

Regression With Special Variables

Junhui Qian

November 29, 2018

Outline

- ▶ **Standardized Scores**
- ▶ Quadratic Terms
- ▶ Interaction Terms
- ▶ Binary Explanatory Variables
- ▶ Binary Choice Models and Some Generalizations

Standardized Scores: An Example

- ▶ When we study how salaries depend on IQ score, we may run the following simple regression,

$$\log(\text{salary}_i) = \beta_0 + \beta_1 \text{IQ}_i + u.$$

- ▶ What is the economic interpretation of β_1 above?
- ▶ To make the coefficient more meaningful, we can define a new variable

$$z_i = \frac{\text{IQ}_i - \overline{\text{IQ}}}{\hat{\sigma}_{\text{IQ}}}$$

and run

$$\log(\text{salary}_i) = \beta_0 + \beta_1 z_i + u.$$

- ▶ Now, what is the economic interpretation of β_1 ?

Standardized Scores

- ▶ Given a variable x_i , the standardized score is defined as

$$z_{x,i} = \frac{x_i - \bar{x}}{\hat{\sigma}_x},$$

where \bar{x} and $\hat{\sigma}_x$ are the sample mean and standard deviation, respectively.

- ▶ The coefficient on a standardized score is interpreted as the increase in y when x is one standard deviation higher.

Standardized Score as Dependent Variable

- ▶ The standardized score can also be the dependent variable in a regression. For example, let

$$z_{y,i} = \frac{y_i - \bar{y}}{\hat{\sigma}_y},$$

and run

$$z_{y,i} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

- ▶ β_1 in this model is interpreted as the increase in y , in terms of standard deviations, with each unit of increase in x_1 , holding other factors fixed.

Standardized Coefficients

- ▶ We may need a model where all variables are standardized, for example,

$$z_{y,i} = \beta_1 z_{x1,i} + \beta_2 z_{x2,i} + u.$$

- ▶ The β_1 is interpreted as the increase in y , in terms of standard deviations, when x_1 is one standard deviation higher, holding other factors fixed.
- ▶ Note that we need no constant term in the above model.
- ▶ The β 's in such models are called “standardized coefficients”.

Outline

- ▶ Standardized Scores
- ▶ **Quadratic Terms**
- ▶ Interaction Terms
- ▶ Binary Explanatory Variables
- ▶ Binary Choice Models and Some Generalizations

Models with Quadratic Terms

- ▶ We have seen models with quadratic terms. For example, in the income determination model,

$$\log(\textit{income}) = \beta_0 + \beta_1 \textit{edu} + \beta_2 \textit{edu}^2 + \beta_3 \textit{expr} + u.$$

- ▶ How do you interpret the parameters β_1 and β_2 ?

The Level-Dependent Partial Effect

In a model with quadratic terms,

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u.$$

- ▶ The partial effect of x on y is dependent on the level of x .
- ▶ To see why,

$$\frac{\partial y}{\partial x} = \beta_1 + 2\beta_2 x.$$

Back to Our Example

We estimate the model and obtain,

$$\log(\text{income}) = 7.55 + 0.0715\text{edu} + 0.00560\text{edu}^2 - 0.00299\text{expr}.$$

We may ask, for a person with 10 years' education, how much income increase would be expected if he receives one more year of education?

Outline

- ▶ Standardized Scores
- ▶ Quadratic Terms
- ▶ **Interaction Terms**
- ▶ Binary Explanatory Variables
- ▶ Binary Choice Models and Some Generalizations

Interaction Terms

- ▶ Recall that linear regression is the first-order approximation of a usually nonlinear relationship:

$$y = f(x_1, x_2) + u \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

- ▶ To get better approximations, we may add quadratic terms,

$$y = f(x_1, x_2) + u \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + u.$$

- ▶ β_3 and β_4 measures nonlinearities in x_1 and x_2 , and β_5 measures the “interaction” effect between x_1 and x_2 .
- ▶ The term $x_1 x_2$ is hence called the “interaction term”.

Interaction Effect

- ▶ There is “interaction effect” when a change in one explanatory variable may affect the slope of another.
- ▶ For example, in a model of housing price,

$$price = \beta_0 + \beta_1 sqft + \beta_2 rooms + \beta_3 sqft \cdot rooms + \beta_4 bath + u.$$

- ▶ The partial effect of *rooms* on *price* is $\beta_2 + \beta_3 sqft$.
- ▶ If $\beta_3 > 0$, an additional room in a large house is more valuable than that in a small house.

Average Partial Effect

In a model with interaction terms,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u.$$

- ▶ β_1 is the partial effect of x_1 on y when $x_2 = 0$.
- ▶ This is not very interesting. Instead, we may be interested in the partial effect of x_1 on y when $x_2 = \bar{x}_2$.
- ▶ We may obtain such an “average partial effect” by plug \bar{x}_2 in $\hat{\beta}_1 + \hat{\beta}_3 x_2$.
- ▶ To obtain average partial effect directly, we can run

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 - \bar{x}_1)(x_2 - \bar{x}_2) + u.$$

- ▶ Now β_1 is the desired average partial effect.

Outline

- ▶ Standardized Scores
- ▶ Quadratic Terms
- ▶ Interaction Terms
- ▶ **Binary Explanatory Variables**
- ▶ Binary Choice Models and Some Generalizations

Binary (Dummy) Variables

- ▶ A binary variable describes qualitative information, in contrast to quantitative information.
- ▶ For example, male or female, urban or rural residence, employed or unemployed, a person buys a car or not, etc.
- ▶ A typical binary variable is defined as

$$female = 0, 1,$$

where 0 corresponds to “not female” and 1 corresponds to “female” .

Binary Variable on the Intercept

- ▶ Binary variables may influence the intercept only. For example, in the following model,

$$\log(\textit{income}) = \beta_0 + \beta_1 \textit{edu} + \beta_2 \textit{expr} + \beta_3 \textit{female} + u,$$

where *female* is a binary variable.

- ▶ For females, the model is

$$\log(\textit{income}) = (\beta_0 + \beta_3) + \beta_1 \textit{edu} + \beta_2 \textit{expr} + u.$$

- ▶ For males, the model is

$$\log(\textit{income}) = \beta_0 + \beta_1 \textit{edu} + \beta_2 \textit{expr} + u.$$

- ▶ The group of males in this model can be called “baseline group”.

Binary Variable on the Slope

- ▶ Binary variables may also influence the slope. For example, in the following model,

$$\log(\textit{income}) = \beta_0 + \beta_1 \textit{edu} + \beta_2 \textit{expr} + \beta_3 \textit{female} \cdot \textit{edu} + u.$$

- ▶ For females, the model is

$$\log(\textit{income}) = \beta_0 + (\beta_1 + \beta_3) \textit{edu} + \beta_2 \textit{expr} + u.$$

- ▶ For males, the model is

$$\log(\textit{income}) = \beta_0 + \beta_1 \textit{edu} + \beta_2 \textit{expr} + u.$$

When There Are Several Categories

- ▶ Binary variables are easily defined when there are two categories. For example, the gender of a person can only be male or female.
- ▶ Other qualitative information may involve more than two categories. For example, to describe the region of the country where people live and work, we may have three categories (coast, middle, west). For another example, to describe seasonality, we have four categories.

When There Are Several Categories

- ▶ Obviously, one binary variable is not enough for more than two categories. The solution is to have more than one binary variables. For example, if there are 3 regions (coast, middle, west) in total, we may consider

$$\log(\text{income}) = \beta_0 + \beta_1 \text{edu} + \beta_2 \text{expr} + \beta_3 \text{west} + \beta_4 \text{middle} + u,$$

where *west* is defined as 1 if the individual is in the west and 0 otherwise.

- ▶ Why not add one more binary variable *coast*?

Testing the Existence of Qualitative Effects

- ▶ We often want to test whether being in some category (such as gender, hukou, race, region, etc.) has any effect on dependent variables, controlling for other factors.
- ▶ With multiple linear regression with binary variables, we may easily achieve this.
- ▶ For example, to test whether or not females are discriminated in work places, we may run

$$\log(\text{income}) = \beta_0 + \beta_1 \text{edu} + \beta_2 \text{expr} + \beta_3 \text{female} + u,$$

and test

$$H_0 : \beta_3 = 0 \quad H_1 : \beta_3 < 0.$$

Outline

- ▶ Standardized Scores
- ▶ Quadratic Terms
- ▶ Interaction Terms
- ▶ Binary Explanatory Variables
- ▶ **Binary Choice Models and Some Generalizations**

Modeling Binary Choices

- ▶ Decision to migrate or not
- ▶ Buy a car or not
- ▶ Rent or buy an apartment
- ▶ To vote for or against a bill
- ▶ For college/MBA graduates, find a job within 3 months
- ▶ For females, to be a housewife or not
- ▶ ∴

Binary Distribution

Let y be the binary choice variable taking values of 0 and 1. We can describe y by binary (or Bernoulli) distribution with parameter p ,

$$f(y) = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$$

Equivalently, we may write

$$f(y) = p^y(1 - p)^{1-y}, \quad y = 0, 1.$$

We have

$$\begin{aligned} \mathbb{E}y &= p \\ \text{var}(y) &= p(1 - p). \end{aligned}$$

Linear Probability Model

Let x be a vector of variables that influences the outcome of y , which takes value of 0 or 1. We may write a **linear probability model** as follows,

$$y = x'\beta + u.$$

- ▶ Let $p = \mathbb{E}(y|x) = x'\beta$. It is the probability of $y = 1$ given x .
- ▶ Conditional on x , y is distributed as $\text{Binary}(p)$.
- ▶ $\text{var}(y|x)$ is then $p(1 - p) = x'\beta(1 - x'\beta)$.

An Example

Suppose y denotes the decision of an individual on whether or not migrates to city. And let x be the income the individual receives from previous job in the countryside. We may construct a simple migration model,

$$y = \beta_0 + \beta_1 x + u.$$

Problem of Linear Probability Model

- ▶ Heteroscedasticity.
- ▶ Nonsense Probability
 - ▶ To see this, observe that after estimation,

$$\hat{y} = \hat{p} = x' \hat{\beta}.$$

- ▶ To make predictions based on the estimated model, it is not guaranteed that $\hat{p} \in [0, 1]$.
- ▶ The nonsense probability comes from linear marginal probability,

$$\frac{\partial p}{\partial x} = \beta.$$

Probit and Logit Model

- ▶ One method to limit p within $[0, 1]$ is to use the cumulative distribution function (cdf) of a distribution, say $F(\cdot)$,

$$p = F(x'\beta) = \mathbb{P}(Z \leq x'\beta).$$

- ▶ If F is the cdf of the standard normal distribution ($N(0,1)$), then the model is called a “probit model”.
- ▶ If $F(s) = \frac{1}{1+e^{-s}}$, then the model is called a “logit model”. Note that in this case, F is the cdf of a logistic distribution.

Marginal Effects on Probability

- ▶ Probit and logit models are nonlinear models with nonlinear marginal effects on probability,

$$\frac{\partial p}{\partial x} = f(x'\beta)\beta, \quad (1)$$

where f is the pdf of the corresponding distribution.

- ▶ When $|x'\beta|$ is large, the marginal probability is small. When $x'\beta = 0$, f reaches maximum and the marginal effects on probability would be the largest.
- ▶ If x contains a binary variable d , the marginal probability is given by

$$\mathbb{P}(y = 1|x_{(d)}, d = 1) - \mathbb{P}(y = 1|x_{(d)}, d = 0).$$

where $x_{(d)}$ contains all other variables.

- ▶ However, the formula in (1) is often accurate enough for binary variables too.

Estimation of Probit and Logit Models

- ▶ We have to estimate the probit and logit models using MLE.
- ▶ The likelihood function is given by

$$p(\beta|Y, X) = \prod_{i=1}^n p_i(\beta|x_i)^{y_i} (1 - p_i(\beta|x_i))^{1-y_i},$$

where

$$p_i(\beta|x_i) = F(x_i'\beta).$$

- ▶ The log likelihood function is given by

$$\begin{aligned} \ell(\beta|Y, X) = & \sum_{i=1}^n \{y_i \log F(x_i'\beta) \\ & + (1 - y_i) \log(1 - F(x_i'\beta))\}. \end{aligned}$$

- ▶ The MLE estimator solves the following maximization problem,

$$\max_{\beta} \ell(\beta|Y, X).$$

Generalizations

- ▶ Ordinal regression
- ▶ Multinomial regression
- ▶ Generalized linear model

Ordinal Regression

- ▶ An ordinal variable is a variable whose value indicates relative ordering.
 - ▶ An example: the health status of individuals may be assigned $1, 2, \dots, 5$, corresponding to “very poor” through “very good”.
- ▶ The ordinal regression assumes that the outcome y is dependent on a latent variable y^* ,

$$y = \begin{cases} 1 & \text{if } y^* \leq \theta_1 \\ 2 & \text{if } \theta_1 < y^* \leq \theta_2 \\ \vdots & \\ k & \text{if } \theta_{k-1} < y^*, \end{cases}$$

where $\theta_1, \dots, \theta_{k-1}$ are coefficients. We may understand $\theta_0 = -\infty$ and $\theta_k = \infty$. And y^* satisfies

$$y^* = x'\beta + u,$$

where u is typically assumed to be normally or logistically distributed.

Ordinal Regression

- ▶ The conditional distribution of y is

$$\begin{aligned}P(y = k|x) &= P(\theta_{k-1} < x'\beta + u \leq \theta_k) \\ &= F(\theta_k - x'\beta) - F(\theta_{k-1} - x'\beta)\end{aligned}$$

- ▶ The log likelihood function for a single observation (y_i, x_i) is

$$\log L(\beta, \theta | x_i, y_i) = \sum_{k=1}^K I\{y_i = k\} \log (F(\theta_k - x_i'\beta) - F(\theta_{k-1} - x_i'\beta))$$

where $I\{\cdot\}$ is an indicator function.

- ▶ When $K = 2$ and let $\theta_1 = 0$, the ordinal regression reduces to logit or probit.

Multinomial Logistic Regression

- ▶ The multinomial logistic (or probit) regression can be applied when the outcome is not ordered.
 - ▶ Example: the selection of locations for a firm.
- ▶ Suppose that there are K categories for y . We use the last one as the benchmark and model the relative probabilities as

$$\begin{aligned}P(y_i = 1) &= P(y_i = K) \exp(x_i' \beta_1) \\ &\dots \\ P(y_i = K - 1) &= P(y_i = K) \exp(x_i' \beta_{K-1})\end{aligned}$$

- ▶ This implies that

$$P(y_i = k) = \frac{\exp(x_i' \beta_k)}{1 + \sum_{j=1}^{K-1} \exp(x_i' \beta_j)}, \quad k = 1, \dots, K - 1.$$

- ▶ When $K = 2$, the model reduces to the logit model.

Generalized Linear Model

- ▶ The GLM generalizes a large class of statistical models, including the classical linear regression and the logit/probit model.
- ▶ In a GLM, the dependent variables y is assumed to be generated from a particular distribution in the exponential family, which includes the normal, binomial, Poisson and gamma distributions, among others. The mean of the distribution depends on the independent variables x through:

$$\mathbb{E}(y|x) = \mu = g^{-1}(x'\beta),$$

where $g(\cdot)$ is a link function that links the linear predictor $\eta = x'\beta$ with μ , $\eta = g(\mu)$.

- ▶ Linear regression: y is normal, $\mu = x'\beta$, g is identity.
- ▶ Logit/probit model: y is binary, $\mu = F(x'\beta)$, $g = F^{-1}$.
- ▶ Poisson regression: y is Poisson, $\mu = \exp(x'\beta)$, $g = \log$.