耐久消費財 保有函數의 推定:
二進數 從屬變數를 利用한 回歸分析

尹 錫 範*
李 會 京**

1. 序 論

本論文에서는 첫째로 單一方程式 模型에서 從屬變數가 兩者擇一(binary choice)의 離散確率變數일 때 이러한 二進的 從屬變數(binary dependent variable)의 變動을 說明하는데 使用되는 몇 가지 模型을 紹介하고 各各의 表記 및 推定方法, 推定量의 性質, 豫測 및 檢定問題등에 關하여 比較 敍述하고자 한다. 둘째, 從屬變數가 離散과 連續의 混合形態일 때 앞에 紹介된 模型이 어떻게 適用될 수 있는가를 살펴보며, 셋째, 選擇對象 및 從屬變數의 數가 增加하여 一般化된 選多型模型(multiple choice model)의 경우, 表記 및 推定方法을 單一方程式 技法을 利用하여 追加로 總覽하고자 한다. 넷째, 本論文에서는 또한 耐久消費財 購入에 關한 調査資料를 利用하여 實際 많이 使用되는 몇 個의 模型을 選擇하여 適用하고 各各의 豫測力을 分析함으로써 各 模型을 比較 檢討하는데 目的이 있다.

從屬變數가 二進的일 때 模型의 表記 및 推定을 爲한 接近方法으로, 線型確率模型(linear probability model), 프로빗模型(probit model), 로짓模型(logit model) 및 곰핏模型(Gompit model)을 들고 各 方法을 比較한다. 또 線型確率模型의 變換에 의한 接近方法을 敍述한 뒤 實際의 行動様式을 說明하고 豫測할 수 있는 模型을 設定하고, 各 方法을 代案的으로 適用하여 그 結果를 比較 檢討함으로써 各 模型의 特徵 및 推定方法의 適合性 與否를 說明하고자 한다. 그리고 理論的인 接近時 처음에는 單純回歸模型에서 시작하여 一般的인 多變量 경우로 擴張하기로 한다.

한편 二進數模型을 一般化한 選多型模型에서는 線型確率模型, 프로빗模型 및 로짓模型에 의한 分析을 通하여 各各의 表記 및 推定方法을 概觀하고자 한다.

* 延世大學校 商經大學 敎授
** 延世大學校 大學院 應用統計學科

實證的인 分析에서는 適用模型을 計算이 容易한 二進數 選擇模型으로만 局限하고, 實際로는 別로 使用되지 않는 곰퓟模型을 除外한 나머지 세가지 模型을 適用하여 調査資料를 分析하기로 한다.

2. 二進數模型의 槪觀

2. 1 線型確率模型(linear probability model)

다음과 같은 單純回歸模型을 假定하자.

$$y_i=\alpha+\beta x_i+u_i \tag{2.1}$$

단, $y_i=\begin{cases}1 & \text{첫째 對象을 선택하면}\\ 0 & \text{그렇지 않으면}\end{cases}$

$$E(u_i)=0$$

$$E(u_iu_j)=\begin{cases}\sigma_i^2 & i=j\\ 0 & i\neq j\end{cases}$$

이때 x_i가 주어졌을 때 y_i의 期待値는

$$E(y_i|x_i)=\alpha+\beta x_i \tag{2.2}$$

가 되며 $y_i=1$일 確率을 p_i, $y_i=0$일 確率을 $(1-p_i)$라 하면 $E(y_i|x_i)$는

$$\begin{aligned}E(y_i|x_i)&=1\cdot p_i+0\cdot(1-p_i)\\&=p_i\end{aligned} \tag{2.3}$$

와 같다. 卽 從屬變數 y_i가 1 또는 0의 값만 取한다는 特性때문에 y의 條件附 期待値는 條件附 確率로 解釋될 수 있다.[1] 따라서 線型確率模型은 다음과 같은 形態를 取하며 이는 그림 1과 같다.[2]

$$E(y_i|x_i)=p_i=\begin{cases}\alpha+\beta x_i & 0<\alpha+\beta x_i<1\\ 1 & \alpha+\beta x_i\geqq 1\\ 0 & \alpha+\beta x_i\leqq 0\end{cases} \tag{2.4}$$

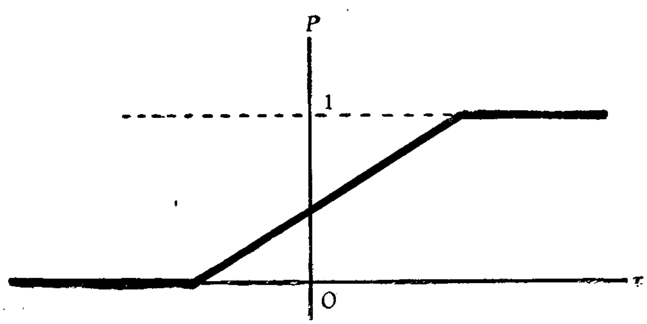


그림 1. 線型確率模型

1) Arthur S. Goldberger, *Econometric Theory* (New York: John Wiley & Sons 1964), p.249.
2) Robert S. Pindyck and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasts*(New York: McGraw-Hill, 1976), pp.239—240.

가. 線型確率模型이 갖는 문제점

1) 不均等分散(heteroscedasticity)

$$\begin{aligned}\mathrm{Var}(u_i)=E(u_i^2)&=(1-\alpha-\beta x_i)^2 p_i+(-\alpha-\beta x_i)^2(1-p_i)\\&=p_i(1-p_i)\\&=E(y_i|x_i)[1-E(y_i|x_i)]\end{aligned} \tag{2.5}$$

式 (2.5)에서와 같이 u_i의 分散은 x_i에 따라 變하며 p_i가 0, 1에 接近할 수록 작아지고 $p_i=\frac{1}{2}$ 일때 제일 커진다.[1]

2) 豫測值의 (0, 1)區間離脫

通常最小自乘法(ordinary least squares method)에 의하여 線型確率模型을 推定하였을 때 그 豫測値가 그림 2와 같이 (0, 1)區間을 벗어날 수 있다.

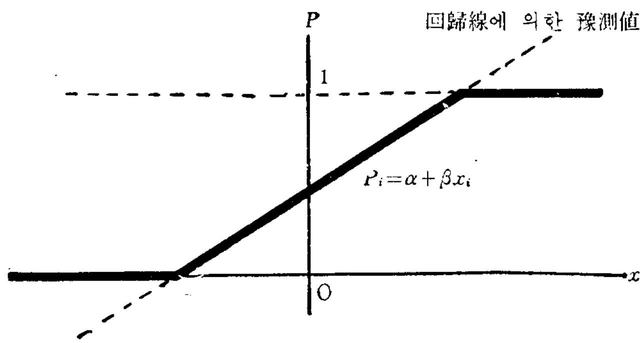


그림 2. 線型確率模型에 의한 豫測

이렇게 되면 첫째 $E(y|x)$를 確率로 說明한 것과 矛盾되며, 둘째 다음의 推定절차에서 나오는 바와 같이 一般化最小自乘法(generalized least squares methed)의 使用이 制限된다.[2]

이러한 점을 해결하기 위하여 x의 어떤 범위에서 豫測確率을 1 또는 0 이라 固定시킨다면, 實際는 일어나지 않을 수 있는데도 불구하고 어떠한 誤差도 없이 틀림없이 일어난다고 豫測하거나 或은 그 반대의 경우를 의미하므로 이러한 推定過程에서 얻어진 豫測値는 偏倚를 갖는다.[3] 또는 制約條件 $0<\hat{y}_i<1$ ($\hat{y}_i$는 y_i의 推定値) 아래서 數理計劃(mathemati-

1) u_i의 確率分布는 다음과 같다.

	u_i	u_i의 確率
$y_i=1$	$1-\alpha-\beta x_i$	p_i
$y_i=0$	$-\alpha-\beta x_i$	$1-p_i$

2) S.M. Goldfeld and R.E. Quandt, *Nonlinear Methods in Econometrics* (Amsterdam: North-Holland, 1972), pp.124—134.

3) 不等式의 制約條件(inequality constraint) 아래서의 이 方法에 의하면 推定量의 分散을 줄일 수는 있어도 不偏性이 保障되지 못하므로 計算의 複雜함을 감안할 때 通常最小自乘推定量이 適切할 수도 있다. R.S. Pindyck and D.L. Rubinfeld, *op. cit.* p.242 參照.

cal programming)의 技法을 使用하여 이 문제를 다룰 수도 있으나 이 推定量이 不偏이라는 보장은 없다.

나. 推　定

一般的인 模型 $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{u}$에서 通常最小自乘推定法(ordinary least squares method), 一般化最小自乘推定法(generalized least squares method)및 最尤推定法(maximum likelihood method)은 다음과 같다.

단, $\boldsymbol{Y}=n\times 1$

$\boldsymbol{X}=n\times k$

$\boldsymbol{\beta}=k\times 1$

$\boldsymbol{u}=n\times 1$

1) 通常最小自乘法

通常最小自乘法이 흔히 使用되는 方法으로 이에 依한 $\hat{y}_i(=x_i\hat{\beta})$은 條件附 確率의 推定値라고 解釋될 수 있으나 앞에서 指摘된 바와 같이 分散項이 不均等하므로 通常最小自乘推定量은 不偏(unbiased) 및 一致(consistent) 推定量은 되어도 有效(efficent) 推定量은 되지 못한다. 이러한 問題點은 다음의 一般化最小自乘推定量에 의하여 解決될 수 있다.

2) 一般化最小自乘法

이 方法을 使用하기 위하여서는 誤差項 u_i의 分散-共分散行列 Ω를 알아야 하나 Ω의 對角項이 $E(y_i)[1-E(y_i)]$로 되어있으므로 未知이다. 골드버거(A.S. Goldberger)는 $E(y_i)[1-E(y_i)]$의 推定値로 通常最小自乘推定値 $\hat{y}_i$에 依한 $\hat{y}_i(1-\hat{y}_i)$을 사용하여 다음과 같이 $\hat{\boldsymbol{\beta}}$을 求하였다.[1]

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{Y} \tag{2.6}$$

단,

$$\hat{\Omega}=\begin{pmatrix} \hat{y}_1(1-\hat{y}_1) & & 0 \\ & \ddots & \\ 0 & & \hat{y}_n(1-\hat{y}_n) \end{pmatrix}$$

이 方法은 推定된 分散의 逆數 $\dfrac{1}{\hat{y}_i(1-\hat{y}_i)}$을 加重値로 擇하는 加重最小自乘法(weighted least squares method)과 마찬가지이다.

一般化最小自乘推定法의 問題點은 앞에서 指摘된 바와 같이 通常最小自乘推定法에 依한 豫測値 $\hat{y}_i$가 (0, 1)區間을 벗어날 수 있으며 이렇게 될 때 $\hat{\Omega}$의 對角項 $\hat{y}_i(1-\hat{y}_i)$이 負의 값을 갖게 되므로 이方法의 適用이 어렵게 된다는 점이다. 이와 같이 $\hat{y}_i$가 (0, 1)區間을 벗어

---

1) 골드버거는 이러한 過程을 二段階 一般化最小自乘法(two-step generalized least squares method)이라고 稱하였다. A.S. Goldberger, *op. cit*, p.250參照. $\hat{y}_i(1-\hat{y}_i)$가 $E(y_i)[1-E(y_i)]$의 一致推定量이 되는 證明은 R. McGillivray, "Estimating the Linear Probability Function," *Econometrica*, Vol. 38 (1970), pp.775—776을 참조할 것.

나는 경우 그러한 $\hat{y}_i$을 豫測結果에서 제거하거나 任意로 0이나 1에 가까운 값을 부여하여 이 問題를 解決할 수도 있으나 이러한 推定過程이 効率的이라고는 할 수 없다.[1]

3) 最尤推定法

$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{u}$에서 $E(y_i)$는

$$E(y_i)=p_i=\boldsymbol{X}_i\boldsymbol{\beta}\text{(단, }\boldsymbol{X}_i\text{은 }\boldsymbol{X}\text{行列의 }i\text{번째 行벡타)} \tag{2.7}$$

이므로 尤度函數는 다음과 같다.

$$L=\prod_{y_i=1}\boldsymbol{X}_i\boldsymbol{\beta}\cdot\prod_{y_i=0}(1-\boldsymbol{X}_i\boldsymbol{\beta}) \tag{2.8}$$

양변에 로그를 取하면

$$\log L=\sum_{y_i=1}\log(\boldsymbol{X}_i\boldsymbol{\beta})+\sum_{y_i=0}\log(1-\boldsymbol{X}_i\boldsymbol{\beta}) \tag{2.9}$$

$\dfrac{\partial\log L}{\partial\hat{\boldsymbol{\beta}}}=0$에서 $\hat{\boldsymbol{\beta}}$을 求할 수 있으나 이 方法 역시 $\boldsymbol{X}_i\boldsymbol{\beta}$가 $(0, 1)$을 벗어날 때 로그가 定義되지 않는다는 점에서 制限을 받는다.

이와 같이 세가지 推定方法 모두 그 豫測値가 $(0, 1)$을 벗어날 수 있다는 점은 模型自體가 갖는 弱點 때문이라고 할 수 있으므로 그 代案으로서 애초부터 期待値 $E(y_i)$가 $(0, 1)$ 區間內로 制限되는 模型을 찾게 되었다. 이러한 條件을 만족시키는 函數로 累積分布函數가 있으며 이중 累積正規分布函數(cumulative normal distribution function), 로지스틱分布函數(logistic distribution function) 및 곰페르츠曲線(Gompertz curve)을 利用한 模型의 表記方法을 다음에서 보기로 한다.

2. 2. 프로빗模型(probit model)

프로빗模型에서는 期待値 $E(y_i)$, 即 p_i를 累積正規分布函數의 形態로 表示하며 이 函數는 새로이 導入되는 指數 I_i를 確率 p_i로 變換시키기 위하여 使用된다. 여기서 I_i는 $\boldsymbol{X}_i$의 函數로

$$I_i=\boldsymbol{X}_i\boldsymbol{\beta} \qquad -\infty<I_i<\infty \tag{2.10}$$

와 같으며 I_i가 增加할 수록 p_i가 커진다고 하면 $y=1$ 또는 $y=0$를 선택하는 基準은 다음과 같이 臨界値 I_i^*에 의한다고 할 수 있다.[2]

$$y_i=\begin{cases}1 & I_i\geqq I_i^*\text{이면}\\ 0 & I_i<I_i^*\text{이면}\end{cases} \tag{2.11}$$

이때 臨界値 I_i^*를 많은 要因들에 의하여 決定되는 一種의 標本平均이라고 看做하면 I_i^*는

1) R.S. Pindyck and D.L. Rubinfeld, *op. cit.*, p.241.
2) A.S. Goldberger, *op. cit.*, p. 250.

中心極限定理에 의하여 正規分布를 한다고 할 수 있다.[1] 프로빗模型에서는 이 I_i^*가 平均 0, 分散 1인 標準正規分布 確率變數라고 假定하므로, 이 模型은 다음과 같이 表記될 수 있다.

$$E(y_i)=p_i=F(I_i)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{I_i} e^{-\frac{S^2}{2}}ds$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{X_i\beta} e^{-\frac{S^2}{2}}ds \qquad (2.12)$$

단, F: 累積正規分布函數

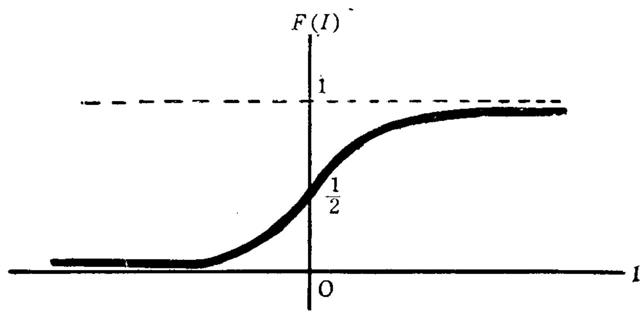


그림 3. 프로빗模型

프로빗模型의 母數 推定方法으로는 最尤推定法과 最小노밋χ^2方法(minimum normit χ^2 method)을 들 수 있다. 두 推定量의 漸近的 性質은 같으나 選多型模型(multiple choice model)의 경우에는 後者가 計算上 더 간편하다.[2]

1) 最尤推定法

式(2.12) 로부터 尤度函數 L은

$$L=F(I_1)\cdots F(I_r)\cdot[1-F(I_{r+1})]\cdots[1-F(I_n)]^{3)} \qquad (2.13)$$

兩邊에 로그를 取하면

$$\log L=\sum_{i=1}^{r}\log F(I_i)+\sum_{i=r+1}^{n}\log[1-F(I_i)] \qquad (2.14)$$

$\frac{\partial \log L}{\partial \hat{\beta}}=0$로 부터 非線型의 正規方程式에 의한 最尤推定値 $\hat{\beta}$을 求할 수 있다.

---

1) H. Theil, *Principles of Econometrics* (New York: John Wiley & Sons, Inc., 1971), p.630.

2) Arnold Zellner and Tong Hun Lee, "Joint Estimation of Relationships Involving Discrete Random Variables," *Econometrica*, Vol.33, No.2(April, 1965), pp.372—394.

3) 標本이 처음 r개는 $(y=1)$, 나머지 $n-r$는 $(y=0)$와 같은 順序로 놓였다고 假定한 것임.

說明變數 한 單位의 變化가 確率 p_i에 미치는 영향을 보려면,[1]

$$\frac{\partial p_i}{\partial x_j}=\frac{\partial F}{\partial I}\cdot\frac{\partial I}{\partial x_j}=f(I)\cdot\beta_j=f(X_i\beta)\cdot\beta_j \tag{2.15}$$

단, f : 標準正規分布의 確率密度函數

x_j: $X_i=(x_1,x_2,\cdots,x_j,\cdots,x_k)$ 中의 한 變數

β_j: x_j의 프로빗 回歸係數

따라서 Δx_j가 p_i에 미치는 영향은 說明變數의 特定한 값에 따라 달라진다.

2) 最小노밋χ^2推定[2]

$y_i=X_i\beta$에서 y_i를 觀察된 比率(r_i)이라고 하면, r_i는 平均 p_i(母比率), 分散$\frac{p_i(1-p_i)}{n_i}$ 인 二項分布 確率變數라고 할 수 있으며, r_i와 p_i는 다음과 같은 관계에 있다.

$$r_i=p_i+u_i \tag{2.16}$$

단, u_i: 平均 0, 分散$\frac{p_i(1-p_i)}{n_i}$인 二項分布

그러면 프로빗模型(또는 노밋模型) 에서 母比率 p_i는 다음과 같이 表記된다.

$$p_i=F(v_i)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{v_i}e^{-\frac{s^2}{2}}ds \tag{2.17}$$

단, $v_i=X_i\beta$

이때 v_i를 노밋 p_i[3]라고 하며 觀察노밋(v_i^o)과 母노밋(true normit)(v_i)의 관계는

$$\begin{aligned}v_i^o&=F^{-1}(r_i)=F^{-1}(p_i+u_i)\\&=F^{-1}(p_i)+u_i/f(p_i)+R_i{}^{4)}\end{aligned} \tag{2.18}$$

단, R_i: $u_i{}^2\frac{d^2F^{-1}}{dp_i{}^2}\Big/$ 2!以上의 高次項

誤差項의 分散은[5]

$$\mathrm{Var}\left(\frac{u_i}{f(p_i)}\right)=\frac{p_i(1-p_i)^{5)}}{n_i(f(p_i))^2} \tag{2.19}$$

따라서 다음과 같은 노밋χ^2를 最小化하는 $\hat{\beta}$을 찾으면 그것이 곧 最小노밋χ^2推定値가 된다.

---

1) Morley Gunderson, "Retention of Trainees: A Study with Dichotomous Dependent Variables," *Journal of Econometrics*, Vol.2, No.1(May, 1974), pp.79—93.

2) A. Zellner and T.H. Lee, *loc. cit.*

3) 프로빗은 노밋에 5을 더한 것과 같다. 노밋에 對한 자세한 것은 J. Berkson, "Estimate of the Integrated Normal Curve by Minimum Normit Chi-square with Particular Reference to Bio-assay," *Journal of the American Statistical Association*, Vol.50(1955), pp.529—549를 참조할 것.

4) 테일러級數展開에 의한 것임.

5) R_i는 無視하여도 v_i推定量의 漸近的 性質은 영향을 받지 않음. A. Zellner and T.H. Lee, *loc, cit.* 참조.

$$\chi^2 = \sum \frac{n_i(f(r_i))^2}{r_i(1-r_i)}(v_i^o - v_i)^2 \tag{2.20}$$

이것은 곧 $\mathrm{Var}\left(\frac{u_i}{f(p_i)}\right)$의 一致推定量 $\frac{r_i(1-r_i)}{n_i(f(r_i))^2}$의 逆數를 加重値로 한 加重最小自乘法과 같아지므로 $\hat{\beta}$을 行列로 表示하면 다음과 같다.

$$\hat{\boldsymbol{\beta}} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}\boldsymbol{v}^o \tag{2.21}$$

단, $\boldsymbol{v}^o$: $(n\times 1)$觀察노멋

$$\hat{\Omega} = \begin{pmatrix} \frac{r_1(1-r_1)}{n_1(f(r_1))^2} & & 0 \\ & \ddots & \\ 0 & & \frac{r_n(1-r_n)}{n_n(f(r_n))^2} \end{pmatrix}$$

2. 3. 로짓模型(logit model)

프로빗模型은 臨界値 I_i^*의 正規分布를 前提로 하나 實際應用에서는 이러한 條件이 區備되는 경우가 드물어 使用이 制限된다. 로짓模型은 로지스틱分布를 前提로 하며 다음과 같이 表記된다.[1)]

$$E(y_i) = p_i = \frac{1}{1+e^{-X_i\beta}} \tag{2.22}$$

式(2.22)를 $X_i\boldsymbol{\beta}$에 對하여 整理하면

$$\log\frac{p_i}{1-p_i} = X_i\boldsymbol{\beta}, \quad -\infty < \log\frac{p_i}{1-p_i} < \infty \tag{2.23}$$

式(2.23)의 左邊 $\log\frac{p_i}{1-p_i}$를 로짓(L_i)이라고 하며 L_i와 p_i의 關係는 그림 4와 같이 表示될 수있다.

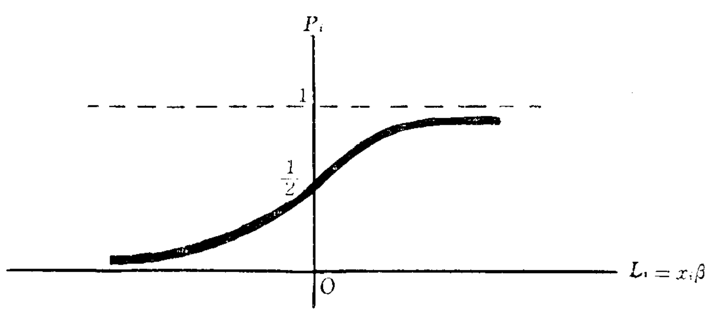


그림 4. 로짓模型

1) 로짓이란 名稱은 로지스틱分布에서 유래하며 버크슨(J. Berkson)에 의하여 命名되었다. J. Berkson, "Application of the Logistic Function to Bio-assay," *Journal of the American Statistical Association*, Vol.39(1944), pp.357—365.

가. 推　　定

1) 一般化最小自乘法

式(2.23)에서 p_i는 未知이므로, X의 分割項(cell)[1] 內의 相對度數 r_i를 p_i代身 使用하여 다음과 같이 母로짓 L_i(true logit)의 推定量 $\bar{L}_i$를 求하고, 이 $\bar{L}_i$를 x에 回歸하여 β 를 推定한다.[2]

$$\bar{L}_i = \log \frac{r_i}{1-r_i} \tag{2.24}$$

이때 各 分割價의 觀察項의 數 n_i가 $n_i \to \infty$함에 따라 r_i는 平均 p_i 分散 $\frac{p_i(1-p_i)}{n_i}$인 正規分布에 接近하며, $\bar{L}_i$는 平均 L_i, 分散 $\frac{1}{n_i r_i(1-r_i)}$인 正規分布에 接近한다.[3] 이와 같이 分散이 不均等하므로 通常最小自乘法보다는 加重最小自乘法의 使用이 有効性을 높일 수 있으며 이때 加重値는 $w_i = n_i r_i(1-r_i)$이다.[4]

即, $$\hat{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}\bar{L} \tag{2.25}$$

단, $\bar{L}$: $(N\times 1)$觀察로짓벡타

$$\hat{\Omega} = \begin{pmatrix} \frac{1}{n_1 r_1(1-r_1)} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n_N r_N(1-r_N)} \end{pmatrix}$$

N : 分割項의 數

說明變數 x의 數와 x의 區劃이 너무 많아지면 分割項의 數가 增加하여 項內의 觀察値 n_i가 적어지므로 相對度數 r_i가 1 또는 0이 될 수 있으며 이러한 경우 로짓은 定義되지 않는다. 따라서 이 推定方法은 充分한 觀察値를 前提로 하고 있으며 경험적으로 보아 n_i가 적어도 5이상 되어야 한다.[5]

한편 로짓 $L_i = \log\frac{p_i}{1-p_i}$의 推定値로 $\bar{L}_i = \log\frac{r_i}{1-r_i}$代身 $\bar{L}_i = \log\frac{r_i + \frac{1}{2n_i}}{1-r_i+\frac{1}{2n_i}}$ 을

1) 가령 說明變數가 x_1, x_2이고 이들 모두가 假變數(dummy variable)라 할 때 分割項은 $(x_1=0,\ x_2=0)$, $(x_1=0,\ x_2=1)$, $(x_1=1,\ x_2=0)$, $(x_1=1,\ x_2=1)$로 4個이며 各 分割項의 觀察値의 數를 n_i, $(y=1)$이 일어난 回數를 m_i라 하면 $r_i = \frac{m_i}{n_i}$이다.

2) r_i는 二項分布 母集團에서 獨立的으로 抽出되었다고 假定한다. H. Theil, "On the Estimation of Relationships Involving Qualitative Variables," *American Journal of Sociology*, Vol.76(1970), pp.103—154.

3) *Ibid.*

4) 加重値 w_i는, r_i가 주어졌을 때 n_i가 增加함에 따라 w_i도 함께 增加하며, n_i가 주어졌다면 r_i가 0 또는 1에 接近함에 따라 w_i는 0에 接近하므로, 이와 같이 r_i의 微小한 變化에 로짓이 크게 움직이는 部分에는 加重値를 작게 附與한다는 點에서 의미가 있다.

5) R.S. Pindyck and D.L. Rubinfeld, *op. cit.*, p.250.

使用할 수도 있으며 이때의 分散은 다음과 같다.[1]

$$\frac{\left(1+\frac{1}{n}\right)\left(1+\frac{2}{n}\right)}{n\left(r+\frac{1}{n}\right)\left(1-r+\frac{1}{n}\right)} \tag{2.26}$$

2) 最小로짓 χ^2 推定法[2]

分割項內의 相對度數 r_i가 平均 p_i, 分散 $\frac{p_i(1-p_i)}{n_i}$ 의 二項分布를 한다고 하면,

$$r_i = p_i + u_i \tag{2.27}$$

라고 할 때 u_i는 平均 0, 分散 $\frac{p_i(1-p_i)}{n_i}$ 의 二項分布를 하며 로짓 L_i는 다음과 같다.

$$\begin{aligned}\log\frac{r_i}{1-r_i} &= \log\frac{p_i}{1-p_i} + \log\left(1+\frac{u_i}{p_i}\right) - \log\left(1-\frac{u_i}{1-p_i}\right)^{3)} \\ &= \log\frac{p_i}{1-p_i} + \frac{u_i}{p_i} + \frac{u_i}{1-p_i} + R_i^{4)} \\ &= \log\frac{p_i}{1-p_i} + \frac{u_i}{p_i(1-p_i)} + R_i \end{aligned} \tag{2.28}$$

分散은 $\mathrm{Var}\left(\frac{u_i}{p_i(1-p_i)}\right) = \frac{1}{n_i p_i(1-p_i)}$ 이므로[5] 最小로짓 χ^2 推定量은

$$\chi^2 = \sum n_i r_i(1-r_i)(\bar{L}_i - L_i)^2 \tag{2.29}$$

이 推定量은 바로 加重値 $n_i r_i(1-r_i)$에 의한 加重最小自乘推定量으로 앞의 方法인 一般化最小自乘法과 一致한다.

3) 最尤推定法

式(2.22)에 의한 尤度函數는

1) 이 修正方法은 小規模 標本에 의한 推定에는 도움이 되나 大規模 標本에는 별 영향을 미치지 않는다. R.S. Pindyck and D.L. Rubinfeld, *ibid.* 및 J. Berkson, "Application of the Logistic Function to Bio-assay," *Journal of the American Statistical Association*, Vol. 39(1944), pp.357—365.

2) J. Berkson, "A Statistically Precise and Relatively Simple Method of Estimating the Bio-assay with Quantal Response, Based on the Logistic Function, *ibid.* Vol.48(1953), pp.565—599. J. Berkson, "Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function," *ibid.*, Vol. 50(1955), pp.130—162. 및 A. Zellner and T.H. Lee. *loc. cit.*

3) $\frac{r_i}{1-r_i} = \frac{p_i}{1-p_i}\left(\frac{1+\frac{u_i}{p_i}}{1-\frac{u_i}{1-p_i}}\right)$을 代入한 것임.

4) 테일러級數展開에 의한 것임.

5) R_i는 推定時 無視하여도 推定値의 漸近性에는 영향을 미치지 않는다. A. Zellner and T. H. Lee, *loc. cit.* 參照.

$$L=\prod_{y_i=1}\frac{1}{1+e^{-X_i\beta}}\prod_{y_i=0}\left(1-\frac{1}{1+e^{-X_i\beta}}\right) \tag{2.30}$$

양변에 로그를 取하면

$$\log L=\sum_{y_i=1}\frac{1}{1+e^{-X_i\beta}}+\sum_{y_i=0}\left(1-\frac{1}{1+e^{-X_i\beta}}\right) \tag{2.31}$$

最尤推定量은 $\frac{\partial \log L}{\partial \hat{\beta}}=0$ 에서 유도되는 非線型 方程式으로부터 구할 수 있다. 이 推定量의 漸近的 性質은 前述한 버크슨-타일(Berkson-Theil)의 一般化 最小自乘法과 같으나[1] $\hat{\beta}$을 非線型方程式으로부터 구하여야 하므로 計算上 어려움이 따른다. 그러나 小規模 標本의 경우 線型確率模型에서 얻어진 $\hat{\beta}$의 推定量은 大部分 最尤推定量과 一致하므로, 複雜한 計算過程을 要하는 最尤推定量 代身 線型模型에서 얻어진 $\hat{\beta}$을 使用하여도 問題가 없다.[2]

나. 로짓表記의 妥當性 檢定

이 檢定을 위한 χ^2 統計量은 觀察된 相對度數 r_i와 로짓模型에 의하여 推定된 確率 $\hat{p}_i$와의 差異에서 얻어질 수 있다.[3]

$$\chi^2=\sum_{i=1}^{N}\frac{n_i(r_i-\hat{p}_i)^2}{\hat{p}_i(1-\hat{p}_i)} \qquad \text{自由度}: N-k \tag{2.32}$$

단, $\hat{p}_i=\frac{1}{1+e^{-X_i\hat{\beta}}}$

N : 分割項의 數

k : 母數의 數

이 統計量은 다음과 같이 一般化最小自乘値의 推定誤差에 加重値行列 $\hat{\Omega}^{-1}$의 對角項이 곱해진 二次型과 같다.

$$\chi^2=(L-X\hat{\beta})'\hat{\Omega}^{-1}(L-X\hat{\beta}) \tag{2.33}$$

2. 4. 곰핏模型[4] (Gompit model)

곰핏模型은 곰페르츠曲線에 根據를 두며, 確率 p_i가

$$p_i=\alpha\delta_2^{x_{2i}}\delta_3^{x_{3i}}\cdots\delta_k^{x_{ki}} \tag{2.34}$$

단, $0<\alpha<1$

$0<\delta_j<1,\ j=2,\cdots,k$

---

1) M.L. Mingche, "A Logit Model of Homeownership," *Econometrica*, Vol.45, No.5(July, 1977), pp.1081—1097.

2) R.S. Pindyck and P.L. Rubinfeld, *op. cit.*, p.251.

3) H. Theil, "On the Estimation of the Relationships Involving Qualitative Variables," *American Journal of Sociology*, Vol. 76(1970), pp.103—154.

4) A. Zellner and T.H. Lee, *loc. cit.*

일 때 곰핏 p_i는 다음과 같이 表記된다.

$$\log\ \log\frac{1}{p_i}=\beta_1+\beta_2 x_{2i}+\cdots+\beta_k x_{ki} \qquad (2.35)$$

단, $\beta_j=\log\delta_j,\ j=2,\cdots,k$

$\beta_1=\log(-\log\alpha)$

곰핏模型의 推定은 다음의 最小곰핏 χ^2方法을 利用한다.

$$\log\ \log\frac{1}{r_i}=\log\ \log\frac{1}{p_i+u_i}$$

$$=\log\ \log\frac{1}{p_i}+\frac{u_i}{p_i\log p_i}+R_i^{1)} \qquad (2.36)$$

그런데,

$$\mathrm{Var}\left(\frac{u_i}{p_i\log p_i}\right)=\frac{p_i(1-p_i)}{n_i(p_i\log p_i)^2}\ \text{이므로} \qquad (2.37)$$

곰핏 χ^2는

$$\chi^2=\sum\frac{n_i(r_i-\log r_i)^2}{p_i(1-p_i)}(g_i-G_i)^2 \qquad (2.38)$$

단, $g_i=\log\ \log\frac{1}{r_i}$

$G_i=\log\ \log\frac{1}{p_i}$

이 推定量은 바로 다음의 一般化最小自乘推定量 $\hat{\beta}$이다.

$$\hat{\beta}=(X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}g \qquad (2.39)$$

단, $g:(n\times1)$觀察곰핏 벡타

$$\hat{\Omega}=\begin{pmatrix}\frac{r_1(1-r_1)}{n_1(r_1\log r_1)^2} & & 0\\ & \ddots & \\ 0 & & \frac{r_n(1-r_n)}{n_n(r_n\log r_n)^2}\end{pmatrix}$$

2. 5. 線型確率模型의 變換에 依한 表記

앞에서 展開되었던 프로빗, 로짓, 곰핏模型은 모두, 線型確率模型에서 얻어지는 豫側値가 (0,1)區間을 벗어나므로 模型의 表記方法으로 不適當하다고 看做하고 그 代案으로 제시되었던 模型들이다. 그러나 豫側値의 問題는 다음과 같이 線型確率模型의 變換을 통하여도 解決可能하다.

가. 오컷(G.H. Orcutt)에 依한 變換[2]

1) 테일러 級數展開에 의한 것임.

2) M. Gunderson, *loc. cit.* 및 Guy H. Orcutt, et. al., *Microanalysis of Socioeconomic Systems: A Simulation Study* (New York: Harper & Row, 1961), pp.229—231.

線型確率模型을 通常最小自乘法으로 推定하여 $\hat{y}$을 구하고 $\hat{y}$의 區間에 따라 觀察値를 分割한 다음, 各 分割區間內의 相對度數 m_i[1] ($\hat{y}=1$이 나타난 比率)를 修正된 推定確率($\hat{p}_0$)로 하였다.

이 方法은 간편할 뿐만 아니라 一般化最小自乘法에 의한 豫測値에도 使用될 수 있으나 $\hat{y}$의 區分이 任意的이라는 問題點이 있다. 또 結果的으로는 $\hat{p}_0$이 (0,1)區間내로 제한되는 曲線의 形態가 되었다 할지라도 基本的인 推定過程에서 이러한 非線型性을 無視하고 있는 點 및 $\hat{p}_0$가 單調(增加)函數라는 保障이 없다는 點등이 이 方法이 지닌 問題點이다.

나. 로짓에 의한 變換[2]

로지스틱分布函數를 使用하여 다음과 같이 線型確率函數의 豫測値가 (0,1)區間內로 制限되게끔 한다.

다음과 같은 函數를 假定하면

$$p_L(x)=\frac{1}{1+e^{-\hat{L}(x)}} \tag{2.40}$$

이때 $\hat{L}(x)$는 다음과 같이 $\hat{y}$의 函數로 求하여 진다.

① 線型確率模型으로부터 $\hat{y}$을 求함.

② $\hat{y}$의 區間設定

③ $\hat{y}$區間에 의한 相對度數 m_i를 求함[3]

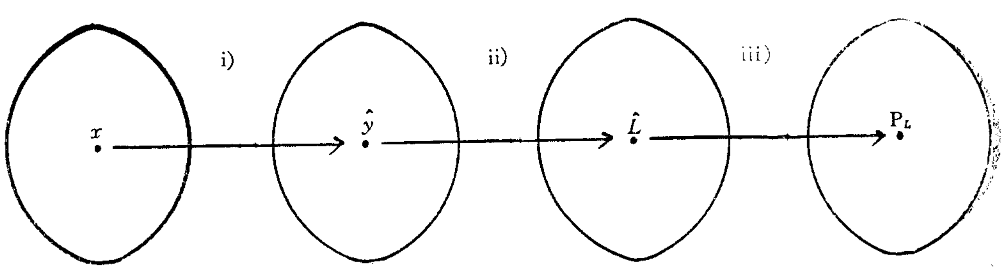
④ 로짓 $\hat{L}_i$의 값을

$$\hat{L}_i=\log\frac{m_i}{1-m_i} \tag{2.41}$$

에서 求한 다음

⑤ $\hat{L}_i$을 $\hat{y}$ 各 區間의 中間값(mid point)에 線型回歸하여 $\hat{L}$와 $\hat{y}$의 關係를 求함.

⑥ 이렇게 얻어진 $\hat{L}$로부터 $p_L(x)$를 求할 수 있다.[4]



1) 相對度數이므로 언제나 (0,1)區間안에 놓인다.
2) M. Gunderson, *loc. cit.*
3) 여기까지는 오컷의 方法과 同一하다.
4) P_L의 線型性이 로짓回歸時 包含되었다.

이 過程은 위와 같이 表示될 수 있다.

단, i) 線型確率函數에 의한 變換

ii) 로짓回歸

iii) 로지스틱變換

다. 워너(S. Warner)에 依한 變換

워너는 確率 $p(x)$을 推定하는 方法으로 다음 式을 提示하였다.

$$p_w(x)=\frac{1}{1+e^{-\hat{D}(x)}} \tag{2.42}$$

단, $\hat{D}(x)=K(\hat{y}-\hat{\beta}_0)-\frac{1}{2}K(\boldsymbol{x}_1+\boldsymbol{x}_2)'\hat{\boldsymbol{\beta}}+\log\frac{N_1}{N_2}$

$$K=\frac{N_1+N_2-2}{\dfrac{N_1N_2-N_1N_2(\boldsymbol{x}_1-\boldsymbol{x}_2)'\hat{\boldsymbol{\beta}}}{N_1N_2}}$$

N_1: $(y=1)$이 일어난 回數

N_2: $(y=0)$ 〃

$\boldsymbol{x}_1$: $(y=1)$集團에 對한 設明變數 $\boldsymbol{x}$의 平均値 벡타

$\boldsymbol{x}_2$: $(y=0)$ 〃

$\hat{\boldsymbol{\beta}}$: 最小自乘에 依한 기울기 벡타

(절편항 $\hat{\beta}_0$는 제외)

라. 세 變換方法의 問題點[1]

위의 方法 모두가 이미 推定된 線型確率函數를 事後的으로 非線型函數로 修正하는데 그 目的이 있으며, 基本的인 推定過程에서는 이러한 非線型性이 無視된다는 데 問題點이 있다. 또 $\hat{y}$을 구하기 위하여 通常最小自乘法을 使用함으로써 不均等에 依한 非有効性의 문제가 그대로 남는다.

2. 6. 連續變量과 離散變量의 混合

從屬變數 y가 1또는 0 이라는 不連續의 두 가지 값만 取하는 경우를 좀 더 擴張하여, 이런 對象의 選擇與否 뿐만 아니라 그 對象을 選擇했을 때 對象의 程度(또는 規模)에 대하여도 前述한 模型의 分析方法을 利用할 수 있다. 例를 들어 自動車購入의 경우, 車의 購入與否 뿐만 아니라 一旦 購入했을 때 그 購入價格(또는 費用)에 對하여도 線型確率模型이나 프로빗模型을 適用시킬 수 있다.

從屬變數 y를 다음과 같이 設定한다고 하자.

1) M. Gunderson, *loc. cit.*

$$y=\begin{cases}\text{價格}\cdots(\text{購入時})\\ 0\cdots\cdots(\text{非購入時})\end{cases} \tag{2.43}$$

이러한 從屬變數를 갖는 模型에 通常最小自乘法을 適用하게 되면, 애초부터 y의 領域이 非負의 값으로 制限되어 있는데다, 만일 非購入者들이 많아 觀察値가 $y=0$로 集中된다면 y가 過小推定될 우려가 있다.[1)]

1. 線型確率模型에 의한 分析[2)]

먼저 1또는 0의 값을 取하는 모든 $y_i{}^*$를 設明變數 x_i에 回歸한 다음, 標本을 $y_i{}^*=1$인 경우로 制限하고 이 '制限된 $y_i{}^*$' $(=y_i{}^{**})$를 다시 x_i에 回歸하여 얻어진 두 結果를 綜合하면 원래의 目的하는 바이었던 0또는 非負의 값인 y에 對한 期待値를 얻을 수 있다.[3)]

卽, $E(y)=\sum y\cdot f(y)$

$$\begin{aligned}
&=\sum_{y=0}0\cdot f(0)+\sum_{y>0}y\cdot f(y)\\
&=0+\sum_{y>0}\{y[f(y|y>0]\cdot f(y>0)\}\\
&=f(y<0)\cdot\sum_{y>0}\{y[f(y|y>0)]\}\\
&=\text{Prob}\{y>0\}\cdot E(y|y>0)
\end{aligned} \tag{2.44}$$

式(2.44)에서 Prob$\{y>0\}$은 첫번째 回歸의 結果이며 $E(y|y>0)$은 두 번째 回歸의 結果로 결국 $\hat{y}=\hat{y}^*\hat{y}^{**}$와 같이 y의 推定値는 두 推定値의 곱으로 나타난다.

2. 프로빗模型에 의한 分析[4)]

토빈(J. Tobin)은 프로빗模型을 利用하여 從屬變數 y가 0 또는 非負의 값인 경우, 接近方法을 다음과 같이 提示하였다.

y_i의 選擇基準은

$$y_i=\begin{cases}0 & I_i<I_i{}^*\text{이면}\\ I_i-I_i{}^* & I_i\geqslant I_i{}^*\text{이면}\end{cases} \tag{2.45}$$

단, I^*: 平均 0, 分散 σ^2인 正規分布確率變數

그러면 各各의 確率은

$$\text{Prob}\{y=0|I\}=\text{Prob}\{I^*>I|I\}$$

1) A.S. Goldberger, *op. cit.*, pp.251—252.
2) 이 方法을 利用한 例는 Janet A. Fisher, "An Analysis of Consumer Goods Expenditures in 1957," *The Review of Economics and Statistics*, Vol. 44(1962), pp.64—71에서 찾을 수 있다.
3) 골드버거(A.S. Goldberger)는 이 方法을 雙型線型確率模型(twin linear probability model)이라고 하였다. A.S. Goldberger, *op. cit.*, pp.251—253 參照.
4) James Tobin, "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, Vol.26(1958), pp.24—36 및 A.S. Goldberger, *op. cit.*, pp.253—255.

$$=1-F\left(\frac{I}{\sigma}\right) \tag{2.46}$$

$$\text{Prob}\{y>y^*\geq 0|I\}=\text{Prob}\{I^*<I-y^*|I\}$$

$$=F\left(\frac{I-y^*}{\sigma}\right) \tag{2.47}$$

이때 母數 推定을 위한 尤度函數는

$$L=\prod_{i=1}^{S}\left[1-F\left(\frac{I_i}{\sigma}\right)\right]\prod_{i=s+1}^{T}\frac{1}{\sigma}f\left(\frac{I_i-y_i}{\sigma}\right) \tag{2.48}$$

兩邊에 로그를 取하면

$$\log L=\log\sum_{i=1}^{S}\left[1-F\left(\frac{I_i}{\sigma}\right)\right]-(T-S)\log\sigma+\sum_{i=s+1}^{T}\log f\left(\frac{I_i-y_i}{\sigma}\right) \tag{2.49}$$

단, $F\left(\frac{I_i}{\sigma}\right)=\frac{1}{\sqrt{2\pi}}\int_{\infty}^{\frac{X_i\beta}{\sigma}}e^{-\frac{u^2}{2}}du$

$$f\left(\frac{I_i-y_i}{\sigma}\right)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{X_i\beta-y_i}{\sigma}\right)^2}$$

따라서 最尤推定値 $\hat{\beta}$, $\hat{\sigma}$은 $\frac{\partial(\log L)}{\partial\hat{\beta}}=0$, $\frac{\partial(\log L)}{\partial\hat{\sigma}}=0$ 의 非線型 正規方程式에서 구하여 진다.

2. 7. 豫　測

지금까지의 二進數 選擇模型은 個人의 行動을 說明하기 위한 것이었다. 그러나 特定한 個個人 보다는 그가 所屬되어 있는 母集團 全體의 行動을 豫測하고자 한다면 個個人의 豫測値를 어떻게 統合할 것인가 하는 問題가 擡頭되게 된다.[1] 例를 들어, 所得을 說明變數로 하는 二進數 模型에서 各 個人의 自動車購入確率이 豫測되었을 때, 所屬集團의 所得分布가 變한다면 그 集團全體의 購入確率이 어떠한 變化를 보일 것인가 하는 問題가 이에 屬한다.

式(2.50)에서와 같이 로짓模型에서는 x_j가 한 單位 變할 때 p_i의 變化量 Δp_i는 p_i의 函數이다.

$$\Delta p_i=\hat{\beta}_j p_i(1-p_i)\Delta x_{ij} \tag{2.50}$$

卽, Δp_i는 모든 사람이 다 같지 않고 各個人의 確率 p_i에 따라 달라지므로 어떤 集團 全體의 確率變化를 豫測하려면 p_i의 分布에 따르는 加重化된 Δp_i을 使用하여야 한다. 웨스틴(R.B.Westin)은 $\boldsymbol{x}$가 平均 μ_x 分散 Ω 인 正規分布를 한다고 假定할 때, p_i의 確率密度函數

1) Richard B. Westin, "Predictions from Binary Choice Models," *Journal of Econometrics*, Vol.2, No.1(1974), pp.1—16.

$f(p)$을 다음과 같이 定義하였다.[1]

$$f(p)=\frac{1}{\sqrt{2\pi}\sigma}\frac{1}{p(1-p)}e^{\left\{-\frac{1}{2\sigma^2}\left[\log\left(\frac{p}{1-p}\right)-\mu\right]^2\right\}} \tag{2.51}$$

단, $\mu=\mu_x\boldsymbol{\beta}$

$\sigma^2=\boldsymbol{\beta}'\Omega\boldsymbol{\beta}$

二進模型에서 $y=1$ 또는 $y=0$ 라는 두 試行結果는 確率 p_i의 베르누이分布를 하므로, 母集團의 경우 試行結果의 分布는 確率 p_i의 分布로 부터 導出되어야 한다.

이때 說明變數 $\boldsymbol{x}$가 變함에 따라 $E(p)$가 어떻게 變하는지를 보기 위하여 웨스틴은 다음의 두가지 方法을 提示하였다. 即, $\boldsymbol{x}$의 變化가 微小할 때는 $E(p)$의 彈力性에 의하고,[2] $\boldsymbol{x}$의 變化가 크며 變化量을 알 때에는 $f(p)$의 母數 μ, σ^2가 變化後 어떻게 달라졌는가를 比較하여 $E(p)$의 움직임을 觀察하였다.

이러한 豫測 結果를 評價하기 위하여 決定係數 R^2 을 使用할 때, 모리슨(D. G. Morrison)은 R^2 의 極大化는 豫測確率 $\hat{p}$가 眞正한 確率 p와 같을 때 이루어 지며,[3] 만일 $f(p)$가 베타分布의 密度函數라고 假定한다면 R^2의 極大値는 다음과 같이 됨을 보였다.

$$R^2=\frac{1}{\alpha+\beta+1} \tag{2.52}$$ [4]

단, α, β는 베타分布의 母數

例를 들어 $\alpha=\beta=1$이라면 p는 均等分布를 하며 이때 R^2의 極大値는 $\frac{1}{3}$이 된다. 이와 같이 모든 豫測 確率이 0 또는 1이 되어 $R^2=1$이 되는 극단적인 경우를 除外하고는 R^2 는 1보다 훨씬 낮은 값을 갖는 경향이 있다.[5]

---

1) $\boldsymbol{x}\to N(\mu_x,\Omega)$, $\boldsymbol{x\beta}\to N(\mu_x\boldsymbol{\beta},\ \boldsymbol{\beta}'\Omega\boldsymbol{\beta})$와 같으므로 $f(p)$는 다음과 같은 過程에서 구하여 진다.

$$f(p)=g(\boldsymbol{x\beta})\cdot|J|$$
$$=g\left(\log\frac{p}{1-p}\right)\cdot\frac{1}{p(1-p)}$$

p는 未知이므로 μ, σ^2의 推定은 먼저 標本確率로 부터 $\hat{\mu}_x$,$\hat{\Omega}$을 구하고 로짓模型으로 부터 $\hat{\beta}$을 구한 다음 $\hat{\mu}=\hat{\mu}_x\hat{\beta}$, $\hat{\sigma}^2=\hat{\boldsymbol{\beta}}\hat{\Omega}\hat{\beta}$에 의하여 μ, σ^2을 推定할 수 있다.

2) 例를 들어 j번째 變量 x_j의 平均 μ_j에 대한 $E(p)$의 彈力性은

$$\frac{\partial E(p)/E(p)}{\partial\mu_j/\mu_j}=\frac{\beta_j\mu_jE[p(1-p)]}{E(p)}$$ 와 같다.

3) 豫測誤差의 分散 $(1-\hat{p})^2\cdot p-\hat{p}^2(1-p)$를 $\hat{p}$에 對하여 偏微分하여 0으로 놓으면 $\hat{p}=p$을 求할 수 있다. Donald G. Morrison, "Upper Bounds for Correlation Between Binary Outcomes and Probabilistic Predictions," *Journal of the American Statistical Association*, Vol. 67, No. 337(March, 1972), pp.68—70, 및 R.S. Pindyck and D.L. Rubinfeld, *op. cit.*, pp.255—256을 참조할 것.

4) 모리슨(D.G. Morrison)은 이 R^2 을 上限(upper bound)이라고 하였으나, 골드버거(A.S.Goldberger)는 만일 p가 베르누이分布를 한다면 $R^2=1$ 이 되므로 R^2의 上限이란 槪念은 잘못된 것이라고 反駁하고 있다. 자세한 것은 A.S. Goldberger, "Correlations Between Binary Outcomes and Probabilistic Predictions," *Journal of the American Statistical Ascociation*, Vol.68, No.341(March, 1973), p.84를 參照할 것.

5) R.S. Pindyck and D.L. Rubinfeld, *op.cit.*, pp.225—226.

3. 選多型 模型(multiple choice model)

지금까지 論議된 模型은 모두 하나의 從屬變數가 0 또는 1의 값을 取하는 單一變數 二進模型(univariate dichotomous model)에 對한 것이었다. 이제 이러한 模型을 擴張하여 三者 以上의 選擇對象에서 하나를 擇하는 경우인 單一變數 選多型模型[1](univariate polytomous model)및 從屬變數가 둘 以上이며 各 從屬變數의 選擇對象이 둘이 되는 多變量 二進模型[2](multivariate dichotomous model), 그리고 從屬變數가 둘 以上, 그 選擇對象이 셋 以上이 되는 가장 一般化한 模型인 多變量 選多型 模型(multivariate polytomous model)에 對하여 살펴 보기로 한다.[3]

3. 1. 線型確率模型에 依한 擴張

가령 選擇對象이 3個라면 單純한 形態의 線型確率模型은 다음과 같다.[4]

$$\begin{aligned} p_{1i} &= \alpha_1+\beta_1 x_i \\ p_{2i} &= \alpha_2+\beta_2 x_i \\ p_{3i} &= \alpha_3+\beta_3 x_i \end{aligned} \qquad (3.1)$$

단, p_{ji} : i번째 사람이 j번째 對象을 擇할 確率

各各의 選擇은 서로 獨立이라고 假定하므로 $p_{1i}+p_{2i}+p_{3i}=1$이 되어 母數 推定値 $\hat{\alpha}_j$, $\hat{\beta}_j$은 다음과 같은 性質을 갖는다.

$$\sum_{j=1}^{3}\hat{\alpha}_j=1 \qquad (3.2)$$

$$\sum_{j=1}^{3}\hat{\beta}_j=0 \qquad (3.3)$$

따라서 3個의 方程式을 모두 推定할 必要가 없으며 2個만 推定하면 나머지 하나는 자연히 求하여 진다.

以上의 單一變數 選多形模型은 變數의 數가 늘어남에 따라 多變量選多形模型이 되며, 多變量 二進模型을 例로 들어 表記하면 다음과 같다.[5]

---

1) 例를 들면, 投票時 選擇對象이 '贊成, 反對, 棄權'으로 주어졌을 경우이다.

2) 二變量 二進模型의 例로, 自動車 購入(y)과 購入에 사용될 豫金의 引出 與否(z)를 들 수 있다. y와 z는 다 같이 0 또는 1의 값을 가지며, y, z의 관계는 一方的이 아니므로 이러한 模型에서는 聯立方程式 偏倚(simultaneous equation bias)의 문제가 提起된다.

3) 本考에서는 選擇對象들이 選擇順位를 갖지 않는 경우만을 고려하였다.

4) R.S. Pindyck and D.L. Rubinfeld, *op. cit.*, pp.256-257.

5) A. Zellner and T.H. Lee, *loc. cit.*

$$\begin{pmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{q}_M \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1 & 0 \cdots 0 \\ 0 & \boldsymbol{x}_2 \cdots 0 \\ \vdots & \vdots \ddots \vdots \\ 0 & 0 \cdots \boldsymbol{x}_M \end{pmatrix} \cdot \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{pmatrix} + \begin{pmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_M \end{pmatrix} \qquad (3.4)$$

단, $\boldsymbol{q}_j$: $(y=1)$이 일어난 相對度數로 表示된 $(T\times 1)$벡타

$\boldsymbol{x}_j$: $(T\times k_j)$觀察値 行列

$\boldsymbol{\beta}_j$: $(k_j\times 1)$母數 벡타

$\boldsymbol{u}_j$: $(T\times 1)$벡타

式 (3.4)는 다음과 같이 表示될 수 있다.

$$\boldsymbol{Q}=\boldsymbol{XB}+\boldsymbol{U} \qquad (3.5)$$

단, $\boldsymbol{Q}=(\boldsymbol{q}_1',\cdots,\boldsymbol{q}_M')'$

$\boldsymbol{B}=(\boldsymbol{\beta}_1',\cdots,\boldsymbol{\beta}_M')'$

$\boldsymbol{U}=(\boldsymbol{u}_1',\cdots,\boldsymbol{u}_M')'$

따라서 一般化最小自乘法에 의한 $\boldsymbol{B}$의 推定量은[1)]

$$\hat{\boldsymbol{B}}=(\boldsymbol{X}'\hat{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\Omega}^{-1}\boldsymbol{Q} \qquad (3.6)$$

단, $\Omega=\begin{pmatrix} \omega_{11} & \omega_{12}\cdots\omega_{1M} \\ \omega_{21} & \omega_{22}\cdots\omega_{2M} \\ \vdots & \vdots \quad\quad \vdots \\ \omega_{M1} & \omega_{M2}\cdots\omega_{MM} \end{pmatrix}$[2)]

3. 2. 프로빗模型에 의한 擴張

한편 이와 같은 多變量 選多型模型은 프로빗 形態로 表示될 수 있으며, 二變量 二進模型을 例로 들 때 두 變量 Y, Z의 結合確率은 다음과 같다.[3)]

$$\begin{aligned} p_1 &= P(Y_i=1,\ Z_i=1)=F(\boldsymbol{X}_i\boldsymbol{\beta},\ \boldsymbol{X}_i\delta,\ \rho) \\ p_2 &= P(Y_i=1,\ Z_i=0)=\Phi(\boldsymbol{X}_i\boldsymbol{\beta})-p_1 \\ &\qquad\qquad =\Phi(\boldsymbol{X}_i\boldsymbol{\beta})-F(\boldsymbol{X}_i\boldsymbol{\beta},\ \boldsymbol{X}_i\delta,\ \rho) \qquad (3.7) \\ p_3 &= p(Y_i=0,\ Z_i=1)=\Phi(\boldsymbol{X}_i\delta)-p_1 \end{aligned}$$

---

1) 아메미야(T. Amemiya)는 이 模型이 q_j 相互間의 結合的인 要素를 고려하지 않고 있으므로 이 模型에 의한 推定量을 制限情報 最小 χ^2 推定量(limited information minimum χ^2 estimator)이라고 부르고 있다. T. Amemiya, "Bivariate Analysis: Minimum Chi-square Methods," *Journal of the American Statistical Association*, Vol.69(Dec., 1974), pp.940—944을 參照할 것.

2) 이 分散-共分散 行列의 導出過程은 A. Zellner and T.H. Lee, *op. cit.*를 參照할 것.

3) T. Amemiya, "The Maximum Likelihood, the Minimum Chi-square and the Nonlinear Weighted Least-squares Estimator in the General Qualitative Response Model," *Journal of the American Statistical Association*, Vol.71, No.354(June, 1976), pp.347—351.

$$=\Phi(X_i\delta)-F(X_i\boldsymbol{\beta},\ X_i\delta,\ \rho)$$

$$p_4=P(Y_i=0,\ Z_i=0)=1-p_1-p_2-p_3$$

단, Φ: 累積標準正規分布函數

F: 相關係數 ρ 인 二變量 累積標準正規分布函數

이의 推定方法으로는 最尤推定法과 最小χ^2推定法[1]이 있으며, $\rho=0$일때 $\boldsymbol{\beta}, \boldsymbol{\delta}$의 最小$\chi^2$推定量은 漸近的 有効性을 갖는다. 그러나 多變量의 경우에서는 Y, Z의 結合確率을 고려하는 全情報 最小χ^2推定量(full information minimum χ^2 estimator)이 언제나 漸近的 有効性을 갖지는 않는다.[2]

3. 3. 로짓模型에 의한 擴張

選擇對象이 3個일 때 i번째 사람의 各 對象의 選擇確率을 p_{1i}, p_{2i}, p_{3i}라 하면 로짓은 다음과 같이 各 確率의 比率로 이루어 진다.[3]

$$\begin{aligned}\log\left(\frac{p_{2i}}{p_{1i}}\right)&=\alpha_{21}+\beta_{21}x_i\\ \log\left(\frac{p_{3i}}{p_{1i}}\right)&=\alpha_{31}+\beta_{31}x_i\\ \log\left(\frac{p_{3i}}{p_{2i}}\right)&=\alpha_{32}+\beta_{32}x_i\end{aligned}\tag{3.8}$$

단, p_{ji}: i번째 사람이 j번째 對象을 擇할 確率

그런데 세번째 方程式은

$$\begin{aligned}\log\left(\frac{p_{3i}}{p_{2i}}\right)&=\log\left(\frac{p_{3i}}{p_{1i}}\cdot\frac{p_{1i}}{p_{2i}}\right)=\log\left(\frac{p_{3i}}{p_{1i}}\right)-\log\left(\frac{p_{2i}}{p_{1i}}\right)\\ &=(\alpha_{31}-\alpha_{21})+(\beta_{31}-\beta_{21})x_i\end{aligned}\tag{3.9}$$

와 같이 되므로 앞의 두개만 推定하면 나머지는 자연히 구하여 진다. 一般的으로 選擇對象이 N개 있을 경우 로짓形態에 의한 方程式은 다음과 같다.[4]

$$\log\frac{p_{ij}}{p_{1j}}=X_i\boldsymbol{\beta}_j \qquad \begin{array}{l} j=2,\cdots,N\\ i=1,\cdots,T\end{array}\tag{3.10}$$

단, T: 觀察値의 數

X_i: i번째 사람의 $(1\times k)$觀察値벡타

---

1) J. Berkson, "Application of the Logistic Fuction to Bio-assay," *ibid*, Vol.39(1944) pp.357—365.
2) T. Amemiya, "Bivariate Probit Analysis: Minimum Chi-Square Methods," *ibid*, Vol. 69(1974), pp.940—944.
3) R.S. Pindyck and D.L. Rubinfeld, *op. cit.*, pp.258—263.
4) P. Schmidt and R.P. Strauss, "The Prediction of Occupation Using Multiple Logit Models," *International Economic Review*, Vol.16, No. 2(June, 1975), pp.471—486.

β_j: j번째 選擇과 關係되는 $(k\times 1)$母數벡타

이와 같은 選多型 로짓模型의 推定方法으로 一般化最小自乘法과 最尤推定法이 있다. 만일 觀察値가 充分하며 個個의 觀察値가 서로 獨立을 維持한다면, 二進模型에서와 같이 各分割項 內의 相對度數 r_{ji}에 의하여 얻어지는 $\log\frac{r_{ji}}{r_{1i}}$를 $\log\frac{p_{ji}}{p_{1i}}$ 代身 使用하여 一般化最小自乘法을 適用할 수 있다.[1)]

한편 式(3.10)을 다음과 같이 變形하여 最尤推定法을 適用할 수 있다.[2)]

$$p_{1i}=\frac{1}{1+\sum_{j=2}^{N}e^{X_i\beta_j}} \tag{3.11}$$

$$p_{mi}=\frac{e^{X_i\beta_m}}{1+\sum_{j=2}^{N}e^{X_i\beta_j}} \qquad m=2,\cdots,N$$

尤度函數 L 은

$$L=\prod_{i\in\theta_1}p_{i1}\prod_{i\in\theta_2}p_{i2}\cdots\prod_{i\in\theta_N}p_{iN}$$

$$=\prod_{i=1}^{T}\left\{\frac{1}{1+\sum_{j=2}^{N}e^{X_i\beta_j}}\right\}\prod_{m=2}^{N}\prod_{i\in\theta_m}e^{X_i\beta_m} \tag{3.12}$$

단, $\theta_j=\{i|\ j$번째 對象을 擇하는 경우$\}$

따라서 最尤推定値는 $\frac{\partial\log L}{\partial\hat{\beta}}=0$ 에서 얻어지는 非線型方程式으로부터 구하여 진다.

이때 0 또는 1의 값을 갖는 또 다른 從屬變數 z가 追加되어 y, z가 相互 因果關係를 갖는다면 y, z에 對한 로짓은 다음과 같이 聯立方程式 形態로 表記될 수 있다.[3)]

$$\log\left[\frac{p(z=1|y)}{p(z=0|y)}\right]_i=V_ib+ay_i \tag{3.13}$$

$$\log\left[\frac{p(y=1|z)}{p(y=0|z)}\right]_i=W_id+cz_i \quad i=1,\cdots,T$$

단, V_i: z에 영향을 주는 $(1\times k_z)$外生變數벡타

W_i: y에 영향을 주는 $(1\times k_y)$外生變數벡타

b, d: 各各 $(k_z\times 1)$, $(k_y\times 1)$인 母數벡타

a, c: 母數로 스칼라(scalar)

1) H. Theil, "On the Estimation of the Relationships Involving Qualitative Variables," *American Journal of Sociology*, Vol. 76(1970), pp.103—154.

2) P. Schmidt and R.P. Strauss, *loc. cit.*

3) P. Schmidt and R.P. Strauss, "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach," *Econometrica*, Vol.43, No.4(July, 1975), pp.745—755.

그런데 $a=c$가 되므로 z, y 의 結合確率은

$$
\begin{aligned}
p(z_i=0,\ y_i=0) &= \frac{1}{1+A} \\
p(z_i=0,\ y_i=1) &= \frac{e^{W_i d}}{1+A} \\
p(z_i=1,\ y_i=0) &= \frac{e^{V_i b}}{1+A} \\
p(z_i=1,\ y_i=1) &= \frac{e^{V_i b+W_i d+a}}{1+A}
\end{aligned}
\tag{3.14}
$$

단, $A=e^{W_i d}+e^{V_i b}+e^{W_i d+V_i b+a}$

와 같이 나타낼 수 있으며 最尤推定法을 위한 尤度函數는 다음과 같다.

$$L=\prod_{j=0}^{1}\prod_{l=0}^{1}\prod_{i\in\theta_{jl}} p(z_i=j,\ y_i=l) \tag{3.15}$$

단, $\theta_{jl}=\{i|z_i=j,\ y_i=l\}$

한편 이와 같은 二變量 二進模型에서 選擇對象이 增加하면 二變量 選多型 模型이 되며 이때의 表記 및 推定은 위의 方法과 비슷한 過程을 거쳐 導出될 수 있다.[1]

4. 二進數 從屬變數模型을 利用한 實證的 分析 : 單位地域農村의 耐久消費財 保有函數 推定

序論에서도 밝힌 바와 같이 實證的 分析 對象은 基本的이며 計算이 容易한 單一變數 二進模型으로 局限하였으며, 한 單位地域의 耐久消費財 保有實態를 線型確率模型(linear probability model), 프로빗模型(probit model), 로짓模型(logit model)에 의하여 分析하고자 하였다. 即 線型確率模型에서 얻어진 推定値를 基礎로 하여 各 模型에서의 確率로서 豫測値를 求하고 各各 模型의 경우 豫測力을 比較함으로써 어느 模型이 實際資料에 잘 附合되는가를 說明하고자 하였다.

4. 1. 資料의 蒐集 및 分析

分析對象資料는 1976年 8月 1日부터 7日 사이에 忠北 報恩郡 炭釜面 九岩里에서 調査된 標本規模 65의 農家實態調査資料를 利用하였다.[2] 調査資料에 包含된 많은 變數中 다음과 같은 理由에서 TV購入與否를 離散從屬變數로 擇하였다. 첫째 TV는 耐久消費財로서 單位地域 農村의 保有函數로 誘導될 수 있는 典型的인 二進數 模型이 될 수 있다는 점 및 둘

1) 表記 및 推定過程, 그리고 應用例는 P. Schmidt and R.P. Strauss, *ibid.*에서 찾을 수 있다.

2) 資料는 金相謙, 朴振根, 鄭暢泳, "傳統的 單位地域社會의 經濟行態分析,"「產業과 經營」第14卷 第1號(延世大學校 產業經營硏究所 : 1977, 8), 39—62面에 使用된 基礎資料로서 筆者가 調査에 直接 參加하였음.

께 調査對象 65家口中 TV購入家口數가 27家口로 購入, 非購入 어느 한 쪽에 치우쳐 있지 않았다는 점에서 였다.

다음 TV購入에 影響을 주리라 期待되는 變數는 全部 考慮하여 說明 및 被說明 變數를 다음과 같이 定義하였다.

$$y=\begin{cases}1 & \text{TV가 있으면}\\ 0 & \text{그렇지 않으면}\end{cases}$$

x_1=家口當 所得(원)

x_2=土地所有(坪)

x_3=常存家族數

$$x_4=\begin{cases}1 & \text{기와집이면}\\ 0 & \text{그렇지 않으면}\end{cases}$$

x_5=建坪(坪)

$$x_6=\begin{cases}1 & \text{家口員中 中卒以上이 있으면}\\ 0 & \text{그렇지 않으면}\end{cases}$$

$$x_7=\begin{cases}1 & \text{선풍기가 있으면}\\ 0 & \text{그렇지 않으면}\end{cases}$$

$$x_8=\begin{cases}1 & \text{라디오가 있으면}\\ 0 & \text{그렇지 않으면}\end{cases}$$

$$x_9=\begin{cases}1 & \text{農牛가 있으면}\\ 0 & \text{그렇지 않으면}\end{cases}$$

$$x_{10}=\frac{x_1}{x_3}=1\text{人當 所得}$$

이에 따라 相關係數 行列은 〈表 1〉 같이 구하여 졌다.

〈表 1〉에 의하면 x_1과 x_{10}, x_2와 x_{10}, x_1과 x_2가 높은 相關關係를 나타내고 있으나, 우선 當然히 多共線性(multicollinearity)이 存在하는 同一한 所得變數 x_1과 x_{10}을 區別하여 回歸模型을 세운 후 그 結果를 檢討하기로 하였다.

4. 2. 回歸模型의 設定

x_1과 x_{10}를 區別한 두개의 線型回歸模型을 通常最小自乘法에 의하여 推定한 結果는 〈表 2〉와 같다.

〈表 1〉 **相關係數 行列**

	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y	1										
x_1	0.4031	1									
x_2	0.5182	0.8935	1								
x_3	0.1907	0.3848	0.4516	1							
x_4	0.1530	0.2020	0.2020	0.1571	1						
x_5	0.2162	0.4867	0.4424	0.0704	0.3529	1					
x_6	0.3821	0.3660	0.3985	0.3298	0.0624	0.3076	1				
x_7	0.5203	0.3867	0.4371	0.1250	0.0474	0.2651	0.2467	1			
x_8	0.2814	0.2973	0.3743	-0.0231	-0.0400	0.4389	0.3329	0.3900	1		
x_9	0.1855	-0.0325	0.0836	0.0811	-0.0538	0.2500	0.0417	0.0722	-0.1174	1	
x_{10}	0.3638	0.8683	0.7446	-0.0660	0.1703	0.5331	0.2704	0.3576	0.2941	-0.0331	1

〈表 2〉 **1次 回歸 結果**

回歸 模型	常數項	x_1	x_2	x_3	x_4
$y^*=f(x_1, x_2, x_3, x_4,$	-0.0180	-0.00000004	0.0001	-0.0311	0.0506
$x_5, x_6, x_7, x_8, x_9)$	(-0.0916)	(-0.7537)	(2.0005)	(-0.8296)	(0.4059)
$y^{**}=f(x_2, x_4, x_5, x_6,$	-0.0418	—	0.00006	—	0.0188
$x_7, x_8, x_9, x_{10})$	(-0.2620)		(1.6705)		(0.1482)

x_5	x_6	x_7	x_8	x_9	x_{10}	R^2	F
-0.0025	0.2419	0.3608	-0.0057	0.1521	—	0.4491	4.9826
(-0.5474)	(1.9576)	(2.8987)	(-0.0424)	(1.2929)			
-0.0043	0.2226	0.3563	0.0523	0.1914	0.0000004	0.4375	5.4450
(-0.8975)	(1.8543)	(2.8531)	(0.4062)	(1.6782)	(0.2625)		

( )속의 數字는 t-統計量임

回歸方程式 y^*와 y^{**}에서 重相關係數는 各各 0.6702, 0.6615인 反面 x_1과 x_2, x_2와 x_{10} 사이의 相關係數는 各各 0.8935, 0.7446으로 나타나 높은 多共線性이 存在함을 보여주고 있다. 이는 農家所得의 大部分이 土地에서 緣由한다는 點에서 充分히 納得이 된다 하겠으며, x_5(建坪)은 推定된 回歸係數가 非有意的일 뿐만 아니라 農村이라는 點을 考慮한다면 그다지 큰 決定要素가 될 수 없으리라고 看做하여 x_2와 x_5을 除外한 나머지 變數들로 다시 回歸分析을 試圖하였다.[1)]

그 結果 x_3과 x_8(라디오 有無)은 어느 경우에도 非有意的으로 나타났으므로 두 變數를

1) 相關係數 行列 및 위 두번의 回歸分析 結果를 參酌하여 x_1(또는 x_{10}), x_6, x_7은 固定하고 x_3, x_4, x_8, x_9을 選擇的으로 考慮함으로써 24個의 代案的인 回歸方程式을 推定하였다.

除外한 나머지 8個의 回歸方程式中 變數의 經濟的 意味, 決定係數, t 및 F統計量 등을 考慮하여 〈表 3〉와 같이 最終的으로 3個의 代案的인 回歸方程式을 確定하였다.

〈表 3〉 **確定 回歸模型의 推定値**

回歸 模型	常數項	所得 x_1	家屋種類 x_4	教育程度 x_6	선풍기有無 x_7	農牛有無 x_9	1人當所得 x_{10}	R^2	F
$y^{(1)}=f(x_1,$	-0.0529	0.00000004	—	0.2427	0.4087	0.1717	—	0.3904	9.6060
$x_6, x_7, x_9)$	(-0.5017)	(1.5363)		(2.1081)	(3.4595)	(1.7086)			
$y^{(2)}=f(x_4, x_6,$	-0.1246	—	0.1052	0.2608	0.4157	0.1766	0.0000001	0.3991	7.8387
$x_7, x_9, x_{10})$	(-1.0147)		(1.0126)	(2.3250)	(3.5349)	(1.7530)	(1.3048)		
$y^{(3)}=f(x_6,$	-0.0697	—	—	0.2629	0.4141	0.1718	0.0000001	0.3887	9.5380
$x_7, x_9, x_{10})$	(-0.6324)			(2.3428)	(3.5212)	(1.7071)	(1.4791)		

(　)속의 數字는 t-統計量임.

〈表 3〉에서 보는 바와 같이 x_7이 y와 正의 方向으로 密接한 關係에 있는 점으로 보아 調査對象 部落의 경우 선풍기와 TV는 補完關係에 있다고 하겠다. 다음 教育程度가 높을 수록 TV를 所有하는 傾向이 强하게 나타났으며 이는 TV에 對한 社會一般의 通念과 一致된다고 하겠다. 한편 x_1과 x_9는 그렇게 높은 有意性을 보이고 있지는 않으나 두 變數 다 富의 程度가 TV所有에 미치는 影響을 肯定的으로 나타내고 있다는 점에서 大體的인 影響力이 把握될 수 있다.

4. 3. 推定方法에 따른 結果의 比較 檢討

가. 豫測基準의 設定

이미 確定된 3個의 回歸方程式에서 얻어진 推定値 $\hat{\beta}$을 基礎로 하여 다음과 같이 세가지 適用模型의 豫測値를 求하고 豫測基準을 定하였다.

1) 線型確率模型

$$\hat{y}_i = \boldsymbol{X}_i\hat{\boldsymbol{\beta}} \tag{4.1}$$

에 의하여 豫測確率 $\hat{y}_i$을 求하고

$$\hat{y}_i^p = \begin{cases} 1 & \hat{y}_i \geqq 0.5\text{이면} \\ 0 & \hat{y}_i < 0.5\text{이면} \end{cases} \tag{4.2}$$

와 같이 豫測基準을 定하였다.

2) 프로빗模型

프로빗模型은 平均 0, 分散 1의 累積正規分布를 前提로 하므로 式 (4.1)에서 얻어진 $\hat{y}_i$을 標準化하여 z_i을 求하였다

$$z_i=\frac{\hat{y}_i-\bar{\hat{y}}}{\sigma_{\hat{y}}} \tag{4.3}$$

단, $\bar{\hat{y}}$: $\hat{y}_i$의 平均

$\sigma_{\hat{y}}^2$: $\hat{y}_i$의 分散

이 z_i를 다음式 (4.4)에 代入하여 프로빗模型에서의 豫測確率 $\hat{\hat{y}}_i$을 求하였다.

$$\hat{\hat{y}}_i=\int_{-\infty}^{z_i}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}s^2}ds \tag{4.4}$$

이 경우 역시 豫測基準을

$$\hat{\hat{y}}_i^p=\begin{cases}1 & \hat{\hat{y}}_i\geqq 0.5\text{이면}\\ 0 & \hat{\hat{y}}_i<0.5\text{이면}\end{cases} \tag{4.5}$$

와 같이 하여 實際行動을 豫測하였다.

3) 로짓模型

로짓模型 $p_i=\frac{1}{1+e^{-X_i\beta}}$에서 $X_i\beta=0$일때 $p_i=\frac{1}{2}$이 된다는 점에서 $\hat{y}_i=X_i\hat{\beta}$ 代身 標準化한 z_i를 使用하여 다음과 같이 로짓模型에서의 豫測確率을 求하였다.

$$\tilde{y}_i=\frac{1}{1+e^{-z_i}} \tag{4.6}$$

이때 標準正規分布變數인 z를 變換하여 얻어진 y의 確率密度函數는 다음과 같다.[1)]

$$\begin{aligned}f(\tilde{y})&=g(z)\cdot|J|\\&=g\left(\log\frac{\tilde{y}}{1-\tilde{y}}\right)\cdot\frac{1}{\tilde{y}(1-\tilde{y})}\\&=\frac{1}{\sqrt{2\pi}\,\tilde{y}(1-\tilde{y})}e^{-\frac{1}{2}\left(\log\frac{\tilde{y}}{1-\tilde{y}}\right)^2}\end{aligned} \tag{4.7}$$

따라서

$$\int_0^{\omega}f(\tilde{y})d\tilde{y}=0.5 \tag{4.8}$$

와 같이 되는 點 ω를 찾아 로짓模型에서의 豫測基準을 다음과 같이 定하였다.

1) $f(\tilde{y})$는 바로 죤슨(N. Johnson)의 S_B分布中 한 形態이다.

即, $z=\varDelta+\delta\log\frac{\tilde{y}}{1-\tilde{y}}$ (단, z는 平均 0, 分散 1인 正規分布) 와 같은 S_B에서 $\varDelta=0$, $\delta=1$이 되면 여기에서 구하고자 하는 $f(\tilde{y})$와 一致한다. 자세한 것은 N. Johnson and S.Kotz, *Continuous Univariate Distributions*-1 (Boston : Houghton Mifflin Company, 1970), pp.23—26을 參照할 것.

$$\tilde{y}_i^p = \begin{cases} 1 & \tilde{y}_i \geqq \omega \text{이면} \\ 0 & \tilde{y}_i < \omega \text{이면} \end{cases} \tag{4.9}$$

그런데 $f(\tilde{y})$는 $\tilde{y}=0.5$를 中心으로 하는 單峰對稱分布의 密度函數이므로[1] 式 (4.8)에서 $\omega=0.5$가 되며 따라서 로짓模型에서의 豫測基準 역시 0.5가 된다.

이와 같은 세가지 推定方法을 各各의 回歸方程式에 適用하여 얻어진 豫測確率은 〈附表1〉과 같다.

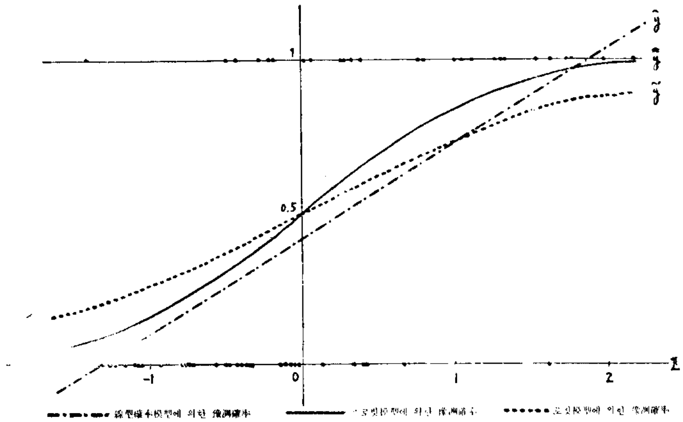


그림5. 豫側確率의 比較

그림 5에 의하면 로짓模型에 의한 豫測確率이 프로빗模型에 의한 것보다 0.5以下에서는 더 크게, 0.5以上에서는 더 작게 豫測되어 완만한 形態를 보이고 있으나, 두 模型의 豫測確率이 모두 0.5에서 交叉하고 있으므로 實際行動을 豫測하는 데는 影響을 미치지 않는다.[2]

나. 豫測誤差의 比較

豫測基準에 의하여 구하여진 $\hat{y}_i^p$, $\hat{\hat{y}}_i^p$, $\tilde{y}_i^p$와 觀察値 y와의 誤差를 平方平均誤差(mean squared error, MSE)에 의하여 比較함으로서 어느 模型이 實際資料에 잘 附合되는가를 보았다. 이때 各各의 平方平均誤差는

$$\text{線型確率模型의 경우}: \frac{\sum(y-\hat{y}_i^p)^2}{65} = MSE_1$$

$$\text{프로빗模型의 경우}: \frac{\sum(y-\hat{\hat{y}}_i^p)^2}{65} = MSE_2 \tag{4.10}$$

---

1) $f(\tilde{y})$의 指數部分 $-\frac{1}{2}\left(\log\frac{\tilde{y}}{1-\tilde{y}}\right)^2$에서 $\left(\log\frac{1-\tilde{y}}{\tilde{y}}\right)^2=\left(-\log\frac{\tilde{y}}{1-\tilde{y}}\right)^2=\left(\log\frac{\tilde{y}}{1-\tilde{y}}\right)^2$와 같이 되므로 $f(\tilde{y})$은 $\tilde{y}=0.5$를 中心으로 對稱을 이루게 된다.

2) $y^{(2)}$의 경우를 例로 들어 그린 것임.

로짓模型의 경우 : $\frac{\sum(y-\hat{y}_{i}^{p})^2}{65}=\text{MSE}_3$

이러한 過程에 의하여 3個의 回歸方程式에서 얻어진 結果는 다음의 〈表 4〉와 같다.

〈表 4〉 **平方平均誤差에 의한 比較**

回歸模型	MSE_1	MSE_2	MSE_3
$y^{(1)}=f(x_1, x_6, x_7, x_9)$	15/65	15/65	15/65
$y^{(2)}=f(x_4, x_6, x_7, x_9, x_{10})$	14/65	12/65	12/65
$y^{(3)}=f(x_6, x_7, x_9, x_{10})$	15/65	14/65	14/65

〈表 4〉에서 보여주는 바와 같이, TV需要를 設明해 주는 3個의 回歸模型 $y^{(1)}$, $y^{(2)}$, $y^{(3)}$ 中 $y^{(2)}$, $y^{(3)}$에 對하여 線型確率模型보다는 프로빗이나 로짓模型을 適用하는 것이 豫測力을 强化시켜 주고 있는 것으로 나타났으며, 그중 $y^{(2)}$가 豫測面에서 보아 가장 良好한 模型設定이라고 할 수 있겠다. $y^{(2)}$에서 프로빗模型의 경우 $\text{MSE}_2=\frac{12}{65}\fallingdotseq 0.18$ 이므로 약 82%가 이 豫測에 의하여 說明된다고 할 수 있으며, 大體로 보아 本研究에 使用된 實證資料에서는 回歸模型을 프로빗이나 로짓形態로 表記하는 것이 線型確率模型에 의한 表記보다 豫測力을 높여 준다고 할 수 있다.

5. 要約 및 結論

本論文에서는 첫째 離散從屬變數模型을 推定하기 위한 諸 分析方法을 紹介하고, 둘째 이의 實證的 分析을 위하여 單位地域農村의 耐久消費財 需要를 說明할 수 있는 3個의 代案的인 回歸模型을 세운 후, 各各에 線型確率模型(linear probability model), 프로빗模型(probit model), 로짓模型(logit model)을 適用하여 豫測確率을 구하고 豫測力을 比較하였다.

그 結果 첫째, 線型確率模型 보다는 變換過程을 거친 프로빗이나 로짓模型이 本研究에 使用된 資料에 더 잘 附合되는 것을 알 수 있었다. 둘째, 推定된 保有函數를 보면 所得의 增加는 TV購入에 正의 影響을 주나 豫想과는 달리 그다지 큰 有意性을 보이지 않았으며 教育水準이 높을수록 TV購入動機가 커짐을 보았고, 特히 선풍기와 TV는 補完財의 關係에 있음이 나타났다. 家屋種類나 農牛有無등의 有意性이 낮게 나타난 것은 一律的인 지붕改良事業 및 耕耘機等의 補給으로 그 原因을 찾을 수 있겠다. 또 實際로 TV를 보고 즐길 수 있다고 생각되는 15才以上으로 構成된 常存家族數는 거의 전혀 TV購入에 影響을 미치지 않는 것으로 나타났다.

實證分析에서의 接近方法이 모두 線型確率模型에서 얻어진 通常最小自乘推定量을 基礎로

하므로 이 推定量이 가지고 있는 問題點, 即 不均等分散에 의한 非有效性의 問題를 그대로 內包하고 있다는 점에서 이 接近方法은 弱點을 가진다 하겠다. 이러한 問題는 充分한 觀察値가 具備될 때 最尤推定法 및 一般化最小自乘法을 使用함으로써 解決될 수 있으며, 이러한 推定方法에 의한 豫測値의 附合與否는 앞으로 解決하여야 할 研究課題라 하겠다.

〈ABSTRACT〉

Holdings of Consumer's Durable Goods: Regressions with Binary Dependent Variables

Suk Bum Yoon
Hoe Kyung Lee*

This study has dual purposes: (1) survey of the specification and estimation of (i) the linear probability model, (ii) probit model, (iii) logit model, and (iv) Gompit model in which qualitative dependent variables are (i) dichotomous and (ii) polytomous, and (2) empirical estimations and comparison of predictability of these models in which holdings of television sets are presented as binary dependent variables.

It is obtained that the predictability defined as revised MSE is greatest in cases of probit and logit models and somewhat less with a linear model. However, the differences are not very remarkable.

* Both are at Yonsei University.

〈附表 1〉 **各模型에 의한 豫測值(I)**

家口番號	觀察值 y	$y=f(x_1, x_6, x_7, x_9)$					
		豫測確率			豫測值		
		$\hat{y}^{1)}$	$\hat{\hat{y}}^{2)}$	$\tilde{y}^{3)}$	$\hat{y}^{p}$	$\hat{\hat{y}}^{p}$	$\tilde{y}^{p}$
1	0	0.1263	0.1762	0.2826	0	0	0
2	0	0.2042	0.2483	0.3367	0	0	0
3	0	0.1353	0.1841	0.2885	0	0	0
4	0	0.1368	0.1841	0.2895	0	0	0
5	0	−0.0309	0.0749	0.1918	0	0	0
6	0	−0.0286	0.0764	0.1930	0	0	0
7	0	−0.0261	0.0778	0.1942	0	0	0
8	0	0.2174	0.2611	0.3457	0	0	0
9	0	0.2188	0.2643	0.3467	0	0	0
10	0	0.2249	0.2676	0.3512	0	0	0
11	0	−0.0169	0.0823	0.1989	0	0	0
12	0	−0.0162	0.0823	0.1992	0	0	0
13	0	−0.0160	0.0823	0.1994	0	0	0
14	0	0.4017	0.4840	0.4890	0	0	0
15	0	−0.0125	0.0838	0.2016	0	0	0
16	0	0.1598	0.2061	0.3050	0	0	0
17	0	0.4005	0.4801	0.4881	0	0	0
18	0	0.1659	0.2119	0.3092	0	0	0
19	1	0.4141	0.4989	0.4990	0	0	0
20	1	0.4146	0.4992	0.4994	0	0	0
21	1	0.0032	0.0918	0.2095	0	0	0
22	0	0.2482	0.2946	0.3685	0	0	0
23	1	0.2482	0.2946	0.3685	0	0	0
24	0	0.4206	0.5080	0.5042	0	1	1
25	1	0.6624	0.7881	0.6891	1	1	1
26	1	0.8367	0.9115	0.7954	1	1	1
27	0	0.2565	0.3050	0.3747	0	0	0
28	0	0.6664	0.7910	0.6919	1	1	1
29	1	0.4314	0.5199	0.5128	0	1	1
30	0	0.2599	0.3085	0.3773	0	0	0
31	0	0.2608	0.3085	0.3779	0	0	0
32	0	0.8527	0.9207	0.8037	1	1	1
33	0	0.2056	0.2483	0.3372	0	0	0
34	1	0.6871	0.8106	0.7059	1	1	1
35	0	0.6189	0.7454	0.6583	1	1	1

家口番號	觀察值 y	$y=f(x_1,x_6,x_7,x_9)$					
		豫 測 確 率			豫 測 値		
		$\hat{y}$[1]	$\hat{\hat{y}}$[2]	$\tilde{y}$[3]	$\hat{y}^p$	$\hat{\hat{y}}^p$	$\tilde{y}^p$
36	1	0.4550	0.5517	0.5319	0	1	1
37	0	0.2126	0.2578	0.3422	0	0	0
38	1	0.8656	0.9264	0.8101	1	1	1
39	1	0.2197	0.2643	0.3474	0	0	0
40	1	0.8725	0.9292	0.8135	1	1	1
41	0	0.2946	0.3483	0.4039	0	0	0
42	1	0.6394	0.7642	0.6730	1	1	1
43	0	0.0609	0.1271	0.2419	0	0	0
44	0	0.0625	0.1271	0.2428	0	0	0
45	0	0.3100	0.3669	0.4159	0	0	0
46	1	0.3165	0.3745	0.4210	0	0	0
47	1	0.7280	0.8438	0.7325	1	1	1
48	0	0.3201	0.3783	0.4239	0	0	0
49	0	0.5030	0.6102	0.5701	1	1	1
50	0	0.3348	0.3974	0.4354	0	0	0
51	1	0.5083	0.6179	0.5743	1	1	1
52	1	0.7513	0.8599	0.7470	1	1	1
53	1	0.7521	0.8621	0.7475	1	1	1
54	0	0.5255	0.6368	0.5878	1	1	1
55	1	0.9346	0.9616	0.8420	1	1	1
56	1	0.9373	0.9625	0.8432	1	1	1
57	1	0.7713	0.8749	0.7590	1	1	1
58	1	0.9496	0.9573	0.8483	1	1	1
59	0	0.5374	0.6517	0.5971	1	1	1
60	1	0.5383	0.6554	0.5978	1	1	1
61	1	0.3668	0.4364	0.4609	0	0	0
62	1	0.9806	0.9656	0.8608	1	1	1
63	0	0.4264	0.5160	0.5088	0	1	1
64	1	0.8851	9.9345	0.8196	1	1	1
65	1	1.2078	0.9938	0.9278	1	1	1

$\sum(y-\hat{y}^p)^2=15$

$\sum(y-\hat{\hat{y}}^p)^2=15$

$\sum(y-\tilde{y}^p)^2=15$

1) $\hat{y}$: 線型確率模型에 의한 豫測確率

2) $\hat{\hat{y}}$: 프로빗模型에 의한 豫測確率

3) $\tilde{y}$: 로짓模型에 의한 豫測確率

各模型에 의한 豫測値(Ⅱ)

家口番號	觀察値 y	$y=f(x_4, x_6, x_7, x_9, x_{10})$					
		豫測確率			豫測値		
		$\hat{y}$	$\hat{\hat{y}}$	$\tilde{y}$	$\hat{y}^p$	$\hat{\hat{y}}^p$	$\tilde{y}^p$
1	0	0.1696	0.2177	0.3136	0	0	0
2	0	0.2582	0.3085	0.3773	0	0	0
3	0	0.0703	0.1359	0.2498	0	0	0
4	0	0.0819	0.1446	0.2568	0	0	0
5	0	−0.0880	0.0548	0.1674	0	0	0
6	0	−0.0842	0.0559	0.1691	0	0	0
7	0	0.0251	0.1075	0.2238	0	0	0
8	0	0.1822	0.2296	0.3223	0	0	0
9	0	0.1684	0.2177	0.3128	0	0	0
10	0	0.2803	0.3336	0.3941	0	0	0
11	0	−0.0947	0.0516	0.1644	0	0	0
12	0	0.0049	0.0951	0.2128	0	0	0
13	0	0.0214	0.1038	0.2217	0	0	0
14	0	0.4514	0.5478	0.5287	0	1	1
15	0	0.0477	0.1210	0.2365	0	0	0
16	0	0.1202	0.1736	0.2807	0	0	0
17	0	0.3654	0.4364	0.4603	0	0	0
18	0	0.1885	0.2358	0.3267	0	0	0
19	1	0.4203	0.5080	0.5040	0	1	1
20	1	0.3570	0.4247	0.4536	0	0	0
21	1	−0.0314	0.0778	0.1941	0	0	0
22	0	0.2803	0.3336	0.3940	0	0	0
23	1	0.2803	0.3336	0.3940	0	0	0
24	0	0.3784	0.4522	0.4705	0	0	0
25	1	0.7633	0.8665	0.7519	1	1	1
26	1	0.7727	0.8729	0.7574	1	1	1
27	0	0.2472	0.2946	0.3691	0	0	0
28	0	0.6271	0.7486	0.6626	1	1	1
29	1	0.5343	0.6480	0.5936	1	1	1
30	0	0.2528	0.3015	0.3733	0	0	0
31	0	0.2886	0.3446	0.4003	0	0	0
32	0	0.9251	0.9474	0.8354	1	1	1
33	0	0.3015	0.3594	0.4102	0	0	0
34	1	0.7307	0.8438	0.7321	1	1	1
35	0	0.5436	0.6591	0.6007	1	1	1

家口番號	觀察値 y	$y=f(x_4,x_6,x_7,x_9,x_{10})$					
		豫測確率			豫測値		
		$\hat{y}$	$\hat{\hat{y}}$	$\tilde{y}$	$\hat{y}^p$	$\hat{\hat{y}}^p$	$\tilde{y}^p$
36	1	0.4957	0.6026	0.5637	0	1	1
37	0	0.2351	0.2843	0.3602	0	0	0
38	1	0.9129	0.9441	0.8300	1	1	1
39	1	0.3249	0.3859	0.4294	0	0	0
40	1	1.0037	0.9700	0.8670	1	1	1
41	0	0.2995	0.3557	0.4087	0	0	0
42	1	0.6658	0.7881	0.6896	1	1	1
43	0	0.0751	0.1401	0.2527	0	0	0
44	0	0.0573	0.1271	0.2421	0	0	0
45	0	0.2913	0.3483	0.4024	0	0	0
46	1	0.3468	0.4129	0.4456	0	0	0
47	1	0.7647	0.8665	0.7528	1	1	1
48	0	0.2446	0.2946	0.3672	0	0	0
49	0	0.5480	0.6591	0.6041	1	1	1
50	0	0.4021	0.4840	0.4894	0	0	0
51	1	0.4993	0.6064	0.5665	0	1	1
52	1	0.6535	0.7764	0.6811	1	1	1
53	1	0.8273	0.9049	0.7880	1	1	1
54	0	0.4037	0.4840	0.4907	0	0	0
55	1	0.9248	0.9474	0.8353	1	1	1
56	1	0.8211	0.9015	0.7847	1	1	1
57	1	0.8007	0.8887	0.7735	1	1	1
58	1	1.0325	0.9756	0.8773	1	1	1
59	0	0.5155	6.6255	0.5791	1	1	1
60	1	0.5160	0.6255	0.5794	1	1	1
61	1	0.2539	0.3050	0.3741	0	0	0
62	1	0.8062	0.8944	0.7765	1	1	1
63	0	0.3987	0.4801	0.4867	0	0	0
64	1	0.8953	0.9370	0.8219	1	1	1
65	1	1.1635	0.9913	0.9156	1	1	1

$\sum(y-\hat{y}^p)^2=14$

$\sum(y-\hat{\hat{y}}^p)^2=12$

$\sum(y-\tilde{y}^p)^2=12$

各模型에 의한 豫測値(Ⅲ)

家口番號	觀察値 y	$y=f(x_6,x_7,x_9,x_{10})$					
		豫測確率			豫測値		
		$\hat{y}$	$\hat{\hat{y}}$	$\tilde{y}$	$\hat{y}^p$	$\hat{\hat{y}}^p$	$\tilde{y}^p$
1	0	0.1161	0.1660	0.2755	0	0	0
2	0	0.2120	0.2546	0.3414	0	0	0
3	0	0.1226	0.1711	0.2798	0	0	0
4	0	0.1357	0.1841	0.2884	0	0	0
5	0	−0.0287	0.0764	0.1924	0	0	0
6	0	−0.0244	0.0778	0.1946	0	0	0
7	0	−0.0198	0.0793	0.1969	0	0	0
8	0	0.2446	0.2912	0.3655	0	0	0
9	0	0.2292	0.2743	0.3541	0	0	0
10	0	0.2369	0.2810	0.3597	0	0	0
11	0	−0.0362	0.0721	0.1887	0	0	0
12	0	−0.0424	0.0694	0.1857	0	0	0
13	0	−0.0239	0.0778	0.1948	0	0	0
14	0	0.4025	0.4840	0.4896	0	0	0
15	0	0.0055	0.0934	0.2102	0	0	0
16	0	0.1785	0.2236	0.3176	0	0	0
17	0	0.4277	0.5160	0.5099	0	1	1
18	0	0.1372	0.1841	0.2894	0	0	0
19	1	0.4630	0.5596	0.5383	0	1	1
20	1	0.4144	0.4990	0.4992	0	0	0
21	1	0.0348	0.1112	0.2263	0	0	0
22	0	0.2367	0.2810	0.3595	0	0	0
23	1	0.2367	0.2810	0.3595	0	0	0
24	0	0.4384	0.5279	0.5185	0	1	1
25	1	0.7262	0.8413	0.7319	1	1	1
26	1	0.8286	0.9082	0.7916	1	1	1
27	0	0.3175	0.3745	0.4216	0	0	0
28	0	0.6915	0.8132	0.7093	1	1	1
29	1	0.4952	0.6026	0.5641	0	1	1
30	0	0.3238	0.3821	0.4266	0	0	0
31	0	0.2460	0.2912	0.3666	0	0	0
32	0	0.8815	0.9332	0.8184	1	1	1
33	0	0.2638	0.3121	0.3800	0	0	0
34	1	0.6898	0.8133	0.7081	1	1	1
35	0	0.6013	0.7258	0.6457	1	1	1

家口番號	觀察値 y	$y=f(x_6, x_7, x_9, x_{10})$					
		豫測確率			豫測値		
		$\hat{y}$	$\hat{\hat{y}}$	$\tilde{y}$	$\hat{y}^p$	$\hat{\hat{y}}^p$	$\tilde{y}^p$
36	1	0.4520	0.5478	0.5295	0	1	1
37	0	0.1895	0.2327	0.3253	0	0	0
38	1	0.8678	0.9279	0.8117	1	1	1
39	1	0.2900	0.3409	0.4002	0	0	0
40	1	0.9656	0.9633	0.8569	1	1	1
41	0	0.2582	0.3050	0.3757	0	0	0
42	1	0.6204	0.7454	0.6597	1	1	1
43	0	0.0362	0.1112	0.2271	0	0	0
44	0	0.0162	0.0985	0.2160	0	0	0
45	0	0.2491	0.2981	0.3689	0	0	0
46	1	0.3112	0.3669	0.4167	0	0	0
47	1	0.7278	0.8438	0.7328	1	1	1
48	0	0.3145	0.3707	0.4193	0	0	0
49	0	0.6285	0.6549	0.6655	1	1	1
50	0	0.3731	0.4443	0.4660	0	0	0
51	1	0.4561	0.6517	0.5328	0	1	1
52	1	0.7210	0.8389	0.7285	1	1	1
53	1	0.7980	0.8925	0.7748	1	1	1
54	0	0.4667	0.5675	0.5414	0	1	1
55	1	0.8811	0.9332	0.8182	1	1	1
56	1	0.8828	0.9345	0.8190	1	1	1
57	1	0.7681	0.8729	0.7576	1	1	1
58	1	1.0018	0.9706	0.8692	1	1	1
59	0	0.4741	0.5754	0.5473	0	1	1
60	1	0.4747	0.5754	0.5477	0	1	1
61	1	0.3250	0.3821	0.4275	0	0	0
62	1	0.8661	0.9279	0.8109	1	1	1
63	0	0.3694	0.4404	0.4629	0	0	0
64	1	0.8741	0.9306	0.8148	1	1	1
65	1	1.1745	0.9929	0.9207	1	1	1

$\sum(y-\hat{y}^p)^2=15$

$\sum(y-\hat{\hat{y}}^p)^2=14$

$\sum(y-\tilde{y}^p)^2=14$

參 考 文 獻

〔1〕 Amemiya, Takeshi, "Bivariate Probit Analysis: Minimum Chi-square Methods," *Journal of the American Statistical Association*, Vol. 69(1974), pp.940-944.

〔2〕 Amemiya, Takeshi, "The Maximum Likelihood, the Minimum Chi-square and the Nonlinear Weighted Least-squares Estimator in the General Qualitative Response Model," *Journal of the American Statistical Association*, Vol. 71(1976), pp.347-351.

〔3〕 Berkson, Joseph, "Application of the Logistic Function to Bio-assay," *Journal of the American Statistical Association*, Vol. 39 (1944), pp.357-365.

〔4〕 Berkson, Joseph, "Approximation to Chi-square by 'Probits' and by 'Logits', *Journal of the American Statistical Association*, Vol. 41(1946), pp.70-74.

〔5〕 Berkson, Joseph, "A Statistically Precise and Relatively Simple Method of Estimating the Bio-assay with Quantal Response, Based on the Logistic Funtion," *Journal of the American Statistical Association*, Vol. 48 (1953), pp.565-599.

〔6〕 Berkson, Joseph, "Estimate of the Integrated Normal Curve by Minimum Normit Chi-square with Particular Reference to Bio-assay," *Journal of the American Statistical Association*, Vol. 50(1955), pp.529-549.

〔7〕 Berkson, Joseph, "Maximum Likelihood and Minimum Chi-square Estimates of the Logistic Function," *Journal of the American Statistical Association*, Vol. 50(1955), pp.130-162.

〔8〕 Finney, D.J., *Probit Analysis*, 3rd ed., Cambridge Univ. Press, Cambridge, 1971.

〔9〕 Fisher, Janet A., "An Analysis of Consumer Goods Expenditure in 1957," *The Review of Economics and Statistics*, Vol. 44(1962), pp.64-71.

〔10〕 Goldberger, Arthur S., *Econometric Theory*, John Wiley & Sons, New

York, 1964.

[11] Goldberger, Arthur S.,"Correlations between Binary Outcomes and Probabilistic Predictions," *Journal of the American Statistical Association*, Vol. 68(1973), p.84.

[12] Goldfeld, Stephen M. and Quandt, Richard E., *Nonlinear Methods in Econometrics*, North-Holland, Amsterdam, 1972.

[13] Gunderson, Morley, "Retention of Trainees: A Study with Dichotomous Dependent Variables," *Journal of Econometrics*, Vol. 2(1974), pp.79-93.

[14] Johnson, Norman L. and Kotz, Samuel, *Continuous Univariate Distributions-1*, Houghton Mifflin Company, Boston, 1970.

[15] Lee, Tong Hun, "Demand for Housing: A Cross-Section Analysis," *The Review of Economics and Statistics*, Vol. 45(1963), pp.190-196.

[16] McFadden, Daniel, "Conditional Logit Analysis of Qualitative Choice Behavior," in Zarembka P. (ed.), *Frontiers in Econometrics*, Academic Press, New York, 1973.

[17] McGillivray, Robert G., "Estimating the Linear Probability Function," *Econometrica*, Vol. 38(1970), pp.775-776.

[18] Mingche, M.L., "A Logit Model of Homeownership," *Econometrica*, Vol. 45(1977), pp.1081-1097.

[19] Morrison, Donald G., "Upper Bounds for Correlations between Binary Outcomes and Probabilistic Predictions," *Journal of the American Statistical Association*, Vol. 67(1972) pp.68-70.

[20] Orcutt, G.H., Greenberger, M., Korbel J. and Rivlin, A.M., *Microanalysis of Socioeconomic Systems; A Simulation Study*, Harper & Row, New York, 1961.

[21] Pindyck, R.S. and Rubinfeld, D.L. *Econometric Models and Economic Forecasts*, McGraw-Hill, New York, 1976.

[22] Schmidt, Peter and Strauss, Robert P., "Estimation of Models with Jointly Dependent Qualitative Variables: A Simultaneous Logit Approach,"

Econometrica, Vol. 43(1975), pp.745-755.

[23] Schmidt, Peter and Strauss, Robert P., "The Prediction of Occupation Using Multiple Logit Models," *International Economic Review*, Vol. 16 (1975), pp.471-486.

[24] Theil, Henri, "A Multinomial Extension of the Linear Logit Model," *International Economic Review*, Vol. 10(1969), pp.251-259.

[25] Theil, Henri, "On the Estimation of Relationships Involving Qualitative Variables," *American Journal of Sociology*, Vol. 76(1970), pp.103-154.

[26] Theil, Henri, *Economics and Information Theory*, Rand McNally, Chicago, 1967.

[27] Theil, Henri, *Principles of Econometrics*, John Wiley & Sons, New York, 1971.

[28] Tobin, James, "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, Vol. 26(1958), pp.24-36.

[29] Westin, Richard B., "Predictions from Binary Choice Models," *Journal of Econometrics*, Vol. 2(1974), pp.1-16.

[30] Zellner, Arnold and Lee, Tong Hun, "Joint Estimation of Relationships Involving Discrete Random Variables," *Econometrica*, Vol. 33(1965), pp. 382-394.