

## 한우 수소 고기 관능평가 데이터에 대한 범주형 자료 분석

이혜정<sup>1</sup> · 조수현<sup>2</sup> · 김재희<sup>3</sup>

<sup>1</sup>덕성여자대학교 정보통계학과 · <sup>2</sup>국립축산과학원, 축산물이용과

접수 2009년 7월 10일, 수정 2009년 8월 29일, 게재확정 2009년 9월 9일

### 요약

국립축산과학원에서 수집한 한우 관능 평가 데이터에 대해 사회 인구학적 요인과 한국 소비자들의 맛 평가에 대한 연관성을 연구하고자 한다. 소비자 거주지역, 연령, 성별, 직업, 월수입과 쇠고기 부위를 설명변수로 맛등급 평가를 반응변수로 이항 다중 로지스틱 모형과 다항 다중 로지스틱 모형을 적합하고 회귀계수별 유의성 검정과 적합도 검정을 실시하였다. 단계별 변수 선택으로 최종 모형을 선택하고 반응변수 범주에 대한 오즈비를 계산하여 관련성을 파악한다. 그 결과 거주 지역, 연령, 월수입과 쇠고기 부위 변수들이 선택되었다. 영남에서 맛을 비교적 높게 평가하는 경향이 있으며 수입이 많고 연령이 높을수록 맛을 까다롭게 평가하는 경향을 보인다. 쇠고기 부위별로는 우둔에 비해서 등심이 다른 부위들 중 맛에 대한 차이가 크다고 볼 수 있다.

주요용어: 로지스틱 회귀, 오즈비, 이탈도, 한우 관능평가 데이터.

### 1. 서론

외국 쇠고기 수입은 한국 고유 축종인 한우 (Hanwoo)의 소비에 큰 영향을 미치며 한우에 대한 맛 경쟁력이 요구되고 있다. 국립축산과학원에서는 2006년 3월부터 전국 소비자들을 대상으로 요리 된 한우를 맛본 후 등급과 더불어 맛과 관련된 변수를 측정하는 대규모 관능평가를 시작하였으며 계속 진행 중에 있다. 국민을 대표할 수 있는 인구학적인 특성을 반영하도록 소비자들을 대상으로 이와 같은 체계적인 조사는 우리나라 소비자들의 맛 등급 판정 특성을 파악하는데 도움이 되며 이러한 조사에 기반한 정보를 한우 개발에 적용할 필요가 있다. 소비자들은 동일한 조건의 쇠고기일지라도 요리방법, 쇠고기 부위에 따라 맛 평가가 달라지며 이와 같은 상황에 대한 이해는 향후 쇠고기 맛에 중요한 정보가 될 수 있다. 쇠고기 맛을 결정하는 가장 중요한 요인으로 연도, 다즙성, 향미가 있으며 이를 바탕으로 만족도에 따라 맛등급 점수를 부여하도록 하였다. 본 연구에서는 소비자들이 한우 맛의 등급을 결정하는데 있어 인구사회학적 관련 요인을 규명하고 유의한 요인들을 탐색하고자 한다. 특히 범주형 다변량 데이터에 대해 로지스틱 회귀모형을 이용하여 한우 맛의 등급 결정에 적합한 모형을 찾고 유의한 요인을 규명하며 이와 같은 형태의 범주형 다변량 데이터분석 응용 예로 보이고자 한다.

<sup>1</sup> (132-714) 서울시 도봉구 쌍문동 419, 덕성여자대학교 정보통계학과 대학원, 대학원생.

<sup>2</sup> (441-748) 경기도 수원시 권선구 오목천동, 국립축산과학원 축산물이용과, 연구사.

<sup>3</sup> 교신저자: (132-714) 서울시 도봉구 쌍문동 419, 덕성여자대학교 정보통계학과, 교수.

E-mail: jaehee@duksung.ac.kr

## 2. 로지스틱 회귀모형

### 2.1. 로지스틱 회귀모형

이항 반응이란 각 개체의 반응이 두 가지, 예를 들어, ‘성공’과 ‘실패’로 측정되는 것으로 이항자료는 범주형 자료에서 가장 흔하게 볼 수 있는 형태로 다음과 같은 이항 로지스틱 모형을 적합할 수 있다. 관련한 모형에 대해서는 박태성 등 (1998), 정광모 등 (2002)와 Agresti (1996)를 참조한다.

종속변수인 이항 반응변수를  $Y$ , 설명변수  $X = (X_1, X_2, \dots, X_p)'$ 가 주어졌을 때  $\pi(x) = P(Y = 1|x)$ 는  $X$  값이  $x$ 일 때의 ‘성공’ 확률로 이항분포에 대한 모수가 된다. 이항 로지스틱 모형은

$$\text{logit} [\pi(x)] = \log \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (2.1)$$

으로 정의된다. 위 식은 다시 쓰면

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \quad (2.2)$$

이 되어 반응확률  $\pi(x)$ 를 추정할 수 있다. 편의상, 성공일 경우  $Y = 1$ , 실패의 경우  $Y = 0$ 으로 놓기로 하자.

예를 들어, 한우 요리를 맛본 후 맛있다 ( $Y = 1$ ), 맛없다 ( $Y = 0$ )라고 정의할 때, 로지스틱 회귀모형을 통해 맛 등급의 표현을 성별, 연령, 거주지역 등의 요인들로 회귀모형을 적합한 후에 ‘맛있다’로 판정할 확률을 식 (2.2)을 이용하여 추정할 수 있다. ‘맛있다’일 오즈는

$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) \quad (2.3)$$

와 같이 구할 수 있다.

범주형 반응변수가 세 수준 이상의 범주를 가지며 반응변수 범주 순서간의 의미가 있을 경우 다음과 같은 누적 로지스틱 회귀모형을 고려할 수 있다.

### 2.2. 다항 다중 일반화 로지스틱 회귀모형

범주가  $1, \dots, J$ 를 가지며 범주의 반응확률을  $\{\pi_1, \dots, \pi_J\}$ 라고 하면,  $\sum_j \pi_j = 1$ 이 된다.  $n$ 개의 서로 독립인 관측값들이 이러한 확률에서 얻어졌을 때,  $J$ 개의 가능한 범주 중에서 각 범주에서 관측된 도수들의 확률분포는 다항분포 (multinomial distribution)를 따르며 명목형 반응변수에 대한 로짓 모형은 임의로 한 기준 범주를 선택한 후 이 범주와 나머지 각 반응 범주와 짝을 지어 로짓을 정의하여 일반화 로지스틱 회귀모형을 정의할 수 있다. 마지막 범주가 기준이 될 때, 기준 범주에 대한 로짓은 다음과 같다.

$$\log \left( \frac{\pi_j}{\pi_J} \right) = \beta_{0j} + \beta_{1j} x_1 + \dots + \beta_{pj} x_p, \quad j = 1, \dots, J - 1 \quad (2.4)$$

즉,  $J - 1$ 개의 로짓방정식으로 이루어졌으며 각 방정식마다 서로 다른 모수를 갖게 되며 기준 범주와 짝지어진 반응범주에 따라 효과가 다양하게 나타난다.  $J - 1$ 개의 로짓방정식들의 모수를 사용하여 다른 범주들의 쌍들에 대해 정의한 로짓을 나타낼 수 있다.

$$\begin{aligned} \log \left( \frac{\pi_a}{\pi_b} \right) &= \log \left( \frac{\frac{\pi_a}{\pi_J}}{\frac{\pi_b}{\pi_J}} \right) = \log \left( \frac{\pi_a}{\pi_J} \right) - \log \left( \frac{\pi_b}{\pi_J} \right) \\ &= (\beta_{0a} - \beta_{0b}) + (\beta_{1a} - \beta_{1b}) x_1 + \dots + (\beta_{pa} - \beta_{pb}) x_p. \end{aligned}$$

위의  $J - 1$ 개의 로짓방정식을 동시에 적합시켜 모수를 구할수 있으며 동시에 적합시키면 어떤 기준 범주가 되는 상관없이 범주들의 쌍에 대한 모수들의 추정값을 얻게 된다. 이와 같은 모형은 Allison (1999)와 Homer 등 (2000)을 참조한다.

추정 회귀계수  $\hat{\beta}$ 이 근사적으로 정규분포를 따른다는 것을 이용해 모수  $\beta$ 에 대한 대표본  $100(1 - \alpha)\%$  신뢰구간은  $\hat{\beta} \pm z_{\alpha/2} ASE$  와 같이 구할 수 있다. 여기서  $ASE = \sqrt{\hat{Var}(\hat{\beta})}$ 이고  $z_{\alpha/2}$ 는 표준정규분포 우측 꼬리의 확률이  $\alpha/2$ 가 되는 분위수를 의미한다.

반응변수에 대한 설명 변수의 효과 유의성 검정으로 로지스틱 회귀계수의 유의성 가설  $H_0 : \beta = 0$ 에 대해  $H_1 : \beta \neq 0$ 에 대한 대표본 검정통계량은

$$z = \frac{\hat{\beta}}{ASE}$$

이고 검정통계량  $z$ 는  $\beta = 0$ 일 경우에 표준정규분포를 따르므로 정규분포를 이용한 검정을 할 수 있다.

로지스틱 회귀모형에 대한 적합성에 대해서는 이탈도 (deviance)를 고려한 우도비 검정과 Wald 검정을 이용할 수 있다. 우도비 검정통계량은

$$U = -2(L_0 - L_1)$$

이며 여기서  $L_0$ 은  $\beta = 0$ 인 경우의 로그우도함수이고  $L_1$ 은 아무런 제한 조건이 없는 완전모형에서 로그우도함수의 최대값이며  $U$ 는 귀무가설 하에서 자유도가 1인 근사적으로 카이제곱 분포를 따른다.

대표본에 대해서는 Wald 통계량

$$z^2 = \left( \frac{\hat{\beta}}{ASE} \right)^2$$

으로 Wald 검정을 할 수 있으며 Wald 통계량은 귀무가설 하에서 자유도 1인 대표본 카이제곱분포를 따른다.

로지스틱 회귀모형의 진단으로 Pearson 잔차를 구하여 오차에 대한 검정을 할 수 있다. 또한 Pearson 카이제곱검정통계량

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

이나 우도비 검정통계량

$$G^2 = 2 \sum O \log \left( \frac{O}{E} \right)$$

을 이용하여 검정을 할 수 있다. 적합된 도수값이 5보다 클 때  $\chi^2$ 이나  $G^2$ 는 카이제곱분포로 근사한다. 여기서  $O$  = 관측값,  $E$  = 귀무가설 하에서 기대값이 된다. 모형에 대한 잔차 자유도  $df$  = 표본 로짓수 - 모형 로짓수 가 된다.  $\chi^2$ 이나  $G^2$ 값이 클수록 적합결여 증거가 된다. 또한 Brown 등 (1977)과 Kritzer (1977)는 두 범주형 변수간의 연관성에 대한 측도를 보여준다.

### 3. 한우데이터에 대한 분석

#### 3.1. 기초분석

우리나라 재래가축 자원 중 가장 경쟁력이 확보되고 경제성이 높은 것으로 확인된 한우에 대한 연구는 축산연구에서 매우 중요한 위치를 차지한다. Lee 등 (2005)와 Lee 등 (2006)은 한우의 주요 경제형질인 육질과 육량에 관련된 유전자 분석을 통해 DNA marker를 찾는 등 한우에 관한 다각적인 연구가 진행되고있다. 농촌진흥청 국립축산과학원에서는 2006년 전 국민을 대상으로 추출한 표본 소비자

650명에게 한우관능평가를 실시하였다. 육질 3등급인 수소 (bull)의 8개 부위 (cut)를 탕, 구이, 스테이크의 세 가지 요리 방법으로 조리한 후 소비자가 시식하여 연도 (tenderness), 다즙성 (juiciness), 향미 (flavor), 전반적인 기호도 (overall acceptability)를 0~100점 사이의 점수로, 맛등급 (taste grade)은 4개의 등급으로 평가하도록 하였다. 한우 수소 (bull)의 8개 부위인 보섭 (B), 채끝 (C), 등심 (D), 꾸리 (G), 목심 (M), 설도 (S), 우둔 (U)과 흥두께에 대해 BBQ 요리방법 (로스구이)으로 요리된 고기의 맛을 평가한 데이터에 대한 범주형 자료 분석을 하고자한다. 여기서 맛등급 값은 1='만족하지 못한다', 2='만족한다', 3='매우 만족한다', 4='극도로 만족한다'를 나타낸다. 우선 한우 맛 데이터의 범주 변수에 대한 기초 통계 분석으로 표 3.1을 보면 각 범주의 빈도와 각 변수와 맛등급 변수에 대한 카이제곱 독립성 검정을 한 결과, 유의수준 5%에서 '지역', '연령', '직업군', '부위'의 변수들이 맛등급과 유의한 관련성을 보였다. 조수현 외 (2007)는 이와 같은 데이터에 대해 사회인구학적 요인이 한우 수소가고기의 부위 및 요리형태별 관능특성에 미치는 영향에 관한 연구 결과를 보여준다.

### 3.2. 이항 다중 로지스틱 회귀모형 단계적 변수선택 결과와 최종모형 적합

한우 요리를 맛본 후 반응변수를 만족한다 ( $Y = 1$ ), 만족하지 못한다 ( $Y = 0$ )라고 정의할 때, 성별, 연령, 거주지역 등의 요인들로 이항 다중 로지스틱 회귀모형을 통해 맛 등급에 대한 회귀모형을 적합한다. 여기서 ( $Y = 1$ )은 맛등급이 2, 3, 4인 경우를 포함하며 ( $Y = 0$ )은 맛등급이 1인 경우만을 의미한다. 로지스틱 회귀모형 (2.1)에서 고려된 변수 '거주 지역', '소비자 성별', '연령', '월수입', '직업', '쇠고기 부위'들에 대해 단계적 변수선택을 하는 과정은 일부의 변수를 포함한 모형의 log-likelihood에서 가장 중요한 변수를 뽑아내는 과정이다. Homer (2000)에는 로지스틱 모형의 분석 예에 대한 설명이 잘 정리되어있다. 표 3.2는 이항 다중 로지스틱 회귀모형에서 단계적 변수선택을 한 결과 변수 '부위', '직업', '지역', '소득', '연령' 순서로 변수 선택이 됨으로써 6개의 설명변수 중 '성별'만 선택되지 않았다. 이들 선택된 변수들 중에서도 '직업'은 다양한 직업이 골고루 고려되지 않은 관계로 제외하기로 하고 유의한 변수들인 '지역', '연령', '수입', '부위'들만 고려한 최종모형을 적합하기로 한다. 최종 적합한 이항 다중 로지스틱 모형은 표 3.3에서 보여준다. 유의수준 1% 에서 적합한 모형이 유의하며 모수추정값과 식 (2.3)을 이용해 구한 오즈비를 살펴보면, 호남과 중부지역 소비자들은 영남지역에 비해 맛의 만족도가 약간 낮고 서울지역 소비자들은 맛의 만족도가 약간 높은 편이다. 20대가 50대에 비해 맛의 만족도가 약간 높은 편이며 30, 40대는 맛의 만족도가 약간 낮은 편이다. 부위에 대한 오즈비를 살펴보면 우둔에 비해 등심은 맛에 대한 만족도가 5.8배로 높게 나타나 부위에 대한 특성이 반영되고 있음을 알 수 있다.

### 3.3. 다항 다중 일반화 로지스틱 회귀모형 단계적 변수선택과 최종모형

맛등급 중 4 번째 등급을 마지막 범주로 하고 6개의 설명변수들로 다항 다중 일반화 로지스틱 회귀모형을 적합하고 변수 선택한 결과는 표 3.4에서 보여준다. 이항 다중 로지스틱 모형에서와 마찬가지로 직업 변수는 제외하고 변수선택 과정을 통해 선택된 지역, 연령, 소득, 부위를 고려하여 식 (2.4)의 다항 다중 일반화 로지스틱 회귀모형을 적합하였다. 회귀계수 유의성 검정에서 보면 수입 변수외에는 대체로 유의하게 나타났으며 모형적합도의 p-값은 이항 다중 로지스틱 회귀모형보다 더 높은 것으로 나타나 모형적합도는 증가하였다. '극도로 만족한다'의 4 번째 등급을 기준으로 1등급, 2등급, 3등급의 맛등급들과의 오즈를 비교한 결과, 모형의 분자가 맛이 덜 한 등급이 비교됨으로써 이항 다중 로지스틱 회귀모형과 반대의 오즈 결과가 나타났지만 오즈비 해석의 결과로 보면 비슷한 경향을 보인다고 할 수 있다. 표 3.5에서 가장 맛등급 차이가 큰 4 등급 기준으로 한 1등급의 모형의 logit 결과를 보면 영남지역에 비해 호남과 중부지역은 만족도가 낮은 오즈가 높고 서울은 영남지역과 비슷한 오즈를 나타낸다. 50대에

비해 그 이하 나이대에서는 오즈비가 1 보다 작아 맛에 대한 만족도가 높은 편으로 나타난다. 평균수입이 500만원 이상인 응답자들에 비해 수입이 그보다 낮은 편인 경우 만족도가 낮은 오즈값이 작게 나타나고 수입자가 맛등급을 더 까다롭게 부여함을 알 수 있다. 부위에 대한 오즈비를 살펴보면 ‘설도’가 ‘우둔’에 비해 맛등급을 낮게 줄 오즈가 1.3배 정도 높은 것으로 보아 응답자들은 설도 부위를 우둔보다 더 맛이 없다고 응답한 경향이 있으며, ‘우둔’에 비해 설도를 제외한 다른 부위는 맛에 대한 만족도가 높은 편으로 나타난다. 마찬가지로 4등급 기준으로 한 2 등급, 3 등급에 대한 경향도 비슷하다.

표 3.1 소비자들의 인구사회학적 빈도와 맛등급과의카이제곱 독립성검정

변수	값	변수명	맛등급				$\chi^2$ 통계량	p-value
			1	2	3	4		
지역	호남	Honam	105	131	101	12	35.54	< 0.001
	충부	Jungbu	251	314	141	29		
	서울	Seoul	69	114	76	21		
	영남	Youngnam	89	153	73	27		
성별	남	gender1	231	350	183	48	3.679	0.298
	여	gender2	283	362	208	41		
연령	20-25세	age1	83	103	66	20	30.975	0.002
	26-30세	age2	42	89	60	12		
	31-39세	age3	169	214	85	22		
	40-49세	age4	132	188	96	25		
	50-59세	age5	88	118	84	10		
직업군	전문직	occ1	22	33	14	8	59.45	< 0.001
	행정관리직	occ2	63	108	50	10		
	기술직	occ3	40	66	32	2		
	영업	occ4	98	82	39	19		
	노동직	occ5	0	11	3	0		
	주부	occ6	134	203	105	19		
	무직	occ7	8	5	1	0		
	학생	occ8	89	122	88	22		
	기타	occ9	60	82	59	9		
평균수입	100이하	income1	48	44	25	2	25.399	0.114
	100-200	income2	123	155	103	18		
	200-300	income3	165	249	127	32		
	300-400	income4	77	148	71	19		
	400-500	income5	53	59	26	9		
	500이상	income6	43	47	33	9		
부위	보섭	cut1	47	95	70	21	154.33	< 0.001
	채끝	cut2	59	58	38	7		
	등심	cut3	43	144	99	19		
	꾸리	cut4	26	48	32	6		
	홍두깨	cut5	61	85	44	8		
	목심	cut6	41	75	39	18		
	설도	cut7	113	101	40	4		
	우둔	cut8	124	106	29	6		

표 3.2 이항 다중 로지스틱 회귀분석 단계적 변수선택 결과

Step	Summary of Backward		Elimination	
	Entered	DF	Score Chi-Square	p-값
1	부위	7	115.2518	< .0001
2	직업	8	27.7987	0.0005
3	지역	3	18.5254	0.0003
4	수입	5	20.0777	0.0012
5	나이	4	13.1913	0.0104

표 3.3 이항 다중 로지스틱 최종모형

변수	모수추정값	ASE	검정통계량	p-value	오즈비	모형적합도
Intercept	0.0915	0.2772	0.109	0.7413		
지역	호남(H)	-0.167	0.1826	0.8363	0.3605	H vs Y 0.846
	중부(J)	-0.4608	0.1581	8.4978	0.0036	J vs Y 0.631
	서울(S)	0.0693	0.1974	0.1231	0.7257	S vs Y 1.072
	영남(Y)	0	.	.	.	
연령	20-25세(1)	0.0285	0.2006	0.0201	0.8872	1 vs 5 1.029
	26-30세(2)	0.5911	0.2321	6.4823	0.0109	2 vs 5 1.806
	31-39세(3)	-0.2866	0.1716	2.7875	0.095	3 vs 5 0.751
	40-49세(4)	-0.0644	0.1791	0.1294	0.719	4 vs 5 0.938
	50-59세(5)	0	.	.	.	
평균수입	100이하(1)	-0.3939	0.2857	1.9008	0.168	1 vs 6 0.674
	100-200(2)	0.1951	0.2291	0.7249	0.3945	2 vs 6 1.215
	200-300(3)	0.3679	0.2213	2.7648	0.0964	3 vs 6 1.445
	300-400(4)	0.6234	0.2411	6.688	0.0097	4 vs 6 1.865
	400-500(5)	0.158	0.2741	0.3322	0.5644	5 vs 6 1.171
	500이상(6)	0	.	.	.	
부위	보쌈(B)	1.3014	0.2095	38.5936	< .0001	B vs U 3.674
	채끝(C)	0.4059	0.2093	3.7636	0.0524	C vs U 1.501
	등심(D)	1.7729	0.2109	70.6783	< .0001	D vs U 5.888
	꾸리(G)	1.1235	0.2607	18.5735	< .0001	G vs U 3.076
	홍두깨(H)	0.738	0.2026	13.2743	0.0003	H vs U 2.092
	목심(M)	1.0673	0.2216	23.1936	< .0001	M vs U 2.908
	설도(S)	0.125	0.1802	0.4814	0.4878	S vs U 1.133
	우둔(U)	0	.	.	.	

표 3.4 다항 다중 일반화 로지스틱 회귀분석 단계적 변수선택 결과

Step	Summary of Backward		Elimination	
	Entered	DF	Score Chi-Square	p-값
1	부위	21	155.4144	< .0001
2	지역	9	40.6757	< .0001
3	직업	24	67.3699	< .0001
4	나이	12	26.2273	0.01
5	수입	15	29.0794	0.0157

#### 4. 결론

한우 관능평가 데이터로 맛등급을 반응변수로 하고 ‘거주 지역’, ‘소비자 성별’, ‘소비자 연령’, ‘직업’, ‘월수입’, ‘쇠고기 부위’의 설명변수로 여러 가지 방법의 로지스틱 회귀분석을 적합한 결과 모든 방법에서 소비자 성별의 설명변수만이 로지스틱 회귀분석과 적합하지 않은 것으로 나타났다. 그 이외의 ‘지역’, ‘연령’, ‘직업’, ‘수입’, ‘부위’의 설명변수는 로지스틱 회귀분석 모형에 적합하며 성별은 맛등급을 높게 줄 오즈에 영향을 미치지 않는다고 할 수 있다. 변수선택 후 이항 다중, 다항 다중 일반화 로지스틱 회귀모형 두 모형에서 모두 ‘거주 지역’, ‘연령’, ‘월수입’, ‘부위’ 변수들이 선택되었다. 그 중 다항 다중 일반화 로지스틱 회귀모형이 모형 적합도가 가장 높았으며 그 결과, 지역에서는 영남이 맛등급을 높게 주는 경향을 보였으며 50대의 연령층에 비해 20대 연령층은 맛등급을 낮게 주고 30대는 맛등급을 높게 주는 경향을 보였다. 이항 다중 로지스틱 모형에서는 20대 맛의 만족도가 50대에 비해 크게 나타났는데 이는 맛등급이 2,3,4인 경우를 만족 그룹으로 묶으면서 만족 비율이 높게 나타나는 경향을 보이게 된 것으로 범주 형태가 바뀌면서 나타난 현상이다. 월수입면에서는 월수입이 낮은 그룹이 월수입이 높은 그룹에 비해 맛등급을 높게 주는 차이를 보였고, 부위에서는 우둔에 비해 다른 모든 부위의 맛등급

표 3.5 다항 다중 일반화 로지스틱 최종모형

변수		모수추정값	ASE	검정통계량	p-value	오즈비	모형적합도
1	Intercept	2.9001	0.6665	18.9361	< .0001		
지역	호남(H)	0.9113	0.3938	5.3539	0.0207	H vs Y 2.488	
	충부(J)	1.1430	0.3173	12.9758	0.0003	J vs Y 3.136	
	서울(S)	0.0874	0.3555	0.0604	0.8058	S vs Y 1.091	
	영남(Y)	0	.	.	.		
연령	20-25세(1)	-0.8317	0.4376	3.6127	0.0573	1 vs 5 0.435	
	26-30세(2)	-1.0192	0.4914	4.3014	0.0381	2 vs 5 0.361	
	31-39세(3)	0.0863	0.4171	0.0428	0.8361	3 vs 5 1.090	
	40-49세(4)	-0.3495	0.4153	0.7083	0.4000	4 vs 5 0.705	
	50-59세(5)	0	.	.	.		
평균수입	100이하(1)	1.6231	0.8326	3.8004	0.0512	1 vs 6 5.069	
	100-200(2)	0.0479	0.4673	0.0105	0.9183	2 vs 6 1.049	
	200-300(3)	-0.2268	0.4354	0.2712	0.6025	3 vs 6 0.797	
	300-400(4)	-0.5250	0.4671	1.2630	0.2611	4 vs 6 0.592	
	400-500(5)	-0.4136	0.5509	0.5638	0.4527	5 vs 6 0.661	
	500이상(6)	0	.	.	.		
부위	보섭(B)	-2.3173	0.4993	21.5383	< .0001	B vs U 0.099	
	채끝(C)	-0.8365	0.5838	2.0530	0.1519	C vs U 0.433	
	등심(D)	-2.3591	0.5064	21.7040	< .0001	D vs U 0.095	
	꾸리(G)	-1.7264	0.6237	7.6626	0.0056	G vs U 0.178	
	홍두깨(H)	-1.0708	0.2026	3.5542	0.0594	H vs U 0.343	
	목심(M)	-2.2636	0.2216	19.6467	< .0001	M vs U 0.104	
	설도(S)	0.3183	0.1802	0.2313	0.6306	S vs U 1.375	
	우둔(U)	0	.	.	.		

표 3.6 다항 다중 일반화 로지스틱 최종모형(계속)

변수		모수추정값	ASE	검정통계량	p-value	오즈비	모형적합도
2	Intercept	2.5763	0.6559	15.4290	< .0001		
지역	호남(H)	0.6526	0.3785	2.9726	0.0847	H vs Y 1.920	
	충부(J)	0.7632	0.3021	6.3836	0.0115	J vs Y 2.145	
	서울(S)	0.0724	0.3327	0.0473	0.8278	S vs Y 1.075	
	영남(Y)	0	.	.	.		
연령	20-25세(1)	-0.8318	0.4232	3.8644	0.0493	1 vs 5 0.435	Deviance/df =1.013
	26-30세(2)	-0.4265	0.4662	0.8369	0.3603	2 vs 5 0.653	
	31-39세(3)	-0.0322	0.4063	0.0063	0.9369	3 vs 5 0.968	
	40-49세(4)	-0.3230	0.4026	0.6436	0.4224	4 vs 5 0.724	
	50-59세(5)	0	.	.	.		
평균수입	100이하(1)	1.4045	0.8244	2.9027	0.0884	1 vs 6 4.073	p-값= 0.3399
	100-200(2)	0.2694	0.4545	0.3511	0.5535	2 vs 6 1.309	
	200-300(3)	0.2224	0.4220	0.2777	0.5982	3 vs 6 1.249	
	300-400(4)	0.1873	0.4486	0.1744	0.6763	4 vs 6 1.206	
	400-500(5)	-0.1899	0.5364	0.1253	0.7233	5 vs 6 0.827	
	500이상(6)	0	.	.	.		
부위	보섭(B)	-1.4144	0.4870	8.4343	0.0037	B vs U 0.243	
	채끝(C)	-0.7139	0.5830	1.4991	0.2208	C vs U 0.490	
	등심(D)	-0.9194	0.4884	3.5440	0.0598	D vs U 0.399	
	꾸리(G)	-0.8950	0.6071	2.1736	0.1404	G vs U 0.409	
	홍두깨(H)	-0.5296	0.5626	0.8859	0.3466	H vs U 0.589	
	목심(M)	-1.4935	0.4986	8.9734	0.0027	M vs U 0.225	
	설도(S)	0.3636	0.6620	0.3016	0.5829	S vs U 1.438	
	우둔(U)	0	.	.	.		

을 높게 줄 오즈가 높았지만 특히 등심이 우둔에 비해 맛등급을 높게 줄 오즈가 월등히 높았다. 결과를 종합해보면, 영남에서 맛을 비교적 높게 평가하는 경향이 있으며 수입이 많고 연령이 높을수록 맛을 까다롭게 평가하는 경향을 보인다는 것을 알 수 있었다. 쇠고기 부위를 고려해보면, 우둔에 비해서 등심이

표 3.7 다항 다중 일반화 로지스틱 최종모형(계속)

변수		모수추정값	ASE	검정통계량	p-value	오즈비	모형적합도
3	Intercept	1.5249	0.6870	4.9271	0.0264		
지역	호남(H)	1.0589	0.3902	7.3647	0.0067	H vs Y 2.883	
	충부(J)	0.6562	0.3198	4.2116	0.0401	J vs Y 1.928	
	서울(S)	0.3670	0.3487	1.1078	0.2926	S vs Y 1.443	
	영남(Y)	0	.	.	.		
연령	20-25세(1)	-0.8871	0.4333	4.1918	0.0406	1 vs 5 0.412	
	26-30세(2)	-0.4884	0.4756	1.0546	0.3044	2 vs 5 0.614	
	31-39세(3)	-0.5835	0.4190	1.9392	0.1638	3 vs 5 0.558	
	40-49세(4)	-0.6494	0.4131	2.4717	0.1159	4 vs 5 0.522	
	50-59세(5)	0	.	.	.		
평균수입	100이하(1)	1.0839	0.8398	1.6658	0.1968	1 vs 6 2.956	
	100-200(2)	0.2629	0.4677	0.3161	0.5739	2 vs 6 1.301	
	200-300(3)	0.0296	0.4366	0.0046	0.9459	3 vs 6 1.030	
	300-400(4)	-0.0404	0.4654	0.0075	0.9308	4 vs 6 0.960	
	400-500(5)	-0.4220	0.5643	0.5594	0.4545	5 vs 6 0.656	
	500이상(6)	0	.	.	.		
부위	보섭(B)	-0.3562	0.5170	0.4747	0.4909	B vs U 0.700	
	채끝(C)	0.1795	0.6127	0.0859	0.7695	C vs U 1.197	
	등심(D)	0.0906	0.5179	0.0306	0.8611	D vs U 1.095	
	꾸리(G)	-0.00054	0.6389	0.0000	0.9993	G vs U 0.999	
	홍두깨(H)	0.1439	0.5953	0.0584	0.8090	H vs U 1.155	
	목심(M)	-0.7680	0.5357	2.0550	0.1517	M vs U 0.464	
	설도(S)	0.7560	0.6620	1.1893	0.2755	S vs U 2.130	
	우둔(U)	0	.	.	.		

다른 부위들 중 맛에 대한 차이가 크다고 볼 수 있다. 각 부위에 대한 오즈비는 각각 다르게 나타나 부위 별로 맛등급 차이가 다르게 나타남을 알 수 있다. 이와 같이 범주형 데이터에 대한 로지스틱 분석은 반응변수에 대한 범주의 연관성을 다양하게 비교할 수 있는 분석 방법으로써 환경, 사회현상 등에 다양한 독립변수의 범주를 고려할 수 있는 통계적 방법으로 이러한 분석을 통해 범주 연관성에 대한 정보를 얻을 수 있다. 다음 연구에서는 맛등급과 관련있는 연속형 변수들도 포함한 로지스틱 회귀모형에 대한 분석도 고려하고자한다.

## 참고문헌

- 박태성, 이승연 (1998). <범주형 자료분석 개론>, 자유아카데미, 서울.
- 정광모, 최용석 (2002). <범주형 자료분석 개론>, 자유아카데미, 서울.
- 조수현, 김진형, 김재희, 성필남, 박범영, 김경의, 서그라운달님, 이종문, 김동훈 (2007). 사회인구학적 요인이 한우 수소고기의 부위 및 요리형태별 관능특성에 미치는 영향. <한국동물자원과학회지>, **49**, 857-870.
- Agresti, A. (1996). *An introduction to categorical data analysis*, Wiley, New York.
- Allison, P. D. (1999). *Logistic regression using SAS*, SAS Institute Inc., New York.
- Brown, M. B. and Benedetti, J. K. (1977). On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika*, **42**, 347-355.
- Homer, D. W. and Lemeshow S. (2000). *Applied logistic regression*, Wiley, New York.
- Kritzer, H. M. (1977). Analyzing measures of association derived from contingency tables. *Sociological Methods & Research*, **5**, 387-418.
- Lee, J. Y. and Lee, Y. W. (2005). A major DNA marker mining of BMS941 microsatellite locus in Hanwoo chromosome. *Journal of the Korean Data & Information Science Society*, **16**, 913-921.
- Lee, J. Y., Lee, Y. W. and Kwon, J. C. (2006). DNA marker mining of BMS1167 microsatellite locus in Hanwoo chromosome. *Journal of the Korean Data & Information Science Society*, **17**, 311-324.



## Categorical data analysis of sensory evaluation data with Hanwoo bull beef

Hyejung Lee<sup>1</sup> · Soohyun Cho<sup>2</sup> · Jaehee Kim<sup>3</sup>

<sup>1</sup>Department of Statistics, Duksung Women's university

<sup>2</sup>Quality Control and Utilization of Animal Products Division, National Institute of Animal  
Science

Received 10 July 2009, revised 29 August 2009, accepted 9 September 2009

### Abstract

This study was conducted to investigate the relationship between the socio-demographic factors and the Korean consumers palatability evaluation grades with Hanwoo sensory evaluation data. The dichotomy logistic regression model and the multinomial logistic regression model are fitted with the independent variables such as the consumer living location, age, gender, occupation, monthly income, and beef cut and the the palatability grade as the dependent variable. Stepwise variable selection procedure is incorporated to find the final model and odds ratios are calculated to find the associations between categories.

*Keywords:* Deviance, Hanwoo sensory evaluation data, logistic regression, odds ratio.

---

<sup>1</sup> Graduate student, Department of Statistics, Duksung Women's university, Seoul 132-714, Korea.

<sup>2</sup> Researcher, Quality Control and Utilization of Animal Products Division, National Institute of Animal Science, Kyunggi 441-748, Korea.

<sup>3</sup> Corresponding author: Professor, Department of Statistics, Duksung Women's University, Seoul 132-714, Korea. E-mail: jaehee@duksung.ac.kr