

유전적 연관성 분석에서
정보충분성을 위한 통계적 방법과
실제 가계도 자료에 적용

연세대학교 대학원
의학전산통계학과
임 현 선

유전적 연관성 분석에서
정보충분성을 위한 통계적 방법과
실제 가계도 자료에 적용

지도 김 동 기 교수

이 논문을 석사 학위논문으로 제출함

2002년 6월 일

연세대학교 대학원

의학전산통계학과

임 현 선

임현선의 석사 학위논문을 인준함

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

연세대학교 대학원

2002년 6월 일

보잘 것 없는 딸을 하염없이 믿어주시는 부모님,
임영산과 한춘지께 이 논문을 바칩니다.

제 목 차 례

표 차례	iii
그림 차례	vi
국문 요약	vii
제 1장 서론	1
제 2장 유전자 연관성 분석	1
2.1 개념	1
2.2 재조합률의 추정과 검정	2
2.2.1 재조합률	2
2.2.2 사상함수(map function)	3
2.3 불완전 침투에서의 연관성	5
2.3.1 불완전 침투의 정의	5
2.3.2 재조합률과 침투율의 추정과 검정	6
2.3.3 불완전 침투로 인한 손실	9
2.4 가계도 자료에서의 정보량 측정	9
2.4.1 정보량 측도	9
2.4.2 Lod score의 기대값과 그와 관련된 방법	10
2.4.3 피셔의 information의 기대값과 분산	11
2.5 모의 실험 기법과 모형	12
2.5.1 모의 실험을 통한 추정	12
2.5.2 모형의 개요	14
2.5.3 프로그램에서의 입력과 출력자료	15
제 3장 가계도 자료의 이해	18
3.1 심혈관질환	18
3.2 심혈관질환을 일으키는 유전자	19

3.3 가계도 자료의 수집기준과 특성	20
제 4장 가계도 자료의 정보충분성 추정 및 비교	21
4.1 표식유전자 종류별 정보충분성 추정 및 비교	22
4.1.1 표식유전자의 대립형질 빈도가 동일한 1개 표식유전자를 사용한 가계도 자료의 경우 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 결과	22
4.1.2 표식유전자의 대립형질 빈도가 다른 1개 표식유전자를 사용한 snp459와 snp461의 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 비교결과	22
4.1.3 2개 표식유전자를 사용한 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 비교결과	23
4.2 가계도 자료 형태 및 자료 수별 정보충분성 추정 및 비교	25
4.2.1 가계도 자료 형태에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교	25
4.2.2 가계도 수에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교	29
제 5장 결론	31
제 6장 토론	32
참고 문헌	33
부 록	36
Abstract	68

표 차례

표 1. 침투율을 고려한 Lod Score의 기대값	8
표 2. 대립형질의 빈도가 동일한 공우성형질일 때 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$	22
표 3. 대립형질의 빈도가 다른 공우성형질일 때 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$	23
표 4. 표식유전자 수에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교	24
표 5. 가계도내 가족의 수에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교	25
표 6. 가계도내 가족 중 질병유무에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교	26
표 7. 가계도 내 부모 질병유무에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교 ...	27
표 8. 가계도 내 자손 질병유무에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교 ...	28
표 9. 가계도 자료의 세대수에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교	28
표 10. 대립형질의 빈도가 동일한 경우 1~35개 가계도의 $E(Z_{\max})$	37
표 11. 대립형질의 빈도가 동일한 경우 36~70개 가계도의 $E(Z_{\max})$	38
표 12. 대립형질의 빈도가 동일한 경우 71~107개 가계도의 $E(Z_{\max})$	39
표 13. 대립형질의 빈도가 동일한 경우 108~145개 가계도의 $E(Z_{\max})$	40
표 14. 대립형질의 빈도가 동일한 경우 1~35개 가계도의 $P(Z_{\max} \geq C)$	41
표 15. 대립형질의 빈도가 동일한 경우 36~70개 가계도의 $P(Z_{\max} \geq C)$	42
표 16. 대립형질의 빈도가 동일한 경우 71~107개 가계도의 $P(Z_{\max} \geq C)$	43
표 17. 대립형질의 빈도가 동일한 경우 108~145개 가계도의 $P(Z_{\max} \geq C)$	44
표 18. 대립형질의 빈도가 다른 경우 1~35개 가계도의 $E(Z_{\max})$	45
표 19. 대립형질의 빈도가 다른 경우 36~70개 가계도의 $E(Z_{\max})$	46
표 20. 대립형질의 빈도가 다른 경우 71~99개 가계도의 $E(Z_{\max})$	47
표 21. snp459에 대한 1~35개 가계도의 $P(Z_{\max} \geq C)$	48
표 22. snp459에 대한 36~70개 가계도의 $P(Z_{\max} \geq C)$	49

표 23. snp459에 대한 71~99개 가계도의 $P(Z_{\max} \geq C)$	50
표 24. snp461에 대한 1~35개 가계도의 $P(Z_{\max} \geq C)$	51
표 25. snp461에 대한 36~70개 가계도의 $P(Z_{\max} \geq C)$	52
표 26. snp461에 대한 71~99개 가계도의 $P(Z_{\max} \geq C)$	53
표 27. 2개 표식유전자인 경우 1~35개 가계도의 $E(Z_{\max})$	54
표 28. 2개 표식유전자인 경우 36~70개 가계도의 $E(Z_{\max})$	55
표 29. 2개 표식유전자인 경우 71~99개 가계도의 $E(Z_{\max})$	56
표 30. 2개 표식유전자, (D19S217 D19S219)인 경우 1~35개 가계도의 $P(Z_{\max} \geq C)$	57
표 31. 2개 표식유전자, (D19S217 D19S219)인 경우 36~70개 가계도의 $P(Z_{\max} \geq C)$	58
표 32. 2개 표식유전자, (D19S217 D19S219)인 경우 71~99개 가계도의 $P(Z_{\max} \geq C)$	59
표 33. 2개 표식유전자, (D19S223, D19S596)인 경우 1~35개 가계도의 $P(Z_{\max} \geq C)$	60
표 34. 2개 표식유전자, (D19S223, D19S596)인 경우 36~70개 가계도의 $P(Z_{\max} \geq C)$	61
표 35. 2개 표식유전자, (D19S223, D19S596)인 경우 71~99개 가계도의 $P(Z_{\max} \geq C)$	62
표 36. 2개 표식유전자, (D19S425, D19S927)인 경우 1~35개 가계도의 $P(Z_{\max} \geq C)$	63
표 37. 2개 표식유전자, (D19S425, D19S927)인 경우 36~70개 가계도의 $P(Z_{\max} \geq C)$	64
표 38. 2개 표식유전자, (D19S425, D19S927)인 경우 71~99개 가계도의 $P(Z_{\max} \geq C)$	65
표 39. 가계도 수에 따른 $E(Z_{\max})$	66
표 40. 가계도 수에 따른 $P(Z_{\max} \geq C)$	67

그림 차례

그림 1. 유전적 거리와 재조합률에 따른 사상함수의 관계	5
그림 2. 재조합률에 따른 Lod score의 기대값과 피셔의 information 비교	10
그림 3. 심혈관질환 관련 유전자, ApoE gene의 타입별 성향	20
그림 4. 가계도 예시	21
그림 5. APOE와 표식유전자 위치	24
그림 6. 가계도 수에 따른 $E(Z_{\max})$	29
그림 7. 가계도 수에 따른 $P(Z_{\max} \geq C)$	30

국문 요약

어떠한 표식유전자가 연관성 분석 시 좀 더 효율적으로 사용될 수 있는지 이전에 두 가계도를 가지고 비교해 보았다(박윤주, 2001). 이 논문에서는 가계도 수를 더 조사하여 대립형질 빈도 및 표식유전자 수별 연관성 정보충분성을 추정 및 비교 검정해 본다. 어떤 가계도의 형태가 연관성 분석을 하기에 충분한 정보를 가지고 있는지 살펴 보고 얼마만큼의 가계도 수를 가지고 분석하면 연관성 분석을 하기에 충분한 정보를 가지고 있는지 알아본다.

그러므로 연관성 분석시 표식유전자의 대립형질 빈도가 동일하고 표식유전자의 수를 여러개 사용한다. 가계도의 자료는 가계도 내 가족의 수가 많고 질병을 앓고 있는 환자가 한명 이상 보유하며 부모가 질병을 앓고 있는 경우가 좋겠다. 가계도 수는 15~20개정도 수집한다. 이처럼 가계도 자료를 구성하여 연관성 분석을 한다면 이전에 하던 방식대로 무조건 많은 자료를 사용하는 것보다 시간적, 비용적인 면에서 좀 더 효율적이라고 본다.

핵심되는 말 : 정보충분성, 가계도 자료의 검정력, 연관성 분석, 불완전 침투, 재조합률, 침투율

제 1장 서론

심혈관질환 가계도 자료에 대한 이해와 특성을 파악하기 위해 자료수집 방법 및 가계도 구성형태를 알아본다. 이 논문의 주 목적인 가계도의 정보충분성을 알아보기 위해 모의실험 프로그램 SIMLINK을 이용하여 각 개인의 질병유전자와 표식유전자의 유전자형을 모의실험하여 연관성 정보충분성을 추정한다. 즉, 질병유전자 APOE의 대립형질의 모집단빈도와 각 대립형질의 조합에서의 발병률, 표식유전자의 대립형질에 관한 정보 등 조사된 정보를 가지고 각 개인의 질병유전자와 표식유전자의 유전자형을 모의실험한다. 이 모의실험을 통하여 어떠한 표식유전자가 연관성 분석 시 좀 더 효율적으로 사용될 수 있는지 이전에 두 가계도를 가지고 비교해 보았다.(박윤주, 2001) 이 논문에서는 가계도 수를 더 조사하여 대립형질 빈도 및 표식유전자 수별 연관성 정보충분성을 추정 및 비교 검토해 본다. 어떤 가계도의 형태가 연관성 분석을 하기에 충분한 정보를 가지고 있는지 살펴보고 얼마만큼의 가계도 수를 가지고 분석하면 연관성 분석을 하기에 충분한 정보를 가지고 있는지 알아본다.

제 2장 유전자 연관성 분석

2.1 개념

유전자 연관성 분석은 염색체에서 유전자 위치 서열을 알아내는 것과 유전자 사이에 유전자 거리를 정하는 것이다. 1980년 이후부터 분자 유전학 등 여러 분야에서 인간 유전자 연관성 분석의 도구인 통계적인 방법에 많은 관심을 가지고 있다. 궁극적인 목적은 인간 DNA의 서열을 밝히는 것이고 유전적 질병원인을 알아내서 병을 치료하는 것이다. 앞으로 중요한 과제는 유전과 환경간의 상호작용을 잘 규명하는 것이다.(Jurg ott, 1999)

2.2 재조합률의 추정과 검정

2.2.1 재조합률

주어진 가계도 정보로부터 유전자 재조합률은 크게 우도함수(likelihood function)와 베이시안(Bayesian)방법을 이용하여 추정할 수 있다.

첫 번째로 우도함수를 이용한 추정은 다음과 같다. 조사된 가계도의 정보를 F 라 하고 두 유전자 사이의 재조합률을 θ 로 정의하면 우도함수는 $L(\theta|F)$ 와 같이 표현할 수 있는데, 이 함수는 가계도에 대한 정보가 주어졌을 때의 재조합률 θ 의 확률에 의해 정의되어질 수 있다.

두 번째로 베이시안 방법에 의한 재조합률 추정과정은 다음과 같다. 우선 θ 의 사전분포(prior distribution)를 미리 가정한 다음 주어진 가계도 자료의 표본분포를 통해 $L(F|\theta)$ 를 계산하고 마지막으로 이 두 정보를 이용하여 다음과 같이 θ 의 사후분포(posterior distribution)를 통한 θ 의 재조합률을 추정할 수 있다,

$$P(\theta|F) = \frac{L(F|\theta)h(\theta)}{\int L(F|\theta)h(\theta)d\theta},$$

$h(\theta)$: θ 의 사전분포,

$P(\theta|F)$: θ 의 사후분포.

위와같이 두 유전자의 재조합률 추정 후 이에 대한 연관성 검정(linkage test)이 이루어지는데 일반적 연관성 검정의 가설은 다음과 같다,

H_0 : 두 유전자는 연관되어 있지 않다. ($\theta = \frac{1}{2}$),

H_1 : 두 유전자는 연관되어 있다. ($\theta < \frac{1}{2}$).

이 가설을 검정하기 위하여 일반적으로 Lod Score가 사용되는데 Lod Score는 아래와 같이 연관이 되어있지 않다고 가정했을 때의 우도 $L(\theta = \frac{1}{2} | F)$

와 연관되어 있다고 가정한 우도 $L(\theta|F)$ 비율(proportion)의 log 값이다,

$$Z(\theta) = \log_{10} \frac{L(\theta|F)}{L(\theta = \frac{1}{2} | F)}.$$

위 식에서 $Z(\theta)$ 를 최대로 하는 값을 최대 Lod Score(maximum lod score)라 하는데 아래와 같이 표현할 수 있다. 그리고 이때의 $\hat{\theta}$ 를 최대우도 추정치라 한다,

$$Z_{\max} = \max LOD = \log_{10} \frac{L(\hat{\theta}|F)}{L(\theta = \frac{1}{2} | F)},$$

$\hat{\theta}$: 최대우도 추정치(maximum likelihood estimate : m.l.e.).

이러한 검정통계량은 표본크기가 클 때 영가설하에서 자유도(degree of freedom)가 1인 χ^2 분포를 따르고 보통 $Z_{\max} \geq 3$ 이면 연관되어 있지 않다는 영가설($H_0: \theta = \frac{1}{2}$)을 기각할 수 있다.(Lander and Bostein, 1989). 즉, 두 유전자가 연관되어 있다고 볼 수 있다.

2.2.2 사상함수(map function)

연관성 측도로 두 loci사이의 관련성은 재조합률(θ)을 이용하지만 loci가 3개이상 서로 관련되어 있을 경우는 사상함수(map function)을 이용한다. 가장 간단한 방법으로 재조합률을 이용하는 모건의 사상함수(Morgan, 1928), $x = \theta$ 이다. 두 loci 사이에 재조합률(θ)가 0.05이라면 유전적 거리는 0.05모건이다. 2개 이상의 loci에서 유전적 거리는 두 loci사이의 유전적 거리의 합으로 계산된다.(Sturtevant, 1913). loci 사이에 다중 교차가 발생할 때 더 이상 모건의 사상함수의 개념으로 설명되지 않는다. 예를들어 세 개의 loci에서 loci 1과 loci 2의 재조합률을 θ_1 라 하고 loci 2와 loci 3

의 재조합률을 θ_2 라 하면 loci 1과 loci 3의 재조합률은 $\theta_1 + \theta_2$ 로 추정되는데 일반적으로 $\theta_1 + \theta_2$ 보다 작은 경우가 많다. 서로 다른 구간에서 교차 발생수가 포와송 확률을 따른다면 헬덴의 사상함수(1919)

$$x = \begin{cases} -1/2 \ln(1 - 2\theta), & 0 \leq \theta < 1/2 \\ \infty, & \text{otherwise} \end{cases}$$

이다. 다시 말하면 $\theta = 1/2[1 - \exp(-2|x|)]$. 예를 들어 재조합률이 $\theta=0.22$ 이면 유전적 거리 $x=0.29$ 모건(29센티모건)으로 계산된다. 참고적으로 모의실험 프로그램 SIMLINK는 헬덴의 사상함수(1919)을 사용한다.

loci간에 서로 간섭(interference)이 있다면 코삼비 사상함수(1944)

$$x = 1/2 \tanh^{-1}(2\theta) = 1/4 \ln \frac{1+2\theta}{1-2\theta}$$

이다. 다시 말하면 $\theta = 1/2 \tanh(2x) = 1/2 \frac{\exp(4x) - 1}{\exp(4x) + 1}$. 예를 들어 재조합률이 $\theta=0.22$ 이면 유전적 거리 $x=0.236$ 모건(23.6센티모건)이다. 위에서 언급한 코삼비 사상함수는 유전학에서 널리 사용되고 있다. 이외에도 여러 가지 사상함수 Carter and Falconer 사상함수(1951), Felsenstein 사상함수(1979), Sturt 사상함수(1976), Binomial 사상함수(Karlin, 1984)등이 있다.

유전적 거리와 사상함수와의 관계를 보면 다음과 같다.

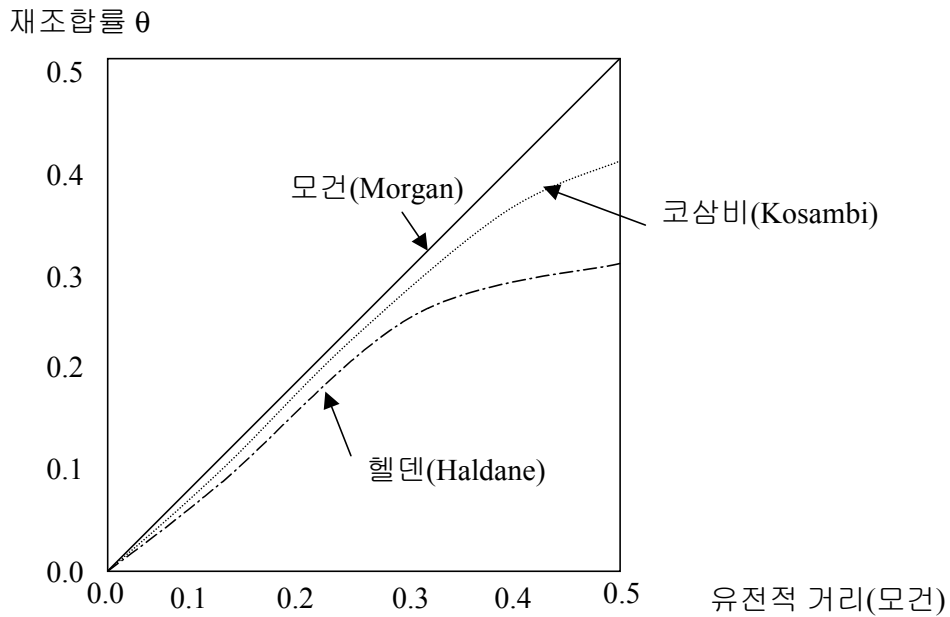


그림 1. 유전적 거리와 재조합률에 따른 사상함수의 관계

2.3 불완전 침투에서의 연관성

2.3.1 불완전 침투의 정의

개체가 질병을 일으키는 유전형을 가지고 있다 해도 발현되지 않는 경우 즉, 다시 말해 질병유전자를 가지고 있다고 하더라도 병이 발병하지 않는 경우가 있는데, 이를 불완전 침투(incomplete penetrance)라 한다. 이와 반대로 질병유전자를 가지면 반드시 발병되는 경우를 완전침투(full penetrance)라 한다(Jurg ott 1999). 또, 침투율(penetrance 또는 penetrance rate)은 다음과 같이 유전형이 주어졌을 때 표현형의 조건부 확률로 정의될 수 있는데 ($P(x|g)$, x : 표현형, g : 유전형) 완전침투는 질병 유전자를 가지고 있으면 반드시 발병이 되므로 이 확률이 1이 된다. 그러나 불완전 침투의 경우 질병유전자를 가지고 있더라도 모두 발병이 되는 것은 아니기 때문에 침투값이 항상 1보다 작아진다. 질병형질이 D, 정상형질이 d라고 정의하고 침투값을 f 라

하면 D가 우성이면 $P(\text{발병}|DD)=P(\text{발병}|Dd)=f$, $P(\text{발병}|dd)=0$ 이다. 즉, $P(\text{발병하지 않음}|DD)=P(\text{발병하지 않음}|Dd)=1-f$, $P(\text{발병하지 않음}|dd)=1$ 이다.

Phenocopy란 설명할 수 없는 환경적인 요인이나 다른 위치의 유전자에 의해 감염된 사람의 수라 한다. Phenocopy rate이란 모든 감염 사람들 중에 phenocopy의 비율이다.

2.3.2 재조합률과 침투율의 추정과 검정

일반적으로 침투율 f 와 재조합률 θ 는 결합우도(joint likelihood) $L(\lambda, \theta)$ 에 의해 추정된다. 이때 대부분 표식유전자의 정보를 이용하여 각 개인의 유전형을 알아내지만 부모의 유전형이 불확실하다면 침투율 f 와 재조합률 θ 의 추정(estimation)은 (f, θ) 의 여러 가지 조합 중 $L(f, \theta)$ 을 최대화(maximize)시키는 조합으로 추정한다(Rossen, R. D. et al, 1980). 일반적으로 연관성 검정을 할 때 최대 Lod Score를 사용하는데 불완전 침투일 경우 최대 Lod Score는 다음과 같이 수정되어 나타낼 수 있다,

$$Z_{\max} = Z(\hat{\theta}) = \log_{10} \frac{L(f, \hat{\theta})}{L(f, \theta = \frac{1}{2})}, \quad \hat{f} : \text{M.L.E.}$$

위와 같은 Z_{\max} 는 침투율 f 를 알고 이 f 와 재조합률 θ 가 독립(independent)일 때는 가능하다. 그러나 실제로 침투율은 대체적으로 추정되어지고 이 추정치는 재조합률과 상관관계(correlation)를 갖기 때문에 침투율 f 와 재조합률 θ 는 (λ, θ) 의 조합을 최대화시키는 최대우도법 $L(\hat{f}, \hat{\theta})$ 로 추정해야 한다(Jurgott, 1999).

$L(\hat{f}, \hat{\theta})$ 를 추정하는 경우는 두가지 방법이 있는데 첫째로는 θ 가 연관되어있다는 대립가설(H_1 of linkage : $H_1: \theta < \frac{1}{2}$)하에서 \hat{f} 를 추정하는 것이고 다음으로는 θ

가 연관되어 있지 않다는 귀무가설(H_0 of absence of linkage : $H_0: \theta = \frac{1}{2}$)하에서 추정하는 것이다. 이런 경우 Z_{\max} 는 아래와 같이 나타내어질 수 있다,

$$Z_2 = \log_{10} \frac{L(\hat{f}, \hat{\theta})}{L(\hat{f}, \theta = \frac{1}{2})}, \quad H_1 \text{ 하에서,}$$

$$Z_1 = \log_{10} \frac{L(\hat{f}, \hat{\theta})}{L(\hat{f}_{\theta = \frac{1}{2}}, \theta = \frac{1}{2})}, \quad H_0 \text{ 하에서.}$$

Z_2 은 $\theta = \frac{1}{2}$ 이라는 제약(restriction)조건이 있기 때문에 자유도 1인 값을 가지고 Z_1 은 $\theta = \frac{1}{2}$ 이라는 제약조건외에 $\hat{f}_{\theta = \frac{1}{2}}$ 이라는 조건이 추가되므로 자유도 2를 가지게 된다. 그러므로 Z_1 이 Z_2 보다 연관성정도를 측정하는데에는 더 효율적이라 할 수 있다. 즉, 같은 Lod score를 가진다면 Z_1 이 Z_2 보다 더욱 강한 연관성정도를 나타낸다고 볼 수 있다.

일반적으로 침투율 f 와 재조합률 θ 의 상관도는 f 의 추정에 영향을 미친다고 알려져 있는데 부모의 유전형을 알 경우(phase-known) 부모의 유전형을 한쪽은 표식유전자와 질병유전자 모두 동형(double homozygote)으로 가정하고 다른 한쪽은 이 두 유전자 모두를 이형(double heterozygote)으로 가정하는 교배방법인 double backcross mating을 고려할 때 f 와 θ 의 상관관계 ρ 는 다음과 같이 추정된다,

$$\rho(\hat{f}, \hat{\theta}) = \frac{f(1-2\theta)\sqrt{\theta-\theta^2}}{\sqrt{1-f+2f^2\theta(1-\theta)(1-2\theta+2\theta^2)}}.$$

이때 f 와 θ 의 범위는 $0 \leq f \leq 1$, $0 \leq \theta \leq \frac{1}{2}$ 이 되는데, $\theta < \frac{1}{2}$ 이면 f 와 θ 는 양의 상관관계를 갖고 $\theta = \frac{1}{2}$ 이면 독립이 된다.

이와 같이 f 와 θ 의 결합 추정치(joint estimate)는 그 추정치가 가지는 실제적인 범위때문에 서로간에 연관성을 가지고 되고 이로 인해서 두 모수를 추정할 경우 그 추정치는 대체적으로 강하게 편향(strongly biased)되어 지게 한다. 또한 이런 우도를 사용한 추정은 그 속성이 매우 복잡하기 때문에 실제 자료에서의 추정에는 많은 어려움이 있다. 예를들어, 질병형질이 D, 정상형질이 d라고 정의하고 질병 loci에서 침투율을 $f=P(\text{발병}|Dd)$ 라 한다. 그리고 phenocopy는 없다고 본다.

그러면 phase-known double backcross mating, $D2/d2 \times d1/d1$ 에 대해 생각해 보자. 다음 표는 질병 loci에서 불완전 침투율 f , 질병 loci와 표식 loci의 재조합률이 r 일 때 침투율을 고려한 Lod Score의 기대값을 추정해 보자.(NR:재조합이 없다. R:재조합이 있다.)

표 1. 침투율을 고려한 Lod Score의 기대값

형질유전자		P(G;r)	표식유전자 x			
(G)			D11	d21	D21	d11
D1/d1	NR	$1/2(1-r)$	f	0	0	$1-f$
d2/d1	NR	$1/2(1-r)$	0	1	0	0
D2/d1	R	$1/2r$	0	$1-f$	f	0
d1/d1	R	$1/2r$	0	0	0	1
$P(x) = \sum P(x G)P(G)$ $P(x)$, under			$1/2(1-r)f$	$1/2(1-rf)$	$1/2rf$	$1/2(1-f+rf)$
			$1/4f$	$1/2(1-f/2)$	$1/4f$	$1/2(1-f+f/2)$
$H_0: r=1/2$ $L(\theta)/L(1/2)$			$2(1-\theta)$	$\frac{1-\theta f}{1-f/2}$	2θ	$\frac{1-f+\theta f}{1-f/2}$

$$\begin{aligned}
 E[Z(\theta, f)] &= \sum_x P(x; r) Z_x(\theta) \\
 &= 1/2(1-r) f \log[2(1-r)] + 1/2(1-rf) \log\left(\frac{1-rf}{1-f/2}\right) \\
 &\quad + 1/2rf \log(2r) + 1/2(1-f+rf) \log\left(\frac{1-f+rf}{1-f/2}\right)
 \end{aligned}$$

로 Lod score의 기대값이 산출된다.

2.3.3 불완전 침투로 인한 손실

완전침투일 때 Lod score의 기대값과 불완전 침투일 때 Lod score의 기대값의 비를 R이라 하는데

$$R = \frac{E[Z(r, f)]}{E[Z(r, 1)]}$$

이 R은 침투율이 f일 때 연관성을 잘 식별하는지 정보를 주는 상대적인 척도이다. R은 침투율 f에 의존적이고 재조합률에는 덜 의존적인 특성을 가지고 있어, 침투율이 낮은 값을 가질 때 연관성을 잘 식별하려면 자료가 많이 필요하다. 예를 들어 $f=0.2$ 이면 $R \approx 0.1$ 이어서 완전침투 하에서 Lod score의 기대값보다 관찰치가 10배가 필요하다.

2.4 가계도 자료에서의 정보량 측정

2.4.1 정보량 측도

첫째는 최대 Lod score로 주어진 가계도 자료에서 연관성 검정하는데 사용된다,

$$Z_{\max} = \begin{cases} n[\log(2) + \hat{\theta} \log(\hat{\theta}) + (1 - \hat{\theta}) \log(1 - \hat{\theta})], & \hat{\theta} > 0 \\ n \log(2) & \hat{\theta} = 0 \end{cases}$$

둘째는 피셔의 information(1925)로 주어진 가계도 자료에서 추정된 재조합률, $\hat{\theta}$ 의 정도를 나타내는 척도이다,

$$I(\hat{\theta}) = - \frac{d^2}{d\theta^2} \ln L(\hat{\theta}).$$

추정량 $\hat{\theta}$ 가 불편추정량이면 $\hat{\theta}$ 의 표본분산,

$$\widehat{\text{var}}(\hat{\theta}) \approx [I(\hat{\theta})]^{-1}$$

이다. 그래서 피셔의 information(1925)이 클수록 표본분산이 작아지고 표본분산이 작아질수록 좋은 정보가 된다.

예를 들어, k가 재조합의 수라 하면 $L(\theta) \propto \theta^k (1 - \theta)^{n-k}$ 이고 $f(\theta)$ 는

$f(\theta) = -\ln L(\theta) = -[k\ln(\theta) + (n-k)\ln(1-\theta) + \text{상수}]$ 이다.

1차 미분을 하면

$$\frac{d}{d\theta} f(\theta) = -\frac{k}{\theta} + \frac{n-k}{1-\theta}$$

이고 재조합률의 추정량 $\hat{\theta} = k/n$ 가 산출된다.

이차 미분을 하면

$$\frac{d^2}{d\theta^2} f(\theta) = -\frac{k}{\theta^2} + \frac{n-k}{(1-\theta)^2} = \frac{n\theta^2 + 2k\theta + k}{\theta^2(1-\theta)^2}$$

이어서 피셔의 information의 추정량 $\mathcal{I}(\hat{\theta}) = n/[\hat{\theta}(1-\hat{\theta})]$ 이 산출된다.

다음은 Lod score의 기대값과 피셔의 information의 관계를 나타내는 그림이다.

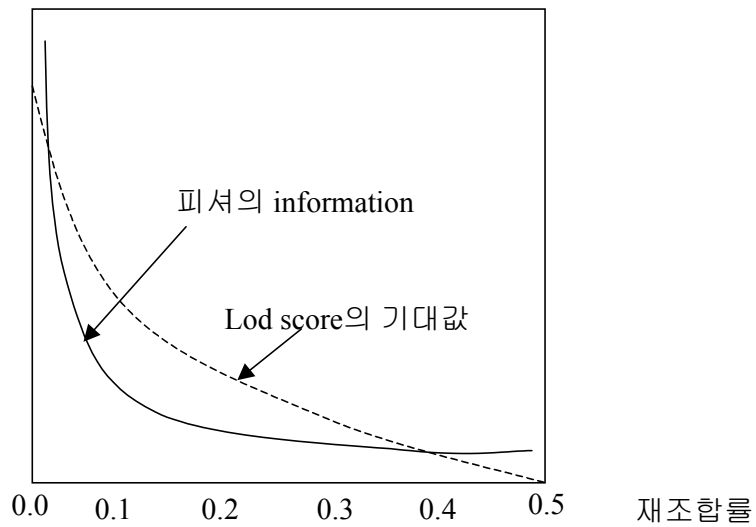


그림 2. 재조합률에 따른 Lod score의 기대값과 피셔의 information 비교

2.4.2 Lod score의 기대값과 그와 관련된 방법

유전통계학에서 널리 쓰이는 Lod score의 기대값은

$$E(Z(\theta)) = \sum_{k=0}^n P(k, r) Z_k(\theta)$$

이다. $Z_k(\theta)$ 의 가중평균이며, Lod score의 기대값의 최대값을 ELOD라 한다. 재조합률의 일치추정량 θ 는 $\theta = r$ 인 Lod score의 기대값의 최대일 때 값이다. 재조합률 θ 이 일치추정량이 아닐 때 Lod score의 기대값의 최대값은 $\theta = r$ 인 Lod score 기대값과 같을 수 없다. 이유는 모수에 대한 틀린 가정이나 표본설계상황시 일어난다. Lod score 기대값의 최대값, $\frac{d}{d\theta} E[Z(\theta)]$ 을 MELOD라 한다. 최대 Lod score의 기대값, EMLOD는

$$E[Z_{\max}] = \sum_i P_i Z_{\max}^{(i)}$$

이다. 이는 $\theta = r$ 이 아닌 $0 \leq \theta \leq 1/2$ 에서 Z_{\max} 을 연관성분석에 적용하므로 좀 더 현실적이다.

대체적으로 $ELOD \leq MELOD \leq EMLOD$ 을 만족한다.

검정력(Power)은 Z_{\max} 가 어떤 상수 c 보다 클 확률을 더한 값으로

$$\text{검정력(Power)} = \sum_i P_i (Z_{\max} > c)$$

이다. 연관성분석에 대한 성공확률을 직접 계산하는 방법이므로 좋은 장점을 가지고 있다.

2.4.3 피셔의 information의 기대값과 분산

c 개로 분류되어지는 가계도에 대한 피셔의 information(Elandt-Johnson, 1971; Rao, 1973)은

$$i(r) = \sum_{i=1}^c q_i^2(r) / p_i(r)$$

이다. 여기서 $q_i(r) = \frac{d}{dr} p_i(r)$, $p_i(r)$ 은 i 번째 범주에서 사건이 일어날 확

를이다. $[q_i(r)]^2/p_i(r)$ 은 피셔의 information에서 i 번째 범주의 기여정도이고 $I(r) = n \times i(r)$ 은 n 개의 관찰치에 대한 피셔의 information의 총합이다. 점근적으로

$$\hat{\theta} \sim N(r, [I(r)]^{-1}) = N(r, [n \times i(r)]^{-1})$$

을 따른다.

모수가 여러개인 경우로 확장시켜 보자.(Elandt-Johnson, 1971)

i 번째 범주에서 일어날 확률을 $p_i = p_i(r_1, r_2)$ 라 하면 피셔의 information 행렬은

$$I(r_1, r_2) = n i(r_1, r_2) = n \begin{pmatrix} a & c \\ c & b \end{pmatrix}$$

이다. 여기서 $a = \sum (\delta p_i / \delta r_1)^2 / p_i$, $b = \sum (\delta p_i / \delta r_2)^2 / p_i$,

$c = \sum (\delta p_i / \delta r_1)(\delta p_i / \delta r_2) / p_i$ 이다.

분산 공분산 행렬은

$$[I(r_1, r_2)]^{-1} = V(\hat{\theta}_1, \hat{\theta}_2) \approx \frac{1}{n(ab - c^2)} \begin{pmatrix} b & -c \\ -c & a \end{pmatrix}$$

이고 $\rho(\hat{\theta}_1, \hat{\theta}_2) \approx -c/\sqrt{ab}$ 이다.

2.5 모의 실험 기법과 모형

2.5.1 모의 실험을 통한 추정

모의실험을 통한 가계도 자료의 검정력을 추정하기 위해서는 몇 가지 가정이 필요한데 첫째로 각 개인의 질병의 표현형, 즉 질병 발병의 여부를 알아야 하고 둘째로는 표식유전자와 질병유전자는 연관되어 있어야 하고 셋째로는 가계도의 구조는 분석시 적합한 구조라고 가정한다(Ploughman and Boehnke, 1989). 이러한 가정을 통해 가계도 자료의 검정력을 다음 단계를 통해 추정할 수 있다.

첫 번째 단계로 알고 있는 질병의 유전형과 표현형의 정보를 이용하여 가계도안의 각 개인에 대한 질병 유전형의 확률 즉, 질병의 유전형과 표현형이 주어졌을 때 개인의 질병유전형의 조건부확률(conditional probability)을 계산한다.

질병의 유전형의 벡터를 $g=(g_1, \dots, g_n)$ 라고 하고 표현형의 벡터를 $x=(x_1, \dots, x_n)$ 이라 할 때, 질병 유전형의 조건부 확률은 $P(g|x)$ 로 다음과 같이 표현할 수 있다.

n명으로 구성된 가계도에서 1, 2는 부모라 하고 자손을 3, ... , n이라 하면 질병의 유전형의 조건부확률은 다음과 같이 나타낼 수 있다,

$$P(g|x) = \prod_{k=1}^n p(g_k|g_1, \dots, g_{k-1}, x) \\ = P(g_1|x)P(g_2|g_1, x_2, \dots, x_n) \prod_{k=3}^n P(g_k|g_1, g_2, x_k)$$

이 조건부확률은 Elston-Stewart 연산(algorithm) (Elston and Stewart, 1971)을 일반화시킨 Lange와 Boehnke(1983)의 연산을 이용한 우도식을 사용하여 계산되어 지는데, 이 우도식은 무작위 교배(random mating)을 가정하고 환경적인 효과(environment effect)로 인한 유전자의 관련이 없다고 가정한다. 이러한 가정이 있을 때 우도식은 아래와 같이 표현되어질 수 있다,

$$L(x) = \sum_{g_1} \dots \sum_{g_n} \prod_i P(x_i|g_i) \prod_j P(g_j) \prod_k P(g_k|g_l, g_m),$$

$P(x_i|g_i)$: i번째 사람의 조건부확률,

(x_i : i번째 사람의 표현형, g_i : I번째 사람의 유전형)

$P(g_j)$: 부모가 가계도 안에 없는 j번째 사람의 유전자형의 사전확률,

$P(g_k|g_l, g_m)$: 자손 k의 부모 l, m의 유전자형이 주어졌을 때 k의 유전형 g_k 의 조건부확률.

두 번째 단계는 첫 번째 단계에서 계산된 유전형의 조건부확률을 이용하여 반복(replicates)된 가계도에서 각 개인의 질병의 유전형을 모의실험한다.

세 번째로 표식유전자와 질병유전자간의 여러 실 재조합률을 가정하고 그 가정된 각 실 조합률에서 표식유전자의 유전형의 모의실험한다.

이때, 가계도안에서 표식유전자 유전형의 대립형질들은 대립형질의 모집단 빈도(population frequency)를 기초로 하여 할당되는데 각 대립형질의 빈도는 HWE(Hardy-Weinberg equilibrium)를 가정한다.

네 번째는 전단계에서 정의된 각 실 재조합률에서 가계도의 Z_{\max} (maximum lod score)를 계산한다.

마지막으로 Z_{\max} 분포의 기초통계량을 이용하여 가계도 정보충분성에 대한 두가지 정보를 측정할 수 있는데, 하나는 $E(Z_{\max})$ (Expected maximum lod score)이고 다른 한가지는 $E_{\max} \geq c$ (c : 특정상수)의 확률이다. 이를 가계도의 검정력이라 하는데 즉, 표식유전자와 질병유전자가 연관되어 있다고 할 경우 주어진 가계도의 자료로서 이것을 맞게 판단할 확률이다. 일반적으로 상염색체(autosomal)일 경우 $Z_{\max} \geq 3$ 이면 연관되어 있다고 보고 성염색체(X-chromosome)일 경우는 $Z_{\max} \geq 2$ 이면 표식유전자와 질병유전자가 연관되어 있다고 판단하는데 경우에 따라서 상수의 기준은 더 낮아질 수 있다 (Jurg ott, 1999).

또, 첫 번째 $E(Z_{\max})$ 는 Z_{\max} 의 표본평균에 의해 추정되고 두 번째 $Z_{\max} \geq c$ 의 확률은 Z_{\max} 의 표본비율 sample proportion에 의해 추정되어진다.

2.5.2 모형의 개요

모의실험 프로그램의 구성

앞 절에서 언급되었던 가계도의 검정력을 추정하기 위하여 SIMLINK(Version 4.12, 1997)프로그램이 사용되는데, 이는 Boehnke(1986)와 Ploughman, Boehnke(1989)이 제안한 이론을 바탕으로 하여 만들어진 프로그램으로 FORTRAN

에 의해 컴파일(compile)되어 MS-DOS상에서 실행이 된다. 이 프로그램은 가계도의 검정력을 추정하기 위해 크게 세 단계를 거치는데 첫 번째로는 질병유전자와 표식유전자의 유전형을 모의실험하고 다음 단계로는 Lod Scores나 Location Score(여러개의 표식유전자를 사용할 경우 사용됨)를 계산하고 마지막으로 이 정보들을 이용해 가계도의 정보 충분성 여부를 검정한다. 여기에서 Lod Scores나 Location Score의 계산은 1988년도에 만들어진 프로그램인 MENDEL(Lange et al., 1988)을 기반으로 하고 있다.

또한, 이 프로그램은 가계도의 검정력을 추정할 때 몇가지의 가정을 필요로 하는데 다음과 같다.

첫 번째로 질병유전자는 두 개의 대립형질로 구성되어야 하고 두 번째로는 가계도에서 각 구성원들의 관계(relationship)와 그들의 질병 표현형을 알아야 하고, 세 번째로 질병유전자의 유전형태(mode of inheritance)를 알아야 한다. 네 번째로는 표식유전자 유전형의 각 대립형질의 빈도는 HWE를 따른다고 가정하고 다섯 번째로 여러 표식유전자를 고려할 경우 표식유전자 사이의 유전적 거리(map distance)는 Haldane(1919)이 제안한 사상함수(mapping function)에 의해서 모의실험 된다고 가정한다. 마지막으로 조사된 가계도에는 일란성 쌍둥이(MZ-twins)가 없다고 가정한다.

2.5.3 프로그램에서의 입력과 출력자료

입력자료

SIMLINK에서의 입력자료는 크게 제어파일(control file), 로쿠스파일(locus file), 가계도파일(pedigree file)로 구성되는데 각 파일은 다음과 같이 구성된다.

첫째로 제어 파일에는 검정력 추정을 위한 일반적인 정보를 기술하게 되는데 우선 첫 번째 열(row)에는 다음과 같은 제어정보(control information)를 기술하게 된다.

모의실험의 반복 횟수, 사용되는 표식유전자의 수, 질병유전자의 침투함수 (penetrance function)의 양상 즉, 질병유전자의 표현형이 질적인가 (qualitative) 양적인가(quantitative)를 판단하고 질적인 양상을 보이면 선형 함수인가 누적정규함수인가를 결정한다. 또, 질병유전자와 표식유전자사이의 실재조합률(true recombination fraction)의 수와 질병유전자의 수에 따른 동질성(homogeneity)와 이형성(heterogeneity) 여부들을 기술한다.

다음으로 두 번째 열에는 실재조합률이나 유전적 거리에 관한 정보를 기술하게 되는데, 단일 표식유전자를 사용하는 경우에는 실재조합률을, 두 개의 표식유전자를 사용하는 경우에는 두 표식유전자 사이의 유전적거리를 언급해 준다. 그러므로 전자의 경우에는 결정된 각 실재조합률하에서 Lod Score를 계산하고 후자의 경우는 유전적 거리를 이용하여 Location Score를 계산하게 된다.

세 번째 열부터는 질병유전자 유전형의 대립형질 각 조합에서의 발병되는 최대, 최소 나이와 이때의 침투율 같은 질병유전자의 침투함수에 따른 정보를 보여주게 된다.

마지막으로 성별의 표시방법과 질병유전자 로쿠스(locus)의 이름, 로쿠스 파일명, 가계도 파일명등을 기술해 주어야 한다.

두 번째로 로쿠스 파일에는 검정력 추정시 사용되는 유전적 정보가 있어야 하는데, 우선 첫 번째 열에는 질병유전자의 다음과 같은 유전적 정보가 언급된다. 질병유전자의 로쿠스의 이름, 이것은 제어파일에서 언급되었던 이름과 동일해야 한다. 또, 질병유전자가 상염색체에 위치하는가 성염색체 위에 위치하는가의 여부와 각 대립형질의 수, 대립형질에 따른 질병유전자의 표현형의 수를 알아야 한다.

두 번째 열부터는 첫 번째 열에서 정의되었던 질병유전자의 각 대립형질의 이름과 모집단빈도를 언급해 주어야 하고 이 대립형질에 따라서 나타날 수

있는 질병유전자의 유전형을 기술해 주어야 한다. 이때 질병유전자의 대립형질 중 질병을 일으키는 대립유전자가 무엇인가와 질병을 일으키는 대립형질이 우성형인지 열성형인지를 고려해야 한다.

표식유전자의 유전적 정보도 위의 경우와 같이 표식유전자의 이름, 표식유전자의 위치(상염색체 또는 성염색체), 각 표식유전자의 대립형질의 수 등을 알아야 한다.

마지막으로 가계도 파일에서는 실제 검정력 추정에서 사용되는 가계도에 대한 정보가 포함되어 있는데, 여러 개의 가계도를 사용할 경우 순차적으로 가계도별로 가계도의 이름과 구성원 수를 언급해 주고 그 다음 열에는 각 구성원의 ID, 부모의 ID, 제어파일에서 언급해 주었던 성별 표시 방법에 따른 성별구분, 구성원의 질병 감염여부, 또는 질병유전자의 보유여부, 불완전 침투일 경우 나이 같은 침투함수의 변수 등을 기술해 준다.

출력자료

제어파일, 로쿠스파일, 가계도파일에 입력된 정보들을 가지고 가계도의 정보충분성에 대한 검정력을 추정하게 되는데, 크게 여섯가지 결과로 나누어 볼 수 있다.

첫 번째 결과에는 모의 실험에 사용된 정보들을 요약해 언급해 주는데, 질병유전자 로쿠스(locus)의 이름, 모의실험 반복횟수, 실재조합률, 실험재조합률(test recombination), 유전적 거리등을 보여준다.

두 번째 결과로는 가정해 준 실 재조합률(또는 유전적 거리)에서 가계도와 각 가계도 합이 최대 Lod Score(또는 Location Score)값의 평균(mean)와 표준오차(standard error), 최대값을 보여준다.

세 번째 결과에서는 실 재조합확률에서 각 가계도와 그들의 합이 최대 Lod Score(또는 Location Score)가 특정상수보다 클 확률 즉, $P(Z_{\max} \geq c)$ 값을 보여준다. 이때 특정상수 c 의 값으로는 0.5, 1.0, 1.5, 2.0, 2.5, 3.0의 값이 사용된다.

네 번째 결과로는 질병유전자가 두 개의 표식유전자 사이에 있을 평균확률을 추정

해 주는데 이 결과는 단일 표식유전자(single genetic marker)를 사용하거나 또는 하나의 질병 유전자를 가정했을 경우에는 생략된다.

다섯 번째 결과는 같은 염색체 상에 존재하며 함께 유전되는 경향이 있는 유전자 집단에 포함되지 않은 비연관 표식유전자(unlinked marker)를 가정하고 각 시험 재조합률에서 Lod Score(또는 Location Score)의 평균, 표준오차, 최소값, 최대값을 추정해 준다. 여기에서는 비연관 표식유전자가 있을 영역의 추정에 초점을 맞추는데 보통 Lod Score 값이 -2.0보다 작으면 비연관 영역에 있다고 추정된다(Morton, 1995) 마지막 결과도 비연관 표식유전자를 가정하고 Lod Score(또는 Location Score)가 특정 상수보다 클 확률을 추정하는데, 여기에서의 특정상수는 -2.0, -1.0, 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0의 값이 사용된다.

제 3장 가계도 자료의 이해

3.1 심혈관질환

심혈관질환, 특히 동맥경화의 진행으로 인한 허혈성 심질환의 발생빈도는 서구국가 뿐 아니라 우리나라에서도 날로 증가하고 있으며 단일 질환군으로 전 국민 의료비의 11%(년간 5,700 억원, 99년 통계청 자료)를 차지하여 국가경제에 큰 영향을 미친다. 또한 심혈관 질환은 발병 후 심각한 합병증 및 지속적인 치료가 요구되는 질환으로 질병의 예방이 중요하다. 고혈압 및 동맥경화성 질환등 심혈관계 질환은 생활 습관이나 환경적인 영향과 함께 유전적 요인에 의해 질병발생률에 차이를 보이고 있어 유전적 위험요인의 규명으로 고위험군을 예측하고 이들에 대한 교육 및 환경적 요인의 조절을 통한 질병발생의 예방이 중요하게 인식되고 있다.

최근 유전체관련 연구가 활성화되면서 유전체관련 정보들이 산업화로 이어져, 현재 세계 각국의 관심이 유전자의 특허권확보에 집중되고 있으며, 우리나라에서도 유전자 특허가 가속화되고 있으나 출원인의 60%이상이 외국인으로 국내 유전체 정보의

보호가 필요한 실정이다. 이러한 인종간에 차이를 보이고 있어 국내 유전자원 및 산업의 보호를 위해서는 국내 연구진에 의한 연구결과의 도출이 필요하며 이러한 정보는 제약업계를 비롯한 산업 분야에까지 확대되어 적용될 수 있을 것이다.

국민보건의 측면에서도 유전체연구의 결과는 심혈관질환의 발생에서의 유전적 요인을 밝혀냄으로써 질병의 발생을 예방, 조기 진단하고 유전체 기능연구로 새로운 치료법의 개발로 질병치료수준의 향상 등을 기대할 수 있다.

3.2 심혈관질환을 일으키는 유전자

심혈관 질환을 일으키는 질병유전자는 여러 종류가 있지만 이중 19번 염색체 위에 있는 APOE(Apolipoprotein E) 유전자는 LDL(low-density) 콜레스테롤 수치를 상승시켜 심혈관질환을 일으키는 대표적인 유전자로 알려져 있다(Carlos Lahoz et al, 2001). 이 APOE 유전자의 타입은 E2, E3, E4의 3종류이며 그 발현은 모계에게서 받은 유전자/부계에게서 받은 유전자가 E2/E2, E3/E3, E4/E4, E2/E3, E3/E4, E2/E4의 6가지 유전자형으로 나타난다. 이들 중에서 E4형을 가진 사람은 그렇지 않은 사람에 비해서 심혈관질환 걸릴 확률이 높다고 보고되고 있다. E4/E4는 아버지 어머니에게서 모두 E4타입의 유전자를 물려받았으므로 심혈관질환 걸릴 확률이 다른 유전자 타입에 비해 매우 높은 편이다. 또한 동맥경화 등의 성인병에 걸리기 쉽고 면역력이 약해서 어려서부터 잔병에 걸릴 가능성이 있다. 아이에게 치매가 걸릴 가능성이 높게 나타나면 부모는 이를 미리 예방할 수 있는 생활습관을 길러주는 것이 중요합니다. E2/E4, E3E4는 부모님 중 한 분으로부터 E4타입의 유전자를 물려받았으며 E4/E4타입보다는 심혈관질환 발생율이 높지 않지만 유전적으로 심혈관질환 발병할 가능성이 있으므로 주의해야 한다. E2/E2, E2E3, E3/E3는 유전적으로 심혈관질환에 걸릴 가능성이 없는 유전자를 가지고 있다. 또한 이러한 유전자타입은 면역기능도 우수하여 잔병치레가 거의 없는 건강체질이다.



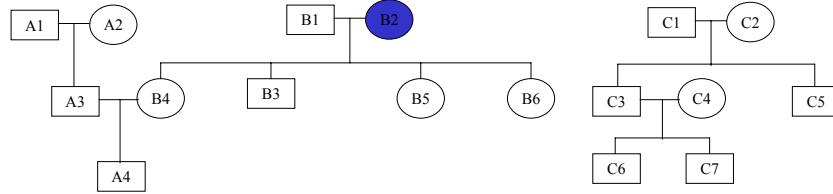
그림 3. 심혈관질환 관련 유전자, ApoE gene의 타입별 성향

APOE의 각 대립형질의 모집단 빈도는 인종에 따라 약간씩 다르지만 대규모의 Framingham 연구에 따르면 E3:0.802, E4:0.119, E2:0.079으로 나타난다(Carlos Lahoz et al, 2001).

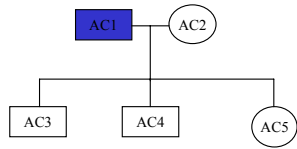
3.3 가계도 자료의 수집기준과 특성

심혈관질환 가계도 자료에 대한 이해와 특성을 파악하기 위해 자료수집기준을 알아보면 다음과 같다. 고혈압에 걸린 환자를 대상으로 나이는 35세에서 75세 사이이고 3회이상 혈압측정과 관련 임상검사로 확진된 경우이며 확진이후 최소 1년이상 약물치료를 받고 있으며 국내 4개 대학병원에 내원한 환자를 대상으로 한다. 자녀가 고혈압이라고 판단한 기준은 수축기혈압이 140mmHg 이상이거나 이완기 혈압이 90mmHg 이상이고 antihypertensive medication 기준으로 잡는다. 가족구성원 검진대상자는 나이가 7세 이상이며 남녀 모두 해당하고 가족 중 부모, 형제, 4촌이내 고혈압환자가 2인 이상인 가계이고 3세대 이상의 대가족으로 구성된다. CEPH형태의 가계도의 핵가족(부모와 2명의 자녀로 구성된 4인이상)이며 환자가 50-60대인 경우 고혈압 자녀나 형제를 검진대상에 우선적으로 포함하고 환자가 30-40대인 경우 부모를 우선 검진대상에 포함한다. 제외가족은 당뇨병나 기타 내분비이상에 의해 유발된 2차성 고혈압환자이거나 BMI(body mass index)가 20이하 35이상인 경우, 일상적인 생활이 불가능한 경우이다. 가계도 자료 형태를 보면 수가 145개이며 총인원은 565명이고 가계도의 예를 그려보면 다음과 같다.

FC046 (10)



FC087 (23)



FC522 (93)

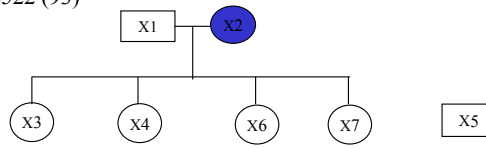


그림 4. 가계도 예시

제 4장 가계도 자료의 정보충분성 추정 및 비교

가계도 자료의 정보충분성을 알아보기 위해 모의실험 프로그램 SIMLINK을 이용한다. 각 개인의 질병유전자와 표식유전자의 유전자형을 모의실험하여 연관성분석시 정보충분성을 추정한다. 즉, 질병유전자 APOE의 우성형질은 E4, 열성형질은 E2, E3이며 E4의 유전확률은 0.119이고 E2, E3의 유전확률은 0.881이다. 이렇게 조사된 정보를 가지고 각 개인의 질병유전자와 표식유전자의 유전자형을 모의실험한다. 19번 염색체(Genebank)위에 있는 유전적으로 떨어진 거리가 다른 표식유전자 D19S217, D19S219, D19S223, D19S596, D19S425, D19S927 6개를 이용하여 심혈관질환을 일으키는 APOE 유전자와의 연관성 여부에 대한 가계도 자료의 검정력을 모의실험으로 추정한다. 모의 실험의 반복횟수는 500번으로 한다. 이 모의실험을 통하여 조사된 가계도 자료가 연관성 분석을 하기에 충분한 정보를 가지고 있는가를 판단하고 어떠한 표식유전자가 연관성 분석 시 좀 더 효율적으로 사용될 수 있는지 비교해 본다.

4.1 표식유전자 종류별 정보충분성 추정 및 비교

4.1.1 표식유전자의 대립형질 빈도가 동일한 1개 표식유전자를 사용한 가계도 자료의 경우 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 결과

표 2. 대립형질의 빈도가 동일한 공우성형질일 때 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$

정보량	개수	평균	표준편차
$E(Z_{\max})$	145	0.070986	0.08966
$P(Z_{\max} \geq C)$	145	0.047076	0.12217

위의 표를 보면 총 145개의 가계도 자료를 이용하여 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 를 각각 구해 평균과 표준편차를 나타낸다. 한 가계도를 이용해서는 연관성 분석시 정보충분성이 만족할 만한 수준이 아님을 보인다.

4.1.2 표식유전자의 대립형질 빈도가 다른 1개 표식유전자를 사용한 snp459와 snp461의 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 비교결과

19번 염색체위에 있는 유전적 거리가 동일하고 각 대립형질의 빈도가 틀린 표식유전자 snp459과 snp461을 이용하여 심혈관질환을 일으키는 APOE 유전자와의 연관성 여부에 대한 가계도의 검정력을 모의실험으로 추정한다. APOE와의 유전적 거리는 두 유전자 모두 1.4cM 떨어져 있고 snp459 유전자는 두 가지의 대립형질로 구성되어 있고 각 대립형질의 빈도는 0.49과 0.51로 구성된다. snp461 유전자도 snp459와 동일하게 두 가지의 대립형질으로 구성되어 있고 대립형질의 빈도는 각각 0.091, 0.909이다(Eden R. Martin et al. 2000). APOE와의 유전적 거리가 두 유전자 모두 1.4cM으로 동일하므로 재조합률을 0.01로 설정한다.

표식유전자 snp459과 snp461를 이용하여 유전적 거리가 동일하고 각 대립형질

의 빈도가 틀린 경우를 나누어 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 값에 차이가 있는지 검정해 보면 다음과 같다. 자료가 정규분포를 따르지 않으므로 비모수검정 윌콕슨 두 표본검정(Wilcoxon two-sample test)을 이용하였다.

표 3. 대립형질의 빈도가 다른 공우성형질일 때 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$

정보량	표식유전자	개수	평균	표준편차	t 의 p값
$E(Z_{\max})$	snp459	99	0.084263	0.10002	0.0001
	snp461	99	0.036010	0.04076	
$P(Z_{\max} \geq C)$	snp459	99	0.055818	0.14283	0.7063
	snp461	99	0.024626	0.06212	

위의 표를 보면 $E(Z_{\max})$ 는 유의수준 5%에서 영가설(평균차이가 없다)를 기각해서(p값=0.001) 대립형질의 빈도가 비슷한 snp459의 경우가 대립형질의 빈도가 다른 snp461의 경우보다 연관성 분석시 더 정보충분성을 만족한다. $P(Z_{\max} \geq C)$ 는 유의수준 5%에서 영가설을 기각하지 못한다(p값=0.7063). 가계도의 정보충분성이 부족하여 검정력이 거의 0에 가까우므로 평균검정차이를 드러내지 못하고 있다.

4.1.3 2개 표식유전자를 사용한 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 비교결과

19번 염색체 위에 있는 D19S217 D19S219, D19S223, D19S596, D19S425, D19S927(Genebank) 6개의 표식유전자를 사용하여 질병유전자 APOE와의 연관성에 대한 검정력을 추정해 본다. 이 절에서는 APOE를 사이에 두고 있는 두 표식유전자를 동시에 고려하여 검정력을 추정해 본다. 표식유전자와 APOE와의 거리는 아래 그림과 같다.

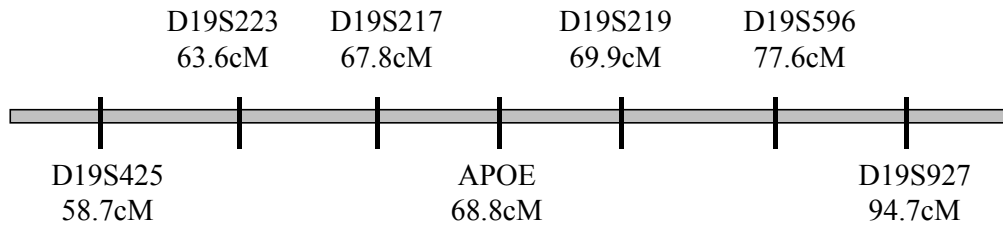


그림 5. APOE와 표식유전자 위치

표식유전자를 3쌍으로 나누어 [(D19S217, D19S219), (D19S223, D19S596), (D19S425, D19S927)] 연관성 분석에 관한 검정력을 추정해 보자.

표식유전자가 1개를 사용한 경우와 표식유전자를 2개를 사용한 경우 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 값에 차이가 있는지 검정해 보면 다음과 같다. 자료가 정규분포를 따르지 않으므로 비모수방법인 윌콕슨 두 표본검정(Wilcoxon two-sample test)를 이용한다. 2개 표식유전자, (D19S217, D19S219)를 이용해서 재조합률이 0.02인 경우를 고려해 본다. 이 경우 검정력은 $P(Z_{\max} \geq 0.5)$ 로 계산된 값이며 표식유전자가 한 개인 경우는 대립형질의 빈도가 동일, 공우성형질을 따른다, 재조합률이 0.0인 경우를 고려해 본다. 이 경우 검정력은 $\sum_{i=0.5}^3 P(Z_{\max} \geq i)$ 로 계산된 값이다.

표 4. 표식유전자 수에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교

정보량	표식유전자 수	개수	평균	표준편차	t 의 p값
$E(Z_{\max})$	1	99	0.085485	0.09774	0.0022
	2	99	0.155818	0.18244	
$P(Z_{\max} \geq C)$	1	99	0.054828	0.13807	0.6949
	2	99	0.117152	0.29955	

위의 표를 보면 $E(Z_{\max})$ 는 유의수준 5%에서 영가설(평균차이가 없다)를 기각해 서(p 값=0.0022) 2개 표식유전자를 고려한 경우가 1개 표식유전자를 고려한 경우보다 연관성 분석시 더 정보충분성을 만족한다. $P(Z_{\max} \geq C)$ 는 영가설을 기각하지 못한다(p 값=0.6949). 가계도 자료의 정보충분성이 부족하여 검정력이 거의 0에 가까우므로 평균검정차이를 드러내지 못하고 있다.

4.2 가계도 자료 형태 및 자료 수별 정보충분성 추정 및 비교

4.2.1 가계도 자료 형태에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교

가계도 자료의 형태를 가족의 수, 질병유무, 부모 질병유무, 자식 질병유무, 세대수 별로 분류하여 연관성 분석시 정보충분성을 비교해 보자. $E(Z_{\max})$ 는 재조합률이 0.0일 때의 값이고 $P(Z_{\max} \geq C)$ 는 재조합률이 0.0인 경우에 $\sum_{i=0.5}^3 P(Z_{\max} \geq i)$ 로 계산된 값이다.

가계도 자료 중 가족의 수에 따라 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 값이 차이가 있는지 검정해 보면 다음과 같다. 자료가 정규분포를 따르지 않으므로 비모수검정 윌콕슨 두 표본검정(Wilcoxon two-sample test)을 이용한다.

표 5. 가계도내 가족의 수에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교

정보량	가계도 종류	개수	평균	표준편차	t 의 p값
$E(Z_{\max})$	가족의 수 < 5	113	0.039257	0.04480	<0.0001
	가족의 수 \geq 5	32	0.183031	0.11602	
$P(Z_{\max} \geq C)$	가족의 수 < 5	113	0.000000	0.00000	<0.0001
	가족의 수 \geq 5	32	0.213313	0.18090	

위의 표를 보면 $E(Z_{\max})$ 나 $P(Z_{\max} \geq C)$ 모두 유의수준 5%에서 영가설(평균 차이가 없다)를 기각해서(p 값 <0.0001) 가족의 수가 5이상인 경우가 가족의 수가 5미만인 경우보다 연관성 분석시 더 정보충분성을 만족한다. 가족의 수를 재그룹하지 않은 상태에서 $E(Z_{\max})$ 와 가족의 수의 피어슨 상관계수를 구하면 0.829(p 값 < 0.0001)이고 $P(Z_{\max} \geq C)$ 와 가족의 수의 피어슨 상관계수를 구하면 0.6721(p 값 < 0.0001)를 나타내므로 가족의 수가 많을수록 연관성 분석시 더 정보충분성을 만족한다.

가계도 내 가족 중 질병이 있는 경우와 없는 경우를 나누어 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 값이 차이가 있는지 검정해 보면 다음과 같다. 자료가 정규분포를 따르지 않으므로 비모수검정 윌콕슨 두 표본검정(Wilcoxon two-sample test)을 이용한다.

표 6. 가계도내 가족 중 질병유무에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교

정보량	가계도 종류	개수	평균	표준편차	t 의 p값
$E(Z_{\max})$	가계도 내 질병이 있는 경우	99	0.085485	0.09774	0.0012
	가계도 내 질병이 없는 경우	46	0.039783	0.05883	
$P(Z_{\max} \geq C)$	가계도 내 질병이 있는 경우	99	0.054828	0.13807	0.8570
	가계도 내 질병이 없는 경우	46	0.030391	0.07634	

위의 표를 보면 $E(Z_{\max})$ 는 유의수준 5%에서 영가설(평균 차이가 없다)를 기각해서(p 값=0.0012) 가계도내 가족 중 한 명이라도 질병이 있는 경우가 가계도 내 가족 중 한명이라도 질병이 없는 경우보다 연관성 분석시 더 정보충분성을 만족한다. $P(Z_{\max} \geq C)$ 는 영가설을 기각하지 못한다(p 값=0.8570). 연관성 분석시 가계도 자료의 정보충분성이 부족하여 검정력이 거의 0에 가까우므로 평균검정차이를 드러내지

못하고 있다.

가계도 내 부모가 질병이 있는 경우와 없는 경우를 나누어 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 값이 차이가 있는지 검정해 보면 다음과 같다. 자료가 정규분포를 따르지 않으므로 비모수검정 윌콕슨 두 표본검정(Wilcoxon two-sample test)을 이용한다.

표 7. 가계도 내 부모 질병유무에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교

정보량	가계도 종류	개수	평균	표준편차	t 의 p값
$E(Z_{\max})$	부모가 질병이 있다	96	0.087448	0.09845	0.0004
	부모가 질병이 없다	49	0.038735	0.05768	
$P(Z_{\max} \geq C)$	부모가 질병이 있다	96	0.028531	0.07428	0.6823
	부모가 질병이 없다	49	0.056542	0.13988	

위의 표를 보면 $E(Z_{\max})$ 는 유의수준 5%에서 영가설(평균차이가 없다)를 기각해서(p값=0.0004) 가계도 내 부모가 질병이 있는 경우가 가계도 내 부모가 질병이 없는 경우보다 연관성 분석시 더 정보충분성을 만족한다. $P(Z_{\max} \geq C)$ 는 영가설을 기각하지 못한다(p값=0.6823). 연관성 분석시 가계도 자료의 정보충분성이 부족하여 검정력이 거의 0에 가까우므로 평균검정차이를 드러내지 못하고 있다.

가계도 내 자손이 질병이 있는 경우와 가계도 내 자손이 질병이 없는 경우를 나누어 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 값이 차이가 있는지 검정해 보면 다음과 같다. 자료가 정규분포를 따르지 않으므로 비모수검정 윌콕슨 두 표본검정(Wilcoxon two-sample test)을 이용한다.

표 8. 가계도 내 자손 질병유무에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교

정보량	가계도 종류	개수	평균	표준편차	t 의 p값
$E(Z_{\max})$	자손이 질병이 있다	5	0.048400	0.04513	0.6020
	자손이 질병이 없다	140	0.071793	0.09083	
$P(Z_{\max} \geq C)$	자손이 질병이 있다	5	0.000000	0.00000	0.3209
	자손이 질병이 없다	140	0.048757	0.12401	

위의 표를 보면 $E(Z_{\max})$ 나 $P(Z_{\max} \geq C)$ 모두 유의수준 5%에서 영가설(평균 차이가 없다)을 기각하지 못한다(p값=0.6020, 0.3209). 자손이 질병이 있는 경우가 5개밖에 안되고 연관성 분석시 가계도 자료의 정보충분성이 부족하여 검정력이 거의 0에 가까우므로 평균검정차이를 드러내지 못하고 있다.

가계도 자료가 3세대에 걸쳐 분포된 경우와 2세대에 걸쳐 있지만 자손의 수가 많은 경우로 나누어 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 의 값이 차이가 있는지 검정해 보면 다음과 같다. 다른 요인에 의한 영향을 제어하고자 가족의 수는 5-7인 자료만 사용한다. 자료가 정규분포를 따르지 않으므로 비모수검정 일콕슨 두 표본검정(Wilcoxon two-sample test)을 이용한다.

표 9. 가계도 자료의 세대수에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교

정보량	가계도 종류	개수	평균	표준편차	t 의 p값
$E(Z_{\max})$	3세대에 걸쳐 분포	4	0.055000	0.06998	0.1542
	2세대에 걸쳐 분포	4	0.230750	0.15371	
$P(Z_{\max} \geq C)$	3세대에 걸쳐 분포	4	0.018000	0.03600	0.1680
	2세대에 걸쳐 분포	4	0.294000	0.26802	

위의 표를 보면 $E(Z_{\max})$ 나 $P(Z_{\max} \geq C)$ 모두 유의수준 5%에서 영가설(평균

차이가 없다)를 기각하지 못하지만(p 값=0.1542, 0.1680) 한쪽 검정시 p 값이 0.0771, 0.0840이므로 2세대에 걸쳐서 자손의 수가 많은 경우가 3세대에 걸쳐 있는 경우보다 연관성 분석시 더 정보충분성을 만족한다고 볼 수도 있지 않을까 한다.

4.2.2 가계도 수에 따른 $E(Z_{\max})$ 와 $P(Z_{\max} \geq C)$ 비교

가계도 자료 형태별 $E(Z_{\max})$ 나 $P(Z_{\max} \geq C)$ 의 특성을 보면 가족의 수가 많고 가계도 내 가족 중 질병이 있으며 부모가 질병이 있는 경우 연관성 분석시 더 정보충분성을 만족한다. 그래서 145개의 가계도 자료 중 가족의 수가 5-7개이고 질병이 있는 가계도 18개를 추출하여 $E(Z_{\max})$ 나 $P(Z_{\max} \geq C)$ 를 구해보면 다음과 같다.

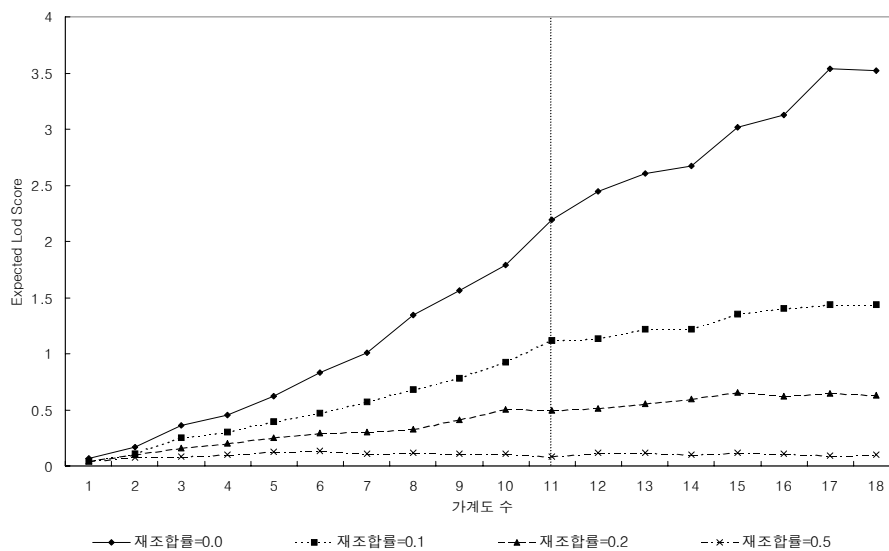


그림 6. 가계도 수에 따른 $E(Z_{\max})$

위의 그림을 보면 재조합률이 0.0일때는 가계도 수가 늘어남에 따라 $E(Z_{\max})$ 의 값이 계속 올라가지만 재조합률이 0.1이상인 경우는 가계도 수가 11이상이면 $E(Z_{\max})$ 의 값이 머물고 있음을 보인다. $E(Z_{\max})$ 의 값을 보고 가계도 수를 정하려

면 11만 사용하여도 연관성 분석시 정보충분성을 어느정도 만족하면서 비용적, 시간적인 면을 고려할 수 있다.

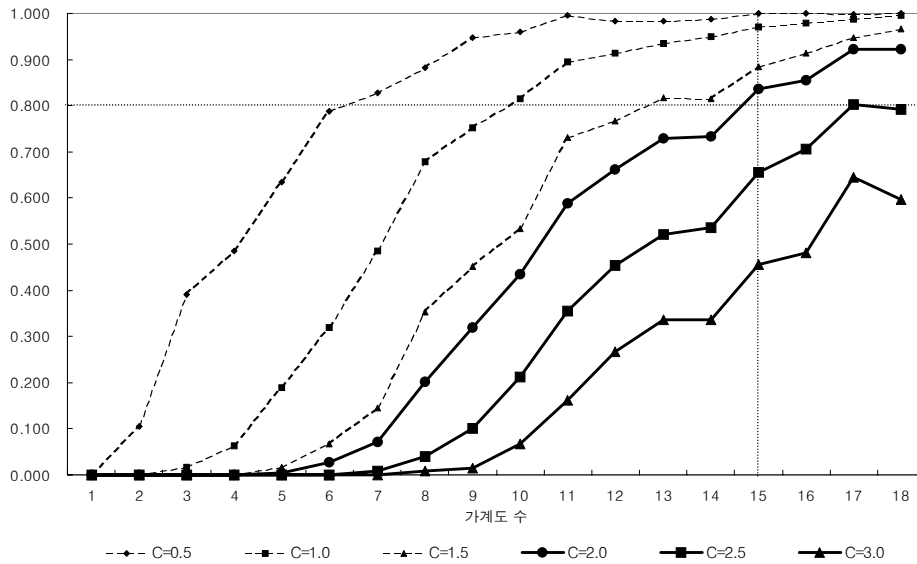


그림 7. 가계도 수에 따른 $P(Z_{\max} \geq C)$

위의 그림을 보면 기각역 C 가 2일 때 $P(Z_{\max} \geq C)$ 가 80%를 만족하려면 가계도 수가 15개정도만 있으면 된다. 기각역 C 가 2.5이거나 3이어도 가계도 수가 20개 정도 이면 $P(Z_{\max} \geq C)$ 가 60~70%이다.

따라서 $E(Z_{\max})$ 과 $P(Z_{\max} \geq C)$ 를 고려해 볼 때 가계도 수가 15개 정도 사용하여도 연관성 분석시 정보충분성을 어느정도 만족하면서 비용적, 시간적인 면을 고려할 수 있다.

제 5장 결론

연관성 분석시 표식유전자가 한 개인 경우 대립형질의 빈도가 비슷할 경우가 대립형질의 빈도가 다른 경우보다 더 정보충분성을 만족한다. 그리고 표식유전자를 두 개를 사용한 경우가 표식유전자를 한 개 사용한 경우보다 연관성 분석시 더 정보충분성을 만족한다.

가계도 형태는 가족의 수가 많이 있는 경우가 월등히 연관성 분석시 더 정보충분성을 만족한다. 가계도 내에 가족 중 질병이 있는 경우가 질병이 없는 경우보다, 부모에게 질병이 있는 경우가 부모에게 질병이 없는 경우보다 연관성 분석시 더 정보충분성을 만족한다. 그 외 자손에게 질병이 있는 경우와 자손에게 질병이 없는 경우, 가족의 수는 비슷하지만 3세대 이상으로 구성된 경우와 2세대 이하에서 넓게 퍼진 경우를 비교해 보았지만 연관성 분석시 정보충분성의 차이를 드러내지 못하고 있다.

표식유전자가 1개이고 대립형질의 빈도수가 동일하고 공우성형질을 따르고 각각역 C를 2.0으로 할 때, 가계도 수를 15개를 가지고 연관성 분석시 정보충분성을 고려해 보는 것이 검정력(Power)을 80%정도 유지하고 비용적, 시간적인 면에서 효율적이라 본다. 각각역 C값이 2.5, 3인 경우에도 가계도 수를 20개 정도를 가지고 연관성 분석을 하면 검정력(Power)이 60~70%정도 유지된다.

그러므로 연관성 분석시 표식유전자의 대립형질 빈도가 동일하고 표식유전자의 수를 여러개 사용한다. 가계도의 자료는 가계도 내 가족의 수가 많고 질병을 앓고 있는 환자가 한명 이상 보유하며 부모가 질병을 앓고 있는 경우가 좋겠다. 가계도 수는 15~20개정도 수집한다. 이처럼 가계도 자료를 구성하여 연관성 분석을 한다면 이전에 하던 방식대로 무조건 많은 자료를 사용하는 것보다 시간적, 비용적인 면에서 좀 더 효율적이라고 본다.

제 6장 토론

이미 주어진 모의실험 프로그램인 SIMLINK를 사용하였기에 많은 제한이 있다. 침투율 함수를 살펴보면 질적형질, 즉 이형형질을 위한 선형함수(piece-wise linear function)만 사용했는데 그 외 누적정규침투함수(cumulative normal penetrance function)나 양적 형질을 위한 혼합정규분포(a mixture of normal distribution)도 고려해 보아야 한다.

표식유전자의 수를 두개까지 적용할 수 있는데 더 많은 표식유전자를 고려할 수 있어야 한다.

사상함수(map function)을 보면 헬덴의 사상함수(1919)를 이용하는데 그 외 사상함수, 코삼비 사상함수(1944), Carter and Falconer 사상함수(1951), Felsenstein 사상함수(1979), Sturt 사상함수(1976), Binomial 사상함수(Karlin, 1984)등을 사용할 방법을 고려해 보아야 한다.

여러 상황을 고려할 수 있는 프로그램 구현이 좀 더 현실상황에 맞게 적용할 수 있으리라 본다.

참고 문헌

- 박윤주 (2001), 유전적 연관성 분석을 위한 가계도 자료의 정보 충분성에 관한 연구, 연세대학교 석사학위논문.
- 통계청 (1999), 사망원인통계연보.
- Boehnke, M. (1986), Estimating the power of a proposed linkage study. *Am. J. Hum.* 39: 513-527.
- Boehnke, M., Ploughman, L. M. (1989), Estimating the power of a proposed linkage study for a complex genetic trait. *Am. J. Hum.* 44: 543-551.
- Carlos Lahoz. et al. (2001), Apolipoprotein E genotype and cardiovascular disease in the framingham heart study. *Atherosclerosis* 154: 529-537.
- Carter, T. C., Falconer, D. S. (1951), Stocks for detecting linkage in the mouse and the theory of their design. *J. Genet.* 50: 307-323.
- Eden, R., Martin. et al. (2000), SNPing away at complex disease : analysis of single-nucleotide polymorphisms around apoe in alzheimer disease. *Am. J. Hum.* 67: 383-394.
- Eun Kyoung Im (2002), Linkage analysis between Apo AI-CIII-AIV cluster and Plasma Lipid Levels in Cardiovascular Disease Family, Brain Korea 21 Project for Medical Sciences. The Graduate School of Yonsei University.
- Felsenstein, J. (1979), A mathematically tractable family of genetic mapping functions with different amounts of interference. *Genetics* 91: 769-775.
- Haldane, J. B. S. (1991), The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* 8:

299-309.

Joseph, Terwilliger, Jurg, Ott (1994), Handbook of Human Genetic Linkage. Johns Hopkins University Press Baltimore.

Jurg, Ott (1999), Analysis of Human Genetic linkage. The Johns Hopkins University Press Baltimore and London.

Karlin, S. (1984), Theoretical aspects of genetic map functions in recombination process. In Human Population Genetics. The Pittsburgh Symposium, edited by A Chakravarti. New York: Van Nostrand Reinhold Company, pp 209-228.

Kenneth, Lange (1997), Mathematical and Statistical Methods for Genetic Analysis. Springer-Verlag New York, Inc.

Kosambi, D. D. (1944), The estimation of map distances from recombination values. Ann. Eugen. 12: 172-175.

Lander, E. S., D. Botstein (1986a), Strategies for studying heterogeneous genetic trait in humans by using a linkage map of restriction fragment length polymorphisms. Proc. Natl. Acad. Sci. USA. 83: 7353-7357.

Lange, K., Boehnke, M. (1983), Extension to pedigree analysis. V. optimal calculation of mendelian likelihoods. Hum. Hered. 33: 291-301.

Lange, K., Boehnke, M. (1988), Weeks D. Programs for pedigree analysis : MENDEL, FISHER and dGENE. Genetic Epidemiology 5: 471-472.

Morgan, T. H. (1928), The Theory of Genes. New Haven. Yale University Press.

Morton, N. E. (1995), Sequential tests for the detection of linkage. Am. J. Hum. 7: 277-318.

Sturt, E. (1976), A mapping function for human chromosomes. Ann. Hum.

Genet. 40: 147-163.

Sturtevant, A. H. (1913), The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J. Exp. Zool.* 14: 43-59.

관련된 인터넷 사이트 주소

http://www.dnantech.com/dna/dna_apoe.asp

<http://linkage.rockefeller.edu/soft/simlink.html>

<http://www.heartgenome.org>

<http://www.ncbi.nlm.nih.gov/genemap99>

부 록

표 10. 대립형질의 빈도가 동일한 경우 1~35개 가계도의 $E(Z_{\max})$

가계도	재조합률			
	0.000	0.100	0.200	0.500
1	0.000	0.000	0.000	0.000
2	0.094	0.073	0.065	0.052
3	0.068	0.043	0.054	0.035
4	0.074	0.043	0.040	0.037
5	0.000	0.000	0.000	0.000
6	0.085	0.079	0.057	0.052
7	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000
10	0.399	0.266	0.171	0.088
11	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000
15	0.090	0.077	0.063	0.038
16	0.094	0.064	0.050	0.040
17	0.096	0.077	0.059	0.046
18	0.097	0.071	0.065	0.049
19	0.095	0.082	0.050	0.044
20	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000
22	0.094	0.070	0.061	0.036
23	0.195	0.119	0.129	0.051
24	0.101	0.080	0.061	0.051
25	0.000	0.000	0.000	0.000
26	0.087	0.076	0.060	0.044
27	0.088	0.065	0.059	0.053
28	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000
31	0.092	0.075	0.061	0.047
32	0.094	0.069	0.059	0.043
33	0.000	0.000	0.000	0.000
34	0.182	0.124	0.104	0.056
35	0.193	0.157	0.119	0.059

표 11. 대립형질의 빈도가 동일한 경우 36~70개 가계도의 $E(Z_{\max})$

가계도	재조합률			
	0.000	0.100	0.200	0.500
36	0.201	0.126	0.101	0.071
37	0.094	0.079	0.060	0.047
38	0.100	0.073	0.060	0.044
39	0.078	0.074	0.054	0.049
40	0.100	0.072	0.063	0.042
41	0.094	0.067	0.055	0.046
42	0.234	0.168	0.117	0.097
43	0.000	0.000	0.000	0.000
44	0.000	0.000	0.000	0.000
45	0.295	0.213	0.117	0.049
46	0.080	0.083	0.057	0.042
47	0.000	0.000	0.000	0.000
48	0.543	0.377	0.221	0.077
49	0.094	0.081	0.060	0.052
50	0.101	0.083	0.064	0.045
51	0.073	0.069	0.061	0.046
52	0.192	0.146	0.089	0.046
53	0.096	0.082	0.070	0.048
54	0.000	0.000	0.000	0.000
55	0.000	0.000	0.000	0.000
56	0.087	0.069	0.066	0.053
57	0.094	0.076	0.070	0.043
58	0.323	0.213	0.127	0.044
59	0.086	0.075	0.060	0.048
60	0.000	0.000	0.000	0.000
61	0.417	0.255	0.171	0.050
62	0.090	0.085	0.060	0.050
63	0.080	0.083	0.058	0.046
64	0.094	0.075	0.067	0.047
65	0.093	0.076	0.065	0.040
66	0.000	0.000	0.000	0.000
67	0.091	0.078	0.056	0.050
68	0.000	0.000	0.000	0.000
69	0.089	0.067	0.064	0.044
70	0.091	0.081	0.058	0.050

표 12. 대립형질의 빈도가 동일한 경우 71~107개 가계도의 $E(Z_{\max})$

가계도	재조합률			
	0.000	0.100	0.200	0.500
71	0.000	0.000	0.000	0.000
72	0.094	0.074	0.058	0.052
73	0.000	0.000	0.000	0.000
74	0.091	0.072	0.061	0.047
75	0.103	0.078	0.060	0.038
76	0.192	0.134	0.114	0.058
77	0.000	0.000	0.000	0.000
78	0.000	0.000	0.000	0.000
79	0.000	0.000	0.000	0.000
80	0.000	0.000	0.000	0.000
81	0.000	0.000	0.000	0.000
82	0.166	0.127	0.109	0.054
83	0.093	0.066	0.062	0.044
84	0.102	0.074	0.063	0.056
85	0.000	0.000	0.000	0.000
86	0.000	0.000	0.000	0.000
87	0.000	0.000	0.000	0.000
88	0.087	0.075	0.063	0.043
89	0.094	0.071	0.057	0.046
90	0.099	0.081	0.060	0.041
91	0.092	0.070	0.056	0.045
92	0.000	0.000	0.000	0.000
93	0.259	0.188	0.144	0.044
94	0.203	0.141	0.085	0.050
95	0.000	0.000	0.000	0.000
96	0.095	0.070	0.064	0.051
97	0.183	0.138	0.122	0.068
98	0.080	0.048	0.040	0.032
99	0.102	0.082	0.059	0.051
100	0.075	0.049	0.041	0.032
101	0.000	0.000	0.000	0.000
102	0.078	0.044	0.047	0.034
103	0.071	0.058	0.043	0.041
104	0.000	0.000	0.000	0.000
105	0.000	0.000	0.000	0.000
106	0.080	0.057	0.035	0.043
107	0.000	0.000	0.000	0.000

표 13. 대립형질의 빈도가 동일한 경우 108~145개 가계도의 $E(Z_{\max})$

가계도	제조합률			
	0.000	0.100	0.200	0.500
108	0.000	0.000	0.000	0.000
109	0.074	0.052	0.035	0.033
110	0.000	0.000	0.000	0.000
111	0.146	0.093	0.070	0.061
112	0.000	0.000	0.000	0.000
113	0.000	0.000	0.000	0.000
114	0.124	0.088	0.043	0.037
115	0.000	0.000	0.000	0.000
116	0.132	0.086	0.044	0.051
117	0.000	0.000	0.000	0.000
118	0.000	0.000	0.000	0.000
119	0.067	0.055	0.030	0.039
120	0.000	0.000	0.000	0.000
121	0.000	0.000	0.000	0.000
122	0.000	0.000	0.000	0.000
123	0.000	0.000	0.000	0.000
124	0.000	0.000	0.000	0.000
125	0.141	0.074	0.053	0.031
126	0.000	0.000	0.000	0.000
127	0.000	0.000	0.000	0.000
128	0.000	0.000	0.000	0.000
129	0.000	0.000	0.000	0.000
130	0.000	0.000	0.000	0.000
131	0.145	0.078	0.039	0.041
132	0.000	0.000	0.000	0.000
133	0.000	0.000	0.000	0.000
134	0.000	0.000	0.000	0.000
135	0.067	0.041	0.036	0.028
136	0.063	0.052	0.043	0.035
137	0.219	0.132	0.092	0.054
138	0.067	0.044	0.028	0.035
139	0.000	0.000	0.000	0.000
140	0.000	0.000	0.000	0.000
141	0.136	0.094	0.073	0.059
142	0.000	0.000	0.000	0.000
143	0.000	0.000	0.000	0.000
144	0.000	0.000	0.000	0.000
145	0.145	0.097	0.074	0.064

표 14. 대립형질의 빈도가 동일한 경우 1~35개 가계도의 $P(Z_{\max} \geq C)$

가계도	제조합률 : 0.000					
	C					
	0.5	1.0	1.5	2.0	2.5	3
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000
10	0.370	0.100	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.000	0.000	0.000
19	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000
23	0.292	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	0.000	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	0.000	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000
31	0.000	0.000	0.000	0.000	0.000	0.000
32	0.000	0.000	0.000	0.000	0.000	0.000
33	0.000	0.000	0.000	0.000	0.000	0.000
34	0.274	0.000	0.000	0.000	0.000	0.000
35	0.278	0.000	0.000	0.000	0.000	0.000

표 15. 대립형질의 빈도가 동일한 경우 36~70개 가계도의 $P(Z_{\max} \geq C)$

가계도	제조합률 : 0.000					
	C					
	0.5	1.0	1.5	2.0	2.5	3
36	0.300	0.000	0.000	0.000	0.000	0.000
37	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000
42	0.148	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	0.000	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000
45	0.316	0.000	0.000	0.000	0.000	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000	0.000	0.000
48	0.400	0.256	0.054	0.000	0.000	0.000
49	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000
51	0.000	0.000	0.000	0.000	0.000	0.000
52	0.284	0.000	0.000	0.000	0.000	0.000
53	0.000	0.000	0.000	0.000	0.000	0.000
54	0.000	0.000	0.000	0.000	0.000	0.000
55	0.000	0.000	0.000	0.000	0.000	0.000
56	0.000	0.000	0.000	0.000	0.000	0.000
57	0.000	0.000	0.000	0.000	0.000	0.000
58	0.354	0.000	0.000	0.000	0.000	0.000
59	0.000	0.000	0.000	0.000	0.000	0.000
60	0.000	0.000	0.000	0.000	0.000	0.000
61	0.384	0.260	0.000	0.000	0.000	0.000
62	0.000	0.000	0.000	0.000	0.000	0.000
63	0.000	0.000	0.000	0.000	0.000	0.000
64	0.000	0.000	0.000	0.000	0.000	0.000
65	0.000	0.000	0.000	0.000	0.000	0.000
66	0.000	0.000	0.000	0.000	0.000	0.000
67	0.000	0.000	0.000	0.000	0.000	0.000
68	0.000	0.000	0.000	0.000	0.000	0.000
69	0.000	0.000	0.000	0.000	0.000	0.000
70	0.000	0.000	0.000	0.000	0.000	0.000

표 16. 대립형질의 빈도가 동일한 경우 71~107개 가계도의 $P(Z_{\max} \geq C)$

가계도	재조합률 : 0.000					
	C					
	0.5	1.0	1.5	2.0	2.5	3
71	0.000	0.000	0.000	0.000	0.000	0.000
72	0.000	0.000	0.000	0.000	0.000	0.000
73	0.000	0.000	0.000	0.000	0.000	0.000
74	0.000	0.000	0.000	0.000	0.000	0.000
75	0.000	0.000	0.000	0.000	0.000	0.000
76	0.286	0.000	0.000	0.000	0.000	0.000
77	0.000	0.000	0.000	0.000	0.000	0.000
78	0.000	0.000	0.000	0.000	0.000	0.000
79	0.000	0.000	0.000	0.000	0.000	0.000
80	0.000	0.000	0.000	0.000	0.000	0.000
81	0.000	0.000	0.000	0.000	0.000	0.000
82	0.236	0.000	0.000	0.000	0.000	0.000
83	0.000	0.000	0.000	0.000	0.000	0.000
84	0.000	0.000	0.000	0.000	0.000	0.000
85	0.000	0.000	0.000	0.000	0.000	0.000
86	0.000	0.000	0.000	0.000	0.000	0.000
87	0.000	0.000	0.000	0.000	0.000	0.000
88	0.000	0.000	0.000	0.000	0.000	0.000
89	0.000	0.000	0.000	0.000	0.000	0.000
90	0.000	0.000	0.000	0.000	0.000	0.000
91	0.000	0.000	0.000	0.000	0.000	0.000
92	0.000	0.000	0.000	0.000	0.000	0.000
93	0.280	0.000	0.000	0.000	0.000	0.000
94	0.288	0.000	0.000	0.000	0.000	0.000
95	0.000	0.000	0.000	0.000	0.000	0.000
96	0.000	0.000	0.000	0.000	0.000	0.000
97	0.268	0.000	0.000	0.000	0.000	0.000
98	0.000	0.000	0.000	0.000	0.000	0.000
99	0.000	0.000	0.000	0.000	0.000	0.000
100	0.000	0.000	0.000	0.000	0.000	0.000
101	0.000	0.000	0.000	0.000	0.000	0.000
102	0.000	0.000	0.000	0.000	0.000	0.000
103	0.000	0.000	0.000	0.000	0.000	0.000
104	0.000	0.000	0.000	0.000	0.000	0.000
105	0.000	0.000	0.000	0.000	0.000	0.000
106	0.000	0.000	0.000	0.000	0.000	0.000
107	0.000	0.000	0.000	0.000	0.000	0.000

표 17. 대립형질의 빈도가 동일한 경우 108~145개 가계도의 $P(Z_{\max} \geq C)$

가계도	재조합률 : 0.000					
	C					
	0.5	1.0	1.5	2.0	2.5	3
108	0.000	0.000	0.000	0.000	0.000	0.000
109	0.000	0.000	0.000	0.000	0.000	0.000
110	0.000	0.000	0.000	0.000	0.000	0.000
111	0.072	0.000	0.000	0.000	0.000	0.000
112	0.000	0.000	0.000	0.000	0.000	0.000
113	0.000	0.000	0.000	0.000	0.000	0.000
114	0.216	0.000	0.000	0.000	0.000	0.000
115	0.000	0.000	0.000	0.000	0.000	0.000
116	0.230	0.000	0.000	0.000	0.000	0.000
117	0.000	0.000	0.000	0.000	0.000	0.000
118	0.000	0.000	0.000	0.000	0.000	0.000
119	0.000	0.000	0.000	0.000	0.000	0.000
120	0.000	0.000	0.000	0.000	0.000	0.000
121	0.000	0.000	0.000	0.000	0.000	0.000
122	0.000	0.000	0.000	0.000	0.000	0.000
123	0.000	0.000	0.000	0.000	0.000	0.000
124	0.000	0.000	0.000	0.000	0.000	0.000
125	0.244	0.000	0.000	0.000	0.000	0.000
126	0.000	0.000	0.000	0.000	0.000	0.000
127	0.000	0.000	0.000	0.000	0.000	0.000
128	0.000	0.000	0.000	0.000	0.000	0.000
129	0.000	0.000	0.000	0.000	0.000	0.000
130	0.000	0.000	0.000	0.000	0.000	0.000
131	0.252	0.000	0.000	0.000	0.000	0.000
132	0.000	0.000	0.000	0.000	0.000	0.000
133	0.000	0.000	0.000	0.000	0.000	0.000
134	0.000	0.000	0.000	0.000	0.000	0.000
135	0.000	0.000	0.000	0.000	0.000	0.000
136	0.000	0.000	0.000	0.000	0.000	0.000
137	0.264	0.000	0.000	0.000	0.000	0.000
138	0.000	0.000	0.000	0.000	0.000	0.000
139	0.000	0.000	0.000	0.000	0.000	0.000
140	0.000	0.000	0.000	0.000	0.000	0.000
141	0.054	0.000	0.000	0.000	0.000	0.000
142	0.000	0.000	0.000	0.000	0.000	0.000
143	0.000	0.000	0.000	0.000	0.000	0.000
144	0.000	0.000	0.000	0.000	0.000	0.000
145	0.066	0.000	0.000	0.000	0.000	0.000

표 18. 대립형질의 빈도가 다른 경우 1~35개 가계도의 $E(Z_{\max})$

표식유전자	snp459	snp461
대립형질의 빈도	0.49	0.091
1	0.000	0.000
2	0.085	0.043
3	0.071	0.009
4	0.066	0.006
5	0.000	0.000
6	0.078	0.043
7	0.000	0.000
8	0.000	0.000
9	0.000	0.000
10	0.410	0.122
11	0.000	0.000
12	0.000	0.000
13	0.000	0.000
14	0.000	0.000
15	0.090	0.044
16	0.082	0.037
17	0.088	0.043
가계도 18	0.087	0.033
19	0.087	0.043
20	0.000	0.000
21	0.000	0.000
22	0.079	0.042
23	0.199	0.084
24	0.093	0.033
25	0.000	0.000
26	0.095	0.044
27	0.092	0.046
28	0.000	0.000
29	0.000	0.000
30	0.000	0.000
31	0.092	0.044
32	0.089	0.039
33	0.000	0.000
34	0.183	0.087
35	0.154	0.107

표 19. 대립형질의 빈도가 다른 경우 36~70개 가계도의 $E(Z_{\max})$

표식유전자	snp459	snp461
대립형질의 빈도	0.49	0.091
36	0.198	0.084
37	0.094	0.040
38	0.094	0.046
39	0.096	0.037
40	0.078	0.044
41	0.103	0.044
42	0.198	0.058
43	0.000	0.000
44	0.000	0.000
45	0.319	0.117
46	0.095	0.042
47	0.000	0.000
48	0.588	0.218
49	0.084	0.040
50	0.092	0.044
51	0.090	0.044
52	0.191	0.076
가계도 53	0.090	0.037
54	0.000	0.000
55	0.000	0.000
56	0.087	0.037
57	0.092	0.040
58	0.300	0.120
59	0.090	0.044
60	0.000	0.000
61	0.413	0.188
62	0.091	0.045
63	0.084	0.038
64	0.080	0.041
65	0.091	0.032
66	0.000	0.000
67	0.088	0.044
68	0.000	0.000
69	0.084	0.044
70	0.097	0.032

표 20. 대립형질의 빈도가 다른 경우 71~99개 가계도의 $E(Z_{\max})$

표식유전자	snp459	snp461
대립형질의 빈도	0.49	0.091
71	0.000	0.000
72	0.096	0.039
73	0.000	0.000
74	0.091	0.043
75	0.101	0.044
76	0.205	0.090
77	0.000	0.000
78	0.000	0.000
79	0.000	0.000
80	0.000	0.000
81	0.000	0.000
82	0.199	0.086
83	0.087	0.044
84	0.074	0.042
가계도 85	0.000	0.000
86	0.000	0.000
87	0.000	0.000
88	0.088	0.041
89	0.092	0.036
90	0.091	0.043
91	0.085	0.032
92	0.000	0.000
93	0.262	0.142
94	0.194	0.079
95	0.000	0.000
96	0.090	0.036
97	0.208	0.080
98	0.064	0.011
99	0.088	0.052

표 21. snp459에 대한 1~35개 가계도의 $P(Z_{\max} \geq C)$

가계도	재조합률 : 0.01					
	C					
	0.5	1.0	1.5	2.0	2.5	3
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000
10	0.362	0.130	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.000	0.000	0.000
19	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000
23	0.298	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	0.000	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	0.000	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000
31	0.000	0.000	0.000	0.000	0.000	0.000
32	0.000	0.000	0.000	0.000	0.000	0.000
33	0.000	0.000	0.000	0.000	0.000	0.000
34	0.258	0.000	0.000	0.000	0.000	0.000
35	0.224	0.000	0.000	0.000	0.000	0.000

표 22. snp459에 대한 36~70개 가계도의 $P(Z_{\max} \geq C)$

가계도	재조합률 : 0.01					
	C					
	0.5	1.0	1.5	2.0	2.5	3
36	0.292	0.000	0.000	0.000	0.000	0.000
37	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000
42	0.106	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	0.000	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000
45	0.350	0.000	0.000	0.000	0.000	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000	0.000	0.000
48	0.432	0.286	0.062	0.000	0.000	0.000
49	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000
51	0.000	0.000	0.000	0.000	0.000	0.000
52	0.280	0.000	0.000	0.000	0.000	0.000
53	0.000	0.000	0.000	0.000	0.000	0.000
54	0.000	0.000	0.000	0.000	0.000	0.000
55	0.000	0.000	0.000	0.000	0.000	0.000
56	0.000	0.000	0.000	0.000	0.000	0.000
57	0.000	0.000	0.000	0.000	0.000	0.000
58	0.324	0.000	0.000	0.000	0.000	0.000
59	0.000	0.000	0.000	0.000	0.000	0.000
60	0.000	0.000	0.000	0.000	0.000	0.000
61	0.384	0.256	0.000	0.000	0.000	0.000
62	0.000	0.000	0.000	0.000	0.000	0.000
63	0.000	0.000	0.000	0.000	0.000	0.000
64	0.000	0.000	0.000	0.000	0.000	0.000
65	0.000	0.000	0.000	0.000	0.000	0.000
66	0.000	0.000	0.000	0.000	0.000	0.000
67	0.000	0.000	0.000	0.000	0.000	0.000
68	0.000	0.000	0.000	0.000	0.000	0.000
69	0.000	0.000	0.000	0.000	0.000	0.000
70	0.000	0.000	0.000	0.000	0.000	0.000

표 23. snp459에 대한 71~99개 가계도의 $P(Z_{\max} \geq C)$

가계도	재조합률 : 0.01					
	C					
	0.5	1.0	1.5	2.0	2.5	3
71	0.000	0.000	0.000	0.000	0.000	0.000
72	0.000	0.000	0.000	0.000	0.000	0.000
73	0.000	0.000	0.000	0.000	0.000	0.000
74	0.000	0.000	0.000	0.000	0.000	0.000
75	0.000	0.000	0.000	0.000	0.000	0.000
76	0.300	0.000	0.000	0.000	0.000	0.000
77	0.000	0.000	0.000	0.000	0.000	0.000
78	0.000	0.000	0.000	0.000	0.000	0.000
79	0.000	0.000	0.000	0.000	0.000	0.000
80	0.000	0.000	0.000	0.000	0.000	0.000
81	0.000	0.000	0.000	0.000	0.000	0.000
82	0.294	0.000	0.000	0.000	0.000	0.000
83	0.000	0.000	0.000	0.000	0.000	0.000
84	0.000	0.000	0.000	0.000	0.000	0.000
85	0.000	0.000	0.000	0.000	0.000	0.000
86	0.000	0.000	0.000	0.000	0.000	0.000
87	0.000	0.000	0.000	0.000	0.000	0.000
88	0.000	0.000	0.000	0.000	0.000	0.000
89	0.000	0.000	0.000	0.000	0.000	0.000
90	0.000	0.000	0.000	0.000	0.000	0.000
91	0.000	0.000	0.000	0.000	0.000	0.000
92	0.000	0.000	0.000	0.000	0.000	0.000
93	0.288	0.000	0.000	0.000	0.000	0.000
94	0.286	0.000	0.000	0.000	0.000	0.000
95	0.000	0.000	0.000	0.000	0.000	0.000
96	0.000	0.000	0.000	0.000	0.000	0.000
97	0.314	0.000	0.000	0.000	0.000	0.000
98	0.000	0.000	0.000	0.000	0.000	0.000
99	0.000	0.000	0.000	0.000	0.000	0.000

표 24. snp461에 대한 1~35개 가계도의 $P(Z_{\max} \geq C)$

가계도	재조합률 : 0.01					
	C					
	0.5	1.0	1.5	2.0	2.5	3
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000
10	0.124	0.008	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.000	0.000	0.000
19	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000
23	0.138	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	0.000	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	0.000	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000
31	0.000	0.000	0.000	0.000	0.000	0.000
32	0.000	0.000	0.000	0.000	0.000	0.000
33	0.000	0.000	0.000	0.000	0.000	0.000
34	0.144	0.000	0.000	0.000	0.000	0.000
35	0.170	0.000	0.000	0.000	0.000	0.000

표 25. snp461에 대한 36~70개 가계도의 $P(Z_{\max} \geq C)$

가계도	재조합률 : 0.01					
	C					
	0.5	1.0	1.5	2.0	2.5	3
36	0.136	0.000	0.000	0.000	0.000	0.000
37	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000
42	0.018	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	0.000	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000
45	0.130	0.000	0.000	0.000	0.000	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000	0.000	0.000
48	0.144	0.134	0.014	0.000	0.000	0.000
49	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000
51	0.000	0.000	0.000	0.000	0.000	0.000
52	0.124	0.000	0.000	0.000	0.000	0.000
53	0.000	0.000	0.000	0.000	0.000	0.000
54	0.000	0.000	0.000	0.000	0.000	0.000
55	0.000	0.000	0.000	0.000	0.000	0.000
56	0.000	0.000	0.000	0.000	0.000	0.000
57	0.000	0.000	0.000	0.000	0.000	0.000
58	0.130	0.000	0.000	0.000	0.000	0.000
59	0.000	0.000	0.000	0.000	0.000	0.000
60	0.000	0.000	0.000	0.000	0.000	0.000
61	0.160	0.150	0.000	0.000	0.000	0.000
62	0.000	0.000	0.000	0.000	0.000	0.000
63	0.000	0.000	0.000	0.000	0.000	0.000
64	0.000	0.000	0.000	0.000	0.000	0.000
65	0.000	0.000	0.000	0.000	0.000	0.000
66	0.000	0.000	0.000	0.000	0.000	0.000
67	0.000	0.000	0.000	0.000	0.000	0.000
68	0.000	0.000	0.000	0.000	0.000	0.000
69	0.000	0.000	0.000	0.000	0.000	0.000
70	0.000	0.000	0.000	0.000	0.000	0.000

표 26. snp461에 대한 71~99개 가계도의 $P(Z_{\max} \geq C)$

가계도	재조합률 : 0.01					
	C					
	0.5	1.0	1.5	2.0	2.5	3
71	0.000	0.000	0.000	0.000	0.000	0.000
72	0.000	0.000	0.000	0.000	0.000	0.000
73	0.000	0.000	0.000	0.000	0.000	0.000
74	0.000	0.000	0.000	0.000	0.000	0.000
75	0.000	0.000	0.000	0.000	0.000	0.000
76	0.150	0.000	0.000	0.000	0.000	0.000
77	0.000	0.000	0.000	0.000	0.000	0.000
78	0.000	0.000	0.000	0.000	0.000	0.000
79	0.000	0.000	0.000	0.000	0.000	0.000
80	0.000	0.000	0.000	0.000	0.000	0.000
81	0.000	0.000	0.000	0.000	0.000	0.000
82	0.142	0.000	0.000	0.000	0.000	0.000
83	0.000	0.000	0.000	0.000	0.000	0.000
84	0.000	0.000	0.000	0.000	0.000	0.000
85	0.000	0.000	0.000	0.000	0.000	0.000
86	0.000	0.000	0.000	0.000	0.000	0.000
87	0.000	0.000	0.000	0.000	0.000	0.000
88	0.000	0.000	0.000	0.000	0.000	0.000
89	0.000	0.000	0.000	0.000	0.000	0.000
90	0.000	0.000	0.000	0.000	0.000	0.000
91	0.000	0.000	0.000	0.000	0.000	0.000
92	0.000	0.000	0.000	0.000	0.000	0.000
93	0.162	0.000	0.000	0.000	0.000	0.000
94	0.126	0.000	0.000	0.000	0.000	0.000
95	0.000	0.000	0.000	0.000	0.000	0.000
96	0.000	0.000	0.000	0.000	0.000	0.000
97	0.134	0.000	0.000	0.000	0.000	0.000
98	0.000	0.000	0.000	0.000	0.000	0.000
99	0.000	0.000	0.000	0.000	0.000	0.000

표 27. 2개 표식유전자인 경우 1~35개 가계도의 $E(Z_{\max})$

표식유전자	D19S217-D19S219	D19S223-D19S596	D19S425-D19S927
거리(cM)	2	30	36
1	0.000	0.000	0.000
2	0.176	0.115	0.110
3	0.156	0.084	0.079
4	0.137	0.083	0.075
5	0.000	0.000	0.000
6	0.167	0.126	0.123
7	0.000	0.000	0.000
8	0.000	0.000	0.000
9	0.000	0.000	0.000
10	0.803	0.422	0.381
11	0.000	0.000	0.000
12	0.000	0.000	0.000
13	0.000	0.000	0.000
14	0.000	0.000	0.000
15	0.171	0.118	0.117
16	0.170	0.112	0.105
17	0.176	0.124	0.121
가계도 18	0.159	0.116	0.112
19	0.158	0.105	0.106
20	0.000	0.000	0.000
21	0.000	0.000	0.000
22	0.160	0.121	0.121
23	0.337	0.203	0.191
24	0.166	0.124	0.123
25	0.000	0.000	0.000
26	0.184	0.129	0.124
27	0.159	0.111	0.106
28	0.000	0.000	0.000
29	0.000	0.000	0.000
30	0.000	0.000	0.000
31	0.164	0.116	0.104
32	0.170	0.121	0.114
33	0.000	0.000	0.000
34	0.356	0.224	0.221
35	0.369	0.229	0.217

표 28. 2개 표식유전자인 경우 36~70개 가계도의 $E(Z_{\max})$

표식유전자 거리(cM)	D19S217-D19S219 2	D19S223-D19S596 30	D19S425-D19S927 36
36	0.364	0.245	0.223
37	0.152	0.121	0.113
38	0.165	0.108	0.101
39	0.173	0.124	0.121
40	0.147	0.102	0.099
41	0.157	0.104	0.104
42	0.444	0.216	0.205
43	0.000	0.000	0.000
44	0.000	0.000	0.000
45	0.510	0.292	0.256
46	0.157	0.117	0.117
47	0.000	0.000	0.000
48	1.053	0.530	0.468
49	0.159	0.111	0.116
50	0.168	0.117	0.109
51	0.157	0.113	0.110
52	0.347	0.217	0.201
가계도 53	0.166	0.114	0.114
54	0.000	0.000	0.000
55	0.000	0.000	0.000
56	0.152	0.110	0.113
57	0.165	0.113	0.113
58	0.520	0.328	0.296
59	0.179	0.131	0.122
60	0.000	0.000	0.000
61	0.728	0.406	0.361
62	0.163	0.111	0.108
63	0.158	0.118	0.116
64	0.168	0.125	0.122
65	0.166	0.113	0.106
66	0.000	0.000	0.000
67	0.166	0.116	0.107
68	0.000	0.000	0.000
69	0.164	0.110	0.109
70	0.170	0.121	0.116

표 29. 2개 표식유전자인 경우 71~99개 가계도의 $E(Z_{\max})$

표식유전자 거리(cM)	D19S217-D19S219 2	D19S223-D19S596 30	D19S425-D19S927 36
71	0.000	0.000	0.000
72	0.173	0.108	0.103
73	0.000	0.000	0.000
74	0.158	0.114	0.114
75	0.162	0.126	0.120
76	0.356	0.227	0.208
77	0.000	0.000	0.000
78	0.000	0.000	0.000
79	0.000	0.000	0.000
80	0.000	0.000	0.000
81	0.000	0.000	0.000
82	0.342	0.206	0.193
83	0.164	0.129	0.115
84	0.170	0.116	0.114
가계도 85	0.000	0.000	0.000
86	0.000	0.000	0.000
87	0.000	0.000	0.000
88	0.156	0.119	0.118
89	0.148	0.118	0.107
90	0.161	0.106	0.098
91	0.173	0.125	0.120
92	0.000	0.000	0.000
93	0.533	0.316	0.306
94	0.333	0.215	0.203
95	0.000	0.000	0.000
96	0.167	0.115	0.111
97	0.362	0.236	0.210
98	0.146	0.076	0.071
99	0.166	0.129	0.115

표 30. 2개 표식유전자, (D19S217 D19S219)인 경우 1~35개 가계도의

$$P(Z_{\max} \geq C)$$

가계도	제조합률 : 0.02					
	C					
	0.5	1.0	1.5	2.0	2.5	3
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000
10	0.734	0.414	0.020	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.000	0.000	0.000
19	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000
23	0.546	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	0.000	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	0.000	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000
31	0.000	0.000	0.000	0.000	0.000	0.000
32	0.000	0.000	0.000	0.000	0.000	0.000
33	0.000	0.000	0.000	0.000	0.000	0.000
34	0.604	0.000	0.000	0.000	0.000	0.000
35	0.590	0.000	0.000	0.000	0.000	0.000

표 31. 2개 표식유전자, (D19S217 D19S219)인 경우 36~70개 가계도의

$$P(Z_{\max} \geq C)$$

가계도	제조합률 : 0.02					
	C					
	0.5	1.0	1.5	2.0	2.5	3
36	0.590	0.000	0.000	0.000	0.000	0.000
37	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000
42	0.540	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	0.000	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000
45	0.588	0.000	0.000	0.000	0.000	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000	0.000	0.000
48	0.660	0.586	0.494	0.000	0.000	0.000
49	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000
51	0.000	0.000	0.000	0.000	0.000	0.000
52	0.554	0.000	0.000	0.000	0.000	0.000
53	0.000	0.000	0.000	0.000	0.000	0.000
54	0.000	0.000	0.000	0.000	0.000	0.000
55	0.000	0.000	0.000	0.000	0.000	0.000
56	0.000	0.000	0.000	0.000	0.000	0.000
57	0.000	0.000	0.000	0.000	0.000	0.000
58	0.612	0.000	0.000	0.000	0.000	0.000
59	0.000	0.000	0.000	0.000	0.000	0.000
60	0.000	0.000	0.000	0.000	0.000	0.000
61	0.646	0.552	0.000	0.000	0.000	0.000
62	0.000	0.000	0.000	0.000	0.000	0.000
63	0.000	0.000	0.000	0.000	0.000	0.000
64	0.000	0.000	0.000	0.000	0.000	0.000
65	0.000	0.000	0.000	0.000	0.000	0.000
66	0.000	0.000	0.000	0.000	0.000	0.000
67	0.000	0.000	0.000	0.000	0.000	0.000
68	0.000	0.000	0.000	0.000	0.000	0.000
69	0.000	0.000	0.000	0.000	0.000	0.000
70	0.000	0.000	0.000	0.000	0.000	0.000

표 32. 2개 표식유전자, (D19S217 D19S219)인 경우 71~99개 가계도의

$$P(Z_{\max} \geq C)$$

가계도	재조합률 : 0.02					
	C					
	0.5	1.0	1.5	2.0	2.5	3
71	0.000	0.000	0.000	0.000	0.000	0.000
72	0.000	0.000	0.000	0.000	0.000	0.000
73	0.000	0.000	0.000	0.000	0.000	0.000
74	0.000	0.000	0.000	0.000	0.000	0.000
75	0.000	0.000	0.000	0.000	0.000	0.000
76	0.576	0.000	0.000	0.000	0.000	0.000
77	0.000	0.000	0.000	0.000	0.000	0.000
78	0.000	0.000	0.000	0.000	0.000	0.000
79	0.000	0.000	0.000	0.000	0.000	0.000
80	0.000	0.000	0.000	0.000	0.000	0.000
81	0.000	0.000	0.000	0.000	0.000	0.000
82	0.556	0.000	0.000	0.000	0.000	0.000
83	0.000	0.000	0.000	0.000	0.000	0.000
84	0.000	0.000	0.000	0.000	0.000	0.000
85	0.000	0.000	0.000	0.000	0.000	0.000
86	0.000	0.000	0.000	0.000	0.000	0.000
87	0.000	0.000	0.000	0.000	0.000	0.000
88	0.000	0.000	0.000	0.000	0.000	0.000
89	0.000	0.000	0.000	0.000	0.000	0.000
90	0.000	0.000	0.000	0.000	0.000	0.000
91	0.000	0.000	0.000	0.000	0.000	0.000
92	0.000	0.000	0.000	0.000	0.000	0.000
93	0.610	0.000	0.000	0.000	0.000	0.000
94	0.548	0.000	0.000	0.000	0.000	0.000
95	0.000	0.000	0.000	0.000	0.000	0.000
96	0.000	0.000	0.000	0.000	0.000	0.000
97	0.578	0.000	0.000	0.000	0.000	0.000
98	0.000	0.000	0.000	0.000	0.000	0.000
99	0.000	0.000	0.000	0.000	0.000	0.000

표 33. 2개 표식유전자, (D19S223, D19S596)인 경우 1~35개 가계도의

$$P(Z_{\max} \geq C)$$

가계도	제조합률 : 0.30					
	C					
	0.5	1.0	1.5	2.0	2.5	3
1	0.650	0.390	0.166	0.062	0.004	0.000
2	0.406	0.126	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000
12	0.406	0.126	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000
15	0.000	0.000	0.000	0.000	0.000	0.000
16	0.000	0.000	0.000	0.000	0.000	0.000
17	0.000	0.000	0.000	0.000	0.000	0.000
18	0.000	0.000	0.000	0.000	0.000	0.000
19	0.000	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000
23	0.000	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000
25	0.304	0.000	0.000	0.000	0.000	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	0.000	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000
31	0.000	0.000	0.000	0.000	0.000	0.000
32	0.000	0.000	0.000	0.000	0.000	0.000
33	0.000	0.000	0.000	0.000	0.000	0.000
34	0.000	0.000	0.000	0.000	0.000	0.000
35	0.000	0.000	0.000	0.000	0.000	0.000

표 34. 2개 표식유전자, (D19S223, D19S596)인 경우 36~70개 가계도의

$$P(Z_{\max} \geq C)$$

가계도	재조합률 : 0.30					
	C					
	0.5	1.0	1.5	2.0	2.5	3
36	0.376	0.000	0.000	0.000	0.000	0.000
37	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000
42	0.104	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	0.000	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000
45	0.332	0.000	0.000	0.000	0.000	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000	0.000	0.000
48	0.352	0.252	0.086	0.000	0.000	0.000
49	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000
51	0.000	0.000	0.000	0.000	0.000	0.000
52	0.338	0.000	0.000	0.000	0.000	0.000
53	0.000	0.000	0.000	0.000	0.000	0.000
54	0.000	0.000	0.000	0.000	0.000	0.000
55	0.000	0.000	0.000	0.000	0.000	0.000
56	0.000	0.000	0.000	0.000	0.000	0.000
57	0.000	0.000	0.000	0.000	0.000	0.000
58	0.366	0.000	0.000	0.000	0.000	0.000
59	0.000	0.000	0.000	0.000	0.000	0.000
60	0.000	0.000	0.000	0.000	0.000	0.000
61	0.362	0.256	0.000	0.000	0.000	0.000
62	0.000	0.000	0.000	0.000	0.000	0.000
63	0.000	0.000	0.000	0.000	0.000	0.000
64	0.000	0.000	0.000	0.000	0.000	0.000
65	0.000	0.000	0.000	0.000	0.000	0.000
66	0.000	0.000	0.000	0.000	0.000	0.000
67	0.000	0.000	0.000	0.000	0.000	0.000
68	0.000	0.000	0.000	0.000	0.000	0.000
69	0.000	0.000	0.000	0.000	0.000	0.000
70	0.000	0.000	0.000	0.000	0.000	0.000

표 35. 2개 표식유전자, (D19S223, D19S596)인 경우 71~99개 가계도의

$$P(Z_{\max} \geq C)$$

가계도	재조합률 : 0.30					
	C					
	0.5	1.0	1.5	2.0	2.5	3
71	0.000	0.000	0.000	0.000	0.000	0.000
72	0.000	0.000	0.000	0.000	0.000	0.000
73	0.000	0.000	0.000	0.000	0.000	0.000
74	0.000	0.000	0.000	0.000	0.000	0.000
75	0.000	0.000	0.000	0.000	0.000	0.000
76	0.342	0.000	0.000	0.000	0.000	0.000
77	0.000	0.000	0.000	0.000	0.000	0.000
78	0.000	0.000	0.000	0.000	0.000	0.000
79	0.000	0.000	0.000	0.000	0.000	0.000
80	0.000	0.000	0.000	0.000	0.000	0.000
81	0.000	0.000	0.000	0.000	0.000	0.000
82	0.312	0.000	0.000	0.000	0.000	0.000
83	0.000	0.000	0.000	0.000	0.000	0.000
84	0.000	0.000	0.000	0.000	0.000	0.000
85	0.000	0.000	0.000	0.000	0.000	0.000
86	0.000	0.000	0.000	0.000	0.000	0.000
87	0.000	0.000	0.000	0.000	0.000	0.000
88	0.000	0.000	0.000	0.000	0.000	0.000
89	0.000	0.000	0.000	0.000	0.000	0.000
90	0.000	0.000	0.000	0.000	0.000	0.000
91	0.000	0.000	0.000	0.000	0.000	0.000
92	0.000	0.000	0.000	0.000	0.000	0.000
93	0.344	0.000	0.000	0.000	0.000	0.000
94	0.324	0.000	0.000	0.000	0.000	0.000
95	0.000	0.000	0.000	0.000	0.000	0.000
96	0.000	0.000	0.000	0.000	0.000	0.000
97	0.356	0.000	0.000	0.000	0.000	0.000
98	0.000	0.000	0.000	0.000	0.000	0.000
99	0.000	0.000	0.000	0.000	0.000	0.000

표 36. 2개 표식유전자, (D19S425, D19S927)인 경우 1~35개 가계도의

$$P(Z_{\max} \geq C)$$

가계도	제조합률 : 0.36					
	C					
	0.5	1.0	1.5	2.0	2.5	3
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.110	0.000	0.000	0.000	0.000	0.000
3	0.079	0.000	0.000	0.000	0.000	0.000
4	0.075	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
6	0.123	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000
10	0.381	0.072	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000
14	0.000	0.000	0.000	0.000	0.000	0.000
15	0.117	0.000	0.000	0.000	0.000	0.000
16	0.105	0.000	0.000	0.000	0.000	0.000
17	0.121	0.000	0.000	0.000	0.000	0.000
18	0.112	0.000	0.000	0.000	0.000	0.000
19	0.106	0.000	0.000	0.000	0.000	0.000
20	0.000	0.000	0.000	0.000	0.000	0.000
21	0.000	0.000	0.000	0.000	0.000	0.000
22	0.000	0.000	0.000	0.000	0.000	0.000
23	0.280	0.000	0.000	0.000	0.000	0.000
24	0.000	0.000	0.000	0.000	0.000	0.000
25	0.000	0.000	0.000	0.000	0.000	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	0.000	0.000
28	0.000	0.000	0.000	0.000	0.000	0.000
29	0.000	0.000	0.000	0.000	0.000	0.000
30	0.000	0.000	0.000	0.000	0.000	0.000
31	0.000	0.000	0.000	0.000	0.000	0.000
32	0.000	0.000	0.000	0.000	0.000	0.000
33	0.000	0.000	0.000	0.000	0.000	0.000
34	0.338	0.000	0.000	0.000	0.000	0.000
35	0.340	0.000	0.000	0.000	0.000	0.000

표 37. 2개 표식유전자, (D19S425, D19S927)인 경우 36~70개 가계도의

$$P(Z_{\max} \geq C)$$

가계도	재조합률 : 0.36					
	C					
	0.5	1.0	1.5	2.0	2.5	3
36	0.344	0.000	0.000	0.000	0.000	0.000
37	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000
42	0.088	0.000	0.000	0.000	0.000	0.000
43	0.000	0.000	0.000	0.000	0.000	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000
45	0.290	0.000	0.000	0.000	0.000	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000	0.000	0.000
48	0.334	0.212	0.066	0.000	0.000	0.000
49	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000
51	0.000	0.000	0.000	0.000	0.000	0.000
52	0.312	0.000	0.000	0.000	0.000	0.000
53	0.000	0.000	0.000	0.000	0.000	0.000
54	0.000	0.000	0.000	0.000	0.000	0.000
55	0.000	0.000	0.000	0.000	0.000	0.000
56	0.000	0.000	0.000	0.000	0.000	0.000
57	0.000	0.000	0.000	0.000	0.000	0.000
58	0.334	0.000	0.000	0.000	0.000	0.000
59	0.000	0.000	0.000	0.000	0.000	0.000
60	0.000	0.000	0.000	0.000	0.000	0.000
61	0.316	0.224	0.000	0.000	0.000	0.000
62	0.000	0.000	0.000	0.000	0.000	0.000
63	0.000	0.000	0.000	0.000	0.000	0.000
64	0.000	0.000	0.000	0.000	0.000	0.000
65	0.000	0.000	0.000	0.000	0.000	0.000
66	0.000	0.000	0.000	0.000	0.000	0.000
67	0.000	0.000	0.000	0.000	0.000	0.000
68	0.000	0.000	0.000	0.000	0.000	0.000
69	0.000	0.000	0.000	0.000	0.000	0.000
70	0.000	0.000	0.000	0.000	0.000	0.000

표 38. 2개 표식유전자, (D19S425, D19S927)인 경우 71~99개 가계도의

$$P(Z_{\max} \geq C)$$

가계도	재조합률 : 0.36					
	C					
	0.5	1.0	1.5	2.0	2.5	3
71	0.000	0.000	0.000	0.000	0.000	0.000
72	0.000	0.000	0.000	0.000	0.000	0.000
73	0.000	0.000	0.000	0.000	0.000	0.000
74	0.000	0.000	0.000	0.000	0.000	0.000
75	0.000	0.000	0.000	0.000	0.000	0.000
76	0.314	0.000	0.000	0.000	0.000	0.000
77	0.000	0.000	0.000	0.000	0.000	0.000
78	0.000	0.000	0.000	0.000	0.000	0.000
79	0.000	0.000	0.000	0.000	0.000	0.000
80	0.000	0.000	0.000	0.000	0.000	0.000
81	0.000	0.000	0.000	0.000	0.000	0.000
82	0.280	0.000	0.000	0.000	0.000	0.000
83	0.000	0.000	0.000	0.000	0.000	0.000
84	0.000	0.000	0.000	0.000	0.000	0.000
85	0.000	0.000	0.000	0.000	0.000	0.000
86	0.000	0.000	0.000	0.000	0.000	0.000
87	0.000	0.000	0.000	0.000	0.000	0.000
88	0.000	0.000	0.000	0.000	0.000	0.000
89	0.000	0.000	0.000	0.000	0.000	0.000
90	0.000	0.000	0.000	0.000	0.000	0.000
91	0.000	0.000	0.000	0.000	0.000	0.000
92	0.000	0.000	0.000	0.000	0.000	0.000
93	0.326	0.000	0.000	0.000	0.000	0.000
94	0.310	0.000	0.000	0.000	0.000	0.000
95	0.000	0.000	0.000	0.000	0.000	0.000
96	0.000	0.000	0.000	0.000	0.000	0.000
97	0.318	0.000	0.000	0.000	0.000	0.000
98	0.000	0.000	0.000	0.000	0.000	0.000
99	0.000	0.000	0.000	0.000	0.000	0.000

표 39. 가계도 수에 따른 $E(Z_{\max})$

가계도 수	재조합률			
	0.000	0.100	0.200	0.500
1	0.069	0.044	0.043	0.043
2	0.167	0.108	0.097	0.078
3	0.359	0.254	0.163	0.083
4	0.457	0.305	0.200	0.101
5	0.621	0.398	0.256	0.129
6	0.828	0.472	0.298	0.131
7	1.005	0.57	0.301	0.109
8	1.342	0.679	0.324	0.121
9	1.564	0.782	0.414	0.107
10	1.788	0.922	0.503	0.108
11	2.194	1.119	0.494	0.085
12	2.444	1.132	0.512	0.118
13	2.606	1.219	0.557	0.115
14	2.671	1.217	0.594	0.097
15	3.013	1.355	0.655	0.114
16	3.125	1.401	0.622	0.109
17	3.537	1.436	0.646	0.096
18	3.522	1.439	0.628	0.100

표 40. 가계도 수에 따른 $P(Z_{\max} \geq C)$

가계도 수	재조합률 : 0.00					
	C					
	0.5	1.0	1.5	2.0	2.5	3
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.106	0.000	0.000	0.000	0.000	0.000
3	0.390	0.016	0.000	0.000	0.000	0.000
4	0.486	0.064	0.000	0.000	0.000	0.000
5	0.634	0.190	0.016	0.004	0.000	0.000
6	0.788	0.320	0.068	0.028	0.000	0.000
7	0.828	0.486	0.146	0.072	0.008	0.000
8	0.882	0.678	0.352	0.202	0.040	0.008
9	0.948	0.752	0.452	0.320	0.100	0.014
10	0.960	0.816	0.534	0.434	0.212	0.068
11	0.996	0.896	0.732	0.588	0.354	0.162
12	0.984	0.914	0.766	0.662	0.454	0.266
13	0.984	0.934	0.818	0.730	0.520	0.336
14	0.988	0.950	0.816	0.734	0.536	0.336
15	1.000	0.970	0.884	0.836	0.656	0.456
16	1.000	0.978	0.914	0.854	0.706	0.482
17	0.998	0.988	0.948	0.922	0.802	0.644
18	1.000	0.996	0.966	0.922	0.792	0.596

Abstract

Statistical methods for informativeness in genetic linkage analysis and an application to pedigree data

Lim, Hyunsun

Dept. of Biostatistics and Computing

The Graduate School

Yonsei University

We estimate the sufficiency of information by calculating expected maximum LOD score and calculating the probability of maximum LOD score more than a certain constant C in H_1 (of linkage). We founded the fact that the statistical power is increased as the frequencies of alleles of genetic marker are similar with each other and the number of genetic marker increases(Park. Yunju, 2001).

In the thesis, on type of pedigree, we found the fact that the statistical power is increased as the family's number of a pedigree increases and cardiovascular disease patient exists and cardiovascular disease patient is parent. Also, we found the fact that the statistical power is over 80% that the

number of pedigree is about to 15. Consequently, we are able to minimize loss of time and cost with the test for sufficiency of information in pedigree by using statistical power since it is examined before genotyping is taken in linkage analysis.

Key words : informativeness, power of pedigree, linkage analysis, incomplete penetrance, recombination fraction, penetrance