

기술통계

정의: 어떤 현상을 수로 측정하여(평균, 분산, 편차, 범위 등) 요약, 설명하는 통계적 방법

기능

1) 요약의 기능 (summarizing)

- 예) 20문항으로 구성된 어떤 설문지를 50명에게 응답하도록 해서 자료를 수집한 경우
- 원 자료(raw data)를 보려면 $20 \times 50 = 1,000$ 개를 검토해야 하는 수고가 따름
 - 원 자료는 그대로 파악해서 대상자의 특성을 이해하기가 매우 어렵다
 - 이것을 평균이나 편차, 범위 등으로 요약해서 보고하면 이해가 쉬울 것임

2) 상관관계(correlation)

- 예1) 광고비를 증가하면 판매량은 어떻게 변화할까?
 즉 광고비와 판매량 간에는 어떤 상관관계가 있을까?
- 예2) 흡연을 많이 하면 과연 폐암으로 발전할 확률이 높은가? 흡연과 폐암의 상관관계?

방법

1) 비율(proportion): 항상 소수로 나타내며 비율은 절대 1보다 클 수 없다.

각 유목(category)의 비율의 합은 항상 1이다. $n_1/N + n_2/N + n_3/N = 1$

2) 백분율(percentage): 퍼센트란 백 당 얼마라는 의미임 $n_1/N \times 100 \%$

퍼센트의 기능은 집단의 사례 수가 다르므로 직접적 비교가 불가능했던 등위점수의 약점을 보완해준다. (예컨대, 달리기에서 2등 했다 -> 몇 명 중 2등이냐에 따라 의미는 다름)

장점: 통계적 소양이 부족해도 쉽게 해석이 가능하다.

단점: 상대적 지위는 명확히 제시해주지만 측정하려는 능력의 정도는 그대로 나타내는 것이 아님 (예컨대, 90%~95%의 차이는 45%~50%의 차이와 같지 않다)

일반적으로 전자(상위 5%)의 능력차이는 후자(중위5%)의 능력차이보다 크다.

3) 비(ratios): A/B (B분의 A) 또는 A:B (A대 B ; B라는 수치에 대한 A라는 수)

예) 여성100명과 남성105명 = 여자와 남자의 구성 비(ratio) = 100: 105 또는 100/105

4) 율(rates): 편리한 수치를 기준으로 하여 그 수치 당 얼마라는 식으로 표현

예) 출산율: 인구 10만명 당 얼마, 신문 보급율: 1000가구 당 얼마 등)

5) 빈도분포(frequency distribution): 원 자료(raw data)의 수치들을 같은 점수 혹은 같은 유목으로 분류해서 일정한 표로 정리하거나 요약하는 방법 또는 <표> 그 자체

	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효 370-374	3	6.0	6.0	6.0
365-369	4	8.0	8.0	14.0
360-364	5	10.0	10.0	24.0
355-359	7	14.0	14.0	38.0
350-354	7	14.0	14.0	52.0
345-349	9	18.0	18.0	70.0
340-344	6	12.0	12.0	82.0
335-339	5	10.0	10.0	92.0
330-334	4	8.0	8.0	100.0
합계	50	100.0	100.0	