

로짓모형을 이용한 통신 서비스품질 평가방법

조재균^{1*} · 안혜숙²

¹동의대학교 경영정보 · 인터넷비즈니스학부 / ²삼성 SDS ERP 사업팀

Evaluation Method of Quality of Service in Telecommunications Using Logit Model

Jae-Gyeun Cho¹ · Hae-Sook Ahn²

¹Division of Management Information & Internet Business, Dong Eui University, Busan, 614-714

²ERP Consulting Team, Samsung SDS, Seoul, 135-082

Quality of Service(QoS) in the telecommunications can be evaluated by analyzing the opinion data which result from the surveyed opinions of respondents and quantify subjective satisfaction on the QoS from the customers' viewpoints. For analyzing the opinion data, MOS(mean opinion score) method and Cumulative Probability Curve method are often used. The methods are based on the scoring method, and therefore, have the intrinsic deficiency due to the assignment of arbitrary scores. In this paper, we propose an analysis method of the opinion data using logit models which can be used to analyze the ordinal categorical data without assigning arbitrary scores to customers' opinion, and develop an analysis procedure considering the usage of procedures provided by SAS(Statistical Analysis System) statistical package. By the proposed method, we can estimate the relationship between customer satisfaction and network performance parameters, and provide guidelines for network planning. In addition, the proposed method is compared with Cumulative Probability Curve method with respect to prediction errors.

Keywords: quality of service, logit model, categorical data analysis

1. 서론

통신에 있어서 서비스품질은 설문조사를 통해 수집된 <표 1>에서 보여지는 것과 같은 서비스 이용자의 주관적 만족도에 관한 데이터(오피니언 데이터(opinion data)라 부름)를 분석함으로써 평가된다. 그리고 오피니언 데이터의 분석을 통하여 서비스 이용자의 만족도와 서비스품질에 영향을 미치는 통신망의 측정 가능한 물리적 성능변수들(예를 들면, 손실, 잡음, 반향, 측음 등)간의 관계를 규명함으로써, 서비스 이용자가 기대하는 수준의 서비스 품질을 제공하기 위한 전략을 수립할 수

있게 된다.

오피니언 데이터의 분석 방법으로는 평균 오피니언 점수법(mean opinion score method, MOS method)과 누적퍼센트곡선 방법(Cumulative Probability Curve method)이 일반적으로 사용된다. 평균 오피니언 점수법은 설문조사 응답자의 주관적인 평가(예를 들면, '매우 나쁘다', '나쁘다', '보통이다', '좋다', '매우 좋다')를 점수 척도를 사용하여 계량화한 후, 평균 오피니언 점수(mean opinion score, MOS)라고 불리는 산술평균을 구함으로써 서비스 이용자의 만족도를 평가하는 방법이다(이영환 외, 1996). 누적퍼센트곡선 방법은 서비스 이용자의 잠재적인 평가

이 연구는 2001년도 동의대학교 학술연구비 지원에 의하여 이루어졌음.

* 연락처: 조재균 교수, 614-714 부산시 부산진구 가야동 산24 동의대학교 경영정보 · 인터넷비즈니스학부, Fax : 051-890-1409, e-mail : jgcho@dongeui.ac.kr

2001년 10월 접수, 1회 수정 후 2002년 3월 게재 확정.

표 1. 설문조사로부터 얻어진 오피니언 데이터의 형태

결합 조건	각 범주의 도수					합계
	매우 나쁘다	나쁘다	보통이다	좋다	매우 좋다	
1	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	$n_{1\cdot}$
2	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}	$n_{2\cdot}$
⋮						
G	n_{G1}	n_{G2}	n_{G3}	n_{G4}	n_{G5}	$n_{G\cdot}$

n_{gi} 는 g 번째 결합조건에서 i 번째 범주에 응답한 응답자수($g=1, \dots, G$; $i=1, \dots, 5$)

행태가 심리척도 상에서 정규분포를 따른다는 가정을 바탕으로, 설문조사 결과 얻어진 점수에 대한 히스토그램을 정규분포에 근사시키고, 데이터의 평균과 표준편차를 조정된 후 해당 서비스 이용자의 만족도 분포를 구하는 방법이다(Cavanaugh et al., 1976). 누적퍼센트곡선 방법에 의해 구해진 만족도 분포는 서비스 이용자가 기대하는 품질 수준의 서비스를 제공하기 위한 전략을 수립함에 있어 효과적으로 활용될 수 있다는 점에서, 누적퍼센트곡선 방법이 평균 오피니언 점수법에 비해 향상된 분석 방법이라 할 수 있다.

그러나, 누적퍼센트곡선 방법도 평균 오피니언 점수법과 마찬가지로 범주들 간에 순서가 존재하는 순서형(ordinal)의 범주형 데이터를 분석하는 가장 일반적인 방법인 점수법(scoring method)을 응용한 것으로, 점수법이 갖는 주관적인 점수 할당으로 인한 근본적인 문제점을 갖는다. 다시 말하면, 누적퍼센트곡선 방법은 설문문항의 각 범주에 주관적으로 점수를 부여함으로써 누적퍼센트곡선을 유도한다. 따라서 각 범주에 부여되는 점수척도가 달라지면 서비스 이용자의 만족도와 성능변수들간의 관계를 추정하기 위한 회귀식의 계수가 변하기 때문에, 성능변수들의 결합조건에 대해 만족도 분포를 추정할 결과가 달라지게 되고, 그로 인하여 이용자 만족도에 대한 회사의 동일한 전략을 달성하기 위해 설정된 성능변수들의 목표치에 대한 일관성이 보장되지 않는다(안혜숙 외, 1999).

본 논문에서는 누적퍼센트곡선 방법의 대안으로서, 설문조사에서 얻은 순서 범주형 자료인 오피니언 데이터를 주관적인 점수를 부여하지 않고 분석할 수 있는 로짓모형(logit model)을 이용한 분석 방법을 제안한다. 로짓모형을 이용한 분석에서는 각 범주에 점수를 부여할 필요가 없기 때문에, 누적퍼센트곡선 방법에서와 같은 주관적인 점수 할당으로 인한 문제점을 극복할 수 있다. 2절에서는 범주형 자료분석을 위해 사용되는 통계적 모형들과 다범주 로짓모형들을 살펴보고, 3절에서는 로짓모형의 분석을 위해 사용 가능한 SAS(Statistical Analysis System) 통계패키지의 절차(procedure)들을 소개한다. 4절에서는 로짓모형을 이용한 오피니언 데이터의 분석 절차를 제안하고 분석 결과의 활용 방법을 설명한다. 5절에서는 제안된 분석 방법을 예증하고, 마지막으로 6절에서는 본 연구에 대한 결론을 맺는다.

2. 로짓모형

2.1 범주형 자료분석을 위한 통계적 모형

범주형 변수는 측정 척도가 여러 범주(category)들의 집합으로 구성되어 있는 변수로 정의된다(Agresti, 1996). 범주형 변수는 2개의 범주를 갖는 변수(dichotomous variable)와 3개 이상의 범주를 갖는 변수(polytomous variable)로 구분되며, 3개 이상의 범주를 갖는 변수는 범주들간에 순서가 존재하지 않는 명목형(nominal) 변수와 범주들간에 순서는 존재하고 범주들간의 절대적인 간격은 분명하지 않은 순서형(ordinal) 변수로 세분된다. 그리고 정수형 변수 중 제한된 개수의 정수값을 갖는 정수형 변수와 연속형 변수 중 작은 개수의 범주들로 그룹핑 되어지는 연속형 변수도 범주형 변수에 포함된다(Agresti, 1990; Stokes, et al., 2000).

종속변수인 반응변수(response variable)가 범주형 변수인 자료를 범주형 자료라 한다. 이때 독립변수인 설명변수는 연속형, 범주형, 또는 연속형과 범주형을 다 포함하는 혼합형일 수 있다. 범주형 자료의 분석을 위해 사용되는 통계적 모형들은 <그림 1>에 나타난 것처럼 분류될 수 있다.

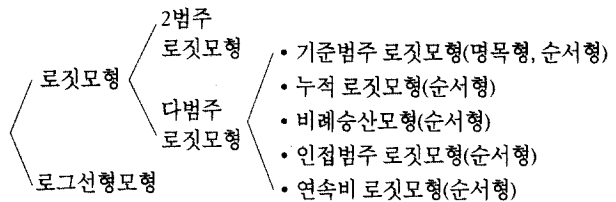


그림 1. 범주형 자료분석을 위한 통계적 모형의 분류.

범주형 자료분석을 위한 대표적인 통계모형들은 로짓모형과 로그선형모형(log-linear model)이다. 로지스틱 회귀모형(logistic regression model)이라고도 불리는 로짓모형은 반응변수가 이항 분포 또는 다항분포를 따른다고 가정할 때 적용할 수 있으며, 로그선형모형은 반응변수가 포아송분포를 따른다고 가정할 때 적용할 수 있는 모형이다. 로짓모형은 반응변수의 범주 수가 2개일 때의 로짓모형인 2범주 로짓모형과 이것을 범주 수가 3개 이상인 경우로 확장했을 때의 로짓모형인 다범주 로짓모형(multicategory logit model)으로 구분될 수 있다. 다범주 로짓모형에는, 명목형의 반응변수를 위한 로짓모형인 기준범주 로짓모형(baseline-category logit model) 또는 일반화 로짓모형(generalized logit model)과 순서형의 반응변수를 위한 모형들인 누적 로짓모형(cumulative logit model), 비례승산모형(proportional odds model), 인접범주 로짓모형(adjacent-category logit model), 연속비 로짓모형(continuation-ratio logit model) 등이 포함된다(Agresti, 1996).

2.2 다범주 로짓모형

본 절에서는 2.1절에서 언급된 다범주 로짓모형들 중 범주형

자료분석을 위해 빈번하게 사용되는 일반화 로짓모형, 누적 로짓모형, 그리고 비례승산모형에 대하여 살펴본다(Agresti, 1990; 변은신, 1996). 다범주 로짓모형의 논의를 위해 사용되는 기호들에 대한 설명은 다음과 같다.

- C : 범주의 수
- G : 성능변수들의 결합조건 수
- p_{gi} : g 번째 결합조건에서 i 번째 범주에 응답한 응답자의 비율($i=1, \dots, C$)
- x_{gi} : g 번째 결합조건에서 성능변수들의 실측값($j=1, \dots, p-1$)
- π_{gi} : g 번째 결합조건에서 i 번째 범주의 반응확률(response probability).
- p_{gi} 의 참값
- $L_r(\boldsymbol{\pi}_g)$: g 번째 결합조건에 대한 로짓($r=1, \dots, C-1$; $\boldsymbol{\pi}_g = (\pi_{g1}, \pi_{g2}, \dots, \pi_{gC})'$)

2.2.1 일반화 로짓모형

명목형 반응변수를 위한 로짓모형으로, 주로 명목형 반응변수에 대해 사용되나 순서형 반응변수에 대해서도 사용될 수 있다. 임의의 한 범주(일반적으로는 마지막 범주)를 기준범주로 하고, 이 기준범주와 나머지 각 범주를 짝을 지어 로짓을 구성한 것을 일반화 로짓이라 한다. 일반화 로짓을 이용한 일반화 로짓모형은 식 (1)과 같다.

$$L_1(\boldsymbol{\pi}_g) \equiv \ln\left(\frac{\pi_{g1}}{\pi_{gC}}\right) = \beta_{01} + \beta_{11}x_{g1} + \beta_{21}x_{g2} + \dots + \beta_{p-1,1}x_{g,p-1} + \epsilon_{g1}$$

$$L_2(\boldsymbol{\pi}_g) \equiv \ln\left(\frac{\pi_{g2}}{\pi_{gC}}\right) = \beta_{02} + \beta_{12}x_{g1} + \beta_{22}x_{g2} + \dots + \beta_{p-1,2}x_{g,p-1} + \epsilon_{g2}$$

$$\vdots$$

$$L_{C-1}(\boldsymbol{\pi}_g) \equiv \ln\left(\frac{\pi_{g,C-1}}{\pi_{gC}}\right) = \beta_{0,C-1} + \beta_{1,C-1}x_{g1} + \beta_{2,C-1}x_{g2} + \dots + \beta_{p-1,C-1}x_{g,p-1} + \epsilon_{g,C-1} \quad (1)$$

식 (1)로부터 결합조건 g 에서 범주 i 에 대한 확률은 다음의 식 (2)를 사용하여 계산되어질 수 있다.

$$\pi_{gi} = \frac{\exp\left(\sum_{j=0}^{p-1} \beta_{ij}x_{gj}\right)}{1 + \sum_{r=1}^{C-1} \exp\left(\sum_{j=0}^{p-1} \beta_{jr}x_{gj}\right)}, \quad i=1, \dots, C-1; g=1, \dots, G \quad (2)$$

2.2.2 누적 로짓모형

순서형 반응변수를 위한 로짓모형으로, 반응범주들이 순서형인 경우에 순서를 고려한 로짓모형을 사용하게 되면, 해석하기가 더 간단하고 보통의 다범주 로짓모형 보다 더 좋은 검

정력을 갖는다. 반응변수 Y 가 g 번째 결합조건에서 r 보다 큰 범주에 속하게 될 누적확률이 $\Pr(Y > r) = \pi_{g,r+1} + \dots + \pi_{gC}$ ($r=1, \dots, C$)일 때, 처음 $C-1$ 개의 누적확률들에 대한 로짓을 누적 로짓이라 하며 다음과 같다.

$$\text{logit} [\Pr(Y > r)] = \ln\left(\frac{\Pr(Y > r)}{1 - \Pr(Y > r)}\right) = \ln\left(\frac{1 - F_{gr}}{F_{gr}}\right) = \ln\left(\frac{\pi_{g,r+1} + \dots + \pi_{gC}}{\pi_{g1} + \dots + \pi_{gr}}\right), \quad r=1, \dots, C-1$$

단, $F_{gr} = \pi_{g1} + \dots + \pi_{gr}$. 누적 로짓을 이용하는 누적 로짓모형은 식 (3)과 같다.

$$L_1(\boldsymbol{\pi}_g) \equiv \ln\left(\frac{1 - F_{g1}}{F_{g1}}\right) = \ln\left(\frac{\pi_{g2} + \dots + \pi_{gC}}{\pi_{g1}}\right) = \beta_{01} + \beta_{11}x_{g1} + \beta_{21}x_{g2} + \dots + \beta_{p-1,1}x_{g,p-1} + \epsilon_{g1}$$

$$L_2(\boldsymbol{\pi}_g) \equiv \ln\left(\frac{1 - F_{g2}}{F_{g2}}\right) = \ln\left(\frac{\pi_{g3} + \dots + \pi_{gC}}{\pi_{g1} + \pi_{g2}}\right) = \beta_{02} + \beta_{12}x_{g1} + \beta_{22}x_{g2} + \dots + \beta_{p-1,2}x_{g,p-1} + \epsilon_{g2}$$

$$\vdots$$

$$L_{C-1}(\boldsymbol{\pi}_g) \equiv \ln\left(\frac{1 - F_{g,C-1}}{F_{g,C-1}}\right) = \ln\left(\frac{\pi_{gC}}{\pi_{g1} + \dots + \pi_{g,C-1}}\right) = \beta_{0,C-1} + \beta_{1,C-1}x_{g1} + \beta_{2,C-1}x_{g2} + \dots + \beta_{p-1,C-1}x_{g,p-1} + \epsilon_{g,C-1} \quad (3)$$

2.2.3 비례승산모형

순서형 반응변수를 위한 로짓모형으로, 식 (3)에 있는 누적 로짓모형의 반응함수 $\ln[(1 - F_{gr})/F_{gr}]$ 에 음수를 취한 것과 동일하며, $\beta_{0r} = -\alpha_r$ ($r=1, \dots, C-1$), $\beta_{1j} = \dots = \beta_{j,C-1} = -\beta_j$ ($j=1, \dots, p-1$)인 누적 로짓모형의 특수한 경우이다. 비례승산모형은 식 (4)와 같다.

$$L_1(\boldsymbol{\pi}_g) \equiv \ln\left(\frac{F_{g1}}{1 - F_{g1}}\right) = \ln\left(\frac{\pi_{g1}}{\pi_{g2} + \dots + \pi_{gC}}\right) = \alpha_1 + \beta_1x_{g1} + \beta_2x_{g2} + \dots + \beta_{p-1}x_{g,p-1} + \epsilon_{g1}$$

$$L_2(\boldsymbol{\pi}_g) \equiv \ln\left(\frac{F_{g2}}{1 - F_{g2}}\right) = \ln\left(\frac{\pi_{g1} + \pi_{g2}}{\pi_{g3} + \dots + \pi_{gC}}\right) = \alpha_2 + \beta_1x_{g1} + \beta_2x_{g2} + \dots + \beta_{p-1}x_{g,p-1} + \epsilon_{g2}$$

$$\vdots$$

$$L_{C-1}(\boldsymbol{\pi}_g) \equiv \ln\left(\frac{F_{g,C-1}}{1 - F_{g,C-1}}\right) = \ln\left(\frac{\pi_{g1} + \dots + \pi_{g,C-1}}{\pi_{gC}}\right) = \alpha_{C-1} + \beta_1x_{g1} + \beta_2x_{g2} + \dots + \beta_{p-1}x_{g,p-1} + \epsilon_{g,C-1} \quad (4)$$

식 (4)에서 보여지듯이 비례승산모형은 '(C-1)개의 식에서 각 독립변수 앞의 계수가 모두 같다'고 가정하고 있으며, 이것을 평행선(parallel line) 가정이라고 한다.

3. 로짓모형의 분석을 위한 SAS 절차

본 절에서는 2절에서 언급된 로짓모형들의 분석을 위해 사용될 수 있는 SAS 통계패키지의 절차들인 CATMOD 절차와 LOGISTIC 절차에 관해 살펴보고, 두 절차에서 제공되는 가설 검정 및 사용상의 차이점 등을 비교해본다(Agresti, 1990; Agresti, 1996; SAS/STAT Users Guide, 1990).

3.1 CATMOD 절차

2절의 식 (1)과 (3)에 있는 일반화 로짓모형과 누적 로짓모형의 분석에 사용될 수 있으며, 일반화 로짓모형을 사용할 경우에는 최우추정법(maximum likelihood method)과 가중최소제곱법(weighted least squares method)에 의해 모수를 추정할 수 있으며, 최우추정법에 의한 모수의 추정치는 뉴턴-라프슨(Newton-Raphson) 알고리즘을 사용하여 계산된다. 그리고 누적 로짓모형을 사용할 경우에는 단지 가중최소제곱법에 의해 모수를 추정하게 된다.

CATMOD 절차에서는 우도비 검정통계량(likelihood-ratio test statistic), 왈드 통계량(Wald statistic), 그리고 잔차제곱합(Residual Sum of Squares)과 같은 검정통계량을 이용하여 다양한 가설 검정을 수행할 수 있다.

식 (1)에 있는 일반화 로짓모형에 대해 최우추정법을 사용하여 모수를 추정할 경우에는 다음과 같은 가설 검정을 수행할 수 있다.

■ 모형 전체의 유의성 검정

우도비 검정통계량이 사용되며, 그 값이 $\chi^2_{\alpha, (G-p)(C-1)}$ 보다 작으면 유의수준 100 $\alpha\%$ 에서 상정한 모형이 타당하다고 판단한다.

■ 모형에 있는 각 독립변수의 유의성 검정

왈드 통계량이 사용되며, 그 값이 $\chi^2_{\alpha, (C-1)}$ 보다 작으면 (C-1) 개의 식 모두에서 해당 독립변수를 제거할 수 있다.

그리고, 식 (1)에 있는 일반화 로짓모형과 식 (3)에 있는 누적 로짓모형에 대해 가중최소제곱법을 사용하여 모수를 추정할 경우에는 다음과 같은 가설 검정을 수행할 수 있다.

■ 모형 전체의 유의성 검정

잔차제곱합이 사용되며, 그 값이 $\chi^2_{\alpha, (G-p)(C-1)}$ 보다 작으면, 유의수준 100 $\alpha\%$ 에서 상정한 모형이 타당하다고 판단한다.

■ 모형에 있는 각 독립변수의 유의성 검정

잔차제곱합이 사용되며, 현재의 모형에 포함되어 있는 독립변수 각각에 대해 계산된 검정통계량값이 $\chi^2_{\alpha, (C-1)}$ 보다 크면, 해당 독립변수는 유의수준 100 $\alpha\%$ 에서 유의하다고 판단한다. CATMOD 절차는 변수선택 알고리즘을 제공하지 않지만, 모형에 변수를 추가하고자 할 때 이러한 가설 검정이 사용될 수 있다.

3.2 LOGISTIC 절차

반응변수가 이항 또는 '순서 척도 반응변수일 때 식(4)에 있는 비례승산모형의 분석에 사용될 수 있다. 최우추정법에 의해 모수를 추정하게 되며, 모수들의 최우추정치들을 계산하기 위하여 반복가중최소제곱(Iteratively Reweighted Least Squares, IRLS) 알고리즘과 동일한 피셔점수법(Fisher scoring method)을 사용한다. 그리고 LOGISTIC 절차는 지정된 변수를 사용하여 모형을 적합시키는 외에도, 전진선택(forward selection), 후진제거(backward elimination), 단계별회귀(stepwise regression) 등의 변수선택 방법을 제공한다.

LOGISTIC 절차에서는 우도비 검정통계량, 왈드 통계량, 그리고 점수 통계량(score statistic)과 같은 검정통계량을 이용하여 다양한 가설 검정을 수행할 수 있다. 식 (4)에 있는 비례승산모형에 대해 다음과 같은 가설 검정을 수행할 수 있다.

■ 평행선 가정의 검정

점수 통계량이 사용되며, 그 값이 $\chi^2_{\alpha, (C-2)(p-1)}$ 보다 작으면 유의수준 100 $\alpha\%$ 에서 평행선 가정이 타당하다고 판단한다.

■ 모형에 있는 독립변수 전체의 유의성 검정

우도비 검정통계량과 점수 통계량이 사용되며, 우도비 검정통계량값이 $\chi^2_{\alpha, p-1}$ 보다 크면 상정한 모형이 유의수준 100 $\alpha\%$ 에서 타당하다고 판단한다.

■ 모형에 있는 각 독립변수의 유의성 검정

왈드 통계량이 사용되며, 식 (4)에 있는 모형의 각각의 독립변수에 대하여 계산된 왈드 통계량값이 $\chi^2_{\alpha, 1}$ 보다 작은 독립변수들 중에서 그 값이 가장 작은 독립변수를 모형에서 제거할 수 있다. LOGISTIC 절차의 후진제거 변수선택 알고리즘은 이러한 방법을 사용하여 모형에 있는 비유의적인 독립변수를 제거하게 된다.

3.3 가중최소제곱법과 최우추정법의 비교 및 SAS 절차에서의 처리

가중최소제곱법은 다항 공분산행렬의 추정을 필요로 하며, 도수 0인 셀이 존재할 경우 공분산행렬의 추정치가 비정칙행렬이 될 수 있다. 어떤 모집단에서 반응함수들의 공분산행렬이 비정칙행렬이 되면, CATMOD 절차는 에러메시지를 출력하고 처리를 멈추게 된다. 이러한 상황을 피하기 위해, 설명변수가 한개인 경우에는 공분산행렬이 비정칙행렬이 되도록 만드는 모집단들을 제거하거나, 또는 MODEL 문에서 ADDCELL 옵션을 사용하여 모든 셀 도수에 0.5와 같은 작은 값을 더해줌 등의 방법이 사용될 수 있다. 그러나 ADDCELL 옵션을 사용하는 방법은 셀 도수가 작은 경우에는 잘못된 결과를 얻을 수도 있는 위험이 있다. 모수 추정과 관련하여, 가중최소제곱법을 사용하면, CATMOD 절차는 관찰된 반응함수들을 항상 계산하게 되고, 도수를 0의 자연대수 ln을 취하게 되면 경고메시지를

출력한 후 처리를 계속하기 위하여 작은 값(즉, 0.5/g번째 결합 조건에서의 총 응답자 수)의 ln을 취하게 된다. 이때, 셀들의 도수가 너무 작으면 모수 추정치의 결과는 유효하지 않게 된다 (Agresti, 1990; SAS / STAT Users Guide, 1990).

최우추정법은 다항 공분산행렬의 추정을 필요로 하지 않는다. 모수 추정과 관련해서는, 최우추정법을 사용하면 관찰된 반응함수들의 계산이 필요없기 때문에 모수 추정치와 모든 예측값들에 대해 유효한 결과를 얻을 수 있다. 그리고 도수 0인 셀이 존재하면 로짓모형 모수들의 최우추정치에 영향을 미칠 수 있으나, 최우추정량은 개개의 셀 도수 보다 덜 빈약한 (sparse) 주변합(marginal total)을 이용하여 구해지기 때문에 최우추정법이 가중최소추정법보다 모수 추정치에 대해 유효한 결과를 얻을 수 있다(Agresti, 1990; SAS/STAT Users Guide, 1990).

일반적으로 최우추정법에 의한 모수의 추정이 가중최소제곱법 보다 더 정확하다고 알려져 있으며, 일반화 로짓합수를 사용하고 그리고 도수 0인 셀들이 존재하면, CATMOD 절차를 사용할 때 가중최소제곱법 보다는 최우추정법을 사용해야 한

다. 그러나 각 셀의 기대도수가 큰 경우에는 최우추정법과 가중최소제곱법에 의한 결과가 비슷해진다(Agresti, 1990).

4. 로짓모형을 이용한 오피니언 데이터 분석 방법

4.1 분석 절차

3절에서의 논의를 바탕으로 순서 범주형의 오피니언 데이터를 분석하기 위한 절차는 다음의 <그림 2>와 같다.

먼저, LOGISTIC 절차를 사용하여 비례승산모형으로 분석한다. LOGISTIC 절차가 지원하는 후진제거 변수선택 방법을 수행하여 유도된 모형이 성능변수들의 주효과를 포함하지 않는다면 수동 변수선택 방법을 수행하여 주효과를 포함하도록 한다. 주효과를 포함하도록 유도된 모형이 평행선 가정을 만족한다면 그 모형을 최종 모형으로 삼고, 만족하지 않는다면 도수가 0인 셀이 존재하는지를 검사한다. 도수가 0인 셀이 존재

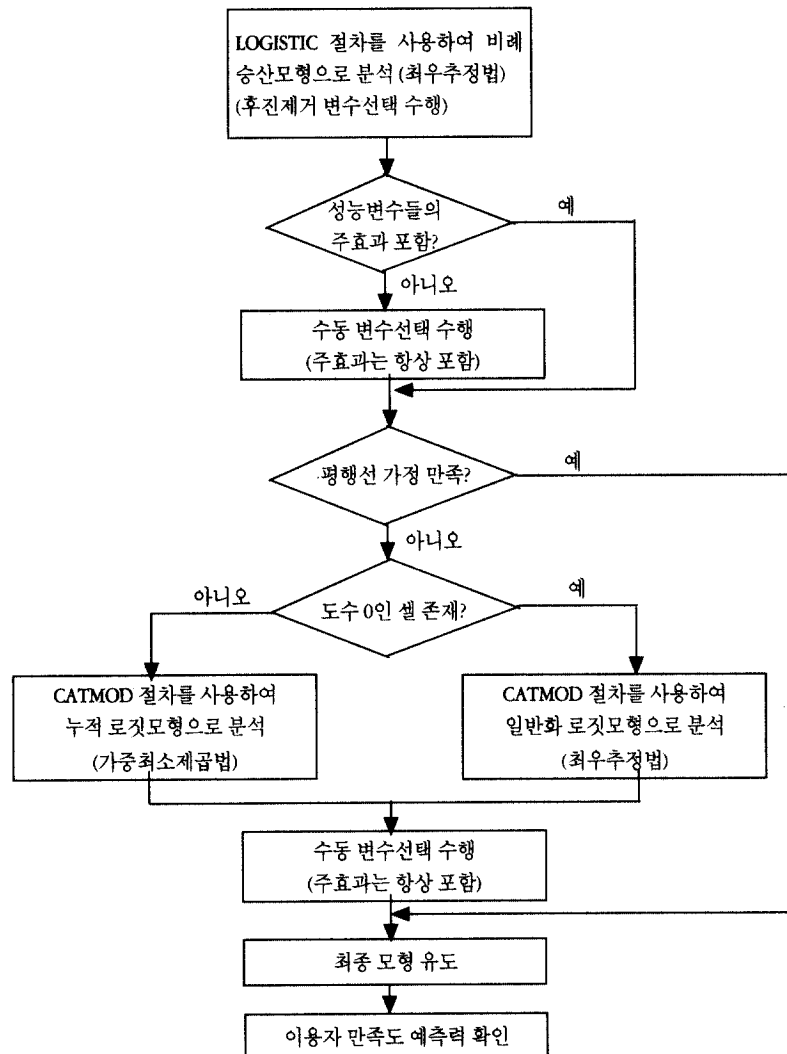


그림 2. 오피니언 데이터 분석 절차.

하지 않는다면 CATMOD 절차를 사용하여 해석이 편리하고 좋은 검정력을 가지는 누적 로짓모형으로 분석하고, 도수가 0인 셀이 존재한다면 CATMOD 절차를 사용하여 최우추정법에 의해 일반화 로짓모형으로 분석한다. 이때, CATMOD 절차는 변수선택 방법을 지원하지는 않기 때문에 수동 변수선택 방법을 수행하여 주효과를 포함하도록 한다. 마지막으로 유도된 최종 모형의 이용자 만족도에 대한 예측력을 확인한다.

4.2 분석 결과의 활용 방법

통신 서비스품질의 평가를 위해 수집된 오피니언 데이터를 4.1절에서 제시한 절차에 따라 로짓모형으로 분석함으로써 얻은 결과는 다음과 같이 이용할 수 있다.

첫째, 성능변수들과 이용자 만족도의 관계를 묘사하는 추정된 로짓모형으로부터 성능변수들의 현재 조건에서 이용자가 ‘매우 나쁘다’, ‘나쁘다’, ‘보통이다’, ‘좋다’, ‘매우 좋다’라고 평가할 확률을 예측할 수 있다. 이것으로부터 자사의 서비스 품질 수준이 이용자의 기대에 부응하고 있는지에 대한 정보를 얻을 수 있다. 만일 평가된 서비스품질에 문제가 있다면 성능변수들의 목표치를 상향 조정함으로써 서비스품질 목표치를 만족시키도록 한다.

둘째, 동일 서비스를 제공하는 경쟁사와의 만족도 비교 결과로부터 자사의 서비스품질의 수준을 파악하고, 경쟁력을 갖기 위해 도달해야 할 이용자의 만족도 수준에 대한 전략이 세워졌다고 하자. 이때 유도된 로짓모형으로부터 이용자의 주관적인 만족도와 망에서 물리적으로 측정가능한 성능변수들과의 관계를 파악할 수 있기 때문에, 회사의 전략을 고려한 적절한 성능변수들의 목표치를 설정하는데 유용하게 이용될 수 있다. 그런데 성능변수들의 조정에 요구되는 노력이나 비용이 다를 수 있으므로, 유도된 로짓모형으로부터 다양한 이용자 만족도 수준에서 등고선을 그려봄으로써 성능변수들간의 상호 관계를 파악할 수 있다. 그리고 이러한 성능변수들의 관계는 망 설계 연구를 위한 자료로 활용될 수 있다.

5. 분석 예제

Bell 연구소에서는 전화망의 전송품질에 영향을 미치는 주요 성능변수들인 회선잡음(circuit noise)과 소리크기 손실(loudness loss)이 이용자의 주관적인 만족도에 미치는 영향을 분석하기 위해 <표 2>와 같은 전송품질 시험을 실시하였다. 이 시험에서는 현장평가의 장점을 유지하면서 성능변수들의 조정을 용이하게 하기 위해, 성능변수들의 값들을 제어함으로써 통화 품질을 조절하는 장치인 SIBYL을 사용하였다. 피시험자들의 전화선은 SIBYL 장치를 거쳐 교환시스템에 연결되도록 하고, 피시험자들의 통화 중 시험자가 랜덤하게 선택한 통화에 대해서 SIBYL에 의해 부여된 시험 조건(즉, 성능변수들의 주어진 결합

표 2. Bell 연구소에서 수행된 전송품질 시험

조건	시험명	HO2 test
시험방법		SIBYL 이용
시험년도		1972년
피시험자 수		74명
주요 성능변수들		회선잡음 소리크기 손실
시험조건 수		12가지
기타 조건	측음로의 소리크기 손실	12dB(± 2dB)
	실내 소음	12dB(A)(± 2dB)

조건)에서 통화를 하도록 하였다. 통화가 완료되면, SIBYL은 방금 완료된 통화가 시험통화였음을 알리는 짧은 파열음 신호를 피시험자에게 보내고, 피시험자는 통화품질에 따라 5등급의 평가척도(‘매우 나쁘다’, ‘나쁘다’, ‘보통이다’, ‘좋다’, ‘매우 좋다’) 중 하나를 선택하고 해당 디지트를 다이얼한다. 이렇게 평가된 오피니언 데이터는 SIBYL에 저장된 후 이용자 만족도 평가모형의 유도를 위한 기초 자료로 이용된다. 시험 결과 수집된 오피니언 데이터는 <표 3>과 같다(Cavanaugh et al., 1976).

본 절에서는 <표 3>의 오피니언 데이터를 이용하여, 4.1절에서 제안한 로짓모형에 의한 분석 절차를 예증하고, 4.2절에서 논의된 분석 결과의 활용 방법을 설명한다.

5.1 분석 결과

<표 3>에서 이용자 만족도에 영향을 미치는 성능변수들은 잡음(N)과 손실(L)이므로, N, L, N · L, N², L² 중에서 최종 모형의 독립변수를 채택하기로 한다. 4.1절에서 제시한 분석

표 3. HO2 SIBYL 시험 데이터

결합조건	각 범주의 도수						합 계
	잡음(N)	손실(L)	매우 나쁘다	나쁘다	보통이다	좋다	
25	5	2	0	13	35	49	99
25	10	7	13	105	418	486	1029
25	20	1	9	20	13	7	50
25	30	16	22	20	2	2	62
32	5	0	4	22	25	19	70
32	10	1	4	19	23	11	58
32	20	4	22	31	4	3	64
32	30	12	27	8	4	1	52
42	5	4	18	19	6	6	53
42	10	7	24	32	17	17	97
42	20	11	19	9	5	2	46
42	30	3	0	1	0	0	4

표 4. 비례승산모형의 분석 결과

Score Test for the Proportional Odds Assumption							
Chi-Square = 81.4431 with 15 DF ($p=0.0001$)							
Model Fitting Information and Testing Global Null Hypothesis BETA=0							
Criterion	Intercept	Intercept and Covariates	Chi-Square for Covariates				
	Only						
AIC	4706.658	3956.771					
SC	4728.374	4005.631					
-2 LOG L	4698.658	3938.771	759.887 with 5 DF ($p=0.0001$)				
Score			581.861 with 5 DF ($p=0.0001$)				
Analysis of Maximum Likelihood Estimates							
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCP1	1	-15.0194	2.3085	42.3307	0.0001	.	.
INTERCP2	1	-13.2115	2.3109	32.6859	0.0001	.	.
INTERCP3	1	-11.4702	2.3101	24.6545	0.0001	.	.
INTERCP4	1	-9.6345	2.3000	17.5476	0.0001	.	.
N	1	0.4472	0.1370	10.6601	0.0011	1.399825	1.564
L	1	0.1478	0.0672	4.8314	0.0279	0.502253	1.159
NL	1	-0.00412	0.00145	8.1312	0.0044	-0.459949	0.996
NN	1	-0.00361	0.00201	3.2228	0.0726	-0.745335	0.996
LL	1	0.00385	0.00112	11.8184	0.0006	0.468502	1.004

절차에 따른 분석 결과와 3.1절과 3.2절에서 언급한 가설 검정 결과는 다음과 같다.

- (i) 비례승산모형으로 분석하기 위해 SAS의 LOGISTIC 절차에서 제공하는 후진제거 변수선택 방법을 적용한 결과는 <표 4>와 같다. <표 4>에서 모형에 있는 독립변수 전체의 유의성을 검정하기 위한 우도비 검정통계량값은 759.887(p -값=0.0001)이고, 점수 통계량값은 581.861(p -값=0.0001)이므로 상정한 비례승산모형이 유의수준 5%에서 타당하다고 판단한다. 모형에 있는 각 독립변수의 유의성을 검정하기 위한 왈드 통계량값들에 대응되는 p -값들이 0.20보다 작기 때문에 유의수준 20%에서 더 이상 제거할 변수들은 없으며, 성능변수들의 주효과 N 과 L 도 모두 포함하고 있다.
- (ii) <표 4>에서 평행선 가정을 검정하기 위한 점수 통계량값이 81.4431이고 p -값이 0.0001이므로, 유의수준 5%에서 평행선 가정은 만족되지 않는다.
- (iii) <표 3>에서 도수가 0인 셀들이 존재하므로, SAS의 CATMOD 절차를 사용하여 최우추정법에 의한 일반화 로짓모형으로 분석한 결과는 <표 5>와 같다.
- (iv) <표 5>에서 3열에 있는 왈드 통계량값들에 의해 현재 모형에 포함된 각 독립변수의 유의성을 검정할 수 있다. 현재 모형에 포함된 모든 변수에 대해 3열의 값이 $\chi^2_{0.20,4}$ 보다 충분히 크기 때문에(즉, 4열에 있는 왈드 통계량값들에 대응되는 p -값들이 0.20보다 작기 때문에), 유의수준 20%에서 더 이상 제거할 변수들은 없다. 그리

표 5. 최우추정법에 의한 일반화 로짓모형의 분석 결과

MAXIMUM-LIKELIHOOD ANALYSIS-OF-VARIANCE TABLE			
Source	DF	Chi-Square	Prob
INTERCEPT	4	56.20	0.0000
N	4	41.67	0.0000
L	4	15.30	0.0041
N*L	4	22.46	0.0002
N*N	4	28.93	0.0000
L*L	4	8.14	0.0865
LIKELIHOOD RATIO	24	40.60	0.0184

고 성능변수들의 주효과 N 과 L 도 모두 포함하고 있으므로, 수동 변수선택은 수행할 필요가 없다.

- (v) 모형 전체의 유의성을 검정하기 위한 즉, 모형의 적합도를 의미하는 우도비 검정통계량값은 40.60(p -값은 0.0184)으로 비교적 컸지만 유의수준 1%에서 상정한 모형을 채택하기로 한다. 최종적으로 유도된 최우추정법에 의한 일반화 로짓모형은 다음과 같다.

$$L_1(\hat{\pi}_g) = \ln(\hat{\pi}_{g1} / \hat{\pi}_{g5})$$

$$= -10.2502 + 0.2367N - 0.00209L$$

$$- 0.00294NL - 0.00006N^2 + 0.0105L^2$$

$$L_2(\hat{\pi}_g) = \ln(\hat{\pi}_{g2} / \hat{\pi}_{g5})$$

$$= -33.8415 + 1.5387N + 0.5349L$$

$$\begin{aligned}
 & -0.0126NL - 0.0172N^2 + 0.00309L^2 \\
 L_3(\hat{\pi}_g) &= \ln(\hat{\pi}_{g3} / \hat{\pi}_{g5}) \\
 &= -24.1796 + 1.2437N + 0.3091L \\
 & -0.00884NL - 0.0153N^2 + 0.00332L^2 \\
 L_4(\hat{\pi}_g) &= \ln(\hat{\pi}_{g4} / \hat{\pi}_{g5}) \\
 &= -8.7720 + 0.5290N + 0.00513L \\
 & + 0.000332NL - 0.00773N^2 + 0.000946L^2
 \end{aligned} \tag{5}$$

단, $\hat{\pi}_g = (\hat{\pi}_{g1}, \hat{\pi}_{g2}, \hat{\pi}_{g3}, \hat{\pi}_{g4}, \hat{\pi}_{g5})$

(vi) 제안된 방법에 의한 이용자 만족도의 예측력을 측정하기 위하여 정밀도를 나타내는 식 (6)과 같은 척도를 사용한다(안혜숙 외, 1999).

$$\text{평균오차(mean error)} = \frac{\sum_{g=1}^G \sum_{i=1}^5 \frac{|d_{gi} - \hat{\pi}_{gi}|}{G \times 5}} \tag{6}$$

단, d_{gi} 는 g 번째 결합조건에서 i 번째 범주에 응답한 응답자의 비율이고, $\hat{\pi}_{gi}$ 는 해당 결합조건에서 π_{gi} 의 추정치이다. 식 (6)을 사용하여 계산된 로짓모형에 의한 분석 결과의 정밀도는 <표 6>과 같다. <표 6>에는 안혜숙 외(1999)에서 보여진 누적퍼센트곡선 방법에 의한 분석 결과의 정밀도도 비교를 위하여 함께 포함되어 있다. <표 6>에서 나타난 것처럼, 평균오차와 예측오차의 범위(=오차의 최대값 - 오차의 최소값) 관점에서 로짓모형에 의한 분석 결과의 정밀도가 누적퍼센트곡선 방법 보다 상당히 만족스러움을 알 수 있다.

표 6. 로짓모형과 누적퍼센트곡선 방법의 정밀도 비교 (단위: %)

	일반화 로짓모형	누적퍼센트곡선 방법
평균오차	3.1670	4.7153
오차의 최소값	0.0050	0.0744
오차의 최대값	21.2370	31.0598

5.2 분석 결과의 활용

위의 식 (5)와 같이 유도된 일반화 로짓모형을 이용하면 성능변수들의 현재 조건에서 이용자가 ‘매우 나쁘다’, ‘나쁘다’, ‘보통이다’, ‘좋다’, ‘매우 좋다’라고 평가할 확률을 예측할 수 있을 뿐만 아니라, 통신망의 설계연구에 이용자의 만족도를 다음과 같이 반영할 수 있다. 동일 서비스를 제공하는 타사와의 만족도 비교 결과 최소한 이용자의 70%가 ‘좋다’ 이상으로 평가하도록 서비스 품질을 향상시켜야 경쟁력을 가질 수 있다는 전략이 세워졌다고 하자. 식 (2)로부터

$$\hat{\pi}_{gi} = \frac{\exp(L_i(\hat{\pi}_g))}{1 + \sum_{r=1}^4 \exp(L_r(\hat{\pi}_g))}, \quad i = 1, 2, 3, 4 \tag{7}$$

이므로, 식 (5)의 $L_i(\hat{\pi}_g)$ ($i = 1, 2, 3, 4$)을 식 (7)에 대입하면 주어진 전략을 반영하는 성능변수들의 목표치는 식 (8)을 만족하는 범위에 있게 된다.

$$\begin{aligned}
 & \hat{\pi}_{g4} + \hat{\pi}_{g5} \\
 &= \frac{\exp(L_4(\hat{\pi}_g)) + 1}{1 + \exp(L_1(\hat{\pi}_g)) + \exp(L_2(\hat{\pi}_g)) + \exp(L_3(\hat{\pi}_g)) + \exp(L_4(\hat{\pi}_g))} \\
 & \geq 0.7
 \end{aligned} \tag{8}$$

식 (8)에서 $\hat{\pi}_{g4} + \hat{\pi}_{g5} = 0.7$ 로 고정시킨 후 성능변수들과 이용자 만족도와의 관계를 등고선으로 그리면 <그림 3>와 같이 나타내어진다. <그림 3>에는 $\hat{\pi}_{g4} + \hat{\pi}_{g5} = 0.5, 0.9$ 인 경우도 함께 포함되어 있다.

<그림 3>의 등고선을 이용하여 현재 망의 상태와 비교하여 성능변수들의 목표치를 세울 수 있으며, 잡음(N)과 손실(L)이 이용자 만족도에 미치는 영향에 관한 상호관계를 파악할 수 있다. 예를 들어, 현재 서비스 중인 자사의 망에 대하여 성능변수들을 측정한 결과 $N=25, L=20$ 이라고 하자. 현재의 조건에서는 이용자 만족도가 약 50% 정도이다. 따라서 이용자 만족도를 70% 정도로 향상시키기 위해서는 성능변수들의 목표치를 상향 조정해야 할 필요가 있으며, $N=25, L=16$ 혹은 $N=22, L=20$ 등이 대안이 될 수 있다. 가능한 여러 가지 대안들 중에서 어떤 것을 선택할 것인가는 기술적으로 구현이 용이한 것 또는 비용이 저렴한 것을 선택하면 될 것이다.

또한 <그림 3>의 등고선으로부터 각 성능변수의 조정이 이용자 만족도 향상에 기여하는 정도를 파악할 수 있다. 예를 들어, 현재 서비스 중인 성능변수들이 $N=15, L=27.5$ 라고 하

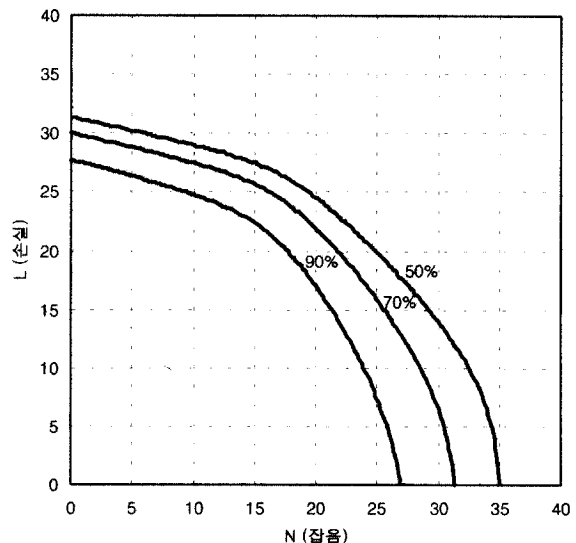


그림 3. 성능변수들과 이용자 만족도간 등고선.

면 이용자 만족도는 약 50%이며, 이용자 만족도를 70% 정도로 향상시키기 위해서는 N 은 5정도 감소시켜야 하는 반면에 L 은 1.8정도 감소시키면 되기 때문에 L 이 N 보다 만족도의 향상에 더 크게 기여함을 알 수 있다. 반면에 현재 서비스 중인 성능 변수들이 $N=30$, $L=14$ 라고 하면 이용자 만족도는 역시 약 50%이며, 이용자 만족도를 70% 정도로 향상시키기 위해서는 N 은 3.7 정도 감소시키면 되지만 L 은 7.5정도 감소시켜야 하기 때문에 N 이 L 보다 만족도의 향상에 더 크게 기여한다고 할 수 있다.

6. 결론

통신 서비스에 대해 이용자가 느끼는 주관적인 서비스품질은 설문조사를 통해 순서 범주형 자료로 계량화된다. 이렇게 얻어진 오피니언 데이터를 분석하기 위한 기존의 누적퍼센트곡선은 설문문항의 각 범주에 주관적으로 점수를 부여함으로써 유도되기 때문에, 이 방법을 사용할 경우 점수척도가 달라지면 어떤 주어진 성능변수들의 상태에서 만족도 분포를 추정할 결과가 달라질 수 있다. 또한 이용자 만족도에 대한 회사의 동일한 전략을 달성하기 위해 설정된 성능변수들의 목표치에 대한 일관성이 보장되지 않는다.

본 논문에서는 누적퍼센트곡선 방법의 대안으로서, 설문조사에서 얻은 순서 범주형 자료인 오피니언 데이터를 주관적인 점수를 부여하지 않고 분석함으로써 주관적인 점수 할당으로 인한 문제점을 극복할 수 있는 로짓모형을 이용한 분석 방법을 제안하고, 분석 절차 및 분석 결과의 활용 방법을 제시하였다. 아울러 범주형 자료 분석을 위해 사용되는 통계적 모형들 및 로짓모형의 분석을 위해 사용 가능한 SAS 통계패키지의 절차들을 상세히 소개하였다.

본 논문에서 제안된 로짓모형에 의한 오피니언 데이터의 분석은, 이용자의 만족도를 평가하고, 회사의 입장에서 이용자 만족도에 대한 전략을 반영하여 성능변수들의 목표치를 설정하는데 일관성이 보장되는 보다 객관적 방법이라 할 수 있다. 그리고 유도된 로짓모형으로부터 이용자의 만족도와 성능변수들 사이의 관계를 등고선으로 나타낼 수 있고, 성능변수의 조정이 이용자 만족도 향상에 기여하는 정도를 쉽게 파악할 수 있다. 또한 분석 예제를 통하여 이용자 만족도에 대한 예측력 관점에서, 제안된 로짓모형을 이용한 방법이 매우 타당함을 볼 수 있었다. 그러나 로짓모형을 이용한 분석 방법은 가설 검정에 의해 유의한 모형을 발견하지 못할 수도 있다는 것이 단점으로 지적될 수 있다.

참고문헌

- Byun, E. S. (1996), *Determination of Engineering Characteristic Values by Quality Function Deployment*, Master's Thesis, Korea Advanced Institute of Science and Technology.
- Ahn, H.-S., Cho, J.-S. and Yum, B.-J. (1999), Evaluation methods for quality of service in telecommunications, *IE Interfaces*, 12(4), 496-505.
- Lee, Y.-H. and Kwon, S.-H. (1996), Methodology on measuring the quality of service in telecommunications, *Journal of the Korean Institute of Telematics and Electronics*, 33A(10), 28-32.
- Agresti, A. (1990), *Categorical Data Analysis*, Wiley, New York.
- Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, Wiley, New York.
- Cavanaugh, J.-R., Hatch, R.-W. and Sullivan, J.-L. (1976), Models for the Subjective Effects of Loss, Noise, and Talker Echo on Telephone Connections, *The Bell System Technical Journal*, 55(9), 1319-1371.
- SAS/STAT User's Guide (1990), *Version 6, 4th ed.*, Vol. 1-2, SAS Institute Inc., Cary, North Carolina.
- Stokes, M.-E., Davis, C.-S. and Koch, G.-G. (2000), *Categorical Data Analysis Using The SAS System*, 2nd ed., SAS Institute inc., Cary, North Carolina.

조 계 규

연세대학교 응용통계학과 학사
한국과학기술원 산업공학과 석사
한국과학기술원 산업공학과 박사
한국전자통신연구원 선임연구원
현재: 동의대학교 경영정보·인터넷비즈니스
학부 조교수
관심분야: 통신 서비스 품질, 통신시스템 구성
요소의 중요도 평가 등

안 해 속

한국과학기술원 산업공학과 학사
한국과학기술원 산업공학과 석사
현재: 삼성 SDS ERP 사업팀
관심분야: 통신 서비스 품질, Analytical CRM 등