

더미(dummy) 변수를 활용한 다중인자 차원 축소(MDR) 방법

이제영¹ · 이호근²

¹영남대학교 통계학과, ²영남대학교 통계학과

(2008년 12월 접수, 2009년 2월 채택)

요약

통계모형의 상호작용 효과를 분석하기 위해 비모수적인 방법인 다중인자 차원 축소(MDR) 방법을 사용하였다. MDR 방법은 사례-대조 데이터에만 적용 할 수 있다. 본 논문에서는 연속형 데이터에도 적용 할 수 있는 더미(dummy) 변수를 활용한 MDR 방법을 소개한다. 아울러 이를 통해 한우의 주요 경제형질인 등심단면적(longissimus muscle dorsi area: LMA), 도체중(carass cold weight: CWT), 일당증체량(average daily gain: ADG)에 영향을 주는 우수 유전자 단일염기다형성(SNP)을 규명한다.

주요용어: MDR, 더미(DUMMY)변수, SNP, 한우 경제형질.

1. 서론

통계모형의 상호작용을 고려한 모형으로 선형모형 같은 표준 통계적 모형을 사용하였다. 그러나 유전자와 같은 범주형 데이터의 경우 변수가 많을 경우 상호작용의 조합이 많아지므로 종종 모수들의 상호작용에 대한 해석과 모형을 결정하는 것이 어려울 수 있다. 그래서 다중인자 차원 축소방법(Ritchie 등, 2001; Chung 등, 2005), 조합 분할 방법(Nelson 등, 2001), 제한된 분할 방법(Culverhouse 등, 2004) 등이 여러 유전자에 대한 상호작용을 결정하는 방법들로 개발되었다. 특히 다중인자 차원 축소(MDR) 방법은 상호작용에 대한 명확한 모형의 가정이 없는 비모수적인 방법으로 적당한 high-order 차수의 데이터로 복잡한 관계를 밝힐 수 있었다. MDR 방법은 사례항목과 대조항목의 비율을 통해 독립변수의 범주를 'high' 집단과 'low' 집단으로 분류한 후 목표변수에 대한 오분류율을 비교하여 분석한다. 그러나 이 MDR 방법은 사례-대조로 이분화 된 데이터에 대해 사용하는 방법으로 연속형 데이터에 대해 적용하는데 문제점이 발생한다. 본 논문에서는 이를 해결하기 위해 범주형 자료로 이루어진 변수를 각 범주를 나타내는 지시 변수들로 표현한 변수인 더미변수를 활용한다. 본 논문에서는 더미변수를 정의하는 방법과 그것을 활용한 회귀분석에서 변수들의 회귀계수의 해석상의 의미를 통해 각 범주들 사이의 상호간의 차이를 비교하여 범주를 'high' 집단과 'low' 집단으로 분류하는 방법을 소개하고, 이를 통해 연속형 데이터에서 적용 할 수 있는 더미변수를 활용한 MDR 방법을 소개한다. 아울러 본 연구에서는 더미변수를 활용한 MDR 방법을 통해 한우의 경제형질인 등심단면적(longissimus muscle dorsi area: LMA), 도체중(carass cold weight: CWT), 일당증체량(average daily gain: ADG)에 영향을 주는 우수 유전자 단일염기다형성(SNPs)을 규명한다.

본 연구는 2009년도 영남대학교 학술연구 조성비에 의한 것임.

¹교신저자: (712-749) 경북 경산시 대동 214-1 영남대학교 통계학과, 교수. E-mail: jlee@yu.ac.kr

2. 더미변수를 활용한 MDR 방법

2.1. 실험자료

본 연구 데이터는 농협중앙회 가축개량 사무소에서 개발되었고 16grand-sire half-sibs families로부터 229두의 수송아지로 구성되어졌다. 한우의 여러 경제형질인 등심단면적(longissimus muscle dorsi area: LMA), 도체중(carcass cold weight: CWT), 일당증체량(average daily gain: ADG)은 모든 F1 자손으로부터 수집되어졌고 한국축산물등급판정소의 규격에 따라 측정되었다.

현재까지 소에서는 도체형질(도체중량, 등지방두께, 등심단면적, 일당증체량, 근내지방도)과 연관이 있는 SNP 마커(marker)들이 일반가축에서 평가되어지거나 적용되고 있다 (Barendse 등, 2004; Page 등, 2004). 따라서 본 연구에서는 EST-based SNP 연관지도 (Snelling 등, 2005)에서 Kim 등 (2003)에 의해 규명 되어진 한우 염색체 6번에 위치한 후보 QTL인 ILSTS035와 같은 거리에 있는 SNP들 중 다형성이 나타난 SNP(19_1, 18_4, 28_2)를 이용하였다.

2.2. 더미변수를 활용한 이분화 방법

더미변수는 범주형 자료로 이루어진 변수를 각 범주를 나타내는 지시 변수들로 표현한 변수를 의미한다. 이는 회귀분석과 같이 수치형 독립변수를 통해 목표변수와와의 관계를 설명하거나 예측하는 분석에서 범주형 독립변수를 통해 목표변수와와의 관계를 설명하기 위해 사용된다. 즉 본 연구에서 사용한 회귀분석의 경우 종속변수와 독립변수가 모두 연속형 자료인 경우에 서로간의 관계를 함수식으로 표현하여 분석하는 방법이며, 이때 독립변수가 범주형 자료인 경우 범주형 독립변수를 기준범주를 제외한 (범주의 수 - 1)개의 더미변수로 정의하여 회귀분석을 시행하게 된다. 더미변수의 정의방법은 해당 범주와 일치하는 경우 1로 나머지를 0으로 코딩하여 준다. 예를들어 실험자료에서 목표변수인 경제형질(Y)과 2개의 SNP의 조합에 의한 9개의 범주(AAAA, AAAT, AATT, ..., GGTT)를 가지는 독립변수에 대한 회귀모형을 만들기 위해 GGTT범주를 기준범주로 하여 더미변수를 정의하면 식 (2.1)과 같이 정의 할 수 있으며, 식 (2.2)와 같이 회귀모형을 설정한다.

$$AAAA = \begin{cases} 1, & \text{범주 AAAA인 경우,} \\ 0, & \text{AAAA가 아닌 경우,} \end{cases} \quad (2.1)$$

$$AAAT = \begin{cases} 1, & \text{범주 AAAT인 경우,} \\ 0, & \text{AAAA가 아닌 경우,} \end{cases}$$

$$GGAT = \begin{cases} 1, & \text{범주 GGAT인 경우,} \\ 0, & \text{AAAA가 아닌 경우,} \end{cases}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_8 X_8 + \epsilon. \quad (2.2)$$

또한 범주형 변수를 더미변수로 표현하면 표 2.1과 같다.

이렇게 정의된 더미변수를 이용한 회귀분석에서 각 더미변수의 회귀계수는 기준집단과 그 더미변수에 해당하는 집단들 간의 종속변수의 평균차이를 나타낸다 (김태근, 2006). 따라서 더미변수의 정의에서 기준을 달리하여 회귀모형을 생성하면 각각의 더미변수 사이의 상호간에 차이를 비교 할 수 있다. 즉, 이 상호간의 차이를 비교하여 종속변수가 상대적으로 높은 집단과 낮은 집단으로 분류하는 방법이며 다음의 절차로 요약 할 수 있다.

표 2.1. 9개의 형태로 나타난 SNP조합 마커에 대한 더미변수의 표현

범주	더미변수							
	X1	X2	X3	X4	X5	X6	X7	X8
AAAA	1	0	0	0	0	0	0	0
AAAT	0	1	0	0	0	0	0	0
AATT	0	0	1	0	0	0	0	0
AGAA	0	0	0	1	0	0	0	0
AGAT	0	0	0	0	1	0	0	0
AGTT	0	0	0	0	0	1	0	0
GGAA	0	0	0	0	0	0	1	0
GGAT	0	0	0	0	0	0	0	1
GGTT	0	0	0	0	0	0	0	0

- 절차 1. 범주형 독립변수를 (범주의 수 - 1)개의 더미변수로 생성한다.
- 절차 2. 기준을 달리하여 범주의 수 만큼의 회귀모형을 생성한다.
- 절차 3. 각 범주에 대한 회귀계수가 유의수준 α 보다 작은 값을 가지는 회귀계수 중 양의 값을 가지는 회귀계수의 개수를 high_count, 음의 값을 가지는 회귀계수의 개수를 low_count라고 지정한다. 이들은 유의수준 α 를 기준으로 유의한 차이를 가지면서 각각 해당 범주가 다른 범주에 비해 상대적으로 높은 종속변수 값을 가지게 되는 수(high_count)와 상대적으로 낮은 종속변수 값을 가지게 되는 수(low_count)를 의미한다.
- 절차 4. high_count와 low_count를 비교하여 high_count가 low_count보다 크거나 같으면 'high'집단으로 작으면 'low'집단으로 분류한다.

위와 같은 절차를 거쳐 3개 이상의 범주로 이루어진 집단을 'high'와 'low' 두 개의 집단으로 분류된다. 2.3절에서는 더미변수를 활용한 이분화 방법을 이용해 MDR 방법을 실험자료에 적용하여 분석하는 절차를 기술한다.

2.3. 더미변수를 활용한 MDR 방법

실험자료의 SNP 마커들은 각각 3개의 범주로 이루어져 있으며, 이들의 조합 마커인 SNP(19_1)*SNP(18_4), SNP(19_1)*SNP(28_2), SNP(18_4)*SNP(28_2)는 각각 9개의 범주로 이루어져있고, 목표변수인 경제형질 데이터는 연속형 자료로 이루어져있다. 따라서 MDR 방법에 적용하기 위해 2.2절에서 소개한 더미변수를 활용한 이분화 방법을 적용하였다. 또한 데이터의 수는 229두로 SNP조합 마커의 각 수준이 9개의 범주임을 감안할 때 각 cell의 빈도가 매우 낮거나 0이 될 가능성이 있다. 이를 해결하기 위해 붓스트랩 방법 (Efron과 Tibshirani, 1993)을 사용하였다. 모든 적용 방법을 그림 2.1과 다음의 절차로 요약 할 수 있다.

- 절차 1. 전체 데이터를 붓스트랩 방법을 통해 10개의 데이터셋을 만든다.
- 절차 2. 10개의 데이터셋 중 9개를 훈련용 데이터셋으로 정하고 1개를 검증용 데이터셋으로 정한다.
- 절차 3. N개의 SNP 마커들로부터 k개의 SNP조합 마커 중 하나를 선택한다.
- 절차 4. 선택된 SNP조합 마커에서 SNP조합 마커 각각의 수준을 기초로 개체들의 조합을 더미변수로 생성하여 2.2절에서 소개한 더미변수기법으로 단순선형 회귀모형을 생성한다. 각각의 회귀모형에서 더미변수에 대해 추정된 회귀계수(기준 변수와 통계적으로 유의한 차이를 가지는)의 부

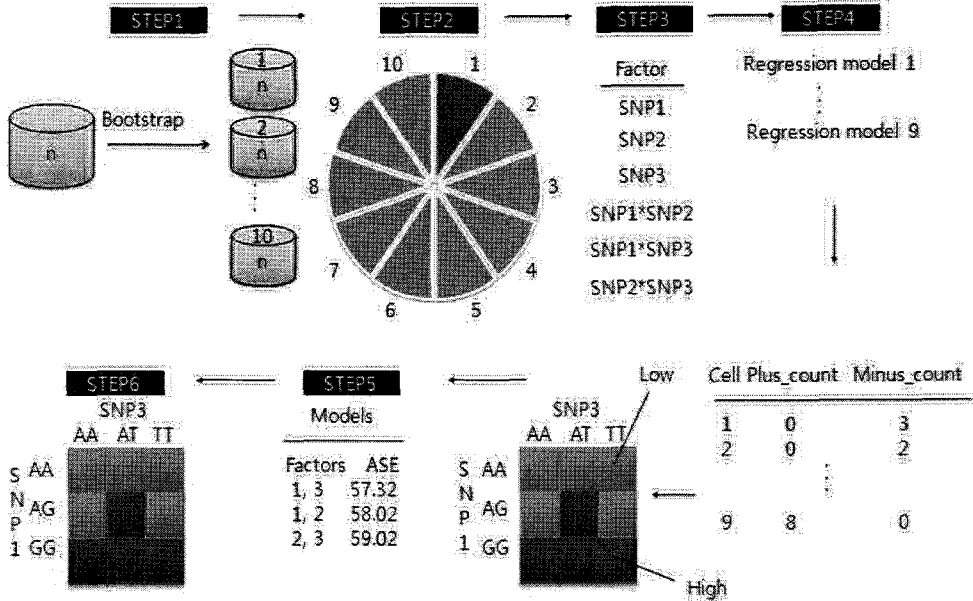


그림 2.1. 한우의 경제형질에 대해 더미변수를 활용한 MDR 방법의 적용과정

호를 중심으로 SNP조합 마커의 각 수준을 상대적으로 양의 영향을 미치는 개체 ‘high’그룹과 음의 영향을 미치는 개체 ‘low’그룹들로 분류한다.

- 절차 5. k 개의 SNP조합 마커 전부에 대한 훈련용 데이터셋에서 식 (2.3)과 같이 평균제곱오차(ASE)를 구하여 그림 2.1에서처럼 ASE의 값이 가장 작은 SNP1과 SNP3의 조합 마커를 선정한다.
- 절차 6. 절차 5에서 선정된 SNP의 조합 마커를 중심으로 검증용 데이터셋에 적용하여 예측한 평균제곱오차(P-ASE)를 구한다.

단,

$$S_{high} = \sum_{i=1}^n I_{high}(i) \sum_{j=1}^{N_i} y_{ij} - \hat{y}_i^2, \quad S_{low} = \sum_{i=1}^n I_{low}(i) \sum_{j=1}^{N_i} y_{ij} - \hat{y}_i^2,$$

$$N_{high} = \sum_{i=1}^n I_{high}(i) \cdot N_i, \quad N_{low} = \sum_{i=1}^n I_{low}(i) \cdot N_i,$$

$$I_{high}(i) = \begin{cases} 1, & i(\text{cell}) \in \text{high group}, \\ 0, & \text{o.w.}, \end{cases} \quad I_{low}(i) = \begin{cases} 1, & i(\text{cell}) \in \text{low group}, \\ 0, & \text{o.w.}, \end{cases}$$

(n : 범주의 수, N_i : i 범주의 관측치 수).

그 다음 위의 과정의 반복(훈련용 데이터셋과 검증용 데이터셋의 변경을 통한 교차검정)에서 나온 10개의 ASE와 P-ASE의 평균을 구해 그 값이 가장 낮은 것을 우수 모델로 정한다 (Bastone 등, 2004). 그리고 앞에서 구한 각각의 ASE를 이용하여 교차검정적합도(CVC)를 구하는데 이것은 10번의 반복(교차검정) 시행을 할 때 각 시행에서 각 모델이 우수 모델로 선택된 횟수이다 (Chung 등, 2005). 따라서

ASE와 P-ASE의 평균이 가장 낮고 CVC가 가장 높은 값을 가지는 모형이 우수 SNP 마커이다. 비록 MDR 방법에 의해 우수 SNP 마커가 선정되었지만, MDR 방법에 의한 통계적 유의성을 확신 할 수 없다. 따라서 순열 검정을 통해 통계적 유의성을 유도하고자 한다. 2.4절에서는 순열 검정 절차에 대해 기술한다.

2.4. 순열 검정

본 연구에서는 다음의 절차에 따라 순열 검정을 시행하였다 (Good, 2000).

- 절차 1. 검정하고자하는 가설 설정 - 2.3절의 절차를 통해 선정된 우수 SNP 마커가 경제 형질에 영향력이 있다.
- 절차 2. 통계량과 기각역 설정 - 분석에 사용할 통계량으로 'high'그룹의 평균을 선택. 선정된 SNP 마커에 의한 'high'그룹의 평균이 그룹간 데이터를 서로 바꾸었을 때 보다 높다면 경제형질에 영향력이 있다고 판단한다.
- 절차 3. 관측치의 통계량 계산 - 각 경제형질 데이터(검정용 데이터)를 더미변수를 활용한 MDR 모형에 의해 선택된 'high'그룹과 'low'그룹으로 이분화 시킨 후 'high'그룹의 평균을 계산한다.
- 절차 4. 관측치의 재배열과 재배열 후의 통계량 계산 - 두 그룹의 데이터를 n 개 만큼 랜덤 추출하여 그룹을 상호 변경한 후 각 'high'그룹의 평균을 구한다. 이 과정을 10,000번 반복한다.
- 절차 5. 결론 - 각 평균을 내림차순으로 정렬한 후 기존의 평균과 비교하여 Monte Carlo의 유의확률을 구한다. 절차 2-4과정을 10번 시행(10개의 훈련용 데이터셋)하여 유희확률들의 평균값을 구하여 유의수준보다 작으면 선정된 우수 SNP 마커가 경제 형질에 영향력이 있다는 가설을 채택한다.

위의 순열 검정을 통해 MDR 방법에 의해 선정된 SNP 마커들에 대해 유의성을 확인하여 결론을 얻게 된다. 3장에서는 2장에서 살펴본 방법들을 적용한 결과를 나타낸다.

3. 더미변수를 활용한 MDR 방법의 적용 결과

3.1. 더미변수와 MDR 방법을 활용한 분석과 결과

국가 후대 검정우 229두를 이용하여 등심단면적과 도체중, 일당증체량에 영향을 주는 SNP들의 주효과와 상호작용효과에 관하여 알아보기 위해 ANOVA모형에 적용한 결과 상호작용에 대한 효과는 모두 통계적으로 유의한 차이가 나타나지 않았으며 하나의 주효과에 대한 것은 SNP(19_1) 마커가 각 경제형질에 대해 통계적으로 유의하게(유의확률 0.047, 0.020, 0.035) 나타났다 (Lee와 Kim, 2009). 하지만 우리는 한우의 경제형질에 영향을 주는 유전자가 개별 SNP유전자가 아닌 여러 개의 복잡한 SNP유전자의 상호작용이 더 영향을 준다고 믿고 있다. 이들을 규명하기 위해 상호작용에 더 효과적이며 연속형 데이터에 적용할 수 있는 더미변수를 활용한 MDR 방법을 적용하였다.

등심단면적, 도체중, 일당증체량에 대해 SNP의 조합에 대하여 1.2절의 절차 2~6의 과정을 10번 반복해서 나온 ASE와 P-ASE의 평균과 10번의 반복과정에서 나온 값을 기준으로 CVC를 구한 결과는 각각 표 3.1, 3.2와 같이 나타났다.

표 3.2는 2.3절의 과정을 10번 반복한 후 각각의 SNP조합 마커에 대한 ASE와 P-ASE의 평균을 구한 값이다. 우수 모델은 평균 ASE와 평균 P-ASE가 가장 낮은 것을 선택하는데 위의 표 3.1을 보면 하나의 SNP 마커에 대한 효과에서는 SNP(19_1) 마커의 ASE가 등심단면적과 도체중, 일당증체

표 3.1. 등심단면적, 도체중, 일당증체량에 대한 SNP 조합의 평균 ASE와 평균 P-ASE 결과

요인의수	SNP 마커	등심단면적		도체중		일당증체량	
		ASE	P_ASE	ASE	P_ASE	ASE	P_ASE
1	SNP(19_1)	59.051	58.501	1089.17	1085.70	.007311	.007269
	SNP(18_4)	60.548	60.039	1102.76	1095.87	.007465	.007427
	SNP(28_2)	60.652	60.139	1103.75	1100.68	.007463	.007423
2	SNP(19_1)*SNP(18_4)	58.351	58.043	1080.12	1070.78	.007243	.007242
	SNP(19_1)*SNP(28_2)	57.793	57.429	1043.67	1031.85	.006990	.006947
	SNP(18_4)*SNP(28_2)	59.110	59.239	1077.68	1075.50	.007302	.007263

표 3.2. 등심단면적, 도체중, 일당증체량에 대한 SNP 조합의 CVC 결과

요인의수	SNP 마커	등심단면적	도체중	일당증체량
1	SNP(19_1)	10	10	10
	SNP(18_4)	0	0	0
	SNP(28_2)	0	0	0
2	SNP(19_1)*SNP(18_4)	0	0	0
	SNP(19_1)*SNP(28_2)	10	10	10
	SNP(18_4)*SNP(28_2)	0	0	0

표 3.3. 표각 형질에 대해 선택된 SNP 조합에 대한 순열 검정 결과

SNP 마커	순열 검정 결과		
	유의확률 값		
	등심단면적	도체중	일당증체량
SNP(19_1)	0.03506	0.04574	0.03579
SNP(19_1)*SNP(28_2)	0.00214	0.01125	0.00255

량에서 각각 59.051, 1089.17, 0.007331로 가장 낮게 나타났으며 SNP조합 마커에 대한 효과에서는 SNP(19_1)*SNP(28_2) 마커의 ASE가 등심단면적과 도체중, 일당증체량에서 각각 57.793, 1043.67, 0.006990으로 가장 낮게 나타났다. 따라서 한우의 종합적인 경제형질에 영향을 주는 요인은 하나의 SNP 마커에 대한 효과보다는 두 개의 SNP조합 마커에 의한 효과가 더 많다는 것을 알 수 있다. 다음으로 표 3.2의 CVC결과를 살펴보면 하나의 SNP 마커에서는 SNP(19_1) 마커가 10으로 가장 높게 나타났으며 SNP조합 마커에서는 SNP(19_1)*SNP(28_2) 마커가 각각 10으로 가장 높게 나타났다. 따라서 앞의 결과와 이 결과를 보면 SNP(19_1)*SNP(28_2) 마커가 가장 우수한 SNP 마커로 선정되었다. MDR 방법에서 선정된 SNP 마커들의 유의성을 확인하기 위해 순열 검정 결과를 3.2절에 나타냈다.

3.2. 순열 검정 결과

등심단면적, 도체중, 일당증체량에 대해 각 선택된 모형의 유의성을 확인하기 위하여 순열 검정('Good (2000)')이 제시한 절차를 통해)을 실시하여 표 3.3과 같이 나타냈다.

등심단면적, 도체중, 일당증체량 모두 하나의 요인에 의한 효과인 SNP(19_1)과 두 개의 조합형태의 효과인 SNP(19_1)*SNP(28_2)가 유의하게 나타났으며, 조합형태의 효과가 더 유의하게 나타났다. 따라서 한우의 종합적인 경제형질에 영향을 주는 요인은 하나의 SNP에 대한 효과보다는 SNP의 조합에 의한 효과가 더 많다는 것을 검증하였다.

4. 결론 및 토의

더미변수를 활용하여 이분형 데이터에만 적용 가능한 MDR 방법의 문제점을 해결하였으며 적용방법 및 실험을 통한 적용가능성을 확인하기 위해 연속형 데이터로 이루어진 한우의 경제형질 데이터에 적용하여 보았다. 더미변수를 활용한 MDR 방법에 적용하여 ASE, P-ASE, CVC, 순열 검정을 통해 분석한 결과 단일 SNP 마커에서는 SNP(19_1) 마커가 한우의 경제형질에 영향을 많이 주었으며, 두 개의 SNP조합 마커에서는 SNP(19_1)*SNP(28_2) 마커가 한우의 경제형질에 영향을 가장 많이 주는 것으로 나타났다. 또한 SNP(19_1) 마커와 SNP(19_1)*SNP(28_2) 마커의 ASE와 P-ASE를 비교한 결과 SNP(19_1) 마커인 하나의 유전자의 효과보다는 SNP(19_1)*SNP(28_2) 마커와 같이 조합에 의한 유전효과가 한우의 경제형질에 더 많은 영향을 주는 인자임을 밝혔다. 이와 같이 더미변수를 활용하여 MDR 방법의 문제점을 해결하는 실험을 성공적으로 수행하였으며, 한우의 경제형질에 영향을 주는 우수 유전자 SNP(19_1)*SNP(28_2) 마커를 규명하였다.

참고문헌

- 김태근 (2006). <u-can 회귀분석>, 인간과 복지, 서울.
- Barendse, W., Bunch, R., Thomas, M., Armitage, S., Baud, S. and Donaldson, N. (2004). The TG5 thyroglobulin gene test for a marbling quantitative trait loci evaluated in feedlot cattle, *Australian Journal of Experimental Agriculture*, **44**, 669-674.
- Bastone, L., Reilly, M., Rader D. J. and Foulkes, A. S. (2004). MDR and PRP: A comparison of methods for high-order genotype-phenotype associations, *Human Heredity*, **58**, 82-92.
- Chung, Y., Lee, S. Y. and Park, T. (2005). Multifactor dimensionality reduction in the presence of missing observations, In *Proceedings of the Autumn Conference Korea Statistical Society*, 31-36.
- Culverhouse, R., Klein, T. and Shannon, W. (2004). Detecting epistatic interactions contributing to quantitative traits, *Genetic Epidemiology*, **27**, 141-152.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman & Hall/CRC, New York.
- Good, P. (2000). *Permutation Test: A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer-Verlag, New York.
- Kim, J. W., Park, S. I. and Yeo, J. S. (2003). Linkage mapping and QTL on chromosome 6 in Hanwoo(Korean cattle), *Asian-Australasian Journal of Animal Sciences*, **16**, 1402-1405.
- Lee, J. Y. and Kim, D. C. (2009). Important SNPs identification from the various economic characters for the high quality Korean cattle, *Journal of Korea Statistical Society*, to appear.
- Nelson, M. R., Kardia, S. L., Ferrell, R. E. and Sing, C. F. (2001). A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation, *Genome Research*, **11**, 458-470.
- Page, B. T., Casas, E., Quaas, R. L., Thallman, R. M., Wheeler, T. L., Shackelford, S. D., Koohmaraie, M., White, S. N., Bennett, G. L., Keele, J. W., Dikeman, M. E. and Smith, T. P. L. (2004). Association of markers in the bovine CAPN1 gene with meat tenderness in large crossbred populations that sample influential industry sires, *Journal of Animal Science*, **82**, 3474-3481.
- Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl F. F. and Moore, J. H. (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer, *American Journal of Human Genetics*, **69**, 138-147.
- Snelling, W. M., Casas, E., Stone, R. T., Keele, J. W., Harhay, G. P., Bennett, G. L. and Smith, T. P. L. (2005). Linkage mapping bovine EST-based SNP, *BMC Genomics*, **6**, 74-84.

Multifactor Dimensionality Reduction(MDR) Analysis by Dummy Variables

Jea-Young Lee¹ · Ho-Guen Lee²

¹Department of Statistics, Yeungnam University; ²Department of Statistics, Yeungnam University

(Received December 2008; accepted February 2009)

Abstract

Multiple genes interacting is a difficult due to the limitations of parametric statistical method like as logistic regression for detection of gene effects that are dependent solely on interactions with other genes and with environmental exposures. Multifactor dimensionality reduction(MDR) statistical method by dummy variables was applied to identify interaction effects of single nucleotide polymorphisms(SNPs) responsible for longissimus mulcle dorsi area(LMA), carcass cold weight(CWT) and average daily gain(ADG) in a Hanwoo beef cattle population.

Keywords: Dummy variables, MDR, SNP, Hanwoo economic traits.

This research was supported by the Yeungnam University research grants in 2009.

¹Corresponding author: Professor, Department of Statistics, Yeungnam University, Kyungsan 712-749, Korea. E-mail: jlee@yu.ac.kr