

▶ 표 4.9 연습문제 4.5의 콘돔사용에 대한 결과

변수	오즈비	95% 신뢰구간
치리그룹(교육, 무교육)	4.04	(1.17, 13.9)
성별(남성, 여성)	1.38	(1.23, 12.88)
사회경제수준(높음, 낮음)	5.82	(1.87, 18.28)
총 섹스 파트너 수	3.22	(1.08, 11.31)

출처: V. I. Rickers et al., *Clin. Pediat.*, 31: 205-210, 1992

- a. 처음 세 개의 예측변수에 대해 (1, 0) 지시변수를 사용하여 적합한 모형의 모수효과를 구하라.
- b. 성별에 대한 오즈비의 추정값 1.38과 이에 대응되는 신뢰구간 중 어느 하나가 잘못된 이유를 설명하라. 만약 제시된 신뢰구간이 옳다면 1.38은 실제로 로그 오즈비가 되며 추정된 오즈비는 3.98이 되는 것을 보여라.

**4.16** <표 4.10>은  $Y$ =일반적인 마취를 동반하는 외과수술 시 마취가 풀릴 때 목의 통증 여부(1=예, 0=아니오)에 대하여  $D$ =수술시간(단위: 분)과  $T$ =기관지 를 안전하게 하는 장치종류(0=후두마스크 기관튜브, 1=기관튜브)의 영향을 알아보기 위한 연구 자료이다.

- a. 이 예측변수들 간의 교호작용항을 포함하는 모형을 적합하라. (i)  $T=1$  과 (ii)  $T=0$ 에 대하여  $D$ 의 효과에 대한 예측식을 구하고 해석하라. 교호작용항이 필요한지 추론하라.

▶ 표 4.10 수술 후 목의 통증에 관한 연습문제 4.16의 자료

환자	D	T	Y	환자	D	T	Y	환자	D	T	Y
1	45	0	0	13	50	1	0	25	20	1	0
2	15	0	0	14	75	1	1	26	45	0	1
3	40	0	1	15	30	0	0	27	15	1	0
4	83	1	1	16	25	0	1	28	25	0	1
5	90	1	1	17	20	1	0	29	15	1	0
6	25	1	1	18	60	1	1	30	30	0	1
7	35	0	1	19	70	1	1	31	40	0	1
8	65	0	1	20	30	0	1	32	15	1	0
9	95	0	1	21	60	0	1	33	135	1	1
10	35	0	1	22	61	0	0	34	20	1	0
11	75	0	1	23	65	0	1	35	40	1	0
12	45	1	1	24	15	1	0				

출처: D. Collett, in *Encyclopedia of Biostatistics*, Wiley, New York, 1998, pp. 350-358. $D$ =수술시간,  $T$ =안전장치 종류

- b. 교호작용이 포함된 모형과 포함되지 않은 두 모형의 예측력을 비교하기 위해 각 모형별로 예측값과 관측값의 상관관계수  $R$ 을 구하여 비교하라.

**4.17** <표 4.11>은 기저편평세포암종 식도암 여부(1=예, 0=아니오)를 반응변수인  $Y$ 로 갖는 로지스틱 회귀모형을 적합한 결과를 보여 주고 있다. 흡연상태( $S$ )는 하루라도 한 갑을 피면 1, 그렇지 않으면 0, 음주소비량( $A$ )는 매일 마시는 평균 주량이며 인종( $R$ )은 흑인인 경우에는 1, 백인인 경우에는 0의 값을 취한다.

- a. 인종과 흡연 간의 교호작용을 설명하기 위해  $R=1$ 인 경우와  $R=0$ 인 경우에 각각 예측식을 구하라. 각각의 경우에 적합된  $YS$  조건부 오즈비를 구하라. 마찬가지로  $S=1$ 인 경우와  $S=0$ 인 경우에 각각 예측식을 구하고 각각의 경우에 적합된  $YR$  조건부 오즈비를 구하라(각각의 인종에 대하여 교호작용항의 추정값은 다른 변수들이 고정되어 있을 때 변수의 고정된 두 수준에서의 로그 오즈비의 차이가 된다).
- b. <표 4.11>에서 인종과 흡연의 모수추정값이 어떤 효과를 나타내는지 명하라. 이 변수들에 대응되는  $P$ -값은 어떤 가설에 대한 것인가?

▶ 표 4.11 연습문제 4.17의 식도암 자료의 결과

변수	효과	$P$ -값
절편	-7.00	<0.01
음주소비량	0.10	0.03
흡연	1.20	<0.01
인종	0.30	0.02
인종×흡연	0.20	0.04

**4.18** <표 4.12>의 자료는 2016년도 일반사회조사 결과이다. 자료 파일을 만들고 로지스틱 회귀분석모형으로 자료를 분석하라. 짧은 보고서로 분석 결과를 요약하여 부록에는 편집된 R 출력 결과를 첨부하라.

- 5.10 이 책의 웹사이트에 있는 Lungs 자료 파일<sup>12</sup>은 중국에서 흡연과 폐암에 대해 수행한 8개 연구를 요약한 것이다. 이 자료를 분석한 후에 분석결과와 해석을 짧게 요약하여 보고서를 작성하라.
- 5.11 연습문제 5.4의 (a)에서 선택한 모형을 참조하라. 이 모형에 대한 적합도를 점검하라. 이 자료 파일로 잔차분석을 수행할 수 있는가? 설명하라.
- 5.12  $x = 0, 10, 20, 30$ 일 때에  $y = 0$ 이고  $x = 70, 80, 90, 100$ 일 때에  $y = 1$ 이라 가정하자.
- $\text{logit}[P(Y=1)] = \alpha + \beta x$  모형에서  $\hat{\beta} = \infty$ 가 되는 이유를 직관적으로 설명하라. 소프트웨어를 이용하여  $\hat{\beta}$ 와 표준오차를 구하라.
  - $x = 50$ 에 대해 한 개의 관측값은  $y = 1$ 이고 다른 한 개의 관측값이  $y = 0$ 인 두 개의 관측값을 자료에 추가하라.  $\hat{\beta}$ 와 표준오차를 구하라. 이 구한 결과가 옳다고 생각하는가? 그 이유는? 두 관측값을  $x = 49.9, y = 1$ 과  $x = 50.1, y = 0$ 로 대체하면 어떻게 되는가?
- 5.13 연습문제 5.4를 참조하자. veg를 반응변수로 사용하여 적어도 하나의 효과에 대하여 ML 추정값이 무한대가 되는 로지스틱 모형을 찾아라. 자료 파일 측면에서 이런 결과가 나온 이유를 설명하라. Firth의 벌점화 로지스틱 회귀모형이나 또는 베이지안 추론을 사용하여 모형을 적합시켜 결과를 구하고 해석하라.
- 5.14 <표 5.8>는 LogXact7 설명서(Cytel Software, 2005, p.171)에 소개된 비전이성 골육종에 대한 연구자료이다. 반응변수는 환자가 3년간 질병에 걸리지 않았는지 여부이다.
- 각 예측변수에 대해서 한 개의 예측변수만을 포함하는 로지스틱모형을 적합하면 모든 예측변수는 다 유의한 효과가 있음을 보여라. 세 개의 모든 예측변수를 포함하는 주효과 로지스틱 회귀모형을 적합하라. 림프구 침윤의 효과에 대한 ML 추정값이 무한대인 이유를 설명하라.
  - Firth의 벌점화 로지스틱 회귀모형이나 베이지안 추론을 사용하여 주효과 모형을 적합시켜 결과를 구하고 해석하라.

<sup>12</sup> 자료 출처: Intern. J. Epidemiol., 21: 197-201 (1992) by Z. Liu.

표 5.8 연습문제 5.14의 자료

림프구 침윤	성별	골형성 병리	무병기간 여부	
			있음	없음
높음	여성	없음	3	0
높음	여성	있음	2	0
높음	남성	없음	4	0
높음	남성	있음	1	0
낮음	여성	없음	5	0
낮음	여성	있음	3	2
낮음	남성	없음	5	4
낮음	남성	있음	6	11

표 5.9 5개 센터의 반응에 대한 치료와 관련된 임상시험

학과	치료	반응	
		성공	실패
1	활성제	0	5
	위약	0	9
2	활성제	1	12
	위약	0	10
3	활성제	0	7
	위약	0	5
4	활성제	6	3
	위약	2	6
5	활성제	5	9
	위약	2	12

출처: Diane Connell, Sandoz 제약 회사.

- 5.15 <표 5.9>는 5개 센터에서 실시한 무작위 임상시험 결과를 보여 준다. 세균성 감염을 치료하기 위해 활성제를 위약과 비교하는 것이 목적이다. 이 자료에 대해  $y = \text{반응}(1=\text{성공}, 0=\text{실패})$ ,  $x = \text{처리}(1=\text{활성}, 0=\text{위약})$ ,  $z = \text{센터}$ 를 나타낸다 고 하자.
- 모형  $\text{logit}[P(Y=1)] = \alpha + \beta x + \beta_k^z$ 에 대해 센터 효과 측면에서 준완전 분리가 발생하는 이유를 설명하라.
  - “no intercept” 옵션을 사용하여 한 센터를 기준으로 하는 대비효과를 사용하지 않고 개별 센터를 나타내는  $\{\beta_k^z\}$ 를 사용하여 모형을 적합하고  $\hat{\beta}_1^z, \hat{\beta}_3^z$ 와 표준오차를 제시하라. 실제 ML 추정값은 얼마인가?
  - 반응과 처리와 관련된  $2 \times 2$  주변분할표의 도수는 모두 양수이므로 빈 칸