

빅데이터의 이해 HW10

- employee_data.sav 데이터에서 소수민족여부와 직종의 교차분석하기 -

영어영문학과(휴먼ICT) 201614010 박도준

1. 가설설정

귀무가설 H_0 : 소수민족 여부에 따라 직종의 분포가 다르지 않다.

귀무가설 H_1 : 소수민족 여부에 따라 직종의 분포가 다르다.

2. 교차분석결과와 해석

케이스 처리 요약

| | 케이스 | | | | | |
|----------------|-----|--------|-----|------|-----|--------|
| | 유효함 | | 결측값 | | 총계 | |
| | N | 퍼센트 | N | 퍼센트 | N | 퍼센트 |
| 소수민족 여부 * 직종구분 | 474 | 100.0% | 0 | 0.0% | 474 | 100.0% |

소수민족 여부 * 직종구분 교차 분석표

| | | | 직종구분 | | | 총계 |
|---------|-----|-------------|-------|-------|-------|--------|
| | | | 사무직 | 관리직 | 경영직 | |
| 소수민족 여부 | 아니오 | 개수 | 276 | 14 | 80 | 370 |
| | | 소수민족 여부 내 % | 74.6% | 3.8% | 21.6% | 100.0% |
| | | 총계의 % | 58.2% | 3.0% | 16.9% | 78.1% |
| | | 잔차 | -7.4 | -7.1 | 14.4 | |
| 예 | | 개수 | 87 | 13 | 4 | 104 |
| | | 소수민족 여부 내 % | 83.7% | 12.5% | 3.8% | 100.0% |
| | | 총계의 % | 18.4% | 2.7% | 0.8% | 21.9% |
| | | 잔차 | 7.4 | 7.1 | -14.4 | |
| 총계 | | 개수 | 363 | 27 | 84 | 474 |
| | | 소수민족 여부 내 % | 76.6% | 5.7% | 17.7% | 100.0% |
| | | 총계의 % | 76.6% | 5.7% | 17.7% | 100.0% |

카이제곱 검정

| | 값 | df | 점근 유의수준(양면) |
|--------------|---------------------|----|----------------|
| Pearson 카이제곱 | 26.172 ^a | 2 | .000 |
| 우도비 | 29.436 | 2 | .000 |
| 선형 대 선형 연결 | 9.778 | 1 | .002 |
| 유효 케이스 N | 474 | | |

a. 0 셀(0.0%)에 5 미만의 개수가 있어야 합니다. 예상되는
최소 개수는 5.92입니다.

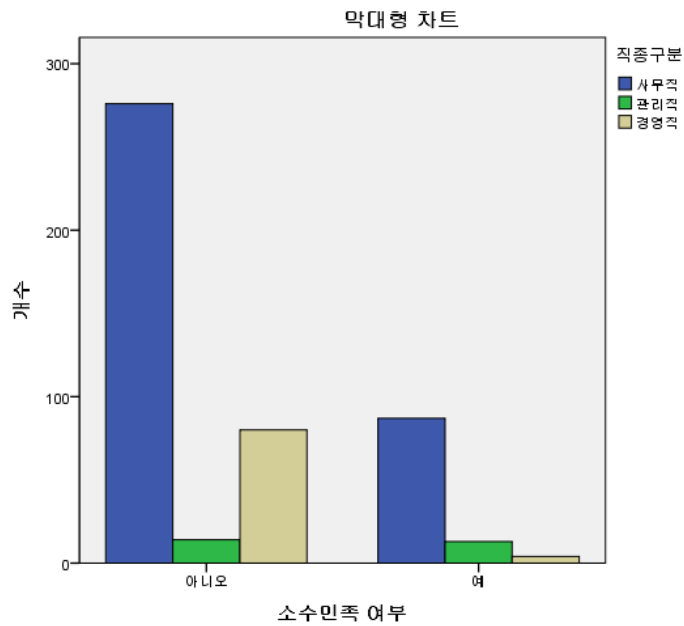
방향성 측도

| | | 값 | 점근 표준 오차 ^a | 근사 T | 근사 유의수준 |
|----------------------|---------|------|-----------------------|----------------|-------------------|
| 람다 | 대칭 | .000 | .000 | . ^b | . ^b |
| | 소수민족 여부 | .000 | .000 | . ^b | . ^b |
| | 종속 | .000 | .000 | . ^b | . ^b |
| | 직종구분 종속 | .000 | .000 | . ^b | . ^b |
| Goodman 및 Kruskal 타우 | 소수민족 여부 | .055 | .018 | | .000 ^c |
| | 종속 | | | | |
| | 직종구분 종속 | .021 | .006 | | .000 ^c |

- a. 귀무가설을 가정하지 않습니다.
b. 점근 표준 오차가 0이므로 계산할 수 없습니다.
c. 카이제곱 근사값 기준

소수민족이 아닌 사원의 경우, 276명(74.6%)이 사무직, 14명(3.8%)이 관리직, 80명(21.6%)이 경영직에 해당한다. 소수민족인 사원의 경우, 87명(83.7%)이 사무직, 13명(12.5%)이 관리직, 그리고 4명(3.8%)이 경영직에 속한다. 카이제곱 검정결과(p 값=0.000)에 의하면 귀무가설이 기각되고 대립가설이 채택되어 소수민족 여부에 따라 직종의 분포가 다르다고 할 수 있다. 기대빈도와의 잔차를 살펴보면, 소수민족이 아닌 사원 중 경영직에 속하는 사원은 14.4명 많고, 소수민족인 사원의 경우, 사무직과 관리직이 각각 7.4명, 7.1명이 많다. 즉, 소수민족이 아닌 사원의 경우 경영직이 많고, 소수민족인 사원의 경우 사무직과 관리직이 상대적으로 많다고 볼 수 있다.

소수민족 여부와 직종구분 모두 명목척도이기에, 연관성의 정도 람다로 나타낼 수 있는데, 예측 오류 감소율(람다)은 소수민족 여부 종속과 직종구분 종속 모두 0.000이다. 람다의 근사유의확률(p 값)은 0으로서 유의수준 5%보다 작기 때문에 연관성이 유의하게 있다고 볼 수 있다.



위 수평누적 막대도표는 소수민족 여부에 따른 직종 분포의 수치적 정보를 그래프로 시각화하여 보여주고 있다. 하지만 전체 사원 474명 중 소수민족이 아닌 사원이 370명, 소수민족인 사원이 104명으로써 약 3배 이상의 인원 수 차이가 있다는 점을 감안하여 볼 때, 위 그래프로 단순한 비교를 내리기는 어렵다고 판단된다.

2. 교차분석 보고서 작성

소수민족 여부 * 직종구분 교차 분석표

개수

| | | 직종구분 | | | 총계 |
|---------|-----|------------|----------|----------|-----|
| | | 사무직 | 관리직 | 경영직 | |
| 소수민족 여부 | 아니오 | 276(283.4) | 14(21.1) | 80(65.6) | 370 |
| | 예 | 87(79.6) | 13(5.9) | 4(18.4) | 104 |
| 총계 | | 363 | 27 | 84 | 474 |

위 표는 소수민족 여부와 직종에 대한 교차표이다. 위 교차분석의 카이제곱 통계량은 26.172이고 유의확률은 0.000이다. 유의확률값이 일반적인 유의수준인 0.05보다 작기 때문에 유의수준 5%에서 귀무가설을 기각하게 된다. 따라서 소수민족 여부에 따라 직종의 분포가 유의하게 다르다고 할 수 있다. 사무직의 경우, 비소수민족 사원이 283.4%, 소수민족 사원이 79.6%이며, 관리직의 경우, 비소수민족 사원이 21.1%, 소수민족 사원이 5.9%이고, 경영직의 경우, 비소수민족 사원이 65.6%, 소수민족의 사원이 18.4%이다. 세 직종 모두 비소수민족이 소수민족에 비해 많은 사원을

고용하고 있다는 사실을 확인할 수 있다.