

# Probit与logit模型的应用

## 以model 1 example为例

刘晶芳 (15320171151900) 余星月 (15320171151888)

### 一、多项 probit/logit模型

模型设置：个体面临的选择有时是多值的，而不仅仅是二值的。例如课堂中Ketchup的例子。对于解释变量只随个体*i*而变，不随方案*j*而变。我们可以应用以下公式：

$$U_{ij} = \alpha + \gamma_j income_i + \varepsilon_{ij}$$

通过推导，我们可以得到

$$P(y_i = j | x_i) = \frac{\exp(x_i' \beta_j)}{\sum_{k=1}^J \exp(x_i' \beta_k)}$$

假设 $\{\varepsilon_{ij}\}$ 为iid且logit模型中服从I型极值分布；probit模型中服从正态分布。

#### (一) 多项 probit 模型

##### (1) model 1 example 综述

数据集来源：链接为<https://l.xmu.edu.cn/mod/folder/view.php?id=37114>

解释变量：income

被解释变量：choice，有四个选择可选，分别为：heinz、hunts、delmonte、stb

描述性分析：

Ketchup数据存储格式为浮点型存储，9个字符的一般格式。

storage variable	display name	value type	format	label
choice	float	%9.0g	choice	the brand purchased
income	float	%9.0g		

数据观测值为800，其中，choice变量的均值为2.1，标准差为0.801，最小值为1，最大值为4，符合我们对choice标签中的定义。Income变量的均值为52.8，标准差12.17，最大值为89，最小值为25。

Variable	Obs	Mean	Std.Dev.	Min	Max
choice	800	2.100	0.801	1	4
income	800	52.80	12.17	24.92	86.92

四个品牌中，可以看出在总样本中，heinz所选择的比重最高，占比51.13%，即有一半人选择该品牌。商店自营品牌以及hunts选择比例相近。

choice	Freq.	Percent	Cum.
stb	177	22.13	22.13
heinz	409	51.13	73.25
hunts	171	21.38	94.63
delmonte	43	5.380	100
Total	800	100	

(2) 多元probit模型（以stb为basicoutcome）

Probit回归结果				
VARIABLES	(1) stb	(2) heinz	(3) hunts	(4) delmonte
o.income	-			
o._cons	0 (0)			
income		0.204*** (0.0135)	0.124*** (0.0128)	0.0936*** (0.0153)
Constant		-9.208*** (0.616)	-5.554*** (0.566)	-5.014*** (0.690)
Observations	800	800	800	800

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.

## (二) 多项logit模型

(1) 被解释变量: choice

解释变量: income

(3) 多项logit模型（以stb 为basic outcome）

Logit回归结果				
VARIABLES	(1) stb	(2) heinz	(3) hunts	(4) delmonte
o.income	-			
o._cons	0 (0)			
income		0.276*** (0.0199)	0.172*** (0.0187)	0.117*** (0.0240)
Constant		-12.54*** (0.907)	-7.658*** (0.822)	-6.396*** (1.071)
Observations	800	800	800	800

Standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.

结果分析：我们考虑choice只随个体而改变而不随方案改变的情况。当仅考虑choice仅受个体变量，即income的效应，我们得到了显著性表现良好的答案。可以看出，相对于超市自营产品，heinz这个品牌受收入的影响相对其他品牌更大。且两种模型的表现一致。

## 二、 条件logit模型

- (1) 模型设置：与上述多项logit模型不同，有些解释变量可能及随个体改变，也随方案改变。这种解释变量既包括随方案与个体而变的变量，也包括随方案而变但不随个体而变的变量。我们将此类问题模型称为条件logit模型。
- (2) 被解释变量：choice  
解释变量：price

VARIABLES	(1) choice
2.brand	2.423*** (0.145)
3.brand	1.718*** (0.159)
4.brand	0.716*** (0.216)
price	-4.518*** (0.285)
Observations	3,200
Standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

结果分析：我们这里考虑choice只受方案而变的情况，即这里仅考虑品牌价格与选择之间的关系，因此只考虑从品牌价格不考虑income。上表显示，如果income的解释变量取值相同，则最优可能选择的是heinz，但此为非线性模型，故不易通过系数来评价边际效应。只能比较品牌选择受价格的影响程度大小。

## 三、 混合probit/logit模型

模型设置 上述分别考虑了解释变量不随方案而变的多项logit/probit模型，以及解释变量随方案而变的条件logit模型。大多数情况下，我们要考虑两者混合的模型。模型表示为：

$$U_{ij} = \alpha + \delta \text{price}_{ij} + \gamma_j \text{income}_i + \varepsilon_{ij}$$

### (一) 混合probit模型

- (1) 被解释变量: choice  
解释变量: income、price
- (2)

混合probit回归				
VARIABLES	(1) brand	(2) heinz	(3) hunts	(4) delmonte
income		0.200*** (0.0177)	0.134*** (0.0197)	0.0333 (0.0243)
price	-3.424*** (0.529)			
Constant		-7.822*** (0.725)	-4.834*** (0.903)	-1.077 (1.012)
Observations	3,200	3,200	3,200	3,200

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## (二) 混合Logit模型

- (1) 被解释变量: choice  
解释变量: income、price
- (2)

Logit回归				
VARIABLES	(1) brand	(2) heinz	(3) hunts	(4) delmonte
income		0.277*** (0.0209)	0.180*** (0.0198)	0.107*** (0.0258)
price	-4.418*** (0.330)			
Constant		-10.89*** (0.946)	-6.305*** (0.872)	-3.832*** (1.169)
Observations	3,200	3,200	3,200	3,200

Standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

结果分析: 我们考虑choice随个体和方案同时改变的情况。以上两个表格, 与R下的结果基本保持一致: 在Probit模型下, 在品牌为delmonte下的income、constant为不显著, 其余为显著; 在Logit模型下结果均为显著。相比于基本超市品牌而言, 顾客选择购买heinz受收入的影响会大于其他两种品牌 (hunts and delmonte)。对比两个表格, Logit和Probit模型的实证结果并不保持一致 (至少对于delmonte而言probit模型表现为不显著), 这表明, 二者之间存在差异。