

로짓모형을 이용한 질적 종속변수의 분석

Application of Logit Model in Qualitative Dependent Variables

신구전문대학 가정과
부교수 : 李 吉 順

*Dept. of Home Economics
Shingu Junior College
Associate Prof. : Kil Soon Lee*

연세대학교 건축공학과
교수 : 俞 浣

*Dept. of Architect Engineering
Yonsei University
Prof. : Wann Yu*

목 차

- | | |
|-----------------|--------------|
| I. 문제의 제기 | IV. 로짓모형의 적용 |
| II. 회귀분석 적용의 문제 | V. 요약 |
| III. 로짓모형의 특성 | 참고문헌 |

= ABSTRACT =

Regression analysis has become a standard statistical tool in the behavioral science. Because of its widespread popularity, regression has been often misused. Such is the case when the dependent variable is a qualitative measure rather than a continuous, interval measure. Regression estimates with a qualitative dependent variable does not meet the assumptions underlying regression. It can lead to serious errors in the standard statistical inference.

Logit model is recommended as alternatives to the regression model for qualitative dependent variables. Researchers can employ this model to measure the relationship between independent variables and qualitative dependent variables without assuming that logit model was derived from probabilistic choice theory.

Coefficients in logit model are typically estimated by the method of Maximum Likelihood Estimation in contrast to ordinary regression model which estimated by the method of Least Squares Estimation.

Goodness of fit in logit model is based on the likelihood ratio statistics and the t-statistics is used for testing the null hypothesis.

I. 문제의 제기

종합과학적 성격을 가진 가정학의 특성상 사회과학적 접근방법을 이용하는 분야에서는 자주 범주화된 질적 변수를 다루게 된다. 변수의 분석은 변수의 측정수준에 맞추어 분석기법이 적용되어야 하므로 질적 변수의 분석에는 명목척도와 서열척도를 분석하는 χ^2 , ϕ , T, Kendall's Tau-a, b, Gamma 등이 이용된다.

그러나 현실적으로 대부분의 연구에서는 몇개의 변수를 동시에 분석하거나 두 변수 이상간의 관계의 강도에 대한 해석을 필요로 하므로, 위의 분석기법들이 갖고 있는 변수 수의 제한이나 해석의 한계성 등의 문제를 극복하기 위하여 변수의 측정수준을 변경하여 보다 높은 수준의 분석방법을 적용하는 경우가 많다.

등간척도 이상의 연속변수들의 관계를 분석하는 여러가지 분석기법 중에서 회귀분석은 가정학은 물론 사회학, 정치학, 경제학 등 인간의 행동을 다루는 분야에서 가장 널리 사용되는 분석기법이다. 회귀분석이 자주 이용되는 이유는 여러개의 변수를 동시에 분석할 수 있고, 설명력이 좋고, 해석이 용이할 뿐만아니라 컴퓨터를 이용한 통계처리 과정이 쉽기 때문이다. 이러한 이유로 연구자들은 자료가 회귀분석의 가정(assumption)과 맞지 않음에도 불구하고 명목척도나 서열척도인 질적 변수들을 등간척도화하여 회귀분석을 시도하곤한다. 회귀분석의 어떤 가정은 그것이 위반되었을 때 심각한 오류를 포함한 결과가 추정되는데, 종속변수가 연속적인 등간변수보다 질적 변수일 때 그런 경우가 많다.

따라서 본 연구에서는 이러한 오류를 방지하고 정확한 분석결과를 추정할수 있도록 첫째, 종속변수가 질적 변수인 경우 회귀분석을 적용했을때 야기되는 문제를 제시하고, 둘째, 질적 변수를 분석하는 대체적 기법인 로짓모형의 기본 이론을 고찰하여 가정학 분야에서 로짓모형의 응용방안을 모색하고자 한다.

II. 회귀분석 적용의 문제

그동안 밝혀진 가정학 분야의 연구에 의하면 한 변수의 변화는 두개 이상의 변수들로 부터 동시에 영향을 받게 된다. 회귀분석은 등간척도로 구성된 두개 이상의 변수들 간의 관련성을 찾아내기 위해 종속변수의 변화를 독립변수의 선형조합으로부터 회귀식을 추정하는 다변량분석기법이다.

회귀모형은 독립변수 X와 종속변수 Y 사이에 선형관계가 성립한다고 가정하므로 변수들 간의 관계를 다음과 같이 나타낼 수 있다.

$$Y_i = a + bX_i \dots\dots\dots (식 2.1)$$

이 식에 의하면 특정의 X값에 반응하는 Y값은 오직 하나뿐이다. 그러나 실제로 관찰하여 보면 실제 관측치들과 이들을 근거로 도출한 직선사이에는 편차가 존재한다. 이러한 편차는 함수관계에 포함되어야 할 변수가 제외된 경우나 사람들의 불규칙적인 행태, 모형의 수학적 형태에 대한 불완전한 설정 등에 기인한다(유지성, 1989). 이러한 편차 즉, Y_i 와 $a + bX_i$ 의 차이를 설명하기 위해 함수에 오차항(error term 또는 stochastic term)이라 불리는 확률변수 e_i 를 첨가함으로써 회귀모형은 다음과 같은 확률적 모형이 된다.

$$Y_i = a + bX_i + e_i \dots\dots\dots (식 2.2)$$

그러므로 Y_i 의 변동은 설명된 변동인 X_i 에서 기인하는 부분과 설명안된 변동인 확률항 e_i 의 영향에서 기인하는 부분에 의해 설명된다.

회귀모형은 아래와 같은 여러 가정에 기초하고 있다(이영준, 1991)

- A. 특정화 과정에서 오차(specification error)가 없다. 즉 X_i 와 Y_i 의 관계는 선형이며 종속변수를 설명하는데 적절한 독립변수는 모두 회귀식에 포함되어 있다.
- B. 측정오차(measurement error)가 없다. 즉 변수 X_i 와 Y_i 는 정확하게 측정되었다.
- C. 오차항 e_i 에 관한 가정
 - 1) 오차항 e_i 의 평균은 0이다. 즉, $E(e_i) = \sum e_i/n = 0$
 - 2) 오차항 e_i 의 분산은 모든 X_i 값에 대해 일정

하다(homoscedasticity). 즉, $E(e_i) = \sigma^2$

3) 오차항 e_i 는 상호독립적이다. 즉 $E(e_i e_j) = 0$, ($i \neq j$).

4) 오차항 e_i 와 독립변수 X_i 는 상호독립적이다. 즉 $E(X_i e_i) = 0$.

5) 오차항 e_i 는 정규분포를 이룬다.

가정 A에서 C까지 모두 위반하지 않는 경우 표본회귀계수(b_i)는 모회귀계수(β_i)의 최우수불편추정치(best unbiased estimates)가 되며 C의 1)이 위반된 경우 그리 큰 문제는 없으나 Y절편 a 는 불편추정치가 아니다. C의 2)가 위반된 경우 최소자승법에 의한 추정치는 여전히 불편추정치이지만 가설검정이나 신뢰구간은 믿을 수 없게 된다. C의 3)을 위반한 경우, 여전히 불편추정치이지만 가설검정이나 신뢰구간은 전혀 옳지 않으며 회귀모형은 실제보다 더욱 좋게 나타난다. C의 4)를 위반한 경우, 최소자승법에 의한 추정치는 불편추정치가 아니며 오차의 분산은 실제보다 작게 나타난다. C의 5)를 위반한 경우 중심극한정리(central limit theorem)를 이용하기 위해 표본의 수를 늘리면 된다.

회귀모형은 독립변수가 서로 정확한 선형관계가 아닐 것 이외에는 그 값에 대해 제한이 없으나 종속변수는 연속변수로 가정하여 Y_i 는 $-\infty$ 에서 $+\infty$ 까지의 값을 자유로 갖는다. 그러나 종속변수가 불연속변수(discrete variable)로 0 혹은 1 값을 갖는 이분화된 변수이거나 범주화된 변수인데 회귀분석을 했을 때는 위의 가정과 맞지 않는 다음과 같은 문제가 발생한다(Aldrich & Nelson, 1984).

첫째, 연속변수가 아닌 질적 변수는 그 기대치가 0과 1사이이므로 비선형이기 쉬운데 선형을 가정하게 되는 경우 진실한 관계를 왜곡하는 체계적 오차가 발생한다.

둘째, 계수의 추정상 이분산(heteroscedasticity)문제가 발생한다. (식 2.2)에서 Y_i 가 (1,0)의 두 값, 즉 첫번째 대안이 선택될 경우 1이고, 두번째 대안이 선택될 경우 0인 두 값만 가진다면 $P_i = \text{Prob}(Y_i=1)$, $1-P_i = \text{Prob}(Y_i=0)$ 이 되므로 Y_i 의 기대값은 다음과 같다.

$$E(Y_i) = 1 \cdot P_i + 0 \cdot (1-P_i) = P_i \dots\dots\dots (\text{식 2.3})$$

따라서 회귀식은 Y_i 가 0~1사이로 제한되는 확

률로 해석되어진다. 이러한 이유로 종속변수가 0 혹은 1인 선형회귀모형은 선형확률모형(linear probability model)이라 불리운다. (식 2.2)에 따르면 Y_i 의 값에 따른 오차항 e_i 는 $Y_i=1$ 일때 $1-a-bX_i$, $Y_i=0$ 일 때는 $-a-bX_i$ 가 된다. 따라서 오차항에 대한 가정에 의하면

$$E(e_i) = (1-a-bX_i)P_i + (-a-bX_i)(1-P_i) = 0 \text{ 이 되며}$$

$$\text{이를 } P_i \text{에 관해 풀면 } P_i = a + bX_i$$

$$1 - P_i = 1 - a - bX_i \text{ 이므로}$$

$$\begin{aligned} e_i \text{의 분산은 } E(e_i) &= (1-a-bX_i)P_i + (-a-bX_i)(1-P_i) \\ &= (1-a-bX_i)(a+bX_i) \\ &= P_i(1-P_i) \\ &= E(Y_i)[1-E(Y_i)] \text{ 이 된다.} \end{aligned}$$

따라서 P_i 가 0이나 1에 가까운 관찰치들은 상대적으로 작은 분산을 가지고 P_i 가 0.5에 가까운 관찰치들은 상대적으로 큰 분산을 가지게 되어 이분산의 문제가 발생한다.

세째, (그림 1-1)에서 $X > X^*$ 이면 e 값은 부(negative)이며 X 의 증가함수이고, $X < X^*$ 이면 e 값은 정적(positive)이므로 e 값과 X 값 간에 상관관계가 있다. 따라서 가정 C의 4)를 위반한다.

또한 X 의 값이 대단히 커지거나 작은 경우 예측된 확률값이 1보다 커지거나 0보다 작아질 수 있는데 이렇게 되면 예측된 확률값이 $0 < P_i < 1$ 의 범주를 벗어나게 되므로 추정된 종속변수가 확률에서 요구되어지는 간격을 만족시킬 수 없다.

이상에서 살펴본 바와 같이 질적인 종속변수를

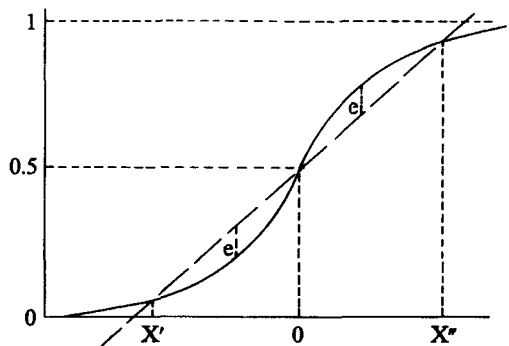


그림 1-1. 선형추정에 따른 오차.

선형으로 가정하여 회귀모형으로 추정하였을 때는 추정치가 불편추정치가 아니며 가설검정이나 신뢰구간도 타당하지 않으므로 추정결과는 유효성을 상실하게 된다.

III. 로짓모형의 특성

회귀모형이 갖는 문제를 해결할 수 있는 대안적 비선형모형으로는 누적확률분포를 이용한 프로빗 모형(probit model)과 로짓모형(logit model)이 넓게 활용되고있다(Hanushek & Jackson, 1977). 두 모형은 매우 유사하여 실증분석의 결과는 그 차이가 0.02 미만인데 프로빗모형은 정산하는데 어려움이 있고 쉽게 일반화시킬 수 없어 실증연구에서는 로짓모형이 더 많이 이용된다(그림 1-2).

로짓모형은 효용극대화이론과 현시적선택이론에 근거하여 선택행위를 확률적 현상으로 가장 잘 설명할 수 있다고 한 Luce(1959)에 의해 제안된 후 McFadden(1973)이 이론적 기초를 세운 1970년대 중반부터 개발이 본격화된 확률선택이론에서 유도되어(로짓모형이 유도되는 자세한 과정은 유완, 음성직, 김설주, 1983; Shin, 1985; 이길순, 1990을 참조) 다음과 같은 다항로짓모형(multinomial logit model)의 일반식으로 정리된다.

$$P_{it} = \frac{e^{V_{it}}}{\sum_{j=1}^J e^{V_{jt}}} \quad (j=1,2,\dots,J) \dots\dots\dots (\text{식 3.1})$$

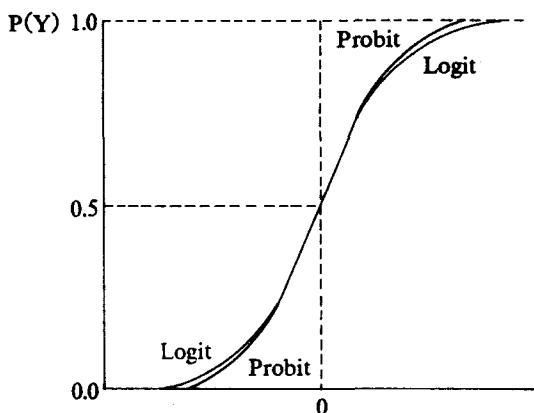


그림 1-2. 로짓과 프로빗 모형의 누적분포.

P_{it} : 가구 또는 개인 t 가 선택대안 중 i 를 선택할 확률

V_{it} : 가구 또는 개인 t 의 사회경제적 특성과 대안 i 의 특성 벡터

V_{jt} : 가구 또는 개인 t 의 사회경제적 특성과 대안 j 의 특성 벡터

로짓모형은 선택대안 상호간의 독립성(independence from irrelevant alternatives: 이하 IIA)이라는 공리(axiom)에 기초를 두고 있다. 이것은 두 선택대안 i, j 의 선택확률 비율은

$$\frac{P_{it}}{P_{jt}} = \frac{V_{it}}{V_{jt}} = V_{it} - V_{jt} \dots\dots\dots (\text{식 3.2})$$

와 같다는 것이다. 즉, 어떤 특정 개인에게 두 선택대안의 선택 확률비는 새로운 선택대안이 나타났을 경우에도 변화가 없다(ratio constant)는 것이다(Yu, 1978).

이 IIA 성질로 로짓모형의 추정과정이 훨씬 단순화되며 모형에 다른 선택대안을 추가하여도 재추정할 필요가 없게 되므로, 모집단을 대상으로 하건 표본을 대상으로 하건 추정 결과가 동일하기 때문에 자료수집 비용도 줄일 수 있어 상당히 유용하다(유완, 음성직, 김설주, 1983). 그러나 이공리는 모형의 적용을 제한하는 요인이 되기도 한다. 즉 선택대안의 집합안에 있는 대안들이 선택자에게 유사하게 인식될 때는 대안의 선택확률비가 변하게 되므로 IIA 공리에 근거한 로짓모형을 적용하면 추정결과가 정확하지 않게 된다.

선택대상의 대안들이 유사한 속성을 갖고 있어 독립성 가정이 불합리하다고 판단될 때는 조건부 로짓모형(conditional logit model 혹은 nested logit model)을 이용할 수 있다. 이것은 다항로짓모형을 이용하여 단계적으로 선택을 추정하는 방법이다. 예를 들면 교통수단으로 자가용과 버스와 지하철 중의 하나를 선택하는 문제에서 이 대안 세가지가 완전히 독립적이지 않다고 믿을만한 이유가 있다면 1단계에서 버스와 지하철만을 대상으로 추정하고 2단계에서는 이 둘을 대중교통으로 묶어서 자가용과의 선택확률을 추정하는 것이다(Sobel, 1980).

로짓모형은 비선형이므로 최소자승법에 의한 선형모형 추정으로는 유효한 효용계수를 추정할 수

없으므로 최우추정법(maximum likelihood estimation)으로 추정한다. 최우추정법이란 모수를 추정하기 위하여 관찰된 확률표본이 있다 할 때 이들 확률표본이 얻어질 수 있는 가능성이 가장 높은 모수의 값 θ 를 찾아 이를 모수의 추정량으로 삼는 방법이다(윤기중, 1984). 모수추정을 위한 연립방정식을 비선형으로 풀기 위하여서는 뉴턴랩스 방법 등을 이용하여 반복연산(iterative algorithms)하여 추정한다. 최우추정법의 가장 큰 결점은 수치분석 방법에 의존하기 때문에 계산비용이 많이 요구되는 점이다. 그러나 최우추정법에 의한 추정결과는 일관성이 있고, 표본크기에 따라 가능한 최선의 추정치를 주는 장점이 있다.

로짓모형의 적합도검정은 우도비검정(likelihood ratio test)을 하는데 우도비 검정통계량은 다음과 같이 정의된다.

$$C = -2\log\left(\frac{L(\beta)}{L(0)}\right) = -2(\log L(\beta) - \log L(0)) \dots\dots\dots (\text{식 3.3})$$

여기서 $L(0)$ 는 절편을 제외한 모든 계수가 0일 때의 우도함수의 최대값이고, $L(\beta)$ 는 추정된 모형의 우도함수 값이다. 이때 우도비검정통계량은 절편을 제외한 모든 계수가 0이라는 귀무가설이 사실일 때 자유도가 K (귀무가설에서 0으로 주어지는 계수의 갯수)인 χ^2 분포를 한다는 것이 알려져 있다. 이렇게 하여 계산된 C 통계량이 $\chi^2(K, \alpha)$ 보다 크면 추정된 모형은 유의수준 α 에서 귀무가설을 기각하게 되므로 모형의 유의성이 인정된다.

또 ρ^2 도모형의 적합도검정에 이용된다.

$$\rho^2 = 1 - \frac{L(\beta)}{L(0)} \quad ; \quad 0 < \rho^2 < 1 \dots\dots\dots (\text{식 3.4})$$

ρ^2 은 회귀모형의 적합도검정에서 이용되는 결정 계수 R^2 와 거의 유사하나 R^2 는 적합도가 좋은 경우 1에 가까우나 ρ^2 의 경우는 0.2~0.3 정도이어서 우수한 모형이라 본다(Demenchich & McFadden 1975).

효용계수의 유의성을 검증하는 방법은 일반적인 선형모형의 경우와 같이 t -검정을 한다.

이상의 추정과정과 적합도검정을 통하여 독립변수 X_i 가 범주형 종속변수 Y_i 에 미치는 영향이 통계적으로 유의적인가를 쉽게 분석할 수 있다. 그

러나 X_i 의 영향의 크기를 해석하는 것은 회귀식보다 복잡하다. 선형회귀식에서 X_i 한 단위 변화의 영향은 계수의 기울기와 동일하지만 비선형인 경우 X_i 의 한 단위의 변화가 Y_i 에 미치는 영향은 X 값이 어디에 있느냐에 따라 달라진다(그림 3.1 참조). 따라서 독립변수의 상대적 영향만(Adrich & Nelson, 1984)을 일반적으로 평가할 수 있다.

한편 종속변수의 응답범주가 둘인 경우(식 3.5)처럼 다항로짓모형을 전환하여 로그를 취하면 Y_i 의 범주가 0~1로 제한되는 문제가 제거되어 보다 간편하게 이용할 수 있다.

$$P_i = \frac{1}{1+e^{-V_i}} = \frac{1}{1+e^{-(a+bX_i)}}$$

따라서 $P_i(1+e^{a+bX_i})=1$,

$$e^{a+bX_i} = \frac{P_i}{1-P_i} \text{ 이 된다.}$$

양변에 로그를 취하면

$$\ln \frac{P_i}{1-P_i} = a+bX_i \dots\dots\dots (\text{식 3.5})$$

가 되어 X 에 대한 함수식으로 바뀐다. 여기서 종속변수는 첫째 대안을 선택하는 확률과 둘째 대안을 선택할 확률비의 로그값이 된다. 따라서 P_i 의 값은 0과 1의 사이이나 $P_i/(1-P_i)$ 는 $0 \sim \infty$ 가 되며, $P_i=0$ 일 때 로그확률비(log odds)는 $-\infty$, $P_i=0.5$ 일 때 로그확률비는 0, $P_i=1$ 일 때 로그확률비는 $+\infty$ 가 된다. 이와 같이 하여 로짓모형은 (0,1) 구간내에서 확률 예측문제가 전 실수 구간에서의 비율예측문제로 전환되며 변형된 종속변수는 X 의 선형함수라고 가정할 수 있다.

로그확률비에 대한 방정식은

$$\ln(\text{odds}) = a + b_1X_1 + b_2X_2 + \dots\dots + b_pX_p \dots\dots\dots (\text{식 3.6})$$

로 다항선형회귀방정식의 형태이다. 이러한 이유로 로지스틱 함수를 로지스틱 회귀식(logistic regression)이라 한다. 로지스틱 회귀분석에서 기본적인 가정은 로그확률비(log odds)가 독립변수와 선형 관계라는 것이며, 독립변수 X 에 대해서는 어떤 가정도 없다. 따라서 독립변수 X 는 불연속변수이거나 연속변수이거나 상관성이 없으며 방정식의 계수는 회귀계수처럼 해석될 수 있다(Afifi & Clark, 1984).

로짓모형은 확률모형이기 때문에 독립변수와 종

속변수의 관계를 설명할 수 있을 뿐만 아니라 대안 i의 선택을 설명하는 한 독립변수가 1% 변화함에 따라서 해당대안의 선택확률이 얼마나 변화하는가를 나타내는 탄력성(elasticity)을 이용하여 대안의 특성이나 가구의 특성 변화에 따른 해당 대안의 선택확률을 구할 수 있어서 실제적인 예측에 이용될 수 있다.

로짓모형을 추정하기 위한 컴퓨터 프로그램은 널리 일반화되어 있지는 않으나 SAS에 범주형 데이터 분석(허명희, 1989)을 위해 개발되어 있으며 그 밖에 QUALL, CHOMP, LIMDEP이 이용되며 특별한 경우 개인이 개발한 프로그램을 이용하기도 한다.

IV. 로짓모형의 적용

그동안 외국에서 로짓모형이 적용되었던 연구는 교통수요분석(Watson, 1974; Demenish & Mcfadden, 1975), 주택수요분석(Friedman, 1975; Lerman, 1975; Boehm, 1982; Shin, 1985 등), 주거이동분석(Cho & Morris, 1986; Cho, 1987; Kim, 1987 등), 대학 및 전공선택(Miller & Radner, 1975; Kim, 1985 등), 직업선택(Boskin, 1974) 등 경제학 관련 분야에서 주로 이루어졌다.

우리나라에서 로짓모형이 적용된 것은 1970년대 말로 특성선택이론(유완, 1979)이 소개된 후 1980년대 초에 주로 교통수단의 선택에 영향을 미치는 변인을 분석하는데 적용되었으며(김형철, 1981; 박천옥, 1981; 조광호, 1982; 조정제, 음성직, 김설주, 1983; 김대훈, 1987 등) 그 이후 경제학과 도시계획 분야의 학자들에 의해 주택수요(김중수, 1984; 국토개발연구원, 1986; 한동근, 1987; 대한주택공사, 1988)의 결정요인 분석에도 이용되었다.

가정학 분야에서는 소비경제학 분야와 주거학 분야에서 로짓모형이 적용된 바 있다. 김기욱(Kim, 1985)은 1960년대 말에서 1970년대 말까지 여자 대학생과 남자 대학생들의 전공선택에 관한 변화 추이 및 성(sex) 차이를 분석하는데 로짓모형을 적용하였다. 이 연구에서 종속변수는 명목변수인

새가지 전공분야, (1) 여학생이 60% 이상되는 여성영역분야(female dominated fields), (2) 여학생이 35~60%를 차지하는 중간영역분야(balanced fields), (3) 여학생이 35% 미만인 남성영역분야(male dominated fields)이다. 전공분야를 선택하는데 영향을 미칠 것으로 기대되는 독립변수는 35세가 되었을 때 무엇을 할 것인가, 결혼예정, 출산예정 자녀수와 같은 직업과 가정에 대한 포부에 관계되는 변수와 어머니의 교육, 소득, 인종, 종교적 행사 참여빈도, 주부의 취업과 같은 개인적인 특성변수가 사용되었다. 종속변수와 독립변수의 관계는 (1) 전공분야를 선택한 여학생의 비율(percent)은 최소자승법으로 계수를 추정하였으며, (2) 명목형(nominal type) 로짓모형으로는 세 전공분야 선택에 영향을 미치는 변수들을 추정하였고, (3) 전공분야의 전통성의 정도에 따른 선택의 경향은 서열형(ordinal type) 로짓모형으로 추정하였다.

조계순과 Morris(Cho & Morris, 1986)는 주택소유 및 주택유형의 선호와, 실제 선택행동과 이에 영향을 미치는 사회인구학적 변수와의 관계를 규명하는데 로지스틱 회귀모형을 적용하였다. 종속변수인 주택소유형태는 소유와 임차 둘로 나누었으며, 주택유형 역시 단독주택과 집합주택이라는 두가지 명목적 대안으로 나누었다. 직업의 종류, 고용상태, 가장의 유형, 결혼상태, 가장의 연령, 가장의 교육수준, 가구원수, 소득 등의 독립변수가 종속변수에 미치는 영향을 개별적으로 로지스틱단순회귀분석하여 각 독립변수의 범주에 따라 주택에 대한 선호가 유사하거나 차이가 있음을 보여 주었다. 뒤이어 가구의 생활사자료(event history data)를 이용하여 처음으로 소유주가 될 확률을 결정하는 요인을 분석한 연구(Cho, 1987)에서는 전의 이변량분석에서 더 나아가 다변량분석을 통하여 독립변수들(교육기간, 결혼년도, 결혼당시연령, 연령, 가족수, 자녀수, 거주기간등)이 종속변수(일정 기간 안에 소유주가 됨: 1, 소유주가 되지못함: 0)에 미치는 영향을 분석하였다.

이길순(1990)은 주택의 소유형태, 주택유형, 주택규모라는 주택의 특성을 선택하는 행동을 분석하는데 로짓모형을 적용하였다. 주택이 이질계의

결합체라는 관점에서 주택의 소유형태와 주택유형이 결합된 선택의 분석에서는 6개의 대안(소유·단독주택, 소유·연립주택, 소유·아파트, 임차·단독주택, 임차·연립주택, 임차·아파트)를 종속변수로 하여 각 대안을 선택할 확률에 미치는 독립변수의 영향을 분석하였다. 이때 사용된 독립변수는 가구의 사회인구학적 변수인 소득, 자산, 가구원수, 가장의 연령, 가장의 교육수준과 주택가격, 방수, 주택규모, 이전소유형태 등의 주택관련변수이었다.

이상의 로짓모형을 적용한 연구들이 대부분 소비자선택이론에 근거한 것은 로짓모형의 논리가 선택행위에 기초를 두고 발전했기 때문이다. 그러나 이 이론적 근거가 로짓모형을 적용하는데 필수적인 조건이 되는것은 아니므로 흥미, 태도, 선호와 같은 인간의 행동을 분석하는 사회과학분야는 물론 자극과 반응의 관계를 분석하는 자연과학분야에서도 계수를 추정하는 통계적 방법으로 이용할 수 있다.

V. 요 약

본 논문은 질적 종속변수를 회귀분석하였을 때 발생하는 문제를 제시하고 이 문제를 해결할 수 있는 대체적 분석기법인 로짓모형을 제안하는데 목적을 두었다. 회귀분석이 제공하는 여러가지 장점때문에 종속변수가 연속적이어서 하는 가정을 무시하고 불연속적인 종속변수를 회귀분석했을 때는 비선형을 선형으로 추정하는 체계적 오차 뿐만 아니라 회귀식에 포함된 확률항에 대한 가정이 맞지않아 가설검정이나 신뢰구간과 같은 기본적 통계량들이 유의성을 잃게된다.

종속변수가 질적인 변수는 선택행동이나 사건의 발생여부에 독립변수들이 어떻게 영향을 미치는 것인가를 분석하는 것이므로 확률모형인 로짓모형을 적용할 수 있다. 로짓모형은 질적종속변수에 미치는 독립변수의 영향을 설명할 수 있을 뿐만 아니라 독립변수들의 변화에 따른 종속변수의 확률을 추정할 수 있어 예측을 위한 연구에도 유용하다. 그러나 비선형모형인 로짓모형에서는 추정된 독립변수의 계수가 선형모형에서의 계수처럼 일정

한 기울기를 의미하지 않으며 독립변수의 크기에 따라 달라지므로 해석에 주의해야 한다. 로짓모형에 로그를 취하여 전환시킨 로지스틱회귀식은 종속변수를 로그확률비로 전환하여 회귀분석이 갖는 문제들을 해결하면서 방정식의 계수를 회귀계수처럼 해석할 수 있어서 사용이 용이하다.

이러한 로짓모형의 특성은 인간과 근접환경의 상호작용을 연구하기 위해 종종 질적 변수를 사용하는 가정학자들에게 연구의 정확성과 깊이를 더하는데 기여할 것이다.

참고문헌

- 1) 국토개발연구원(1986). 서울주택시장분석. 서울 : 국토개발연구원.
- 2) 김대훈(1987). 개별수단선택모형의 예측력 비교. 석사학위논문, 연세대학교 대학원.
- 3) 김중수(1984). 주택소유형태에 관한 계량분석. 한국개발연구, 여름호.
- 4) 김현식(1987). Housing Tenure and Residential Mobility : A Multi-Choice Logit Approach. 국토연구, VII.
- 5) 김형철(1981). 통행시간의 가치 산정. 석사학위논문, 연세대학교 대학원.
- 6) 대한주택공사(1988). 임대주택의 지역별 배분모형연구. 서울 : 대한주택공사.
- 7) 박천옥(1981). 선택대상의 상대적인 효과가 선택행위에 미치는 영향. 석사학위논문, 연세대학교 대학원.
- 8) 여영종(1984). 다단계 주거선택 모형. 주택금융, 17(6).
- 9) 유 완(1979). 특성선택이론. 국토계획학회지 14 (1).
- 10) 유 완(1982). 선택행위의 특성모델. 국토계획학회지 17(2).
- 11) 유 완·음성직·김철주(1983). 선택이론을 이용한 도시활동 분석기법개발. 연세대학교 건축공학과.
- 12) 유지성(1989). 계량 경제학원론. 서울 : 박영사.
- 13) 윤기중(1987). 통계학. 서울 : 법문사.
- 14) 이길순(1991). 로짓모형을 이용한 주택선택 결정요인의 분석. 박사학위논문, 연세대학교 대학원.

- 15) 이영준(1991). 다변량분석. 서울 : 석정.
- 16) 장상희 · 홍동식(1985). 사회통계학. 서울 : 박영사.
- 17) 조광호(1982). "교통수단 선택모형의 비교분석. 석사학위논문. 연세대학교 대학원.
- 18) 조정제 · 음성직 · 김설주(1983). 지역간 교통수단의 선택행위분석. *국토계획 학회지* 18(1).
- 19) 한동근(1987). 가구특성에 따른 주거선택에 관한 연구. 석사학위논문. 서울대학교 환경대학원.
- 20) 허명희(1989). SAS 범주형데이터분석. 서울 : 자유아카데미.
- 21) 홍두승(1987). 사회조사분석. 서울 : 다산출판사.
- 22) Affi. A.A. & Clark Virginia.(1984). Computer aided multivariate analysis. Belmont : Lifetime Learning Publication.
- 23) Aldrich, John. H. & Nelson, Forrest. D.(1985). Linear probability, logit, and probit models. Beverly Hills : Sage Publication, Inc.
- 24) Boehm, Thomas. P.(1982). A hierarchical model of housing choice. *Urban Studies*, 19.
- 25) Boskin, M.(1974). A conditional logit model of occupational choice. *Journal of Political Economy*, 82.
- 26) Cho Jaesoon. & Morris. E.(1986). Logit models for housing preferences, demographic variable and actual housing conditions. *Housing and Society* 13(2).
- 27) Cho Jaesoon.(1987). Becoming a first-time home owner. Ph.D. Dissertation. Iowa State University.
- 28) Derrick, Frederik, W.(1979). Notes on the use of qualitative variables in regression. *Home Economics Research Journal* 7(4).
- 29) Domenish, T.A. & McFadden. D.(1975). Urban travel demand. Amsterdam : North Holland Publishing Company.
- 30) Flenberg, Stephen, E.(1980). The analysis of cross-classified categorical data. Cambridge : The MIT Press.
- 31) Friedman, Joseph.(1975). Housing location and the supply of local public services. Ph.D. Dissertation. University of California, Berkeley.
- 32) Hanushek, E.A. & Jackson, J.E.(1977). Statistical methods for social scientists. New York : Academic Press.
- 33) Kenneth, Train.(1986). Qualitative choice analysis : The theory and an application to automobile demand. Cambridge : The MIT Press.
- 34) Kim Jeong-Ho.(1987). Residential mobility and housing consumption adjustment. Ph.D. Dissertation. University of Michigan.
- 35) Kim Kee-Ok.(1985). Determinants of choice of college major among woman and man. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.
- 36) Lerman, S.(1975). A Disaggregate behavior model of urban mobility decision. Ph.D. Dissertation. M.I.T.
- 37) Luce, R. Duncan.(1959). Individual choice behavior. New York : John Wiley & Sons.
- 38) McFadden, Daniel.(1973). Conditional logit analysis of qualitative choice behavior. Berkeley : Institute of Urban and Regional Development.
- 39) Miller, L. & Radner, R.(1975). Demand and supply in U.S higher education. *American Economic Review*, 60.
- 40) Shin Jung-Chuel.(1985). Residential location choice. Ph.D. Dissertation. Stanford University.
- 41) Sobel, Kenneth. L.(1980). Travel demand forecasting by using the nested multinomial logit model. *Transportation Research Record*, 775.
- 42) Wann Yu.(1978). Contributions to applications of linear logit model in transportation research. Ph.D. Dissertation. University of Pennsylvania.
- 43) Watson, P.L.(1974). Homogeneity of models of transport mode choice. *Journal of Regional Science* 14(2).