Quantitative Methods 2

Tutorial 12

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	$\text{Var}(\varepsilon_i \| x_i) = \sigma^2$	Heteroskedasticity
MLR4	$E\left(\varepsilon_i \varepsilon_j \| x_i, x_j\right) = 0$	Autocorrelation
MLR5	No exact linear relationship among x_i	(Perfect) multicollinearity
MLR6	$\varepsilon_i \| x_i \sim N(0, \sigma^2)$	Unreliable hypothesis testing

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

Coefficients - Marginal effects

- Multiple linear regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\frac{d(Y)}{d(X_2)} = \beta_2$$

- Polynomial regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2^2 + \varepsilon$$

$$\frac{d(Y)}{d(X_2)} = \beta_2 * 2X_2$$

- Log-transformed linear model

$$\ln(Y) = \beta_0 + \beta_1 X + \beta_2 \ln(Z) + \varepsilon$$

  - Interpretation: Elasticity/Semi-elasticity

Tutorial 12

Dummy dependent variable

- Linear probability model:

$$D_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$
$$E(D_i|X_i) = P_i = \beta_0 + \beta_1 X_i$$

  - $\varepsilon_i | x_i \nsim N(0, \sigma^2)$
  - $\mathrm{Var}(\varepsilon_i | x_i) = \sigma_i^2$
  - $\widehat{D}_i$ not necessarily within [0,1]
  - β_1 constant

- Logit/probit model

$$E(D_i|X_i) = P_i = F(\beta_0 + \beta_1 X_i)$$

Tutorial 12

- Logit
  - $P_i = \frac{1}{1+\exp(\beta_0+\beta_1 X_i+\cdots+\beta_k X_k)}$
  - Interpretation:
    - change in odds (%) $= \exp(\beta_k) - 1$
    - change in odds $= \exp(\beta_k)$
    - change in log(odds) $= \beta_k$

- Probit
  - $P_i = \Phi(\beta_0 + \beta_1 X_i + \cdots + \beta_k X_k)$
  - $\Phi(I)$: (cumulative standard normal distribution)