

한우 맛 등급 판별방법 비교 연구

김재희¹ · 서그러운달님²

¹덕성여자대학교 정보통계학과, ²덕성여자대학교 정보통계학과

(2008년 7월 접수, 2008년 9월 채택)

요약

본 연구에서는 한우의 연도, 다즙성, 향미, 전반적인 기호도에 관한 소비자 관능평가 데이터를 기반으로 소비자 맛 등급을 판별하는 방법을 비교 연구하였다. 판별방법으로는 선형판별, 이차판별, 정준판별, 비모수적 판별방법을 사용하였고, 재대입 방법에 의한 분류율을 측정하여 판별방법들의 분류능력을 비교하였다.

주요용어: 맛 등급, 판별분석, 오분류율.

1. 한우 맛 등급 판별의 필요성

현재 한우소비량은 타 육류(돼지고기, 닭고기)에 비하여 상대적으로 감소하는 추세에 있으며, 외국산 소고기에 비해 상대적으로 가격이 비싼 한우의 경우 앞으로도 지속적인 소비의 감소세가 예상되고 있다. 본 연구에서는 한우고기에 대한 등급평가를 기존의 도체의 질과, 양 중심으로 한 마리의 소는 하나의 등급을 가지는 방식을 보완하여 한우고기의 연도(Tenderness), 다즙성(Juiciness), 향미(Flavor) 세 변수와 전반적인 맛을 평가하는 전반적인 기호도(Overall acceptability)를 중심으로 한우 맛 등급(grade) 평가방법을 개발하고자 한다. 이미 호주와 미국의 경우 Thompson (2002) 등에 의해 각 근육별로 요리방법별 소고기의 등급을 소비자의 맛 평가 자료에 기반을 둔 자료가 사용되고 있으며, 우리나라 또한 축산과학원에서 실제 소비자를 대상으로 한 관능평가 자료를 이용하여 기존의 등급 시스템을 보완하려는 연구가 진행 중이다.

우리나라의 소고기 등급 판정은 크게 육량등급과 육질등급으로 나누어 판정을 내린다. 육량등급은 육량 지수에 의해 판정되며 한 소에서 얼마나 많은 정육을 생산해 낼 수 있는가에 대한 정보로 맛에 관한 의미가 없다. 육질등급은 도축시 등심 부분을 보고 근내지방도, 육색, 지방색에 대한 부분과 소의 나이나 도축당시의 상태변화에 따른 차이를 알 수 있는 조직감, 성숙도에 따라 매긴다. 한 마리의 소에서 등심이 1등급이라고 그 외의 다른 부위도 모두 1등급이라 할 수 없고, 또한 요리 방법에 따라 사용되어야 하는 고기의 특성이 매우 다양한 만큼 기존의 방식을 좀 더 보완할 필요가 있다. 우리나라 사람들이 직접 맛을 보고 평가한 소비자 맛 점수를 기반으로 요리방법별로 각 부위의 맛 등급을 제공하는 차별화된 전략으로 한우의 소비심리를 자극할 수 있을 것이다.

본 연구에서는 소고기의 맛에 대한 소비자 관능평가 데이터를 기반으로 맛 등급 판별함수를 구하고자 한다. 판별함수 형태로는 선형, 이차, 정준, 비모수적 판별함수를 사용하였으며, 판별함수의 분류능력은 재대입 방법에 의한 분류율로써 비교해 보았다.

본 연구는 2008학년도 덕성여자대학교 연구비 지원을 받았습니니다.

¹교신저자: (132-714) 서울시 도봉구 쌍문동 419 덕성여자대학교 정보통계학과, 교수.

E-mail: jaehee@duksung.ac.kr

²(132-714) 서울시 도봉구 쌍문동 419 덕성여자대학교 정보통계학과, 석사과정.

E-mail: star4023@duksung.ac.kr

표 2.1. 분석에 사용된 표본 수

	요리방법			전체
	구이	스테이크	탕	
수소	1681	672	1952	4305
거세우	1703	665	2057	4425
전체	3384	1337	4009	8730

표 2.2. 소비자 맛 등급 별 기초통계량 표

맛등급 grade	외적기준	명수	비율	연도		다즙성		향미		전반적인 기호도	
				Mean	(sd)	Mean	(sd)	Mean	(sd)	Mean	(sd)
1	만족하지 못한다	2474	28.35%	33.98	(18.25)	43.34	(19.81)	44.64	(19.35)	36.64	(15.10)
2	만족한다	3381	38.75%	59.68	(17.91)	62.83	(16.59)	62.78	(15.79)	62.69	(12.98)
3	매우 만족한다	2098	24.04%	78.64	(13.74)	77.95	(13.74)	76.45	(13.82)	80.77	(9.88)
4	극도로 만족한다	773	8.86%	89.89	(11.99)	90.15	(11.19)	87.59	(12.60)	92.45	(9.29)

2. 한우 맛 등급 평가 데이터

본 연구에 사용된 데이터는 2006년도와 2007년도에 걸쳐 축산 과학원에서 시행한 관능평가 자료로 한우 시험장에서 수소(Bull) 10마리, 거세우(Steer) 5마리를 도축하여 총 10개의 대분할로 부위를 나누고, 각 부위를 구이, 스테이크, 탕의 3가지 방법으로 요리한 후 소비자가 직접 고기의 연도, 다즙성, 향미, 전반적인 기호도에 대하여 점수를 기록하고 최종적으로 소비자가 직접 자신이 먹은 고기의 등급을 1등급부터 4등급까지 정하도록 하였다.

소비자 관능평가는 소고기 시료에 관련된 요인이외에는 다른 요인에 의해서 영향을 받지 않도록 관능평가 protocol (Gee 등, 1998)에 따라 시료준비, 제시순서 및 평가 방법 등 관련된 모든 공정을 세부화 하여 가능한 동일한 조건으로 진행할 수 있도록 하였다. 관능평가는 조리방법에 따라 session 별로 수행되었으며 1명의 소비자가 총 7개의 시료를 평가하였다. 소비자들에게 동일한 평가기준을 주기 위하여 첫 번째 표준시료를 제외한 나머지 시료는 Latin Square 배열 방법에 의하여 제공하였다. 관능평가를 수행한 소비자는 층화추출방법을 사용하여 서울, 경기, 호남, 영남 지방의 인구밀도와 비슷한 비율로 추출하였으며, 모든 지역의 성비는 비슷하게 추출되도록 디자인 되었다.

관능평가에 사용된 변수는 연도, 다즙성, 향미, 전반적인 기호도로 각각 100mm로 준비된 선척도법을 이용하여 수집하였으며, 연도, 다즙성, 향미, 전반적인 기호도에 대한 항목척도는 다음과 같다: 연도 = 매우 질기다(0)~매우 연하다(100), 다즙성 = 매우 건조하다(0)~매우 다즙하다(100), 향미 = 대단히 싫어한다(0)~대단히 좋아한다(100), 전반적인 기호도 = 대단히 싫어한다(0)~대단히 좋아한다(100).

표 2.1에서는 관능평가에 사용된 소의 종류와 요리방법별 사용된 시료의 수를 볼 수 있으며, 표 2.2는 관능평가로 측정된 맛변수들의 기초통계량을 보여준다. 소비자 맛등급은 등급 값이 높을수록 만족도가 높음을 의미한다.

3. 한우 맛 등급 다변량 판별분석

한우 맛을 결정하는데 중요한 역할을 하는 연도, 다즙성, 향미 3가지 변수들로부터 소비자 등급을 예측할 수 있는 선형, 이차, 정준 판별함수를 만들고 판별방법별로 재대입방법에 의한 분류율을 계산하여 소비자 등급을 구하는 가장 적절한 판별함수를 찾고자 한다. 판별분석에 관한 이론은 Hand (1981)과 Huberty (1994)를 참고 하였으며, 응용부분에 관하여서 Kim과 Kim (2005)를 참고하였다.

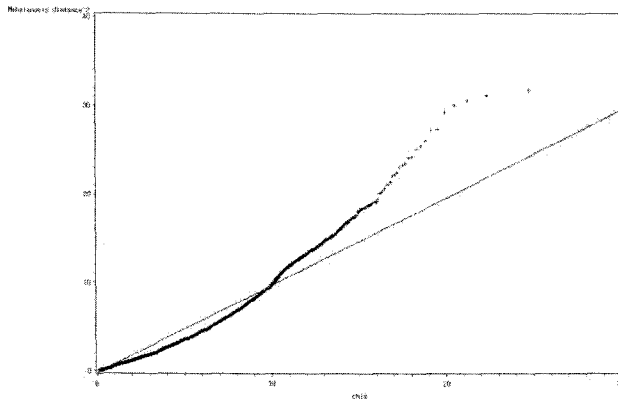


그림 3.1. 자유도가 3인 카이제곱 그림

3.1. 다변량 정규분포 및 공분산 행렬 동일성 검정

다변량 판별분석을 시행하기에 앞서 김재희 (2008)에 따라 판별분석에 필요한 기본가정을 살펴보면 다음과 같다. 첫째, 각 설명변수들은 정규분포를 이루어야 하며, 모든 설명변수들이 다변량 정규분포를 이루어야 한다. 둘째, 각 집단의 공분산 행렬(covariance matrix)이 동일한 경우 선형 판별함수를 얻게 되고, 공분산 행렬이 동일하지 않으면 이차 판별함수를 얻게 된다.

연도, 다즙성, 향미 세 변수가 삼변량 정규분포를 따르는지를 보기 위해 카이제곱 분포의 분위수와 마할라노비스 거리의 순서통계량을 이용해 그림 3.1의 카이제곱 그림(chi-square plot)을 그려 보았다. 그림 3.1을 보면 카이제곱분포의 분위수가 커질수록 다변량 정규성에서 벗어남을 볼 수 있다.

다변량 판별분석의 두 번째 가정인 공분산 행렬의 동일성 검정을 한 결과 카이제곱 통계량값이 1343.69로 유의확률값이 0.0001 이하로 매우 작아 공분산 행렬에 대한 차이가 유의하다. 그러나 Klecka (1980)에 의하면 비록 판별분석의 기본가정이 충족되지 않는다 해도 이것이 판별분석에 큰 영향을 미치는 것이 아니라, 단지 분석의 효율성과 정확도를 다소 감소시킬 뿐이라고 하였으므로 다변량 정규분포와 공분산 행렬의 동일성의 충족여부를 비중 있게 고려하지 않고 선형 판별분석을 수행하고자 한다.

3.2. 다변량 분산분석과 그룹간 거리

한우 맛등급 판별을 위해 사용된 변수들의 일변량 분산 분석결과 모든 변수가 유의확률값이 0.0001 이하로 매우 유의하며 맛 변수들에 대해 다변량 분산분석을 시행한 결과 Wilks Lambda 값이 0.372, 유의확률값이 0.0001 이하로 맛등급간 평균벡터의 차이가 매우 유의함을 알 수 있다.

그룹간의 거리를 구하기 위해 합동공분산을 이용하여 계산된 상호 중심까지의 마할라노비스 거리는 표 3.1과 같으며 표 3.1을 보면 매우 만족한다(grade = 3)와 극도로 만족한다(grade = 4)의 거리가 상당히 가깝다는 것을 알 수 있다.

3.3. 선형 판별함수

p_1, p_2, \dots, p_g 는 관측벡터 \mathbf{X} 가 각각의 그룹 G_1, G_2, \dots, G_g 로부터 발생할 사전확률이고, 각 집단의 공분산 행렬을 모두 같다고 가정한다. 각 그룹으로 부터의 확률표본의 개수는 n_1, n_2, \dots, n_g 이며 전체 표본의 크기는 $N = \sum_{i=1}^g n_i$ 이다.

표 3.1. 각 집단 중심까지의 마할라노비스 거리

소비자 맛등급	1	2	3	4
1	0	3.00	9.17	15.34
2	3.00	0	1.68	4.81
3	9.17	1.68	0	0.83
4	15.34	4.81	0.83	0

표 3.2. 선형 판별함수의 분류결과

소비자맛등급	분류 결과				행합
	만족하지 못한다	만족한다	매우 만족한다	극도로 만족한다	
1 만족하지 못한다	1947 (78.7%)	473 (19.1%)	45 (1.8%)	9 (0.4%)	2474
2 만족한다	718 (21.2%)	1714 (50.7%)	832 (24.6%)	117 (3.5%)	3381
3 매우 만족한다	39 (1.9%)	373 (17.8%)	1058 (50.4%)	628 (29.9%)	2098
4 극도로 만족한다	7 (0.9%)	19 (2.5%)	138 (17.9%)	609 (78.8%)	773
열합	2711	2579	2073	1363	8726
사전확률	0.25	0.25	0.25	0.25	

Rencher (1995)를 보면 확률밀도함수를 아는 경우 최적 판별규칙은

$$p_i f(\mathbf{X} | G_i) \geq p_j f(\mathbf{X} | G_j), \quad j = 1, 2, \dots, g \tag{3.1}$$

를 만족하는 경우 \mathbf{X} 를 G_i 그룹에 분류한다. \mathbf{X} 가 다변량 정규분포를 따르는 경우 $p_i f(\mathbf{X} | G_i)$ 를 최대화하는 선형함수

$$L'_i(\mathbf{X}) = \ln p_i - \bar{\mathbf{X}}'_i \mathbf{S}_{pl}^{-1} \mathbf{X} - \frac{1}{2} \bar{\mathbf{X}}'_i \mathbf{S}_{pl}^{-1} \bar{\mathbf{X}}_i \tag{3.2}$$

를 얻게 된다. 여기서 p_i 는 G_i 에서 발생할 수 있는 사전확률, \mathbf{S}_{pl} 은 공통 공분산 행렬의 추정량이다. 판별규칙은 새로운 관측벡터 \mathbf{X} 를 $L'_i(\mathbf{X})$ 를 최대로 하는 그룹에 분류하는 것이다.

연도, 다즙성, 향미의 3개 변수를 이용하여 추정된 선형 판별함수

$$\begin{aligned} G_1(\mathbf{X}) &= -5.386 + 0.053 \text{Tndr} + 0.085 \text{Jcy} + 0.118 \text{Flvr} \\ G_2(\mathbf{X}) &= -12.054 + 0.125 \text{Tndr} + 0.107 \text{Jcy} + 0.157 \text{Flvr} \\ G_3(\mathbf{X}) &= -19.045 + 0.177 \text{Tndr} + 0.127 \text{Jcy} + 0.186 \text{Flvr} \\ G_4(\mathbf{X}) &= -25.084 + 0.201 \text{Tndr} + 0.151 \text{Jcy} + 0.218 \text{Flvr} \end{aligned} \tag{3.3}$$

를 사용하여 소비자 등급을 판별하고 실제 분류율이 어느 정도 되는지 계산해 보았다. 분류함수의 능력을 판단하기 위해서는 오분류(misclassification)의 확률로서 오류율(error rate)과 정확한 분류율(correct classification rate)을 이용한다. 재대입(resubstitution)분류, 표본분할(partitioning the sample), 교차타당성(cross validation)방법 등을 이용하여 오분류율을 계산할 수 있으나 본 연구에서는 재대입분류에 의한 오분류율을 이용하였다. 재대입 분류에 의한 분류결과는 표 3.2와 같으며, 정확한 분류율은 $(1947 + 1714 + 1058 + 609)/8726 = 61.1\%$ 이다.

3.4. 이차 판별함수와 분류율

p_1, p_2, \dots, p_g 는 관측벡터가 각각 그룹 G_1, G_2, \dots, G_g 로부터 발생할 사전확률이고 각 집단의 공분산 행렬이 모두 같다고 할 수 없는 경우를 고려해 보자.

표 3.3. 이차 판별함수의 분류결과

소비자맛등급	분류 결과				행합
	만족하지 못한다	만족한다	매우 만족한다	극도로 만족한다	
1 만족하지 못한다	1941 (78.5%)	471 (19.0%)	47 (1.9%)	15 (0.6%)	2474
2 만족한다	721 (21.3%)	1625 (48.1%)	807 (23.9%)	228 (6.7%)	3381
3 매우 만족한다	60 (2.9%)	356 (17.0%)	795 (37.9%)	887 (42.3%)	2098
4 극도로 만족한다	11 (1.4%)	26 (3.4%)	95 (12.3%)	641 (82.9%)	773
열합	2733	2478	1744	1771	8726
사전확률	0.25	0.25	0.25	0.25	

\mathbf{X} 가 다변량 정규분포를 따르며 각 그룹의 공분산행렬이 다를 경우, 표본으로 부터의 추정량을 이용하면 다음의 이차형식함수

$$Q_i(\mathbf{X}) = \ln p_i - \frac{1}{2} \ln |\mathbf{S}_i| - \frac{1}{2} \bar{\mathbf{X}}_i' \mathbf{S}_i^{-1} \bar{\mathbf{X}}_i - \frac{1}{2} \mathbf{X}' \mathbf{S}_i^{-1} \mathbf{X} + \bar{\mathbf{X}}_i' \mathbf{S}_i^{-1} \mathbf{X} \quad (3.4)$$

를 얻는다. 여기서 $i = 1, 2, 3, 4$ 에 대해 p_i 는 i 번째 그룹에 대한 사전확률, $\bar{\mathbf{X}}_i$ 는 i 번째 그룹의 표본평균 벡터, \mathbf{S}_i 는 i 번째 그룹의 표본공분산 행렬을 나타낸다.

식 (3.4)에서 사용된 추정값은 (3.5)와 같다.

$$\mathbf{X} = \begin{pmatrix} \text{Tndr} \\ \text{Jcy} \\ \text{Flvr} \end{pmatrix}, \quad \bar{\mathbf{X}}_1 = \begin{pmatrix} 33.98 \\ 43.34 \\ 44.64 \end{pmatrix}, \quad \bar{\mathbf{X}}_2 = \begin{pmatrix} 59.68 \\ 62.83 \\ 62.78 \end{pmatrix}, \quad \bar{\mathbf{X}}_3 = \begin{pmatrix} 78.64 \\ 77.95 \\ 76.45 \end{pmatrix}, \quad \bar{\mathbf{X}}_4 = \begin{pmatrix} 89.89 \\ 90.15 \\ 87.59 \end{pmatrix},$$

$$\mathbf{S}_1 = \begin{pmatrix} 332.67 & 154.37 & 93.64 \\ 154.37 & 392.37 & 172.47 \\ 93.64 & 172.47 & 374.52 \end{pmatrix}, \quad \mathbf{S}_2 = \begin{pmatrix} 320.61 & 132.46 & 80.20 \\ 132.46 & 275.12 & 103.77 \\ 80.20 & 103.77 & 249.39 \end{pmatrix},$$

$$\mathbf{S}_3 = \begin{pmatrix} 188.77 & 93.15 & 50.76 \\ 93.15 & 188.77 & 73.23 \\ 50.76 & 73.23 & 190.99 \end{pmatrix}, \quad \mathbf{S}_4 = \begin{pmatrix} 143.68 & 71.74 & 48.88 \\ 71.74 & 125.29 & 73.18 \\ 48.88 & 73.18 & 158.79 \end{pmatrix}. \quad (3.5)$$

2차 판별함수에 의한 분류결과는 표.3.3과 같으며, 정확한 분류율은 $(1941 + 1625 + 795 + 641)/8726 = 57.3\%$ 이다.

3.5. 정준 판별분석

정준 판별분석(canonical discriminant analysis)은 주성분분석과 정준상관계수를 결합한 차원축소기법이라고 할 수 있다. 그룹 간에 가능한 가장 큰 다중 상관계수를 갖게 되도록 변수들의 선형조합을 구하게 되며, 이와 같이 얻어진 선형조합을 첫 번째 정준변수(the first canonical variable)라 하고 여기서 가장 큰 다중 상관계수를 첫 번째 정준상관계수(the first canonical correlation)라 한다. 두 번째 정준변수는 첫 번째 정준변수와는 서로 독립이며 그 다음으로 큰 다중 상관계수를 갖게 되도록 변수들의 선형조합을 구한다. 첫 번째 정준변수는 그룹간의 차이를 최대한 반영하는 변수가 되며 이러한 이유로 정준변수는 판별함수라고 부르기도 한다 (Rencher, 1992; 허명희, 1999).

정준 판별분석이 갖는 장점으로선 선형판별이나 이차판별과 같이 정규성 가정이 필요하지 않으므로 정규성을 만족하지 못하는 경우에 발생한 손실을 줄일 수 있는 방법이다. 정준 판별함수(canonical discriminant function) $Y = \mathbf{a}'\mathbf{X}$ 는 그룹내 제곱합행렬(within-group sum of squares matrix) \mathbf{W} 와 그룹

표 3.4. 정준 판별함수의 분류결과

소비자맛등급	분류결과				행합
	만족하지 못한다	만족한다	매우 만족한다	극도로 만족한다	
1 만족하지 못한다	1942 (78.5%)	478 (19.3%)	45 (1.8%)	9 (0.4%)	2474
2 만족한다	704 (20.8%)	1730 (51.2%)	814 (24.1%)	133 (3.9%)	3381
3 매우 만족한다	37 (1.8%)	375 (17.9%)	1031 (49.1%)	655 (31.2%)	2098
4 극도로 만족한다	7 (0.9%)	18 (2.3%)	137 (17.7%)	611 (79.0%)	773
열합	2690	2601	2027	1408	8726
사전확률	0.25	0.25	0.25	0.25	

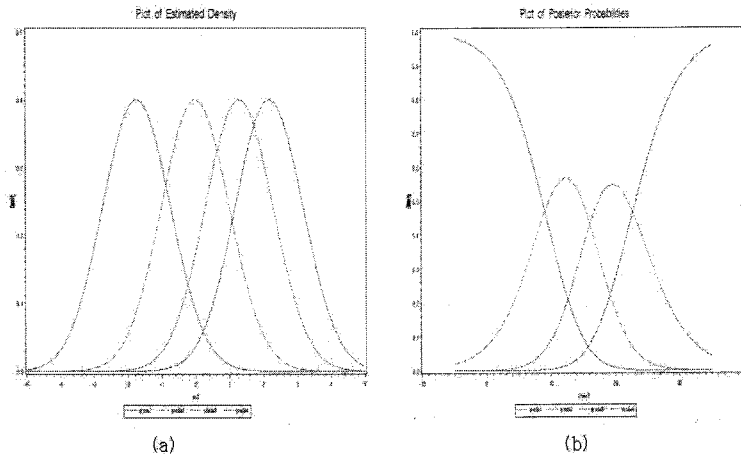


그림 3.2. 연도, 다중성, 향미를 이용한 정준 판별결과((a)확률밀도함수그림과 (b)사후 확률그림)

간 제곱합행렬(between-group sum of squares matrix) **B**에 대해

$$\frac{\sum_{ij} (\bar{Y}_i - \bar{Y})^2}{\sum_{ij} (Y_{ij} - \bar{Y}_i)^2} = \frac{\mathbf{a}'\mathbf{B}\mathbf{a}}{\mathbf{a}'\mathbf{W}\mathbf{a}} \tag{3.6}$$

를 최대화하도록 **a**를 구하게 되며 $\mathbf{W}^{-1}\mathbf{B}$ 의 고유벡터로 얻게 된다. $r = \min(g - 1, p)$ 개의 정준 판별함수를 얻을 수 있으며 계수 벡터 **a**는

$$\mathbf{a}_s = \mathbf{D}\mathbf{w}\mathbf{a} \tag{3.7}$$

로 표준화하며 여기서 $\mathbf{D}_W^2 = \text{diag}(\mathbf{W})$ 이다. r 개의 정준계수벡터 $\mathbf{a}_i, i = 1, \dots, r$ 에 대응되는 정준 판별함수는 \mathbf{a}_i 와 **X**의 선형결합

$$\mathbf{Y}_i = a_{i1}\mathbf{X}_1 + a_{i2}\mathbf{X}_2 + \dots + a_{ip}\mathbf{X}_p, \quad i = 1, \dots, r \tag{3.8}$$

로 얻을 수 있으며, 이에 따른 분류방법은 정준 판별함수 중 r 개의 정준 판별함수를 사용하는 경우 모든 $i \neq k$ 에 대하여 $r \leq g$ 일 때, \mathbf{X}_0 가 식 (3.9)를 만족하는 경우 G_k 에 분류하게 된다.

$$\sum_{j=1}^r \{\mathbf{a}'_j(\mathbf{X}_0 - \bar{\mathbf{X}}_k)\}^2 < \sum_{j=1}^r \{\mathbf{a}'_j(\mathbf{X}_0 - \bar{\mathbf{X}}_i)\}^2. \tag{3.9}$$

표 4.1. 선형 판별함수의 분류결과

소비자맛등급	분류결과				행합
	만족하지 못한다	만족한다	매우 만족한다	극도로 만족한다	
1 만족하지 못한다	1985 (80.4%)	452 (18.3%)	24 (1.0%)	7 (0.3%)	2468
2 만족한다	503 (14.9%)	2044 (60.7%)	748 (22.2%)	75 (2.2%)	3370
3 매우 만족한다	10 (0.5%)	374 (17.9%)	1112 (53.1%)	598 (28.6%)	2094
4 극도로 만족한다	6 (0.8%)	7 (0.9%)	124 (16.0%)	636 (82.3%)	773
열합	2504	2877	2008	1316	8705

맛 변수들에 대한 정준 판별함수 (3.10)을 살펴보면 소비자 맛등급을 구분하는데 있어 연도(Tndr)가 가장 큰 공헌을 하며, 그 다음으로 향미(Flvr), 다즙성(Jcy) 순으로 영향을 미치고 있다는 것을 알 수 있다.

$$Y = 0.994 \times ZTndr + 0.340 \times ZJcy + 0.493 \times ZFlvr, \tag{3.10}$$

여기서 변수 앞에 Z는 해당 변수의 표준화를 의미한다.

정준 판별함수에 의한 분류결과 표 3.4를 보면 정준 판별함수를 이용한 판별함수의 정확한 분류율은 $(1942 + 1730 + 1031 + 611)/8726 = 60.9\%$ 이다.

그림 3.2는 연도, 다즙성, 향미를 이용한 정준 판별분석결과에 대한 (a)확률밀도함수그림(density plot)과 (b)사후확률그림(posterior probability plot)이다.

4. 전반적인 기호도에 의한 일변량 판별분석

연도, 다즙성, 향미 등의 의미를 포함하여 포괄적인 맛점수인 전반적인 기호도(OAll)만을 사용한 일변량 판별분석을 하고자 한다. 전반적인 기호도 변수에 대해 일변량 분산분석을 시행한 결과 F통계량 값이 6302.36, 유의확률값이 0.0001 이하로 소비자등급별 전반적인 기호도의 평균차이가 매우 유의하다.

4.1. 선형 판별함수

전반적인 기호도를 사용한 일변량 선형 판별함수를 구하면

$$\begin{aligned} G_1(OAll) &= -4.170 + 0.228 OAll \\ G_2(OAll) &= -12.210 + 0.390 OAll \\ G_3(OAll) &= -20.267 + 0.502 OAll \\ G_4(OAll) &= -26.554 + 0.574 OAll \end{aligned} \tag{4.1}$$

으로 식 (4.1)를 이용한 일변량 선형 판별함수의 분류결과는 표 4.1에 나와있으며, 선형 판별함수의 정확한 분류율은 $(1985 + 2044 + 1112 + 636)/8705 = 66.36\%$ 이다.

4.2. 이차 판별함수

맛등급별 집단의 공분산에 대한 동일성 검정 결과 카이제곱 통계량 값이 518.195, 유의확률값이 0.0001 이하로 공분산 행렬이 동일하다고 보기 어렵다. 공분산 행렬의 비동일성을 고려하여 이차 판별함수를

표 4.2. 이차 판별함수의 분류결과

소비자맛등급	분류결과				행합
	만족하지 못한다	만족한다	매우 만족한다	극도로 만족한다	
1 만족하지 못한다	1979 (80.2%)	454 (18.4%)	27 (1.1%)	8 (0.3%)	2468
2 만족한다	491 (14.6%)	1959 (58.1%)	842 (25.0%)	78 (2.3%)	3370
3 매우 만족한다	10 (0.5%)	323 (15.4%)	1152 (55.0%)	609 (29.1%)	2094
4 극도로 만족한다	6 (0.8%)	6 (0.8%)	115 (14.9%)	636 (83.6%)	773
열합	2486	2742	2136	1341	8705

표 4.3. 소비자 맛등급별 띠풍

소비자 맛등급	bandwidth	등급중심점
만족하지 못한다	0.06606197	24.0
만족한다	0.07140569	61.5
매우 만족한다	0.06264937	80.5
극도로 만족한다	0.04164038	93.5

구하면

$$\begin{aligned}
 Q_1(OAll) &= -9.99 + 0.16 OAll - 0.002 OAll^2 \\
 Q_2(OAll) &= -27.28 + 0.37 OAll - 0.003 OAll^2 \\
 Q_3(OAll) &= -70.51 + 0.82 OAll - 0.005 OAll^2 \\
 Q_4(OAll) &= -102.65 + 1.07 OAll - 0.006 OAll^2
 \end{aligned} \tag{4.2}$$

을 얻는다.

표 4.2의 이차 판별함수의 분류결과를 보면 이차 판별함수의 정확한 분류율은 $(1979 + 1959 + 1152 + 636)/8705 = 65.78\%$ 으로 전체의 65.8%를 정확하게 분류한 것을 알 수 있다. 분류결과를 보면 선형 판별함수는 소비자 맛 등급이 1, 2등급인 경우, 이차 판별함수는 소비자 맛 등급이 3, 4등급인 경우를 더 잘 판별하는 경향이 있다는 것을 알 수 있다.

4.3. 비모수적 판별분석

그룹에 대한 확률밀도함수에 대한 비모수적 추정에 근거하여 비모수적 판별분석을 할 수 있다. Silverman (1986), Härdle (1991)은 확률밀도함수 추정에 대해 자세히 다루고 있으며 커널함수(kernel function)를 이용한 방법이나 k -이웃점 방법(k -nearest-neighbor method) 등을 이용해 각 그룹에 대한 비모수적 확률밀도함수를 추정하여 분류의 기준으로 사용할 수 있게 된다. 커널함수로 uniform, normal, Epanechnikov, biweight, triweight 커널함수 등을 사용할 수 있다.

본 연구에서는 Gaussian 커널을 사용한 비모수적 판별분석을 하였으며, 분석에 사용된 띠풍(bandwidth)은 Silverman (1986)의 엄지 법칙(rule of thumb)으로 추정 하였다. 소비자 맛등급별로 띠풍을 구하는 식은

$$\hat{h} = 1.06 \min \left(\hat{\sigma}, \frac{R}{1.34} \right) n^{-\frac{1}{5}} \tag{4.3}$$

으로 R 은 데이터의 1사분위수와 3사분위수의 차이, $\hat{\sigma}$ 은 표준편차, n 은 데이터 수를 의미한다. 표 4.3에서는 소비자 맛등급별 띠풍과 그룹 중심점을 보여준다.

표 4.4. 전반적인 기호도 변수만을 사용한 비모수적 판별함수의 분류결과

소비자맛등급	분류 결과				행합
	만족하지 못한다	만족한다	매우 만족한다	극도로 만족한다	
1 만족하지 못한다	1931 (78.2%)	510 (20.7%)	19 (0.8%)	8 (0.3%)	2468
2 만족한다	432 (12.8%)	2301 (68.3%)	579 (17.2%)	80 (2.4%)	3370
3 매우 만족한다	15 (0.7%)	477 (22.8%)	977 (46.7%)	631 (30.1%)	2094
4 극도로 만족한다	6 (0.8%)	9 (1.2%)	106 (13.7%)	652 (84.3%)	773
열합	2384	3297	1681	1371	8705

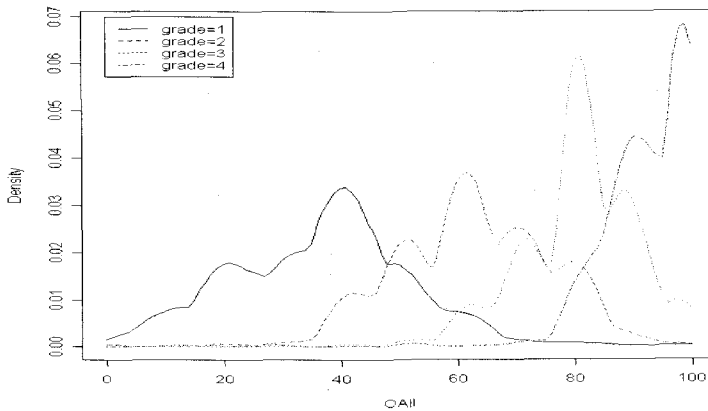


그림 4.1. OALL을 사용한 소비자 맛등급별 GAUSSIAN 커널 이용한 확률밀도함수그림

표 4.3에서 그룹별 띠폭을 h 라 하고, x 는 구하고자 하는 위치의 데이터라고 할 때, Venables와 Repley (2002)와 같이 커널함수 K 를

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right) \tag{4.4}$$

으로 정의한다. 커널함수 K_h 를 이용한 확률밀도함수는

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \tag{4.5}$$

이며 본 연구에서 사용된 Gaussian 커널함수는

$$K_h(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) \tag{4.6}$$

이다.

표 4.3의 등급 중심점에 가까운 등급으로 전반적인 기호도를 판별한 결과 표 4.4를 얻는다. Gaussian 커널함수를 이용한 비모수 판별함수의 정확한 분류율은 $(1931 + 2301 + 977 + 652)/8723 = 67.19\%$ 이다.

소비자 등급별로 Gaussian 커널함수를 이용한 확률밀도함수를 그려본 결과 그림 4.1과 같은 분포를 가지는 것으로 나타났다. 그림 4.1을 보면 등급이 오르면서 전반적인 기호도의 값이 커지는 경향을 보이고 있다. 또한 1,2등급의 분포와 3,4등급의 분포가 비슷한 형태를 가지는 것을 알 수 있다.

표 5.1. 판별방법별 정확한 분류율과 오분류율

사용변수	판별방법	정확한분류율(%)	1→4 오분류율(%)	1→3 오분류율(%)	2→4 오분류율(%)
연도 다즙성 향미 (삼변량)	선형	61.1	0.4	1.8	3.5
	이차	57.3	0.6	1.9	6.7
	정준	60.9	0.4	1.8	3.9
전반적인 기호도 (일변량)	선형	66.4	0.3	1.0	2.2
	이차	65.8	0.3	1.1	2.3
	비모수	67.2	0.3	0.8	2.4

5. 결론

본 연구는 등급 판정사에 의해 정해진 한우 등급을 더 세분화된 맛에 대한 정보를 제공하기 위해 소비자 등급에 기반을 둔 한우 맛 등급 판별방법을 비교 연구 해 보았다. 소비자 등급을 구분해 내기 위한 판별 방법들은 대부분 비슷한 수치로 소비자 등급을 판별해 내는 것을 알 수 있다. 하지만 맛 등급을 판별할 때 중요한 요소는 정확한 맛 등급을 분류하는 것이고, 이와 더불어 소비자들에게 이 시스템에 대한 신뢰를 주기 위해 실제 맛있는 소고기를 맛있는 소고기로 판별하는 경우를 최소화해야 할 것이다.

판별방법별 정확한 분류율과 1등급을 4등급으로 판별한 경우(1→4), 1등급을 3등급으로 판별한 경우(1→3), 2등급을 4등급으로 판별한 경우(2→4) 각각에 대한 분류율을 표 5.1에 정리해 두었다. 표 5.1을 살펴보면 정확한 분류율을 기준으로 전반적인 기호도 변수 하나로 비모수적 판별방법을 사용하는 것이 최상의 판별방법이라고 할 수 있다. 비모수적 판별방법을 사용하면 분류율이 좋고 분포에 대한 가설이나 그 외에 다른 가설에 대한 검증 없이 사용할 수 있다는 점에서 유용하나 락, 커널함수 선택 문제와 해석의 어려움이 있다. 그다음으로 분류율이 높은 전반적인 기호도 변수만으로 선형 판별을 한 경우 식이 간단하고 비모수적 판별함수에 비해 일반인들이 이해하기 쉬운 방법이라는 장점이 있다.

이번에는 맛 관련 세 변수를 모두 포함한 판별방법을 살펴보고자 한다. 세 변수들은 고기의 맛을 좌우하는 중요한 변수라는 것이 기존의 연구들에 의해 이미 밝혀져 있으며 (Thompson, 2002; 조수현 등, 2007) 전반적인 기호도 만을 사용한 경우보다는 데이터에 대한 민감도가 낮을 것이다. 다변량 정규분포를 만족하지 않고, 공분산 행렬이 동일하다는 가정을 기각하므로 선형 판별함수를 사용한 방법보다는 연도, 다즙성, 향미 3가지 변수를 사용한 정준 판별방법을 추천하고자 한다. 선형 판별함수나 이차 판별함수를 사용하기 위해 필요하였던 다변량 정규분포에 대한 가정이 없는 만큼 기본 가정을 만족하지 못함에서 오는 오류도 최소화 할 수 있다. 또한 일변량 비모수적 판별함수에 비해 전체적인 분류율을 낮았으나 (1→4)오분류율과 (1→3)오분류율이 비슷한 수준이고, (2→4)오분류율의 경우의 차이가 크지 않다는점, 데이터에 대한 민감도가 상대적으로 낮다는 점으로 볼 때 세 변수를 사용한 정준 판별분석이 가장 적절한 판별방법이라 할 수 있을 것이다.

감사의 글

본 연구에 귀중한 한우 관련 데이터를 제공해 주신 축산과학원 안종남 과장님과 조수현박사님께 감사드립니다.

참고문헌

김재희 (2008). <SAS를 이용한 다변량 통계 분석>, 교우사.

- 조수현, 김진형, 김재희, 성필남, 박범영, 김경의, 서그려운달님, 이종운, 김동훈 (2007). 사회인구학적 요인이 한우 수소고기의 부위 및 요리형태별 관능특성에 미치는 영향, <한국 동물자원과학회지>, **49**, 857-871.
- 허명희 (1999). <사회과학을 위한 다변량자료분석>, 자유아카데미.
- Gee, A., Porter, M., Coffey, D. and Polkinghorne, R. (1998). *Design and Protocol for Steak Grilling Trials*, MSA, Sydney.
- Hand, D. J. (1981). *Discrimination and Classification*, Wiley John & Sons, New York.
- Härdle, W. (1991). *Smoothing Techniques with Implementation in S*, Springer-Verlag, New York.
- Huberty, C. J. (1994). *Applied Discriminant Analysis*, Wiley John & Sons, New York.
- Kim, J. and Kim, J. (2005). Statistical discriminant analysis on the driving ability of the brain-injured, *Journal of Korean Data & Information Science Society*, **16**, 19-31.
- Klecka, W. R. (1980). *Discriminant Analysis*, Sage Publications, London.
- Rencher, A. C. (1992). Interpretation of canonical discriminant functions, canonical variates, and principal components, *The American Statistician*, **46**, 217-225.
- Rencher, A. C. (1995). *Methods of Multivariate Analysis*, Willey, New York.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall/CRC, London.
- Thompson, J. (2002). Managing Meat Tenderness, *Meat Science*, **62**, 295-308.
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*, Springer, New York.

Comparison of Discriminant Analyses for Consumers' Taste Grade on Hanwoo

Jaehee Kim¹ · Gureoundalnim Seo²

¹Dept. of Statistics, Duksung University; ²Dept. of Statistics, Duksung University

(Received July 2008; accepted September 2008)

Abstract

This paper presents the comparison of four methods, linear, quadratic, canonical and non-parametric discriminant analyses to discriminate the consumers' taste grade with sensory variables, such as tenderness, juiciness, flavor, and overall acceptability based on Consumer Sensory Survey. The classification ability of each method is measured and compared by the resubstitution error rate.

Keywords: Taste grade, discriminant analysis, resubstitution error rate.

This research was supported by the research fund of Duksung University in 2008.

¹Corresponding author: Professor, Dept. of Statistics, Duksung University, 419 Ssangmun-dong,

Dobong gu, Seoul 132-714, Korea. E-mail: jaehee@duksung.ac.kr

²Graduate student, Dept. of Statistics, Duksung University, 419 Ssangmun-dong, Dobong gu, Seoul 132-714, Korea. E-mail: star310@duksung.ac.kr