \beta_3 D_1 + \varepsilon$$