

한우 거세우 고기 관능평가 데이터의 로지스틱 회귀분석

이혜정¹ · 김재희²

¹덕성여자대학교 정보통계학과, ²덕성여자대학교 정보통계학과

(2010년 5월 접수, 2010년 8월 채택)

요약

국립축산과학원에서는 2006년 부터 2008년 까지 전국 소비자들을 대상으로 한우 거세우 표본 시료에 대한 관능 평가 조사를 실시하여 데이터를 수집하였으며 본 연구에서는 한우 관능 평가 데이터에 대해 사회 인구학적 요인과 한국 소비자들의 맛 평가에 대한 연관성을 탐구하고자 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 실시한다. 단계별 변수 선택으로 최종 모형을 선택하고 반응변수 범주에 대한 오즈비를 계산하여 맛등급과 설명변수들 간의 관련성을 파악한다. 또한 맛과 관련 있는 연속형 변수를 설명변수로 포함한 경우에 대해서도 이항 다중 로지스틱 모형과 다항 다중 로지스틱 모형을 적합하고 비교한다. 그 결과 거주 지역, 연령, 월수입과 쇠고기 부위 변수들이 선택되었으며 영남지역에서 맛에 대한 오즈가 큰 편이며 수입이 많고 연령이 높을수록 맛에 대한 오즈가 작은 편이었다. 요리법으로는 탕에 대한 구이의 오즈비가 큰 편이며 쇠고기 부위별로는 우둔에 비해서 등심이 다른 부위들 보다 맛에 대한 차이가 크다고 볼 수 있다. 연속형 변수로는 연도가 맛등급에 큰 영향을 미치는 변수로 나타났다.

Keywords: 로지스틱 회귀, 오즈비, 이탈도, 카이제곱검정, 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 요리된 한우를 맛본 후 등급과 더불어 맛과 관련된 변수를 측정하는 대규모 관능평가를 실시하였으며 국민을 대표할 수 있는 인구학적인 특성을 반영하도록 소비자들을 대상으로 이와 같은 체계적인 조사는 우리나라 소비자들의 맛등급 판정 특성을 파악하는데 도움이 되며 이러한 조사에 기반 한 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.

본 연구는 2010년 덕성여자대학교 연구비 지원을 받았습니다.

²교신저자: (132-714) 서울시 도봉구 쌍문동 419, 덕성여자대학교 정보통계학과, 교수.

E-mail: jaehlee@duksung.ac.kr

2. 한우 관능평가 데이터에 대한 로지스틱 회귀분석

2.1. 시료준비 및 소비자 관능평가

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

본 연구에 사용된 데이터는 2006년도에서 2008년도에 축산과학원에서 시행한 관능평가 결과자료를 이용하였다. 한우시료는 축산과학원 한우시험장과 가락동 공판장에서 구입한 20~30개월령의 1++, 1+, 1등급 거세한우를 도축하여 각 두수에서 11개 부위(업진, 보섭, 채끝, 등심, 어깨삼각근, 흉두께, 목심, 설도, 우둔, 양지, 도가니)를 발골하여 소비자 관능평가 시료로 사용하였다. 각 부위별 시료들은 관능평가 session별로 구이(BBQ: 로스 구이), 그릴(grill: steak 요리), 탕(soup)의 3가지 방법으로 조리하여 소비자들에게 제공되었으며 각 소비자들은 동일한 조건으로 준비된 쇠고기 시료에 대하여 소비자가 직접 먹어보고 고기의 연도, 다즙성, 향미, 전반적인 기호도를 평가하였다. 관능평가는 1점의 쇠고기 시료에 대하여 연도, 다즙성, 향미 및 전반적인 기호도를 평가한 후 최종적으로 1~4등급으로 구분된 최종 만족도를 선택하도록 하였다. 소비자 맛등급(palatability grade)은 소비자들이 실제 쇠고기 시료를 먹은 후 육질을 평가하여 등급을 매긴 4개의 최종 만족도등급(1 = 만족하지 못한다, 2 = 만족한다, 3 = 매우 만족한다, 4 = 대단히 만족한다)으로 구분하였으며 등급값은 높을수록 만족도가 높음을 의미하였다.

소비자 관능평가는 쇠고기 시료에 관련된 요인이외에는 다른 요인에 의해서 영향을 받지 않도록 준비된 관능평가 protocol (Gee 등, 1998)에 의하여 시료준비, 제시순서 및 평가 방법 등 관련된 모든 공정을 세부화 하여 가능한 동일한 조건으로 진행할 수 있도록 하였다. 관능평가는 탕, 구이 및 그릴형태로 수행되었고 쇠고기 시료들은 각각의 조리방법에 따라 session별로 총 2,500여명의 소비자를 대상으로 수행되었으며 1명의 소비자가 총 7개의 시료를 평가하였다. 평가하는 모든 소비자들에게 동일한 평가기준을 주기 위하여 7개의 시료 중에서 첫 번째 시료를 표준시료(등심)로 제시하였고, 라틴방격법(Latin Square) 배열 방법에 의하여 나머지 6개의 시료들을 순서대로 제공하였다. 관능평가를 수행한 소비자는 층화추출방법을 사용하여 서울, 경기, 호남, 영남 지방의 4개 지역이 인구밀도와 비슷한 비율로 추출하였으며, 각 지역별 성비는 비슷하게 추출되도록 디자인 되었다. 관능평가에 사용된 데이터 변수는 연도(Tndr), 다즙성(Jcy), 향미(Flvr), 전반적인 기호도(OAll)에 대하여 각각 100mm로 준비된 선척도법을 이용하여 수집하였으며, 연도, 다즙성, 향미, 전반적인 기호도에 대한 항목적도는 다음과 같았다: 연도 = 매우 질기다(0)에서 매우 연하다(100), 다즙성 = 매우 건조하다(0)에서 매우 다즙하다(100), 향미 = 대단히 싫어한다(0)에서 대단히 좋아한다(100), 전반적인 기호도 = 대단히 싫어한다(0)에서 대단히 좋아한다(100).

우선 한우 맛 데이터의 범주 변수에 대한 기초 통계 분석으로 표 2.1을 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 등 (2007)은 이와 같은 데이터에 대해 사회인구학적 요인이 한우 수소고기의 부위 및 요리형태별 관능특성에 미치는 영향에 관한 연구 결과를 보여 준다.

2.2. 로지스틱 회귀모형

한우(Hanwoo)는 한국의 토종 축종으로서 호주의 앵거스(Angus), 일본의 와규(Wagyu)와 같이 그 나라

표 2.1. 소비자들의 인구사회학적 빈도와 맛등급과의 카이제곱 독립성검정

변수	값	변수명	맛등급				χ^2 통계량	p-value
			1	2	3	4		
요리법	구이	BBQ	1936	3433	2984	2220	773.27	< .0001
	스��이크	Grill	1519	1622	1173	686		
	탕	Soup	2836	3746	2451	1031		
거주지역	호남	Honam	1284	1858	1391	788	150.01	< .0001
	중부	Jungbu	2251	3025	2267	1305		
	서울	Seoul	1720	2103	1465	865		
	영남	Youngnam	1036	1815	1485	979		
성별	남	gender1	2174	2979	2238	1161	33.02	< .0001
	여	gender2	4101	5796	4348	2771		
연령	20-25세	age1	1090	1331	901	563	177.88	< .0001
	26-30세	age2	789	978	718	391		
	31-39세	age3	1869	2423	1676	1039		
	40-49세	age4	1676	2532	1948	1201		
	50-59세	age5	867	1537	1365	743		
직업군	전문직	occ1	212	300	196	116	333.78	< .0001
	행정관리직	occ2	874	1100	756	396		
	기술직	occ3	283	404	345	192		
	영업	occ4	1249	1646	1104	513		
	노동직	occ5	65	135	95	64		
	주부	occ6	1895	3057	2552	1722		
	무직	occ7	39	64	64	53		
	학생	occ8	1322	1572	1094	634		
	기타	occ9	356	510	391	246		
평균수입	100이하	income1	585	801	585	339	60.19	< .0001
	100-200	income2	1421	2122	1541	913		
	200-300	income3	1563	2374	1807	1113		
	300-400	income4	1289	1747	1326	773		
	400-500	income5	679	964	677	413		
	500이상	income6	708	715	609	361		
부위	업진	cut1	103	283	279	194	2793.19	< .0001
	보쌈	cut2	639	1057	941	670		
	채끝	cut3	439	863	800	532		
	등심	cut4	418	1533	1819	1278		
	꾸리	cut5	391	476	320	133		
	홍두깨	cut6	728	953	498	182		
	도가니	cut7	250	325	232	138		
	목심	cut8	566	569	288	173		
	설도	cut9	986	1081	635	368		
	양지	cut10	415	549	294	123		
	우둔	cut11	1357	1112	502	146		
쇠고기 먹는 횟수	매일	eathabit 1	18	16	13	6	76.10	< .0001
	주4-5회	eathabit 2	70	125	75	31		
	주2-3회	eathabit 3	685	900	697	497		
	주1회	eathabit 4	1452	2166	1715	1027		
	2주한번	eathabit 5	1596	2244	1737	1037		
	월1회	eathabit 6	2415	3277	2313	1328		
	안먹음	eathabit 7	44	46	33	7		

의 대표 축종이다. 각 나라마다 고유 축종에 대한 연구와 개발은 축산연구의 중요 관심이 되고 있다. 이 절에서는 이산형 변수인 한우 맛등급을 반응 변수로 하여 한우 맛과 관련한 설명변수들과의 관계를 범주형 자료 분석 기법을 통한 통계적 모형 연구를 하고자한다. 소비자가 직접 부여한 맛등급(반응 변수)과

소비자의 특성을 나타내는 변수로 ‘지역’, ‘연령’, ‘직업군’, ‘수입’과 ‘쇠고기 먹는 회수’ 등을 고려하고 한우의 맛과 관련한 변수로는 ‘요리법’과 ‘부위’를 고려하고자 한다. 이와 같은 변수들을 고려한 이유로는 거주지역의 음식문화, 나이에 따른 음식 경험, 쇠고기 먹는 회수에 따른 고기맛 평가에 대한 차별화, 수입에 따른 음식문화 차이 등을 생각할 수 있다. 또한 한우 고기 맛에 가장 큰 영향을 주는 쇠고기 부위와 같은 부위라고 요리법에 따라 맛이 차이남을 고려하여 ‘요리법’과 ‘부위’ 변수를 고려하기로 한다.

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 박태성과 이승연 (1998), 정광모와 최용석 (2002)과 Agresti (1996)를 참조한다. Homer와 Lemeshow (2000)에는 로지스틱 모형의 분석 예에 대한 설명이 잘 정리되어있다.

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. 반응 변수를 한우 요리를 맛본 후 반응변수로 만족한다 ($Y = 1$), 만족하지 못한다 ($Y = 0$)라고 정의한다. 여기서 ($Y = 1$)은 맛등급이 2, 3, 4인 경우를 포함하며 ($Y = 0$)은 맛등급이 1인 경우만을 의미한다. 종속변수가 다항 반응변수인 경우는 $Y = 1, 2, 3, 4$ 를 가지며 만족하지 못한다 $Y = 1$ 를 기준 범주로 정한다.

소비자들의 맛등급에 영향을 주는 이산형 설명 변수로 성별, 연령, 거주지역, 직업군, 수입과 쇠고기 먹는 회수 등을 고려하고자한다. 또한 쇠고기의 맛과 맛등급은 연도, 다즙성, 향미와 관련이 많다고 알려져 있다 (Thompson, 2002). 여기서 연도, 다즙성, 향미는 맛등급과 같이 소비자들이 한우 쇠고기 맛을 평가한 후 직접 부여한 값으로 맛등급과 관련성이 높은 변수들이다. 그러므로 이러한 연속형 설명 변수를 포함한 로지스틱 모형을 고려하여 다음과 같은 4개의 모형을 고려해 보고자 한다.

[모형 1]은 소비자들이 부여한 맛등급에 관해 이항 반응 변수(만족한다/만족하지 못한다)에 대해 소비자들의 사회인구학적 상황과 요리법과 부위 등의 이산형 변수들만 고려한 로지스틱 모형이다. 이러한 경우의 설명변수들은 고정된 상황을 반영하는 변수들로 예를 들어 성별, 나이, 직업 등과 같이 소비자의 상황을 나타내는 변수들만의 적합이다. [모형 2]는 [모형 1]에서 고려된 이산형 변수와 연도, 다즙성, 향미의 연속형 변수를 설명변수로 포함한 로지스틱 모형이다. [모형 3]은 소비자들이 부여한 4개의 맛등급을 반응변수로한 다항 로지스틱 회귀모형으로 설명변수가 이산형 변수인 경우만을 포함한 모형이고 [모형 4]는 다항 반응변수와 [모형 3]에서 포함된 이산형 설명 변수와 연속형 설명 변수를 포함한 모형이다.

[모형 1] 이산형 설명변수만을 고려한 이항 로지스틱 모형

$$\text{logit} [\pi(x)] = \log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_8 x_8, \quad (2.1)$$

여기서 $x_1 =$ 요리법, $x_2 =$ 거주지역, $x_3 =$ 성별, $x_4 =$ 연령대, $x_5 =$ 직업군, $x_6 =$ 평균수입, $x_7 =$ 쇠고기 부위와 $x_8 =$ 쇠고기 먹는 횟수이다. 식 (2.1)을 $\pi(x)$ 로 다시 표현하면

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_8 x_8)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_8 x_8)} \quad (2.2)$$

이 되어 반응확률 $\pi(x)$ 를 추정할 수 있다. 위의 설명변수들로 로지스틱 회귀모형을 적합한 후에 쇠고기 맛에 대해 ‘만족한다’로 판정할 확률을 식 (2.2)를 이용하여 추정할 수 있다. ‘(맛에 대해) 만족한다’일 오즈는

$$\frac{\pi(x)}{1 - \pi(x)} = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_8 x_8) \quad (2.3)$$

표 2.2. 로지스틱 모형 비교 (변수 선택 유의수준 0.002)

	모형 1	모형 2	모형 3	모형 4
반응 변수	이항변수 0 = 만족하지못한다(기준), 1 = 만족한다.		다항변수 1 = 만족하지못한다(기준), 2 = 만족, 3 = 매우 만족, 4 = 대단히 만족	
선택된 설명변수	부위, 요리법, 지역, 연령, 수입, 고기먹는 회수	부위, 요리법, 지역, 연령, 수입, 연도, 다즙성, 향미	부위, 요리법, 지역, 연령, 수입, 성별, 고기먹는 회수	부위, 요리법, 지역, 연령, 수입, 연도, 다즙성, 향미
모형적합도 이탈도(p-값)	1.1550 (< .0001)	0.5560 (1.000)	0.9781 (0.9978)	0.5285 (1.000)
AIC	25502.643	14077.831	63959.295	40193.083
R ²	0.1065	0.4297	0.1655	0.6685
특성있는 오즈비				
요리법	BBQ vs Soup: 2.137	BBQ vs Soup: 0.956	BBQ vs Soup(4): 4.976	BBQ vs Soup(4): 1.590
부위	등심 vs 우둔: 9.645	등심 vs 우둔: 2.432	등심 vs 우둔(4): 36.802 업진 vs 우둔 (4): 44.268	등심 vs 우둔(4): 5.023 업진 vs 우둔(4): 7.398
지역	서울 vs 영남: 0.592	서울 vs 영남: 0.707	서울 vs 영남(3): 0.533	서울 vs 영남(4): 0.538
연령	20대 vs 50대: 0.556	20대 vs 50대: 0.808	20대 vs 50대(3): 0.492	20대 vs 50대(4): 2.292
수입	평균250 vs 500이상: 1.503	100이하 vs 500이상: 1.588	250 vs 500이상(4): 1.553	250 vs 500이상(2): 1.443
고기먹는 회수	(주1회) vs (안먹음): 1.471		(주2-3회) vs (안먹음)(4): 3.793	

와 같이 구할 수 있다. 단계적 변수선택(stepwise selection method)을 이용하면 표 2.2에서 보여주는 5개의 변수가 선택되었다. 모수추정값과 오즈비(표 생략)를 살펴보면, 탕에 비해 구이의 요리법이 만족도를 높게 줄 오즈가 2배 가량 높았으며 소득이 가장 높은 층을 기준으로 다른 소득층들이 만족도를 높게 줄 오즈가 모두 높았다. 영남지역에 비해 서울지역에서 만족도를 높게 줄 오즈가 0.59로 서울지역 소비자들의 만족도가 영남지역에 비해 낮은 편임을 알 수 있다. 부위에서는 업진 부위가 우둔에 비해 맛등급을 높게 줄 오즈가 7.9배, 등심은 9.6배가 높은 것이 특징이다.

[모형 2] 이산형 변수와 연속형 변수의 이항 다중 로지스틱 회귀모형

$$\text{logit} [\pi(x)] = \log \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1x_1 + \dots + \beta_8x_8 + \beta_9\text{tndr} + \beta_{10}\text{jcy} + \beta_{11}\text{flvr}, \quad (2.4)$$

여기서 이산형 변수는 (2.1) 모형에서 사용된 변수들이며 연속형 변수는 tndr = 연도, jcy = 다즙성, flvr = 향미이다. 식 (2.4)를 $\pi(x)$ 로 다시 표현하면

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1x_1 + \dots + \beta_8x_8 + \beta_9\text{tndr} + \beta_{10}\text{jcy} + \beta_{11}\text{flvr})}{1 + \exp(\beta_0 + \beta_1x_1 + \dots + \beta_8x_8 + \beta_9\text{tndr} + \beta_{10}\text{jcy} + \beta_{11}\text{flvr})} \quad (2.5)$$

이 되어 반응확률 $\pi(x)$ 를 추정할 수 있다.

모형 (2.1)에 ‘연도’, ‘다즙성’, ‘향미’ 3개의 연속형 변수들을 추가하여 이항 다중 로지스틱 모형 모형 (2.4)를 적합한 결과 모형적합도가 [모형 1] 보다 높다. 이는 연속형 맛 변수인 ‘연도’, ‘다즙성’, ‘향미’가 쇠고기 맛에 대한 만족도에 대한 설명력이 높기 때문이라고 볼 수 있다. 단계별 변수선택 결과는 표 2.2에 있으며 모수 추정 결과는 표 2.3에서 보여준다. 오즈비 결과 표 2.4를 살펴 보면 ‘거주지역’에서 ‘영남’ 지역을 기준으로 ‘호남’, ‘중부’, ‘서울’ 지역의 맛등급을 높게 줄 오즈가 모두 낮았으며 50대의 연령층에 비해 20대, 30대, 40대 연령층에서 만족도를 높게 줄 오즈가 더 낮았다. ‘수입’에서는 500만원 이상의 소득층에 비해 다른 소득층들이 만족도를 높게 줄 오즈가 더 높으므로 소득이 가장 높은 층이 만족도를 낮게 주는 경향이 있다고 할 수 있다. ‘부위’에서는 ‘우둔’을 기준으로 다른 부위가 모두

표 2.3. 이산형 변수와 연속형 변수의 이항 다중 로지스틱 최종모형

변수		모수추정값	ASE	검정통계량	p-value	모형 적합도
Intercept		-5.9301	0.1459	1651.0560	< .0001	deviance
Cooking	BBQ	-0.0462	0.0549	0.7085	0.4000	
Cooking	Grill	-0.2515	0.0625	16.2133	< .0001	
Cooking	Soup	0	.	.	.	
loc2	호남(H)	-0.0409	0.0713	0.3295	0.5659	
loc2	충부(J)	-0.2695	0.0641	17.6906	< .0001	
loc2	서울(S)	-0.3366	0.0678	24.6654	< .0001	
loc2	영남(Y)	0	.	.	.	
age	20-25세(1)	-0.2067	0.0788	6.8859	0.0087	Value/DF
age	26-30세(2)	-0.1942	0.0845	5.2868	0.0215	= 0.5560
age	31-39세(3)	-0.2136	0.0706	9.1584	0.0025	
age	40-49세(4)	-0.3028	0.0698	18.8288	< .0001	p-value
age	50-59세(5)	0	.	.	.	= 1.000
income	100이하(1)	0.4852	0.1007	23.2374	< .0001	
income	100-200(2)	0.4305	0.0822	27.4490	< .0001	
income	200-300(3)	0.3487	0.0810	18.5438	< .0001	
income	300-400(4)	0.2970	0.0834	12.6852	0.0004	
income	400-500(5)	0.3963	0.0939	17.8328	< .0001	
income	500이상(6)	0	.	.	.	
cut	업진(A)	0.7074	0.1538	21.1683	< .0001	
cut	보섭(B)	0.4643	0.0831	31.2008	< .0001	
cut	채끝(C)	0.5802	0.0902	41.3650	< .0001	
cut	등심(D)	0.8821	0.0864	104.3270	< .0001	
cut	꾸리(G)	0.4164	0.1023	16.5869	< .0001	
cut	홍두깨(H)	0.4203	0.0819	26.3385	< .0001	
cut	도가니(K)	0.3651	0.1203	9.2052	0.0024	
cut	목심(M)	0.2013	0.0919	4.7987	0.0285	
cut	설도(S)	0.2769	0.0766	13.0540	0.0003	
cut	양지(Y)	0.4763	0.1045	20.7855	< .0001	
cut	우둔(U)	0	.	.	.	
Tndr		0.0617	0.00134	2130.8858	< .0001	
Jcy		0.0288	0.00144	401.1194	< .0001	
Flvr		0.0323	0.00136	568.4475	< .0001	

만족도를 높게 줄 오즈가 더 높았지만 ‘업진’과 ‘등심’이 약 2배이상으로 높은 편으로 나타났다. 부위에 대한 오즈비를 살펴 보면 [모형 1]에서는 우둔에 대한 등심의 오즈비가 약 9배인데 반해 [모형 2]에서는 약 2배로 줄어든 이유는 만족도와 관련이 높은 연도, 다즙성, 향미 부분에서 설명하는 부분이 커졌기 때문으로 여겨진다.

[모형 3] 이산형 변수들만의 다항 다중 로지스틱 회귀모형

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \beta_{0j} + \beta_{1j}x_1 + \cdots + \beta_{8j}x_8, \quad j = 1, \dots, J-1 \quad (2.6)$$

한우 쇠고기 맛 맛등급은 4개의 범주 $J = 4$ 를 가지며 범주의 반응확률을 $\{\pi_1, \dots, \pi_4\}$ 라고 하면, $\sum_j \pi_j = 1$ 이 된다. n 개의 서로 독립인 관측값들이 이러한 확률에서 얻어졌을 때, $J = 4$ 개의 가능한 범

표 2.4. 이산형 변수와 연속형 변수의 이항 다중 로지스틱 최종모형 오즈비 결과

Odds Ratio Estimates		
변수		오즈비
Cooking	BBQ vs Soup	0.956
Cooking	Grill vs Soup	0.782
loc2	Honam vs Youngnam	0.955
loc2	Jungbu vs Youngnam	0.758
loc2	Seoul vs Youngnam	0.707
age	1 vs 5	0.808
age	2 vs 5	0.828
age	3 vs 5	0.800
age	4 vs 5	0.733
income	1 vs 6	1.588
income	2 vs 6	1.524
income	3 vs 6	1.413
income	4 vs 6	1.332
income	5 vs 6	1.472
cut	A vs U	2.062
cut	B vs U	1.598
cut	C vs U	1.791
cut	D vs U	2.432
cut	G vs U	1.526
cut	H vs U	1.529
cut	K vs U	1.427
cut	M vs U	1.235
cut	S vs U	1.342
cut	Y vs U	1.608
Tndr		1.064
Jcy		1.029
Flvr		1.033

주 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 회귀모형 [모형 3]을 정의할 수 있다. 마지막 범주가 기준이 될 때, 기준 범주에 대한 로짓은 다음과 같다. 즉, 한우 맛등급인 경우 $J - 1 = 3$ 개의 로짓방정식으로 이루어지며 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.

$$\begin{aligned} \log\left(\frac{\pi_a}{\pi_b}\right) &= \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right) = \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right) \\ &= (\beta_{0a} - \beta_{0b}) + (\beta_{1a} - \beta_{1b})x_1 + \cdots + (\beta_{8a} - \beta_{8b})x_8. \end{aligned}$$

위의 $J - 1 = 3$ 개의 로짓방정식을 동시에 적합시켜 모수를 구할수 있으며 동시에 적합시키면 어떤 기준범주가 되든 상관없이 범주들의 쌍에 대한 모수들의 추정값을 얻게 된다. 이와 같은 모형은 Allison (1999)와 Homer와 Lemeshow (2000)을 참조한다. 다항 다중 로지스틱 모형에서는 반응변수가 마지막 변수를 기준으로 각 변수와의 비를 나타내기 때문에 오즈비가 한 범주당 $J - 1 = 3$ 개가 나온다. 단계

별 변수선택 결과는 표 2.2에서 보여준다. 오즈비 결과(표 생략)를 살펴 보면 탕에 비해 구이의 요리법이 만족도를 높게 줄 오즈가 4개 맛등급 그룹에서 높았으며 특히 맛등급 4인 경우에 오즈비가 5배 정도이다. ‘거주지역’에서 ‘영남’ 지역을 기준으로 ‘호남’, ‘중부’, ‘서울’ 지역의 맛등급을 높게 줄 오즈가 모두 낮았으며 50대의 연령층에 비해 20대, 30대, 40대 연령층에서 맛등급을 높게 줄 오즈가 더 낮았다. ‘수입’에서는 500만원 이상의 소득층에 비해 다른 소득층들이 맛등급을 높게 줄 오즈가 더 높으므로 소득이 가장 높은 층이 맛등급을 낮게 주는 경향이 있다고 할 수 있다. ‘부위’에서는 우둔을 기준으로 다른 부위가 모두 맛등급을 높게 줄 오즈가 더 높았지만 맛등급 4인 경우 우둔에 비한 업진 오즈비가 44.2이고 우둔에 비한 등심 오즈비가 36.8로 매우 높게 나타났다. 즉 맛등급이 높은 그룹에서 부위에 대한 오즈비 값이 크게 나타남을 알 수 있다. ‘쇠고기 먹는 횟수’에서도 4등급 그룹에서 가장 큰 차이를 보이지만 다른 등급 대부분에서도 ‘쇠고기를 안먹는’ 범주를 기준으로 다른 범주들에서 맛등급을 높게 줄 오즈가 더 높게 나타났다.

[모형 4] 이산형 변수와 연속형 변수의 다항 다중 로지스틱 회귀모형

$$\log\left(\frac{\pi_j}{\pi_J}\right) = \beta_{0j} + \beta_{1j}x_1 + \cdots + \beta_{8j}x_8 + \beta_9\text{tndr} + \beta_{10}\text{jcy} + \beta_{11}\text{flvr}, \quad j = 1, \dots, J-1. \quad (2.7)$$

다항 다중 로지스틱 회귀모형에서도 ‘연도’, ‘다즙성’, ‘향미’의 연속형 변수와 ‘거주지역’, ‘성별’, ‘연령’, ‘월수입’, ‘부위’, ‘요리법’, ‘쇠고기 먹는 횟수’의 이산형 변수들을 고려하여 다음의 다항 다중 로지스틱 회귀모형[모형 4]을 적합한 후 단계별 변수선택을 한 결과 표 2.2에서 보여준다. 다른 두 범주간의 로짓은 다음과 같이 나타낼 수 있다:

$$\begin{aligned} \log\left(\frac{\pi_a}{\pi_b}\right) &= \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right) = \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right) \\ &= (\beta_{0a} - \beta_{0b}) + (\beta_{1a} - \beta_{1b})x_1 + \cdots + (\beta_{8a} - \beta_{8b})x_8 \\ &\quad + \beta_9(\text{tndr}_a - \text{tndr}_b) + \beta_{10}(\text{jcy}_a - \text{jcy}_b) + \beta_{11}(\text{flvr}_a - \text{flvr}_b). \end{aligned}$$

단계별 변수선택 결과는 표 2.2에 있으며 모수 추정 결과는 표 3.5에서 보여준다. 오즈비 결과 표 3.6을 살펴 보면 ‘거주지역’에서 ‘영남’ 지역을 기준으로 ‘호남’, ‘중부’, ‘서울’ 지역의 맛등급을 높게 줄 오즈비가 모두 낮았으며 50대의 연령층에 비해 20대, 30대, 40대 연령층에서 맛등급을 높게 줄 오즈가 더 낮았다. ‘수입’에서는 500만원 이상의 소득층에 비해 다른 소득층들이 만족도를 높게 줄 오즈가 더 높으므로 소득이 가장 높은 층이 맛등급을 낮게 주는 경향이 있다고 할 수 있다. ‘부위’에서는 ‘우둔’을 기준으로 다른 부위가 모두 맛등급을 높게 줄 오즈가 더 높았지만 ‘업진’과 ‘등심’이 약 2배 이상으로 높은 편으로 나타났다. 부위에 대한 오즈비를 살펴 보면 [모형 3]에서는 우둔에 대한 등심의 오즈비가 약 37배 인데 반해 [모형 4]에서는 약 5배로 줄어든 이유는 맛등급과 관련이 높은 연도, 다즙성, 향미 부분에서 설명하는 부분이 커졌기 때문으로 여겨진다.

3. 결론

거세우에 대해 실시한 한우 관능평가 데이터에서 맛등급을 반응변수로 하고 거주 지역, 소비자 성별, 소비자 연령, 직업, 월수입, 쇠고기 부위, 쇠고기 요리법, 쇠고기 먹는 횟수 등의 이산형 설명변수와 연도, 다즙성, 향미 등의 연속형 설명변수를 고려하여 4가지 로지스틱 회귀모형을 적합하였다. 이 중 ‘직업’ 변수는 고려되지 않은 직업군이 있는 관계로 통계적 유의성에 상관없이 제외시키고 모형을 적합하였으며 4개 모형에 대한 요약은 표 2.2에 있다. 관측 케이스가 많고 변수들의 유의수준이 높아 신뢰수준(유의수준 0.002)을 높여 변수 선택을 하였다.

표 2.5. 이산형 변수와 연속형 변수의 다항 다중 로지스틱 최종모형

변수		grade	df	모수추정값	ASE	검정통계량	p-value
Intercept		4	1	-28.3380	0.3918	5232.3547	< .0001
Intercept		3	1	-13.8148	0.2312	3569.8650	< .0001
Intercept		2	1	-5.2417	0.1480	1253.6740	< .0001
Cooking	BBQ	4	1	0.4635	0.0886	27.3893	< .0001
Cooking	BBQ	3	1	-0.0176	0.0699	0.0632	0.8014
Cooking	BBQ	2	1	-0.0522	0.0549	0.9066	0.3410
Cooking	Grill	4	1	0.2190	0.1074	4.1609	0.0414
Cooking	Grill	3	1	-0.2024	0.0821	6.0857	0.0136
Cooking	Grill	2	1	-0.2544	0.0625	16.5586	< .0001
Cooking	Soup	4	0	0	Z	Z	Z
Cooking	Soup	3	0	0	Z	Z	Z
Cooking	Soup	2	0	0	Z	Z	Z
loc2	Honam	4	1	-0.2083	0.1128	3.4086	0.0649
loc2	Honam	3	1	-0.0773	0.0911	0.7192	0.3964
loc2	Honam	2	1	-0.0562	0.0717	0.6149	0.4330
loc2	Jungbu	4	1	-0.6348	0.0992	40.9575	< .0001
loc2	Jungbu	3	1	-0.3888	0.0809	23.0754	< .0001
loc2	Jungbu	2	1	-0.2594	0.0642	16.3057	< .0001
loc2	Seoul	4	1	-0.6201	0.1067	33.7890	< .0001
loc2	Seoul	3	1	-0.4900	0.0862	32.3238	< .0001
loc2	Seoul	2	1	-0.3178	0.0679	21.9143	< .0001
age	1	4	1	0.8294	0.1273	42.4488	< .0001
age	1	3	1	0.0053	0.1012	0.0027	0.9583
age	1	2	1	-0.2389	0.0791	9.1278	0.0025
age	2	4	1	0.4507	0.1376	10.7241	0.0011
age	2	3	1	-0.0272	0.1091	0.0620	0.8034
age	2	2	1	-0.2350	0.0854	7.5677	0.0059
age	3	4	1	0.2532	0.1102	5.2810	0.0216
age	3	3	1	-0.1908	0.0893	4.5633	0.0327
age	3	2	1	-0.2336	0.0705	10.9755	0.0009
age	4	4	1	-0.1219	0.1057	1.3302	0.2488
age	4	3	1	-0.3726	0.0869	18.3783	< .0001
age	4	2	1	-0.2886	0.0696	17.1825	< .0001
age	5	4	0	0	Z	Z	Z
age	5	3	0	0	Z	Z	Z
age	5	2	1	0.4249	0.0763	31.0370	< .0001
gender	1	3	1	0.3206	0.0598	28.7285	< .0001
gender	1	2	1	0.0838	0.0463	3.2766	0.0703
income	1	4	1	0.0484	0.1650	0.0860	0.7693
income	1	3	1	0.2655	0.1310	4.1045	0.0428
income	1	2	1	0.4828	0.1011	22.7884	< .0001
income	2	4	1	0.1876	0.1350	1.9306	0.1647
income	2	3	1	0.2842	0.1070	7.0525	0.0079
income	2	2	1	0.4392	0.0831	27.9213	< .0001
income	3	4	1	0.1832	0.1324	1.9150	0.1664
income	3	3	1	0.2176	0.1053	4.2722	0.0387
income	3	2	1	0.3670	0.0820	20.0393	< .0001
income	4	4	1	0.2333	0.1367	2.9122	0.0879
income	4	3	1	0.2227	0.1084	4.2187	0.0400
income	4	2	1	0.3023	0.0844	12.8240	0.0003
income	5	4	1	0.1774	0.1536	1.3335	0.2482
income	5	3	1	0.2195	0.1222	3.2268	0.0724
income	5	2	1	0.4095	0.0944	18.8124	< .0001

계속

cut	A	4	1	2.0012	0.2361	71.8670	< .0001
cut	A	3	1	1.1301	0.1883	36.0301	< .0001
cut	A	2	1	0.6420	0.1549	17.1788	< .0001
cut	B	4	1	1.2762	0.1567	66.3234	< .0001
cut	B	3	1	0.7877	0.1114	49.9947	< .0001
cut	B	2	1	0.4076	0.0833	23.9425	< .0001
cut	C	4	1	1.3080	0.1641	63.5655	< .0001
cut	C	3	1	0.9069	0.1186	58.4424	< .0001
cut	C	2	1	0.5257	0.0905	33.7487	< .0001
cut	D	4	1	1.6141	0.1521	112.6483	< .0001
cut	D	3	1	1.2347	0.1104	125.1800	< .0001
cut	D	2	1	0.8108	0.0868	87.3402	< .0001
cut	G	4	1	0.9603	0.2008	22.8760	< .0001
cut	G	3	1	0.7437	0.1395	28.4226	< .0001
cut	G	2	1	0.3812	0.1025	13.8331	0.0002
cut	H	4	1	0.7487	0.1766	17.9813	< .0001
cut	H	3	1	0.5903	0.1156	26.0636	< .0001
cut	H	2	1	0.4074	0.0814	25.0431	< .0001
cut	K	4	1	1.1194	0.2166	26.7175	< .0001
cut	K	3	1	0.6003	0.1610	13.9088	0.0002
cut	K	2	1	0.3285	0.1206	7.4169	0.0065
cut	M	4	1	0.9663	0.1916	25.4375	< .0001
cut	M	3	1	0.3700	0.1329	7.7460	0.0054
cut	M	2	1	0.1931	0.0918	4.4273	0.0354
cut	S	4	1	1.0327	0.1613	40.9777	< .0001
cut	S	3	1	0.5279	0.1094	23.2718	< .0001
cut	S	2	1	0.2645	0.0765	11.9545	0.0005
cut	Y	4	1	1.2692	0.2117	35.9461	< .0001
cut	Y	3	1	0.6368	0.1450	19.2864	< .0001
cut	Y	2	1	0.4377	0.1032	18.0020	< .0001
Tndr		4	1	0.1725	0.00334	2668.9084	< .0001
Tndr		3	1	0.1060	0.00200	2816.6856	< .0001
Tndr		2	1	0.0536	0.00138	1503.3319	< .0001
Jcy		4	1	0.1039	0.00343	920.3538	< .0001
Jcy		3	1	0.0541	0.00208	677.6890	< .0001
Jcy		2	1	0.0246	0.00145	286.2080	< .0001
Flvr		4	1	0.0999	0.00287	1212.3059	< .0001
Flvr		3	1	0.0609	0.00192	1000.6168	< .0001
Flvr		2	1	0.0277	0.00137	410.3713	< .0001

이항 다중 로지스틱 모형과 다항 다중 로지스틱 모형에서 모두 이산형 변수만 고려한 모형보다 연속형 변수와 이산형 변수가 고려된 모형이 더 모형적합도가 높았는데 특히 포함된 연속형 변수가 맛등급과 상관성이 높기 때문이다. 그러나 설명 변수가 추가되면 모형적합도는 당연히 높아지므로 모수 절약의 원칙을 고려한 적절한 모형 선택이 필요하므로 AIC 값을 고려하여도 연속형 변수들이 포함된 모형이 선택된다. 즉 포함된 연속형 변수는 모형의 설명력을 증가시키는 주요 설명변수로 기여함을 알 수 있다. 각 모형에서의 오즈비를 살펴보면, 대체적으로 ‘영남’ 지역에 비해 다른 지역들이 맛등급을 높게 줄 오즈가 낮았으며 50대 연령층을 기준으로 20, 30대 연령층이 맛등급을 높게 줄 오즈가 더 낮았다. 소득이 가장 높은 그룹이 맛등급을 낮게 주는 경향이 있으며 ‘우둔’에 비해 다른 모든 부위가 더 맛있는 경향이 나타났다. 특히 ‘업진’과 ‘등심’ 부위가 ‘우둔’에 비해 맛등급 오즈의 차이가 가장 크게 나타났다. 본 연구에서의 분석과 같이 이산 반응 변수에 대해 적절한 로지스틱 모형적합을 통해 모형을 찾아 설명하고 해석하므로써 한우 소비자들에 대한 의미있는 정보를 얻을 수 있는데 이번 통계 분석에 대한 의의가 있다고 할 수 있다.

표 2.6. 이산형 변수와 연속형 변수의 다항 다중 로지스틱 회귀분석 최종모형 오즈비 결과

Odds Ratio Estimates			
변수		맞등급	오즈비
Cooking	BBQ vs Soup	4	1.590
Cooking	BBQ vs Soup	3	0.983
Cooking	BBQ vs Soup	2	0.949
Cooking	Grill vs Soup	4	1.245
Cooking	Grill vs Soup	3	0.817
Cooking	Grill vs Soup	2	0.775
loc2	Honam vs Youngnam	4	0.812
loc2	Honam vs Youngnam	3	0.926
loc2	Honam vs Youngnam	2	0.945
loc2	Jungbu vs Youngnam	4	0.530
loc2	Jungbu vs Youngnam	3	0.678
loc2	Jungbu vs Youngnam	2	0.772
loc2	Seoul vs Youngnam	4	0.538
loc2	Seoul vs Youngnam	3	0.613
loc2	Seoul vs Youngnam	2	0.728
age	1 vs 5	4	2.292
age	1 vs 5	3	1.005
age	1 vs 5	2	0.787
age	2 vs 5	4	1.569
age	2 vs 5	3	0.973
age	2 vs 5	2	0.791
age	3 vs 5	4	1.288
age	3 vs 5	3	0.826
age	3 vs 5	2	0.792
age	4 vs 5	4	0.885
age	4 vs 5	3	0.689
age	4 vs 5	2	0.749
gender	1 vs 2	4	1.529
gender	1 vs 2	3	1.378
gender	1 vs 2	2	1.087
income	1 vs 6	4	1.050
income	1 vs 6	3	1.304
income	1 vs 6	2	1.621
income	2 vs 6	4	1.206
income	2 vs 6	3	1.329
income	2 vs 6	2	1.551
income	3 vs 6	4	1.201
income	3 vs 6	3	1.243
income	3 vs 6	2	1.443
income	4 vs 6	4	1.263
income	4 vs 6	3	1.249
income	4 vs 6	2	1.353
income	5 vs 6	4	1.194
income	5 vs 6	3	1.245
income	5 vs 6	2	1.506
cut	A vs U	4	7.398
cut	A vs U	3	3.096
cut	A vs U	2	1.900
cut	B vs U	4	3.583
cut	B vs U	3	2.198
cut	B vs U	2	1.503
cut	C vs U	4	3.699

계속

cut	C vs U	3	2.477
cut	C vs U	2	1.692
cut	D vs U	4	5.023
cut	D vs U	3	3.438
cut	D vs U	2	2.265
cut	G vs U	4	2.612
cut	G vs U	3	2.104
cut	G vs U	2	1.464
cut	H vs U	4	2.114
cut	H vs U	3	1.805
cut	H vs U	2	1.503
cut	K vs U	4	3.063
cut	K vs U	3	1.823
cut	K vs U	2	1.389
cut	M vs U	4	2.628
cut	M vs U	3	1.448
cut	M vs U	2	1.213
cut	S vs U	4	2.809
cut	S vs U	3	1.695
cut	S vs U	2	1.303
cut	Y vs U	4	3.558
cut	Y vs U	3	1.890
cut	Y vs U	2	1.549
Tndr		4	1.188
Tndr		3	1.112
Tndr		2	1.055
Jcy		4	1.110
Jcy		3	1.056
Jcy		2	1.025
Flvr		4	1.105
Flvr		3	1.063
Flvr		2	1.028

감사의 글

본 연구에 귀중한 한우 관련 데이터를 제공해 주신 국립 축산 과학원 안중남 과장님, 김동훈 과장님과 조수현 박사님께 감사드립니다.

참고문헌

- 박태성, 이승연 (1998). <범주형 자료분석 개론>, 자유아카데미, 서울.
- 정광모, 최용석 (2002). <범주형 자료분석 개론>, 자유아카데미, 서울.
- 조수현, 김진형, 김재희, 성필남, 박범영, 김경의, 서그러운달님, 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.
- Gee, A., Porter, M., Coffey, D. and Polkinghorne, R. (1998). *Design and Protocol for Steak Grilling Trials*, MSA, Sydney.
- 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 17, *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 17, *Journal of the Korean Data & Information Science Society*, **17**, 311-324.
- Thompson, J. (2002). Managing meat tenderness, *Meat Science*, **62**, 295-308.

Logistic Regressions with Sensory Evaluation Data about Hanwoo Steer Beef

Hyejung Lee¹ · Jaehee Kim²

¹Department of Statistics, Duksung Women's University

²Department of Statistics, Duksung Women's University

(Received May 2010; accepted August 2010)

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 from 2006 to 2008 by National Institute of Animal Science. 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, beef cut and the the palatability grade as the categorical dependent variable and tenderness, flavor and juiciness as the continuous 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.

This work was supported by a grant (2009) from Duksung Women's University.

²Corresponding author: Professor, Department of Statistics, Duksung Women's University, 419Ssangmun-Dong, Dobong-Gu, Seoul 132-714, Korea. E-mail: jaehee@duksung.ac.kr