



电子科技大学 经济与管理学院
School of Management and Economics of UESTC

计量经济学

Econometrics

任课老师：李亚静

电子科大经管学院



电子科技大学 经济与管理学院

School of Management and Economics of UESTC

第九讲 虚拟应变变量

(教材第13章)



主要内容

- ❖ 虚拟应变量的概念
- ❖ 线性概率模型
- ❖ Logit模型
- ❖ Probit模型



引例

❖ 关于大学**毕业生去向**的研究

- **就业 VS 深造**：如何选择？哪些因素影响这些选择，又怎么来研究？
- 从就业来看，是选择**一线城市**，还是**二线城市**，或者**三线城市**？
- 从深造来看，是选择**国外名校**，还是**香港高校**，还是国内**一流大学**，或者国内**高水平大学**，或者国内**一般高校**？



虚拟应变量的概念

- ❖ 虚拟变量既可以作为解释变量，又可以作为被解释变量(应变量)
- ❖ 虚拟应变量模型又称为概率模型、离散选择模型
 - 二元选择模型
 - 多元选择模型

第九讲 虚拟应变量

虚拟应变量的概念

二元
选择

高中分科
大学录取
期末考试
日常购物
.....

职业选择
高考志愿
交通工具
食堂选择
.....

多元
选择

第九讲 虚拟应变量

线性概率模型(LPM)

❖ 直接采用OLS估计虚拟应变量模型

$$D_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{其中, } D_i = \begin{cases} 1 \\ 0 \end{cases}$$

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

$$\Rightarrow E(D_i | X_i) = p_i \times 1 + (1 - p_i) \times 0 = p_i$$

$$\Rightarrow p_i = \Pr[D_i = 1] = \beta_0 + \beta_1 X_i$$

第九讲 虚拟应变量

线性概率模型(LPM)

❖ 例：研究女性是否参与劳动力市场([Table 13-1](#))

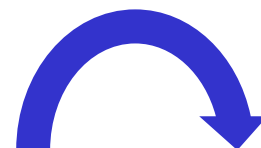
$$D_i = \beta_0 + \beta_1 M_i + \beta_2 S_i + \varepsilon_i$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.284301	0.435743	-0.652452	0.5196
M	-0.381780	0.153053	-2.494430	0.0190
S	0.093012	0.034598	2.688402	0.0121

R-squared	0.363455	Mean dependent var	0.600000
Adjusted R-squared	0.316304	S.D. dependent var	0.498273
S.E. of regression	0.412001	Akaike info criterion	1.159060
Sum squared resid	4.583121	Schwarz criterion	1.299180
Log likelihood	-14.38590	Hannan-Quinn criter.	1.203885
F-statistic	7.708257	Durbin-Watson stat	2.550725
Prob(F-statistic)	0.002247		

变量，已婚
否则取0

受教育
年限



第九讲 虚拟应变量

表13-1 女性参与劳动力市场的数据

样本观测期	D_i	M_i	A_i	S_i	\hat{D}_i
1.0	1.0	0.0	31.0	16.0	1.20
2.0	1.0	1.0	34.0	14.0	0.63
3.0	1.0	1.0	41.0	16.0	0.82
4.0	0.0	0.0	67.0	9.0	0.55
5.0	1.0	0.0	25.0	12.0	0.83
6.0	0.0	1.0	58.0	12.0	0.45
7.0	1.0	0.0	45.0	14.0	1.01
8.0	1.0	0.0	55.0	10.0	0.64
9.0	0.0	0.0	43.0	12.0	0.83
10.0	1.0	0.0	55.0	8.0	0.45
11.0	1.0	0.0	25.0	11.0	0.73
12.0	1.0	0.0	41.0	14.0	1.01
13.0	0.0	1.0	62.0	12.0	0.45
14.0	1.0	1.0	51.0	13.0	0.54
15.0	0.0	1.0	39.0	9.0	0.17
16.0	1.0	0.0	35.0	10.0	0.64
17.0	1.0	1.0	40.0	14.0	0.63
18.0	0.0	1.0	43.0	10.0	0.26
19.0	0.0	1.0	37.0	12.0	0.45
20.0	1.0	0.0	27.0	13.0	0.92
21.0	1.0	0.0	28.0	14.0	1.01
22.0	1.0	1.0	48.0	12.0	0.45
23.0	0.0	1.0	66.0	7.0	-0.01
24.0	0.0	1.0	44.0	11.0	0.35
25.0	0.0	1.0	21.0	12.0	0.45
26.0	1.0	1.0	40.0	10.0	0.26
27.0	1.0	0.0	41.0	15.0	1.11
28.0	0.0	1.0	23.0	10.0	0.26
29.0	0.0	1.0	31.0	11.0	0.35
30.0	1.0	1.0	44.0	12.0	0.45

注：本表格数据文件名为WOMEN13。



线性概率模型(LPM)

❖ LPM模型存在的问题:

- \hat{D}_i 的取值可能不在0和1之间
- 调整的判定系数 \bar{R}^2 不能准确度量模型的拟合优度
- (随机干扰项不服从正态分布且存在异方差)

$$\bar{R}_p^2 = ?$$

请仔细阅读教材p237上小节13.1.2，参考古扎拉蒂教材

二、0-1选择模型及其求解

❖ 1、0-1选择模型

假设被解释变量 Y 是虚拟变量，并且只取1或0两个值， X 为解释变量，建立的简单线性回归模型为：

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

$$Y_i = \begin{cases} 1, & Y_i \text{ 具有某种特征} \\ 0, & Y_i \text{ 不具有某种特征} \end{cases}$$

第九讲 虚拟应变量

二、0-1选择模型及其求解

❖ 1、0-1选择模型

由于被解释变量是虚拟变量，要反映它的取值特征，必须要同概率结合起来。用 p 表示被解释变量具有某种性质的发生概率，则 $1-p$ 表示被解释变量不具有某种性质的发生概率。因此，被解释变量的概率分布为：

被解释变量 Y_i ：	1	0
概 率 p_i ：	p_i	$1-p_i$

二、0-1选择模型及其求解

❖ 1、0-1选择模型

现在我们来看看被解释变量的数学期望：

$$\begin{aligned} E(Y_i) &= 1 \times p_i + 0 \times (1 - p_i) \\ &= p_i \end{aligned}$$

即被解释变量的数学期望恰好为被解释变量发生的概率。

概率的取值总是在 $[0, 1]$ 之间，所以，原则上讲，被解释变量的数学期望也应该是一个不大于1不小于0的数。

二、0-1选择模型及其求解

❖ 2、0-1选择模型的估计

以上的0-1选择线性模型虽然在外在形式上，同普通线性回归模型十分相似，但由于被解释变量是虚拟变量，因而在运用普通最小二乘法进行估计的过程中，会遇到不少新的问题。

第九讲 虚拟应变量

二、0-1选择模型及其求解

❖ 2、0-1选择模型的估计

(1) 随机误差项不服从正态分布。在被解释变量是虚拟变量时，随机误差项随之呈离散分布。

$$\varepsilon_i = Y_i - \alpha + \beta X_i \quad i = 1, 2, \dots, n$$

$$\varepsilon_i = \begin{cases} 1 - \alpha - \beta X_i & Y_i = 1 \\ -\alpha - \beta X_i & Y_i = 0 \end{cases}$$

二、0-1选择模型及其求解

❖ 2、0-1选择模型的估计

(2) 存在异方差

$$\begin{array}{ll} \text{随机项 } \varepsilon_i: & 1 - \alpha - \beta x_i \quad -\alpha - \beta x_i \\ \text{概 率:} & p_i \quad 1 - p_i \end{array}$$

$$\text{➤ } E(\varepsilon_i) = (1 - \alpha - \beta x_i) p_i + (-\alpha - \beta x_i) (1 - p_i) = p_i - \alpha - \beta x_i$$

$$Var(\varepsilon_i) = E[\varepsilon_i - E(\varepsilon_i)]^2$$

二、0-1选择模型及其求解

❖ 2、0-1选择模型的估计

(2) 存在异方差

$$Var(\varepsilon_i) = (\hat{p}_i \times (1 - \hat{p}_i))$$

如何得出？

异方差虽然不影响估计量的线性和无偏性特征，但会使由普通最小二乘法导出的估计量失去有效性。

二、0-1选择模型及其求解

❖ 2、0-1选择模型的估计

(3) $0 \leq E(Y_i) \leq 1$ 并不总是能保证成立。

(4) 拟合优度不可能很好

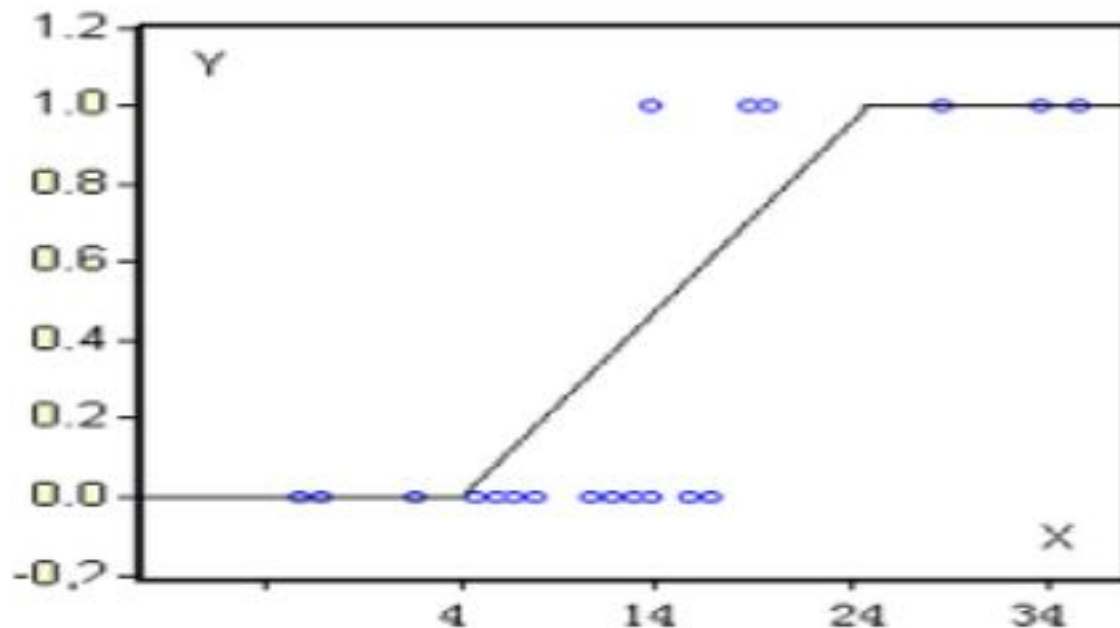
由于被解释变量是虚拟变量，发生就是1，不发生就是0，由 (X, Y) 绘制成的散点图其散点要么落在x轴上，要么落在平行于x轴的另一条直线上。这样，无论用什么样的函数方程去拟合这些数据，都不可能获得一个拟合优度系数较大的值。所以，对于被解释变量是虚拟变量的回归模型，不宜用来评价模型的好坏。

第九讲 虚拟应变量

二、0-1选择模型及其求解

❖ 2、0-1选择模型的估计

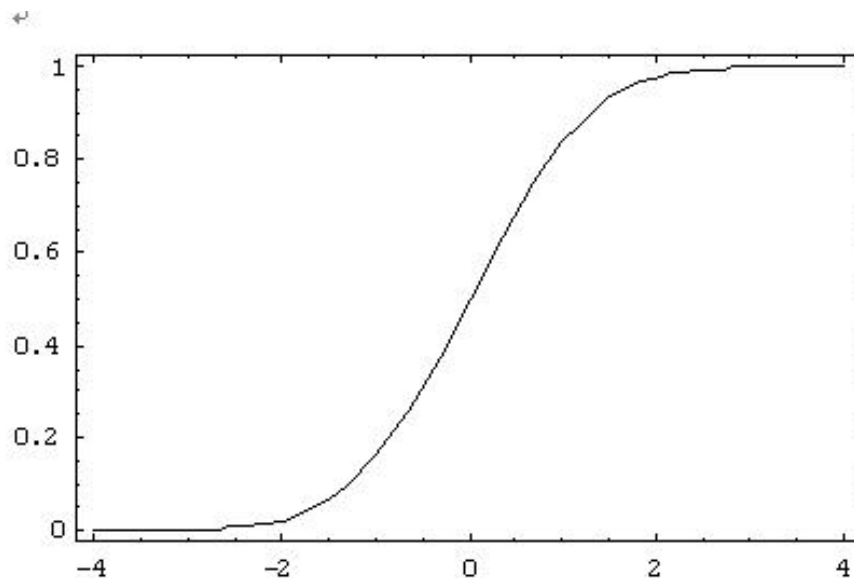
(4) 拟合优度不可能很好



第九讲 虚拟应变量

思考题

❖ 女性工作的概率会随受教育年限而**线性**增加吗？



累积
分布
函数
CDF

❖ 这种模型如何估计？能避免LPM模型的问题吗？

第九讲 虚拟应变量

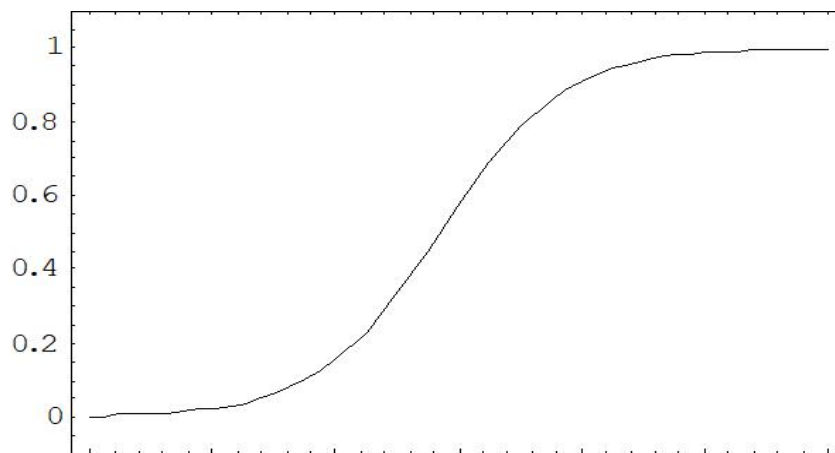
- ❖ 由于线性概率模型的上述缺点，希望能找到一种变换方法：
- ❖ (1) 使解释变量 X_i 所对应的所有预测值（概率值）都落在 $(0, 1)$ 之间。
- ❖ (2) 同时对于所有的 X_i ，当 X_i 增加时，希望 Y_i 也单调增加或单调减少。
- ❖ 累积概率分布函数 $F(z_i)$ 满足要求。采用累积正态概率分布函数的模型称作**Probit模型**。用**正态分布的累积概率作为Probit模型的预测概率**。
- ❖ **logistic函数**也能满足这样的要求。采用**logistic函数的模型称作logit模型**。

第九讲 虚拟应变量

- ❖ Logistic函数是一种S函数，由皮埃尔·弗朗索瓦·韦吕勒（1844?）提出且命名。
- ❖ 广义Logistic曲线可模仿一些情况人口增长（P）的S形曲线。

❖ Logistic 函数形式：

$$Y(X) = \frac{1}{1 + e^{-X}}$$



logistic函数曲线

0-1选择模型及其求解

❖ 3、LOGIT模型

(1) 定义与解释

LOGIT模型是McFadden于1973年首次提出。其采用的是logistic概率分布函数。其形式是：

$$D_i = \frac{1}{1 + e^{-y_i}} = \frac{1}{1 + e^{-[\beta_0 + \beta_1 X_i + \varepsilon_i]}}$$

第九讲 虚拟应变量



Logit模型

- ❖ 不直接对虚拟应变量 D 建模，而是引入一个无约束的变量 y ，令

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

- ❖ 第一种方式：
$$D_i = \frac{1}{1 + e^{-y_i}} = \frac{1}{1 + e^{-[\beta_0 + \beta_1 X_i + \varepsilon_i]}}$$

Logistic分布CDF→Logit模型

累计分布函数CDF (Cumulative Distribution Function)



Logit模型

Logit模型:

$$\diamond \mathbf{L} = \ln\left(\frac{P_i}{1 - P_i}\right) = \alpha + \beta X_i + \varepsilon_i$$

Logit模型标准范式: $L: \Pr(D_i = 1) = \beta_0 + \beta_1 X_i + \varepsilon_i$

$\Pr(D_i = 1)$ 刻画的是 $D_i = 1$ 的概率

二、0-1选择模型及其求解

❖ 3、LOGIT模型

(2) 求解方法--极大似然估计法

$$\begin{aligned} L(\alpha, \beta) &= \prod_{i=1}^n p_i^{Y_i} (1 - p_i)^{1-Y_i} \\ &= \prod_{i=1}^n (F(z_i))^{Y_i} (1 - F(z_i))^{1-Y_i} \\ &= \prod_{i=1}^n (F(\alpha + \beta x_i))^{Y_i} (1 - F(\alpha + \beta x_i))^{1-Y_i} \end{aligned}$$

二、0-1选择模型及其求解

❖ 3、LOGIT模型

(2) 求解方法

$$\begin{aligned}\ln L(\alpha, \beta) &= \ln \prod_{i=1}^n (F(\alpha + \beta X_i))^{Y_i} (1 - F(\alpha + \beta X_i))^{1-Y_i} \\ &= \sum_{i=1}^n Y_i \ln F(\alpha + \beta X_i) + \sum_{i=1}^n (1 - Y_i) \ln(1 - F(\alpha + \beta X_i))\end{aligned}$$

第九讲 虚拟应变量

❖ 3、LOGIT模型

(2) 求解方法

$$\begin{aligned}\frac{\partial \ln L(\alpha, \beta)}{\partial \alpha} &= \sum_{i=1}^n Y_i \times \frac{F'(\alpha + \beta X_i)}{F(\alpha + \beta X_i)} \\ &\quad - \sum_{i=1}^n (1 - Y_i) \times \frac{F'(\alpha + \beta X_i)}{1 - F(\alpha + \beta X_i)} = 0\end{aligned}$$

$$\begin{aligned}\frac{\partial \ln L(\alpha, \beta)}{\partial \beta} &= \sum_{i=1}^n Y_i \times \frac{F'(\alpha + \beta X_i) X_i}{F(\alpha + \beta X_i)} \\ &\quad - \sum_{i=1}^n (1 - Y_i) \times \frac{F'(\alpha + \beta X_i) X_i}{1 - F(\alpha + \beta X_i)} = 0\end{aligned}$$

第九讲 虚拟应变量

Logit模型

$$D_i = \frac{1}{1 + e^{-y_i}} \Rightarrow \frac{1}{e^{y_i}} = \frac{1}{D_i} - 1 \Rightarrow e^{y_i} = \frac{D_i}{1 - D_i}$$

$$\Rightarrow \ln\left(\frac{D_i}{1 - D_i}\right) = y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

机会比率
(odds ratio)

Logit模型标准范式: $L: \Pr(D_i = 1) = \beta_0 + \beta_1 X_i + \varepsilon_i$

第九讲 虚拟应变量

❖ LOGIT模型的估计:

方法：极大似然估计法

$$\begin{aligned} L(\alpha, \beta) &= \prod_{i=1}^n p_i^{Y_i} (1 - p_i)^{1-Y_i} \\ &= \prod_{i=1}^n (F(z_i))^{Y_i} (1 - F(z_i))^{1-Y_i} \\ &= \prod_{i=1}^n (F(\alpha + \beta x_i))^{Y_i} (1 - F(\alpha + \beta x_i))^{1-Y_i} \end{aligned}$$

第九讲 虚拟应变量

线性概率模型(LPM)

❖ 例：研究女性是否参与劳动力市场(Table 13-1)

$$D_i = \beta_0 + \beta_1 M_i + \beta_2 S_i + \varepsilon_i$$

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.284301	0.435743	-0.652452	0.5196
M	-0.381780	0.153053	-2.494430	0.0190
S	0.093012	0.034598	2.688402	0.0121

R-squared	0.363455	Mean dependent var	0.600000
Adjusted R-squared	0.316304	S.D. dependent var	0.498273
S.E. of regression	0.412001	Akaike info criterion	1.159060
Sum squared resid	4.583121	Schwarz criterion	1.299180
Log likelihood	-14.38590	Hannan-Quinn criter.	1.203885
F-statistic	7.708257	Durbin-Watson stat	2.550725
Prob(F-statistic)	0.002247		

变量，已婚
否则取0

受教育
年限

第九讲 虚拟应变量

Logit模型

❖ 例：研究女性是否参与劳动力市场(Table 13-1)

Equation Estimation

Specification Options

Equation specification

Binary dependent variable followed by list of regressors, OR a linear explicit equation like

labor c m s

Binary estimation ☐ Probit ☒ Logit ☐ Extreme value

Estimation settings

Method: BINARY - Binary Choice (Logit, Probit, Extreme Value)

Sample: 1 30

确定 取消

虚拟变量，
取1，否则

教育
年限

第九讲 虚拟应变量

❖ LOGIT模型的统计推断:

因为估计方法是**最大似然法**，所以统计性质一般是大样本性质，即估计的标准误差是**渐近性的**。

对系数的统计显著性使用（标准正态）**Z统计量**。

模型的拟合程度的度量使用McFadden的 R^2 ，记为 R^2_{McF} 。

模型的总显著性检验使用似然比（LR）统计量。
 $LR \sim \chi^2(k)$ ， k 为解释变量个数。

Logit模型

Sample: 1 30

Included observations: 30

Convergence achieved after 5 iterations

Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-5.895933	3.324732	-1.773356	0.0762
M	-2.586110	1.180162	-2.191317	0.0284
S	0.690368	0.315828	2.185898	0.0288
McFadden R-squared	0.342412	Mean dependent var	0.600000	
S.D. dependent var	0.498273	S.E. of regression	0.399177	
Akaike info criterion	1.085128	Sum squared resid	4.302237	
Schwarz criterion	1.225248	Log likelihood	-13.27693	
Hannan-Quinn criter.	1.129954	Deviance	26.55385	
Restr. deviance	40.38070	Restr. log likelihood	-20.19035	
LR statistic	13.82685	Avg. log likelihood	-0.442564	
Prob(LR statistic)	0.000994			
Obs with Dep=0	12	Total obs	30	
Obs with Dep=1	18			

第九讲 虚拟应变量

Logit模型

❖ 参数估计值的解释：以受教育年限S变量为例

➤ 1、采用机会比率：

$$\ln(P_i/1-P_i)=0.69$$

$$P_i/1-P_i=\exp(0.69)=1.99$$

受教育年限每增加1年，女性参加工作的机会是原来的1.99倍

Sample: 1 30

Included observations: 30

Convergence achieved after 5 iterations

Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z
C	-5.895933	3.324732	-1.773
M	-2.586110	1.180162	-2.187
S	0.690368	0.315828	2.187

第九讲 虚拟应变量

Logit模型

❖ 参数估计值的解释

❖ 2、采用偏导数：

$$\frac{\partial \hat{P}_i}{\partial X_i} = \frac{\partial [1 / (1 + e^{-y_i})]}{\partial X_i} = (1 + e^{-y_i})^{-2} e^{-y_i} \frac{\partial y_i}{\partial X_i}$$

$$= \frac{1}{1 + e^{-y_i}} \frac{e^{-y_i}}{1 + e^{-y_i}} \hat{\beta}_1 = \hat{\beta} \hat{P}_i (1 - \hat{P}_i)$$

✓ 受教育年限每增加1年，女性参加工作的概率增加0.17 (即 $0.69 * 0.6 * 0.4$)



Logit模型

❖ 参数估计值的解释

➤ 3、采用经验法则： $\hat{D}_i = 0.5$

✓ 受教育年限每增加1年，女性参加工作的概率增加0.17 (即 $0.69*0.5*0.5$)

4、Probit模型

- ❖ 不直接对虚拟应变量 D 建模，而是引入一个无约束的变量 y ，令

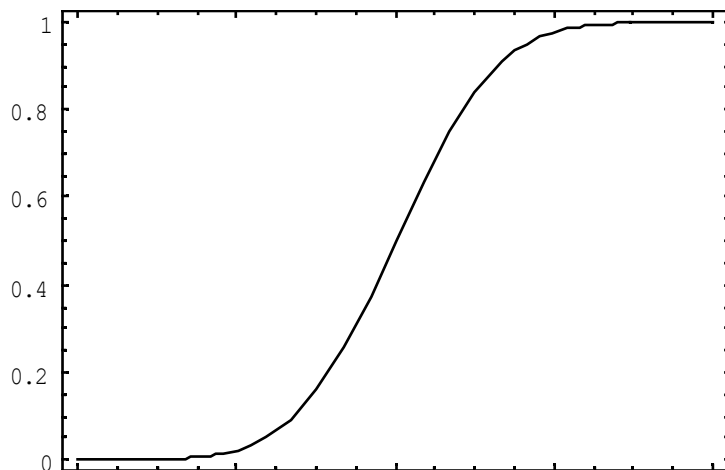
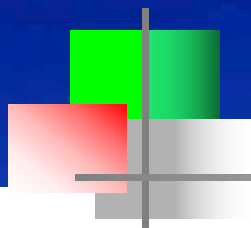
$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

- ❖ 第二种方式：
$$D_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y_i} e^{-s^2/2} ds$$

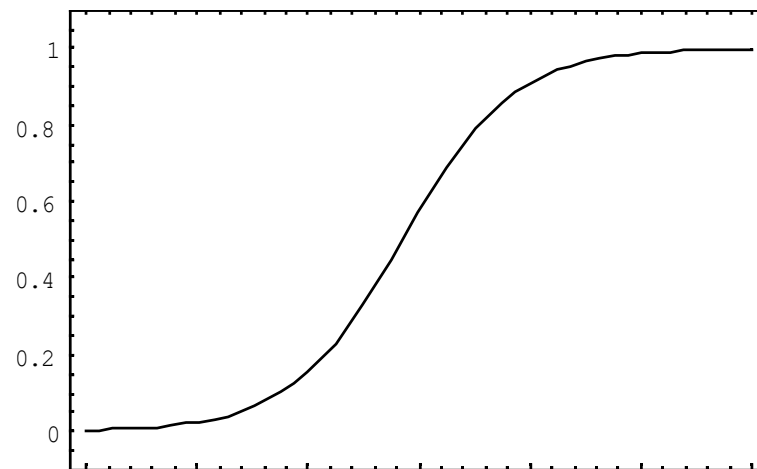
正态分布CDF→Probit模型

Probit模型标准范式：
$$Z_i = \Phi^{-1}(D_i) = \beta_0 + \beta_1 X_i + \varepsilon_i$$

第九讲 虚拟应变量



累积正态概率分布曲线



logistic函数曲线

怎样估计Probit模型？

极大似然法

第九讲 虚拟应变量

❖ PROBIT模型

(2) 求解方法

在正态分布假定下，构造似然函数：

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \\ &= \prod_{i=1}^n (F(z_i))^{y_i} (1 - F(z_i))^{1-y_i} \\ &= \prod_{i=1}^n (F(\beta_0 + \sum_{j=1}^k \beta_j x_{ji}))^{y_i} (1 - F(\beta_0 + \sum_{j=1}^k \beta_j x_{ji}))^{1-y_i} \end{aligned}$$

第九讲 虚拟应变量

Probit模型

❖ 例：研究女性是否参与劳动力市场(Table 13-1)

Equation Estimation

Specification Options

Equation specification

Binary dependent variable followed by list of regressors, OR a linear explicit equation like

labor c m s

Binary estimation ☒ Probit ☐ Logit ☐ Extreme value

Estimation settings

Method: BINARY - Binary Choice (Logit, Probit, Extreme Value)

Sample: 1 30

确定 取消

虚拟变量，
取1，否则

教育
年限

Probit模型

Sample: 1 30

Included observations: 30

Convergence achieved after 5 iterations

Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-3.439215	1.870689	-1.838475	0.0660
M	-1.443938	0.622521	-2.319499	0.0204
S	0.397604	0.171111	2.323661	0.0201
McFadden R-squared	0.341228	Mean dependent var	0.600000	
S.D. dependent var	0.498273	S.E. of regression	0.401104	
Akaike info criterion	1.086722	Sum squared resid	4.343884	
Schwarz criterion	1.226842	Log likelihood	-13.30083	
Hannan-Quinn criter.	1.131548	Deviance	26.60167	
Restr. deviance	40.38070	Restr. log likelihood	-20.19035	
LR statistic	13.77903	Avg. log likelihood	-0.443361	
Prob(LR statistic)	0.001018			
Obs with Dep=0	12	Total obs	30	
Obs with Dep=1	18			

Logit

Variable	Coefficient	Std. Error	z-Statistic	Prob.
	-5.895933	3.324732	-1.773356	0.0762
	-2.586110	1.180162	-2.191317	0.0284
	0.690368	0.315828	2.185898	0.0288

这两组
系数不
能直接
相比

Probit

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-3.439215	1.870689	-1.838475	0.0660
M	-1.443938	0.622521	-2.319499	0.0204
S	0.397604	0.171111	2.323661	0.0201

标准正态分布的方差为1，而标准逻辑分布的方差为 $\pi^2/3$ 。因此将Probit系数乘以 $\pi\sqrt{3}$ （约1.81）就会近似得到Logit系数（相反约为0.55）

第九讲 虚拟应变量

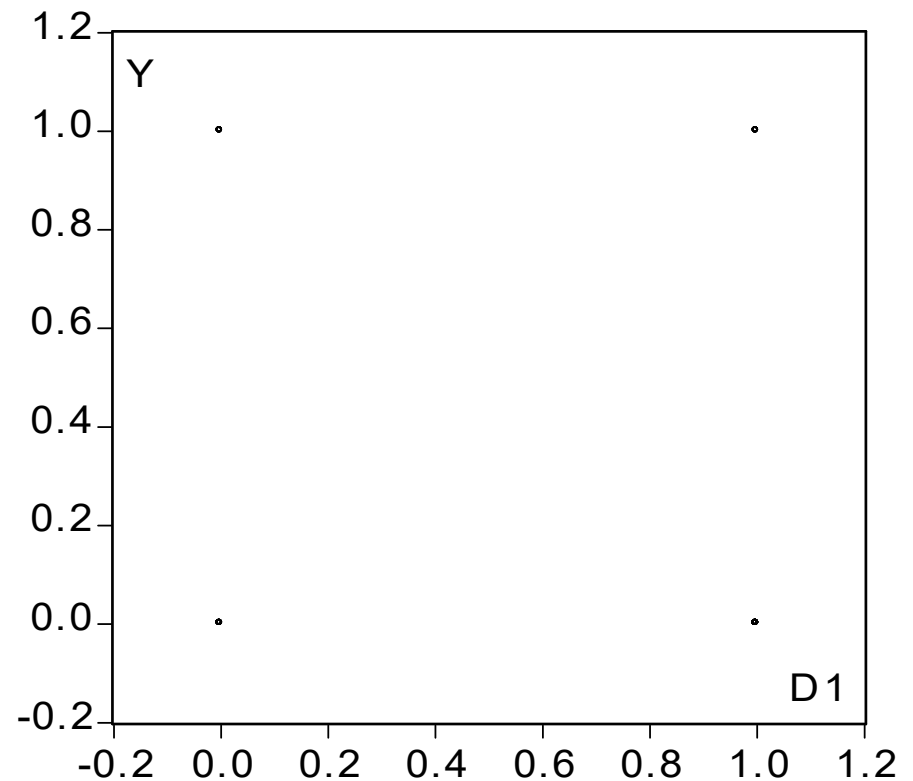
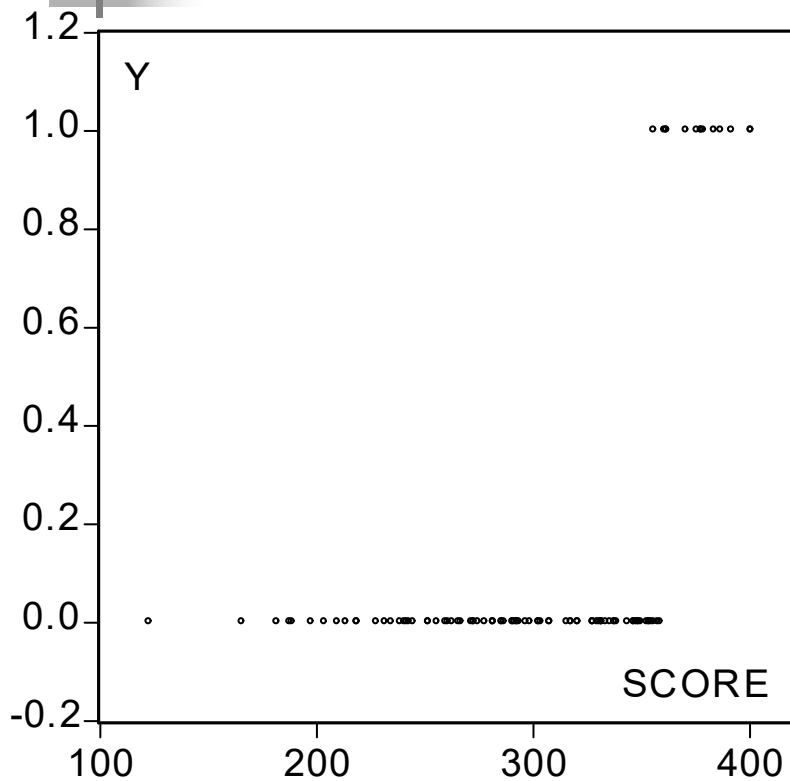
- ❖ 例1：南开大学国际经济研究所1999级研究生考试分数及录取情况见数据表 ($N=95$)。定义变量 Y ：考生录取为1，未录取为0；SCORE：考生考试分数；虚拟变量D1：应届生为1，非应届生为0。研究应届生与非应届生是否会影响考生的录取？

表 2

数据表⁴⁾

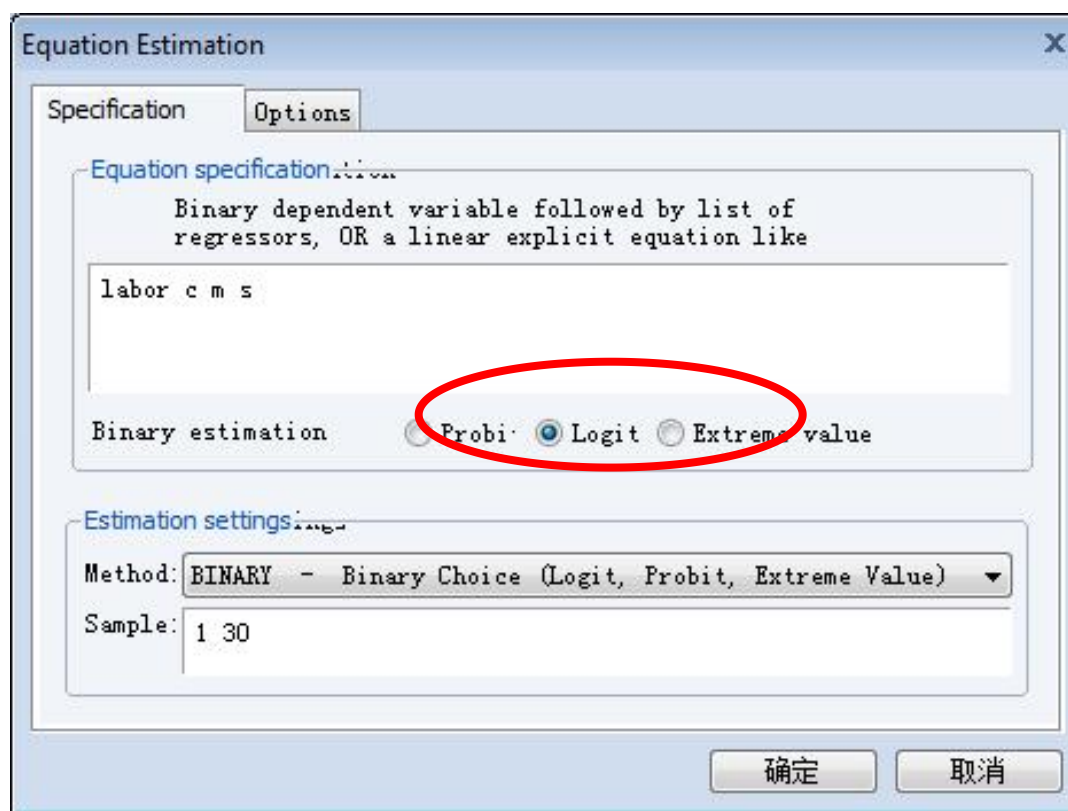
obs ⁴⁾	Y ⁴⁾	SCORE ⁴⁾	D1 ⁴⁾	obs ⁴⁾	Y ⁴⁾	SCORE ⁴⁾	D1 ⁴⁾	obs ⁴⁾	Y ⁴⁾	SCORE ⁴⁾	D1 ⁴⁾
1.	1.	401.	1.	34.	0.	332.	1.	67.	0.	275.	0.
2.	1.	401.	0.	35.	0.	332.	1.	68.	0.	273.	0.
3.	1.	392.	1.	36.	0.	332.	1.	69.	0.	273.	1.
4.	1.	387.	0.	37.	0.	331.	1.	70.	0.	272.	1.
5.	1.	384.	1.	38.	0.	330.	1.	71.	0.	267.	0.
6.	1.	379.	0.	39.	0.	328.	1.	72.	0.	266.	1.
7.	1.	378.	0.	40.	0.	328.	1.	73.	0.	263.	1.
8.	1.	378.	0.	41.	0.	328.	1.	74.	0.	261.	1.
9.	1.	376.	1.	42.	0.	321.	1.	75.	0.	260.	0.
10.	1.	371.	0.	43.	0.	321.	1.	76.	0.	256.	0.
11.	1.	362.	0.	44.	0.	318.	1.	77.	0.	252.	0.
12.	1.	362.	1.	45.	0.	318.	0.	78.	0.	252.	1.
13.	1.	361.	1.	46.	0.	316.	1.	79.	0.	245.	1.
14.	0.	359.	1.	47.	0.	308.	0.	80.	0.	243.	1.
15.	0.	358.	1.	48.	0.	308.	1.	81.	0.	242.	0.
16.	1.	356.	1.	49.	0.	304.	0.	82.	0.	241.	0.
17.	0.	356.	1.	50.	0.	303.	1.	83.	0.	239.	1.
18.	0.	355.	1.	51.	0.	303.	1.	84.	0.	235.	0.
19.	0.	354.	1.	52.	0.	299.	1.	85.	0.	232.	0.
20.	0.	354.	0.	53.	0.	297.	1.	86.	0.	228.	1.
21.	0.	353.	1.	54.	0.	294.	0.	87.	0.	219.	1.
22.	0.	350.	0.	55.	0.	293.	1.	88.	0.	219.	1.
23.	0.	349.	0.	56.	0.	293.	1.	89.	0.	214.	1.
24.	0.	349.	0.	57.	0.	292.	0.	90.	0.	210.	1.
25.	0.	348.	1.	58.	0.	291.	1.	91.	0.	204.	1.
26.	0.	347.	1.	59.	0.	291.	1.	92.	0.	198.	0.
27.	0.	347.	1.	60.	0.	287.	1.	93.	0.	189.	1.
28.	0.	344.	1.	61.	0.	286.	1.	94.	0.	188.	1.
29.	0.	339.	1.	62.	0.	286.	0.	95.	0.	182.	1.
30.	0.	338.	0.	63.	0.	282.	1.	96.	0.	166.	1.
31.	0.	338.	1.	64.	0.	282.	1.	97.	0.	123.	0.
32.	0.	336.	1.	65.	0.	282.	0.

第九讲 虚拟应变量



第九讲 虚拟应变量

Logit模型



得Logit模型估计结果如下

变量

Dependent Variable: Y
Method: ML - Binary Logit
Date: 09/16/03 Time: 10:00:00
Sample: 1 97
Included observations: 97
Convergence achieved in 10 iterations
Covariance matrix computed

因为D1的系数没有显著性。说明“应届生”和“非应届生”不是决定是否录取的重要因素。

Variable			z-Statistic	Prob.
C	2.4576	124.5164	-1.947194	0.0515
SCORE	0.677061	0.348030	1.945409	0.0517
D1	-0.476605	2.984542	-0.159691	0.8731

Mean dependent var	0.144330	S.D. dependent var	0.353250
S.E. of regression	0.115377	Akaike info criterion	0.143907
Sum squared resid	1.251316	Schwarz criterion	0.223537
Log likelihood	-3.979482	Hannan-Quinn criter.	0.176105
Restr. log likelihood	-40.03639	Avg. log likelihood	-0.04102
LR statistic (2 df)	72.11382	McFadden R-squared	0.90060
Probability(LR stat)	2.22E-16		

Obs with Dep=0	83	Total obs	97
Obs with Dep=1	14		

第九讲 虚拟应变量

❖ 得Logit模型估计结果如下

$$❖ \quad p_i = F(y_i) = \frac{1}{1 + e^{-(-243.7362 + 0.6794x_i)}}$$

❖ 注意：表达式中指数的写法。

其它虚拟应变量模型

❖ 多元Logit模型 (multinomial logit model)

- 被解释变量存在多种离散且无序的选择
- 如择业的选择、交通工具的选择
- 采用极大似然估计法，Eviews中有示例程序

❖ 序次Logit模型 (ordered logit model)

- 被解释变量存在多种离散且有序的选择
- 如债券的信用评级、舆论调查结果
- 采用极大似然估计法，Eviews中可实现菜单操作



本讲小结

- ❖ 当被解释变量是虚拟变量时，能否采用线性概率模型(OLS估计方法)?
- ❖ Logit和Probit模型可用于估计被解释变量是二分应变量(二元选择)的问题
- ❖ 在Logit模型中，斜率系数表示什么意思?
- ❖ 在被解释变量是多元选择问题中，可以采用多元Logit模型或序次Logit模型

第九讲 虚拟应变量



作业

❖ 第13章习题: 2、4、5



小组作业

- ❖ 提交时间：
- ❖ 第十三周周五
- ❖ 小组作业可选题—2023软件学院班级.docx