

```
> summary(fit.negbin)
              Estimate Std. Error   z value   Pr(>|z|)
(Intercept)  -4.05251    1.17143   -3.459    0.000541
width         0.19207    0.04406    4.360    1.3e-05
---
(Dispersion parameter for Negative Binomial(0.9046) family taken to be 1)
# D in our notation is 1/0.9046 = 1.11
Residual deviance: 195.81 on 171 degrees of freedom
```

이 모형에 대응되는 음이항 GLM은

$$\log(\hat{\mu}) = -4.05 + 0.192x$$

이고, $\hat{\beta}$ 의 표준오차는 0.044이며, $\hat{D}=1.1$ 이다. 포아송 GLM에서 추정된 분산 $\hat{\mu}$ 과 비교할 때, 음이항모형에서 분산 추정값은 $\hat{\mu} + 1.1\hat{\mu}^2$ 이다. 적합된 값들은 비슷하지만 음이항모형에서 분산 추정값이 더 크고, 결과적으로 $\hat{\beta}$ 의 표준오차도 더 커지는데 이는 포아송 GLM에서 고려하지 못한 과대산포를 반영하기 때문이다. <그림 3.3>에서 표본평균값이 상대적으로 큰 값을 가짐에도 불구하고 관측값의 일부가 0의 값을 갖는 현상이 보이는데, 이는 과대산포를 반영하는 것이다.

포아송 모형 하에서, 너비 효과를 나타내는 (β)의 95% 월드 신뢰구간은 $0.164 \pm 1.96(0.020)$, 즉 $(0.125, 0.203)$ 이다. 음이항모형 하에서는 $0.192 \pm 1.96(0.044)$, 즉 $(0.105, 0.278)$ 이고, 이 결과는 프로파일 가능도 신뢰구간과 유사하다. 포아송 GLM에서 β 의 신뢰구간이 비현실적으로 짧은 걸 볼 수 있는데 이는 과대산포를 반영하지 않았기 때문이다.

연습 문제

- 7.1** 최근 일반사회조사 자료에서 X =성별(남, 녀), Y =사후 세계에 대한 믿음(예, 아니오)이라 할 때, <표 7.12>은 독립성 로그선형모형의 적합 결과이다.
- 이탈도 통계량은 0.82이고, 자유도가 1이다. 이 결과를 해석하라.
 - $\{\hat{\lambda}_j^Y\}$ 를 구하고, $\hat{\lambda}_1^Y - \hat{\lambda}_2^Y$ 를 해석하라.
 - 포화모형에서 $\{\hat{\lambda}_{ij}^{XY}\}$ 에 대한 출력 결과이다. 오즈비를 추정하라.

| | Estimate | Std Error |
|-------------------------|----------|-----------|
| genderfemales:beliefyas | 0.1368 | 0.1507 |

- 이 책의 웹사이트에 있는 파일 Postlife에 인종(흑인, 백인, 그 외)과 사후 세계에 대한 믿음을 교차분류한 자료가 있다. 이 자료에 대해 독립모형을 적합하고 $\hat{\lambda}_1^Y - \hat{\lambda}_2^Y$ 를 해석하라.

▶ 표 7.12 사후 세계에 대한 믿음에 대한 연습문제 7.1의 출력 결과

| | Estimate | Std. Error |
|---------------|----------|------------|
| Intercept | 4.5849 | 0.0752 |
| genderfemales | 0.2192 | 0.0599 |
| beliefyas | 1.4165 | 0.0752 |

- 7.2** <표 2.9>에 있는 자료에서 D = 피고인의 인종, V = 학생자의 인종, P = 사형 선고 여부를 나타낸다. <표 7.13>은 로그선형모형(DV , DP , PV)의 적합 결과이다.
- V 의 각 수준에서 D 와 P 간의 조건부 오즈비를 추정하고 그 의미를 해석하라.
 - 이 모형에 대한 적합도 검정을 시행하고 결과를 해석하라.
 - 이 책의 웹사이트에 있는 파일 DeathPenalty를 이용하여, 아래 출력 결과를 구하라. P 를 반응변수로 취하여 대응되는 로지스틱 모형을 제시하라. 자료를 그룹화한 파일을 만들어 모형을 적합하고 D 와 V 효과에 대한 추정값과 로그 선형모형에서의 추정값이 어떻게 관련되는지를 설명하라. 둘 중 어떤 모형이 주어진 자료를 분석하는 데 더 적절한지 설명하라.

표 7.13 사형선고에 대한 연습문제 7.2의 출력 결과

| Coefficients: # not showing intercept and main effect terms | | | | |
|---|----------|------------|---------|----------|
| | Estimate | Std. Error | z value | Pr(> z) |
| Dwhite:Vwhite | 4.59497 | 0.31353 | 14.656 | < 2e-16 |
| Dwhite:Pyes | -0.86780 | 0.36707 | -2.364 | 0.0181 |
| Vwhite:Pyes | 2.40444 | 0.60061 | 4.003 | 6.25e-05 |
| Residual deviance: 0.37984 on 1 degrees of freedom | | | | |

7.3 <표 7.14>는 플로리다주의 고속도로교통안전국에서 제공된 1988년도의 교통사고 자료이다. 변수들은 안전벨트 착용 여부, 사고 당시 차에서 튕겨나갔는지 여부와 사망 여부로 분류되어 있다.

- 이 자료에 적합한 로그 선형모형을 찾고 변수들 간의 연관성을 해석하라.
- 표본이 충분히 크기 때문에 모형이 잘 적합되지 않더라도 적합성 검정통계량의 값이 크게 나온다. 위의 (a)에서 적합된 모형에서 비유사성 지수를 계산하고 결과를 해석하라.
- 동질연관성모형에 대한 베이지안 분석을 시행하고 조건부 오즈비에 대한 사후신뢰구간을 구하고 결과를 해석하라. 이 해석이 빈도론적 추론 결과와 어떻게 다른지 설명하라.

표 7.14 연습문제 7.3의 자료

| 안전벨트 착용 여부 | 차에서 튕겨 나감 | 사망 | |
|---------------|--------------|---------|------|
| | | 예 | 아니오 |
| 예 | 예 | 1,105 | 14 |
| | 아니오 | 411,111 | 483 |
| 아니오 | 예 | 4,624 | 497 |
| | 아니오 | 157,342 | 1008 |

출처: Florida Department of Highway Safety and Motor Vehicles.

7.4 이 책의 2판 웹사이트 www.stat.ufl.edu/~aa/intro-cda/data에서 MBTI 자료 파일은 MBTI 단계 II의 국가표본자료로 Myers-Briggs 적성검사에 대한 4가지 이항변수로 교차분류된 것이다. 즉, 외향성/내향성(E/I), 감각형/직관형(S/N), 사고형/감정형(T/F)과 판단형/인식형(J/P)에 의하여 교차분류한 자료이다. 동질적 연관성을 갖는 로그 선형모형을 적합하고 적합도 검정을 시행하라. 이 적합 결과를 바탕으로, (i) S/N과 J/P 간의 조건부연관성이 가장 크고, (ii) E/I와 T/F 간 또

는 E/I와 J/P 간의 조건부연관성은 크지 않음을 보여라.

- 7.5 <표 7.5>의 자동차 사고 자료에 대하여 다음 물음에 답하라.
- <표 7.7>에서 모형 (GL, GS, LS, GL, LI, ST)에서 추정된 오즈비를 이용하여 농촌지역에서 안전벨트를 매지 않은 여성이 부상당할 가능성이 가장 높게 예측되는지를 설명하라.
 - 다음과 같이 2단계에 걸쳐 모형을 적합해 보자. 1단계에서는 삼차원 $G \times L \times S$ 분할표에서 S 를 반응변수로 간주하는 로지스틱 모형을 적합한다. 2단계에서는 사차원 분할표에서 위의 세 변수를 I 에 대한 예측변수로 간주하여 로지스틱 모형을 적합한다. 이와 같은 복합적인 모형이 합리적인 이유를 설명하고 모형을 적합한 후에 그 결과를 해석하라.

7.6 일반사회조사로부터 얻어진 <표 7.15>의 자료에서 R =얼마나 자주 종교 집회에 참석하는지 여부(1=일 년에 기껏해야 몇 번, 2=적어도 일 년에 여러 번), P =정치성향(1=진보적, 2=온건적, 3=보수적), B =14~16세의 십대들에 대한 피임 허용 여부(1=동의함, 2=동의하지 않음), S =남성과 여성의 혼전 성관계에 대한 의견(1=경우에 따라 또는 전혀 나쁘지 않음, 2=항상 또는 거의 항상 나쁨)과 같다.

- 변수 하나만 포함한 경우, 두 변수 모두 포함한 경우, 세 변수 모두 포함한 경우에 대응되는 로그 선형모형을 적합하는 과정을 통하여 모형의 복잡성을 살펴보아라. 이 중 선택된 모형에서 조건부 오즈비를 추정하고 결과를 해석하라.
- 로그 선형모형 (BP, BR, BS, PS, RS)에 대한 독립성 그래프를 그리고, 조건부독립성을 나타내는 패턴에 대하여 논하라. 각 쌍의 변수들에 대해서 적합된 주변적 연관성과 조건부연관성이 동일한 경우가 있는지 밝혀라.

표 7.15 연습문제 7.6의 자료(웹사이트의 BPRS 파일)

| | | 혼전 성관계(S) | | | | | | | |
|-------------|----------------------|-----------|----|----|----|----|----|----|----|
| | | 1 | | | | 2 | | | |
| | | 1 | | 2 | | 1 | | 2 | |
| 정치성향 (P) | 종교 행사 참석(R) 피임(B) | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| | 1 | 99 | 15 | 73 | 25 | 8 | 4 | 24 | 22 |
| | 2 | 73 | 20 | 87 | 37 | 20 | 13 | 50 | 60 |
| | 3 | 51 | 19 | 51 | 36 | 6 | 12 | 33 | 88 |

2장 분할표

이, 사회·경제적 지위 등과 같은 요인들은 배우자의 흡연 여부와 연구참여자가 폐암 환자인지 아닌지의 여부에 영향을 미치므로 이러한 요인들을 통제할 수 있도록 시도해야 한다. 만약 이러한 변수들을 통제하지 않으면, 그 연구 결과는 극히 제한적으로 활용될 수밖에 없다. 가령, 비흡연자의 배우자들이 흡연자의 배우자들보다 젊은 경향이 있다고 가정하자. 그리고 젊은이들이 폐암에 걸릴 확률이 더 작다고 가정해 보자. 그러면, 비흡연자의 배우자들의 폐암확률이 낮게 나온 결과는 간접흡연 효과가 없기 때문이 아니고 단순히 그들의 평균연령이 낮기 때문일 수 있다.

제어변수를 포함하는 분석은 이변량분석이 아니라 다변량분석을 사용해야 한다. 기본개념을 설명하기 위해 하나의 범주형 제어변수 Z 를 고려해 보자. 이 경우 삼차원 분할표는 세 변수의 수준의 조합에 대한 도수를 나타낸다.

2.7.1 부분분할표

삼차원 분할표에서의 이차원 **부분분할표**(partial table)는 제어변수 Z 의 각 수준에서 X 와 Y 를 분류한 표이다. 이 표는 Z 의 값을 일정하게 유지시켜 Z 의 효과를 제거함으로써 XY 의 관계를 보여 준다.

부분분할표에서의 연관성을 **조건부연관성**(conditional association)이라고 하는데, 이는 변수 Z 의 값이 어떤 수준에서 고정되어 있다는 조건 하에서 Y 에 미치는 X 의 효과를 설명하기 때문이다. 부분분할표에서의 조건부연관성은 부분분할표를 결합해서 얻은 XY **주변분할표**(marginal table)에서의 연관성과 다르다. 다음 예제를 통해 이를 확인해 보자.

2.7.2 예제: 사형판결과 인종

〈표 2.9〉는 살인죄를 범한 피고들에 대한 사형선고 언도 여부에 인종이 미치는 영향을 조사해서 얻은 $2 \times 2 \times 2$ 분할표이다. 〈표 2.9〉에서 분류된 674명의 대상자들은 미국 플로리다주에서 12년 동안 살인죄로 기소된 피고인들이다. Y 는 “사형판결”을 나타내는 변수이고, X 는 “피고의 인종”을 나타내는 변수이며, Z 는 살해된 “피해자의 인종”을 나타내고 있다. 여기서 피해자의 인종을 제어변수로 간주하여, 피고의 인종, 즉 살인범의 인종이 사형판결에 미치는 영향을 조사해 보고자 한다. 〈표 2.9〉는 피해자의 인종의 각 수준에서 피고의 인종과 사형판결에 관한 2×2 부분분할표이다.

피해자의 인종과 피고의 인종의 각 결합 분류가 〈표 2.9〉에 나타나 있고, 〈그림 2.4〉는 피고 중에서 사형판결을 받은 비율을 보여 준다. 피해자가 백인일 때는 백인

표 2.9 피해자의 인종과 피고의 인종에 의해 분류된 사형선고판결

| 피해자의 인종 | 피고의 인종 | 사형선고 | | “예”의 비율 |
|---------|--------|------|-----|---------|
| | | 예 | 아니오 | |
| 백인 | 백인 | 53 | 414 | 11.3 |
| | 흑인 | 11 | 37 | 22.9 |
| 흑인 | 백인 | 0 | 16 | 0.0 |
| | 흑인 | 4 | 139 | 2.8 |
| 합 | 백인 | 53 | 430 | 11.0 |
| | 흑인 | 15 | 176 | 7.9 |

출처: M. L. Radelet and G. L. Pierce, *Florida Law Rev.* 43: 1-34(1991).

피고보다 흑인 피고에 대하여 사형판결의 비율이 $22.9\% - 11.3\% = 11.6\%$ 만큼 더 높은 것을 알 수 있다. 피해자가 흑인일 때는 흑인 피고의 사형판결이 $2.8\% - 0.0\% = 2.8\%$ 백인 피고보다 더 높았다. 따라서 피해자의 인종을 고정시켰을 때, 사형판결이 “예”일 확률은 백인 피고보다 흑인 피고가 더 높게 나왔다.

〈표 2.9〉의 아래 부분은 피고의 인종과 사형판결에 관한 주변분할표를 나타내고 있다. 이 표는 〈표 2.9〉에서 피해자의 인종의 두 수준에서 각 칸에 있는 도수를 합하여 얻어지는데, 예를 들어 $(11 + 4 = 15)$ 처럼 두 개의 부분분할표를 결합한 것이다. 전체적으로 백인 피고인 중에서 11.0%, 흑인 피고인 중에서 7.9%가 사형판결을 받은 것을 알 수 있다. 즉 피해자의 인종을 무시하게 되면, 사형판결을 받은 비율은 백

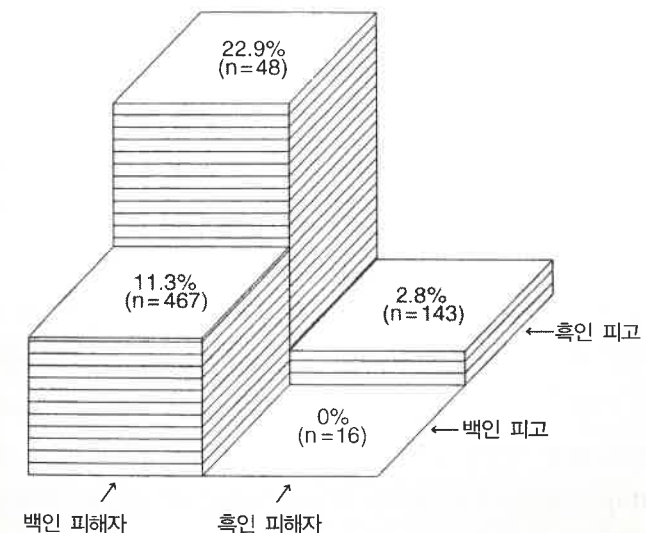


그림 2.4 사형선고를 받은 인종의 비율