Generalized Linear Models

Statistical Research Methods I

Seongsoo Choi (최성수)

How can we extend the linear regression model?

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

- Relaxing
  - linearity: $y_i = f(\beta_0 + \beta_1 x_i + \varepsilon_i)$ where $f(\cdot)$ may be a nonlinear function $\Longrightarrow$ Nonlinear probability models
  - the assumptions about ε: $E(\varepsilon) = 0$ and $E(\varepsilon|x) = 0$
    - iid (independently and identically distributed) observations: multilevel structures (e.g., i is embedded in j or i is repeatedly observed several times, t)
    - then, $\varepsilon_{ij} = u_j + \epsilon_i$ and u_j may not behave like ε_i (e.g., $E(u) \neq 0$ and $E(u|x) \neq 0$) $\Longrightarrow$ Panel models (FE, RE models)
  - β_0 and β_1 may vary across observations: $\beta_{0,i}$ and $\beta_{1,i}$ $\Longrightarrow$ mixed effects models or hierarchical linear model (HLM)

Relaxing Linearity

- The issue of relaxing the linearity assumption arises when we want to explain a categorical variable as a dependent variable
  - dichotomous or polychotomous outcomes (ordered or unordered)
- For now, let's have a look at a dichotomous outcome: e.g., whether someone attained a college degree ($y = 1$) or not ($y = 0$)
- Now the outcome we want to predict is the *probability* of attaining a college degree: $Pr(y = 1)$

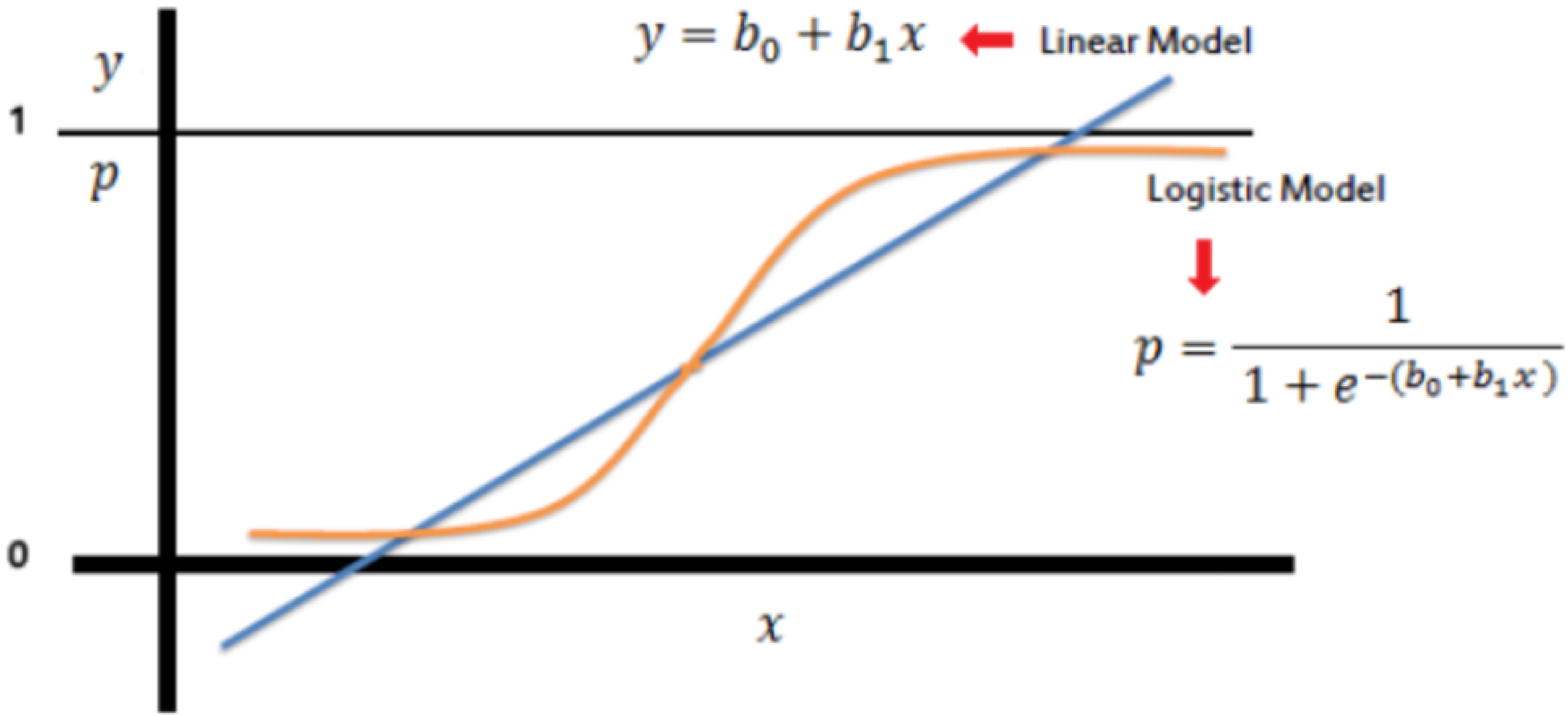
Relaxing Linearity

- There are two possible options
  - Run a linear regression model: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ (linear probability model, LPM)
    - y can be interpreted as probability, but what the model predicts ($\hat{y}_i$) may fall beyond the legitimate range (between 0 and 1)
  - Nonlinear probability model: limit the range of $\hat{y}_i$ between 0 and 1 by applying a function that makes the RHS fall within the probability range no matter what value x has
    - $y_i = f(\beta_0 + \beta_1 x_i + \varepsilon_i)$ where $0 \geq f(x) \geq 1$ for any value of x
    - $f(\cdot)$ is called link function:
      - e.g., inverse logit function (logit model), the CDF of the standard normal function (probit model)

Logit model (or Logistic regression model)

$$Pr(y_i = 1) = \Lambda(\beta_0 + \beta_1 x_i) = \frac{exp(\beta_0 + \beta_1 x_i)}{1 + exp(\beta_0 + \beta_1 x_i)}$$

LPM and NLPM (logit model)



Logit model (or Logistic regression model)

$$Pr(y_i = 1) = \Lambda(\beta_0 + \beta_1 x_i) = \frac{exp(\beta_0 + \beta_1 x_i)}{1 + exp(\beta_0 + \beta_1 x_i)}$$

- We can make the RHS a linear form by rearranging

$$log\left(\frac{Pr(y_i = 1)}{1 - Pr(y_i = 1)}\right) = \beta_0 + \beta_1 x$$

- Now the dependent variable is the log odds of $y = 1$, so we can interpret β_1 like
  - one unit increase in x is associated with a β_1 increase in the log odds of $y = 1$

Logit model (or Logistic regression model)

- If you are not happy with "log odds", you can throw "log" away by taking exponential on both hand sides

$$\frac{Pr(y_i = 1)}{1 - Pr(y_i = 1)} = exp(\beta_0 + \beta_1 x)$$

- Then, now the dependent variable is the odds of $y = 1$, but its relationship with x is not additive but multiplicative
  - one unit increase in x is associated with a $exp(\beta_1)$ times increase in the odds of $y = 1$
  - If $exp(\beta) = 1.3$, a unit increase in x is associated with 1.3 times (or 30%) increase in the odds of $y = 1$
  - If $exp(\beta) = 0.7$, a unit increase in x is associated with 0.7 times increase (or 30% decrease) in the odds of $y = 1$

Logit model (or Logistic regression model)

- We have to go back and forth between probability (most intuitive), odds, and log odds (least intuitive)
- This also means we have to go back and forth between a nonlinear form of RHS (least convenient) and a linear form (the most convenient)
- In case of the linear regression model, we don't have this annoying situation
  - The linear regression model is a special case where $f(x) = x$ while the logit model is a case where $f(x) = \Lambda(x)$
- In Stata, the command is *logit*

Other nonlinear probability models

- There are several other differences in NLPM, but we won't cover them
- NLPM for polychotomous outcomes (multinomial logit model for unordered outcomes and ordered logit model for ordered outcomes) are a little bit more complicated, but basically they are straightforward extensions of the logit model
- Count models: Poisson model, negative binomial regression model, zero-inflated poisson model, etc.
- Event history models: discrete-time or continuous-time, simple transition vs. competing events, etc.

Wrapping up the course: we have learned

- random variables, probabilities, distribution functions (PDF and CDF), etc.
- data exploration through graphic approaches and descriptive statistics
- estimation and statistical inferences; population-sample and sampling distribution
- how to gauage the association between two variables
- linear regression model:
    - assumptions, OLS estimation, statistical inferences
    - modeling, model building, confounding and mediation
    - dummy variables, nonlinearity and interactions (moderation)
    - model selection (F-test); ANOVA
    - generalization of linear regression model (e.g., logit model)