

2.11.

<표 2.11>은 2008년과 2012년 미국 대통령 선거 투표 결과를 분류한다. 모비율 오즈비에 대한 95% 신뢰구간을 추정하고 구한 후 해석하라.

2012년 선거 (오바마) 2012년 선거 (롬니)

2008년 오바마	802	53
2008년 맥케인	34	494

오즈비의 95% 신뢰구간을 구하려면 로그 오즈비를 이용하여 신뢰구간을 계산한다. 우선 오즈비를 구한다.

1.오즈비(두 사건이 발생할 확률의 비율을 의미)

$$OR = \frac{(802 \times 494)}{(53 \times 34)}$$

2.95% 신뢰구간

신뢰구간 공식: $\log(OR) \pm Z(1.96) \cdot SE$

$$SE = \sqrt{\frac{1}{802} + \frac{1}{53} + \frac{1}{34} + \frac{1}{494}}$$

3.계산결과

계산 결과, 오즈비는 약 **219.86**이며, 95% 신뢰구간은 **140.89~343.09**이다.

4.해석

오즈비가 219.86이라는 것은, 2008년 오바마 후보에게 투표했던 사람들이 2012년에도 오바마에게 투표할 확률이, 2008년 맥케인에게 투표했던 사람들이 2012년에 롬니에게 투표할 확률보다 약 220배 더 높다는 것을 의미한다.

95% 신뢰구간은 140.89~343.09까지로, 이 범위는 오즈비의 불확실성을 나타내며, 오즈비가 이 범위 내에 있을 가능성이 95%라는 의미이다. 신뢰구간이 1보다 크므로, 두 사건 간의 관계는 통계적으로 유의미하다고 볼 수 있다.

2.17

인종	민주당	공화당	무소속
백인	871	821	336
흑인	347	42	83

a. 카이제곱 독립성 검정 결과:

- 카이제곱 통계량: 184.32
- p-value: 9.44×10^{-41} (매우 작은 값)
- 자유도: 2

인종	민주당	공화당	무소속
백인	988.04	700.07	339.89
흑인	229.96	162.93	79.11

- 기대 빈도:
- 해석:

p-value가 매우 작기 때문에, 귀무가설을 기각할 수 있다. 즉, 정당지지 성향과 인종은 독립적이지 않다는 결론을 내릴 수 있다.

b. 표준화 잔차 분석:

인종	민주당	공화당	무소속
백인	-3.72	4.57	-0.21
흑인	7.72	-9.47	0.44

표준화 잔차가 절대값 2 이상인 경우, 그 셀에서 기대 빈도와 관측 빈도 사이에 유의미한 차이가 있다고 보는데,

- 백인: 공화당에 대한 잔차가 4.57로, 예상보다 백인 공화당 지지자가 많다.
- 흑인: 민주당에 대한 잔차가 7.72, 공화당에 대한 잔차가 -9.47로, 예상보다 흑인 민주당 지지자는 많고, 공화당 지지자는 매우 적다.

c. 카이제곱을 두 개의 부분항으로 분할:

민주당과 공화당만을 포함하여 인종과 정당 지지 간의 관계를 검정하면, 흑인의 민주당 지지 성향이 매우 높고, 백인의 공화당 지지 성향이 높다는 점에서 연관성을 확인할 수 있다

2.21

1991년 일반사회조사에서 조사된 다음의 4X4 분할표는 흑인들에 대해서 직업 만족도와 수입에 대해서 조사한 결과를 보여 준다.

수입	매우불만족	약간만족	적절히만족	매우만족
<5000	2	4	13	3
5000~15000	2	6	22	4
15000~25000	0	1	15	8
>25000	0	3	13	8

a. 카이제곱 독립성 검정 결과:

카이제곱 통계량: 11.52

p-value: 0.2415

자유도: 9

기대 빈도:

<표준화 잔차>

수입	매우불만족	약간만족	적절히만족	매우만족	수입	매우불만족	약간만족	적절히만족	매우만족
< 5000	0.85	2.96	13.33	4.87	< 5000	1.25	0.60	-0.09	-0.85
5000 15000	1.31	4.58	20.60	7.52	5000 15000	0.61	0.67	0.31	-1.28
15000 25000	0.92	3.23	14.54	5.31	15000 25000	-0.96	-1.24	0.12	1.17
> 25000	0.92	3.23	14.54	5.31	> 25000	-0.96	-0.13	-0.40	1.17

해석: p-value가 0.2415로 귀무가설을 기각할 수 없다. 이는 직업 만족도와 수입 간에 통계적으로 유의미한 관계가 없다는 것을 의미한다.

표준화 잔차의 절대값이 2를 넘는 값이 없으므로, 큰 차이가 보이지 않는다는 것을 의미한다. 특정 패턴이 드러나지 않으며, 수입과 직업 만족도 간에 강한 상관관계는 발견되지 않았다.

b. 질적인 변수로 변환 후 검정:

수입에 대해 (3, 10, 20, 35) 점수, 직업만족도에 대해 (1, 3, 4, 5) 점수를 할당한 후 검정을 수행하면, 두 변수 간의 관계가 질적인 형태로 다뤄져 상관관계 분석이 가능해진다. 이는 카이제곱 검정이 아닌 상관계수와 같은 방식으로 분석되며, 연속적인 값의 차이를 반영하므로 결과가 달라질 수 있다.

이 방식이 (a)와 다른 이유는, (a)는 범주형 데이터를 비교하여 독립성을 검정하는 반면, (b) 방식은 변수를 연속적 또는 순위형 변수로 간주하고 분석하여 차이를 더 정밀하게 반영하기 때문이다.

2.23

<표 2.15>은 후두암을 치료하는 데 있어 수술과 방사선 치료를 비교하는 연구 결과이다.

	치료됨	치료안됨
수술	21	2
방사선치료	15	3

a. 검정 수행 및 해석:

주어진 R 코드를 통해 Fisher의 정확검정을 수행한 결과 양측 검정에서 p-value는 0.6384이다. 여기서 p-value가 0.05보다 크므로 귀무가설을 기각할 수 없다. 즉, 수술과 방사선 치료의 효과 차이가 통계적으로 유의미하지 않다는 결론을 내릴 수 있다.

b. 중앙 P-값 및 해석:

중앙 P-값을 계산하면, 이 값은 일반 P-값보다 좀 더 보수적인 검정을 할 수 있게 해준다. 이 경우 p-value가 여전히 0.6384로 크기 때문에 귀무가설을 기각하기 어렵다.