

[기획]

퍼펙트 論述

수리논술 실제-통계

잘못 활용맨 전혀 다른 해석 통계오류 짚어내는 훈련해야



이승중 1318논술연구소 수리논술팀장

통계적 추론은 많은 경우 정보의 획득에 있어 매우 편리한 기반을 제공한다. 하지만 잘못 활용할 경우 실제와 전혀 다른 해석을 낳아서 목적했던 정보를 얻는 것은 요원한 일 이 될 수도 있다.

예를 들어보자. 65세 이상 노인 인구가 자치단체 내에 10만명 정도 있는 한 지방자치단체에서 노인들의 한 달 평균 용돈이 5만 원 미만이면 노인들을 위한 복지 예산을 증액하기로 결정하고, 자치단체 내에 있는 노인들 100명을 무작위 추출하여 한 달 용돈을 조사하였다. 다음은 그 결과이다.

조사대상 : 65세 이상 노인 평균 용돈 : 4만8천원 표준편차 : 4만원

이 자료를 바탕으로 지방자치단체는 노인들의 평균 용돈이 5만원 미만인 것으로 판단하고 노인들에 대한 복지 예산을 증액하기로 결정하였다. 그렇다면 이 자료는 얼마나 믿을 만한가? 그리고 지방자치단체의 결정은 적절하다고 할 수 있는가?

지방자치단체의 결정에 대한 적절성은 순수하게 수리적으로만 논할 수 있는 부분은 아니다. 그러나 자료의 신뢰성에 대해서는 통계적 기법을 통해 판단해 볼 수 있고, 역시 이를 통해서 자치단체 결정의 적절성에 대해 간접적으로 이야기할 수는 있을 것이다.

이 경우 조사 대상 노인들의 평균은 표본평균에 해당하고, 모집단이 정규분포를 따르지 않더라도 모집단의 크기가 충분히 크다면 표본평균은 정규분포를 따르는 확률변수로 볼 수 있고, 표본평균의 평균은 모집단의 평균과 동일하고, 표준편차는 모집단의 표준편차를 표본의 크기의 제곱근으로 나눈 값과 동일하다.

따라서 모집단의 평균을 m, 표준편차를 sigma라고 하면, 표본평균 x는 N(m, sigma/sqrt(n))을 따른다. 여기서 모집단의 표준편차를 알 수 없으므로 표본의 표준편차로 대체하면 z는 N(m, 4000)을 따르게 된다.

따라서 z = (x-m)/(sigma/sqrt(n)) 이라 하면 z는 표준정규분포를 따르게 되고, P(-1.96 <= z <= 1.96) = 0.95이므로 P(40160 <= m <= 55840) = 0.95가 되고, 이는 자치단체 내의 노인들의

평균 용돈을 신뢰도 95%로 추정한다면 4만160원에서 5만5천840원 사이에 분포할 것으로 추정할 수 있다는 것을 의미한다.

실제 자치단체 내 노인들의 한 달 평균 용돈이 5만원 미만으로 결정되는 범위에서 추정했을 경우의 신뢰도는 약 69% 정도밖에 되지 않는다. 이는 조사 결과를 바탕으로 자치단체 내 노인들의 한 달 평균 용돈이 5만원 미만이라고 단정적으로 이야기한다면 이 말은 대략 69% 정도의 신뢰도를 갖는다는 것을 의미한다. 따라서 이 조사를 통해 지방자치단체에서 노인들의 평균 용돈이 5만원 미만인 것으로 단정적으로 판단하기는 어려울 것으로 보인다.

또한 표준편차가 4만원으로 매우 큰 것은 노인들 간의 용돈의 격차가 크다는 것을 반영하는 것이므로 복지예산 편성하더라도 용돈이 부족한 노인들에게 적절히 배분될 수 있도록 예산을 집행하는 것이 예산을 편성하는 것만큼이나 또는 그 이상으로 중요하다는 것을 의미한다.

여기서 통계적 해석은 너무도 쉽고도 자연스럽게 오류를 범하기 쉬운 분야이다.

통계적 오류의 아주 큰 특징은 매우 그럴 듯해서 잘 짜인 논리적 사고체계를 갖고 있지 않다면 그냥 지나치기 쉽다는 것이다. 이러한 통계적 오류를 정확히 짚어내고 올바르게 재해석하기 위해서는 훈련이 필요하다.

통계적 오류의 다양한 예를 통해서 어떻게 통계적 오류를 찾아낼 수 있고, 정확한 재해석을 할 수 있는지 살펴보자.

다음은 중앙대학교에서 2007학년도 수시 1학기 인문계 논술고사에서 출제되었던 문제이다.

Table with 4 columns: 가해자, 피해자, 백인, 흑인. Rows show percentages for 백인 (97%/467명) and 흑인 (3%/16명).

이제 한 인권주의자가 위에 발표된 통계를 보완하여 아래의 <표 2>를 제시하였다.

Table with 4 columns: 피해자, 가해자, 사형, 기타, 사형 선고율(%). Rows show percentages for 백인 and 흑인.

이제 한 인권주의자가 위에 발표된 통계를 보완하여 아래의 <표 2>를 제시하였다.

Table with 4 columns: 가해자, 피해자, 사형, 기타, 사형 선고율(%). Rows show percentages for 백인 and 흑인.

이제 한 인권주의자가 위에 발표된 통계를 보완하여 아래의 <표 2>를 제시하였다.

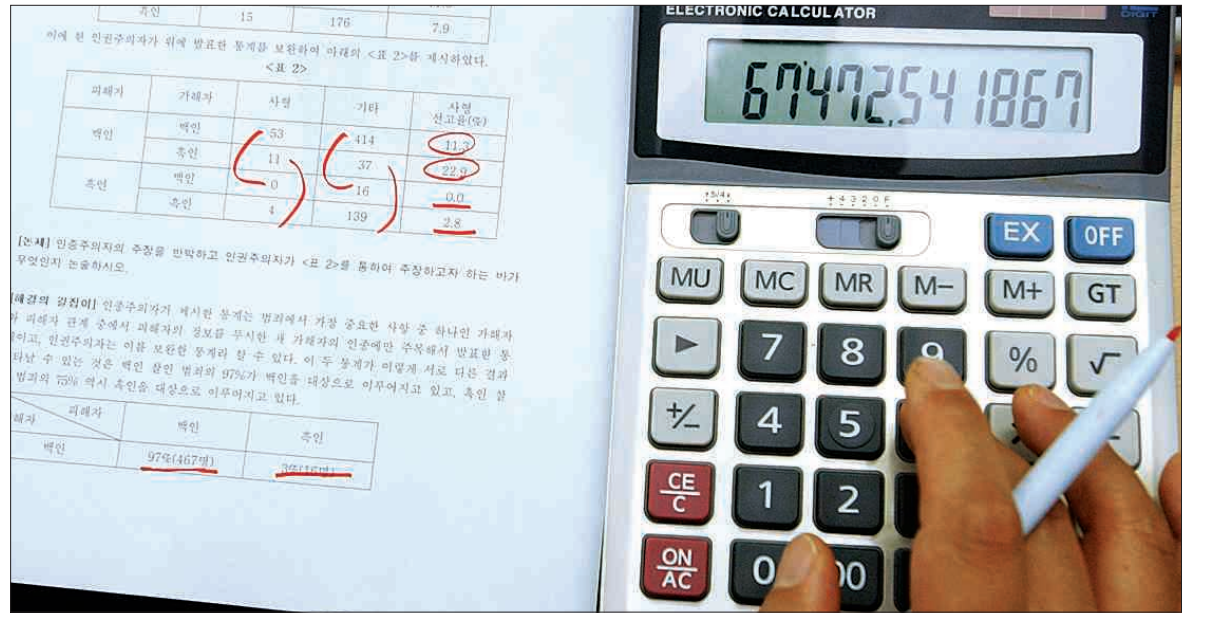
Table with 4 columns: 가해자, 피해자, 사형, 기타, 사형 선고율(%). Rows show percentages for 백인 and 흑인.

이제 한 인권주의자가 위에 발표된 통계를 보완하여 아래의 <표 2>를 제시하였다.

Table with 4 columns: 가해자, 피해자, 사형, 기타, 사형 선고율(%). Rows show percentages for 백인 and 흑인.

이제 한 인권주의자가 위에 발표된 통계를 보완하여 아래의 <표 2>를 제시하였다.

Table with 4 columns: 가해자, 피해자, 사형, 기타, 사형 선고율(%). Rows show percentages for 백인 and 흑인.



[논제] 인권주의자의 주장을 반박하고 인권주의자가 <표 2>를 통하여 주장하고자 하는 바가 무엇인지 논술하시오.

[해결의 길잡이] 인권주의자가 제시한 통계는 범죄에서 가장 중요한 사항 중 하나인 가해자와 피해자 관계 중에서 피해자의 정보를 무시한 채 가해자의 인종에만 주목해서 발표된 통계이고, 인권주의자는 이를 보완한 통계라 할 수 있다. 이 두 통계가 이렇게 서로 다른 결과가 나타날 수 있는 것은 백인 살인 범죄의 97%가 백인을 대상으로 이루어지고 있고, 흑인 살인 범죄의 75% 역시 흑인을 대상으로 이루어지고 있기 때문이다.

Table with 4 columns: 가해자, 피해자, 백인, 흑인. Rows show percentages for 백인 (97%/467명) and 흑인 (3%/16명).

이 통계를 활용하면 인종과 살인 범죄의 연관성을 주장하기에는 무리가 있다. 즉, 살인 범죄에 대한 인종간의 연관성이 없음에도 불구하고 이 둘이 서로 연관성이 있다고 무리하게 주장하고 있는 인권주의자의 결론에 오류가 발생하고 있다는 뜻이다.

많은 사회적 현상은 통계적 수치에 의해 표현되고, 사람들은 이를 통해서 사회적 현상을 분석한다. 통계 자료는 사회적 현상을 있는 그대로 반영함에도 불구하고 그 자체로 오류를 범하게 될 여지를 다분히 갖고 있다.

이들 자료 자체가 갖고 있는 정확성의 문제가 아니다. 아무리 정확한 자료라 하더라도 이를 해석하는 과정에서 현상을 있는 그대로 정확히 받아들이지 못하면 사회적 현상은 왜곡되어 나타나게 된다는 것을 의미한다. 사회적 현상의 정확한 분석을 위해서는 정확하고도 충분한 양의 자료를 수집하는 것만큼이나 수집된 자료를 어떻게 해석하느냐가 매우 중요하다.

자료를 올바르게 해석하는 것이 얼마나 의미 있는 과정인지를 볼 수 있는 하나의 재미있는 사례를 살펴보자. 1898년 미국과 스페인이 쿠바에서 벌인 전쟁에서 미 해군의 전사율은 1천명당 9명이었다. 같은 기간 뉴욕의 사망률은 1천명당 16명이었다. 전쟁터에 나가는 것보다 대도시에서 죽을 확률이 더 높게 나온 것이다. 해군 징병관들은 이를 근거로 해군에 입대하는 것이 뉴욕에 사는 것보다 안전하다고 선전했다. 어쩌서 그런 일이 벌어졌 것일까?

해군은 대부분 육체적으로 건장한 청년들로 구성돼 있는데 반해 뉴욕 시민 중에는 갓난아기와 노인, 환자 등이 포함돼 있다. 그래서 뉴욕의 사망률이 더 높게 나올 수도

있지만 그것은 의미 있는 수치는 아니다. 굳이 비교하려면 해군 징병들과 같은 연령대의 청년층 사망률을 따져야 한다.

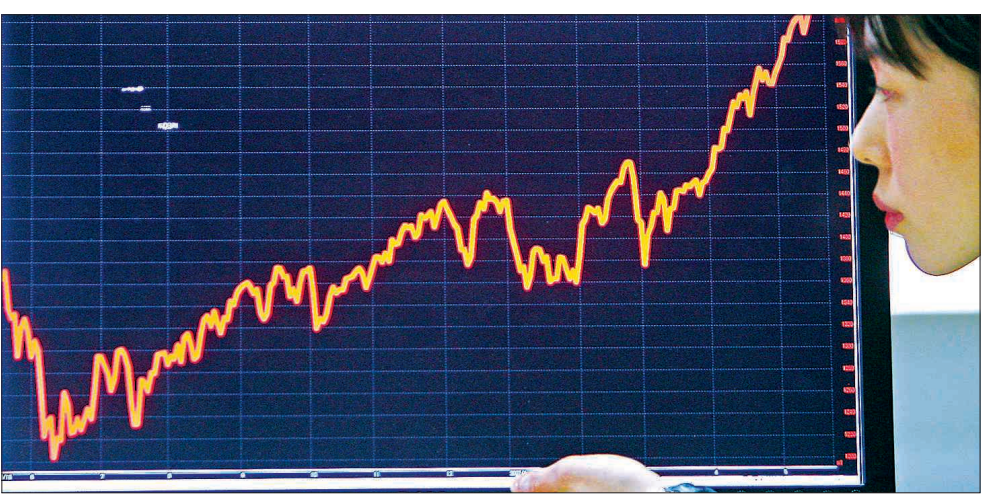
다만 예를 들어 보자. 훈련 중에 강을 건너야 하는 병사들이 지도에 표시되어 있는 대로 평균 수심이 1.3m라는 사실만 믿고 도강을 한다면 어떻게 될지 생각해 보자. 평균은 바로 데이터들이 모여 있는 특성을 나타내는 대푯값이다. 그러나 평균은 숫자들이 모여 있는 중심점을 나타내는 유용한 정보이지만 분포 전체의 모양을 보여 주는 것은 아니다. 때로는 흩어져 있는 정도를 나타내는 산포도가 더 중요한 경우가 많다. 따라서 흩어진 정도를 모르거나 무시한다면 잘못된 판단을 하게 된다. 이 경우도 그러한 경우에 해당될 수 있다. 대형하 강물의 수심이 고르다면 큰 문제가 없을 수도 있지만, 강물의 수심이 고르지 않고, 얇은 곳은 매우 낮고, 깊은 곳은 매우 깊어서 전체적인 평균이 1.3m인 것이라면 강물의 깊은 곳을 지날 때는 병사들이 큰 위험에 처할 수도 있다.

이와 같이 자료를 해석함에 있어서 어떠한 부분들이 주요하게 고려되어야 하고, 놓치기 말아야 하는 것들에는 어떤 것이 있는지 등을 우선 확인하고 점검해 보아야 할 필요성이 있다.

[실전 문제] 이혼율 선정 방식으로는 여러 가지가 있을 수 있지만, 그 중 '배우자가 있는 사람의 이혼율(有配偶者離婚率)'은 특정 연도 말을 기준으로 혼인 부부의 수를 분모로, 특정 연도 중에 이혼한 부부의 수를 분자로 하여 산정한 수치를 천분율로 나타내는 방식이다. 예를 들면, 2002년 말 현재 혼인한 부부의 수가 1천101만1천902쌍이고 2002년 중에 이혼한 부부의 수는 14만5천300쌍이므로, 2002년 이혼율의 나라의 이혼율은 1.3%라는 것이다. 매년 이혼율의 변화 추이를 비교할 때와 이혼율은 성향에 대해 절대적인 판단을 내릴 때로 나누어, 각각 배우자가 있는 사람의 이혼율의 장점을 단점을 설명하여라.

* 출제 의도 및 key-point 통계 수치에 시간적 요소가 개입하게 되어 변화의 정도나 누적적인 계산으로 변형하게 되면, 합수의 미분과 적분의 개념과 그 맥을 같이하게 된다. 사회적인 현상에서 이를 이해하고 이를 수치화할 수 있는지를 기본적인 수준에서 확인하고자 했다. 배우자가 있는 사람의 수와 이혼하는 사람의 수를 비교하여 산정하는 방식이 가질 수 있는 오류에 대해서 생각해 본다면 어렵지 않게 해결될 수 있는 문제이다.

동영상 강의 www.nonsul.1318hi.com



대인동삼일부동산
모텔 목욕탕 및 각종 상가건물
을 빨리 팔고 싶으시거나 임대료를 내고 싶으시면 저희 부동산으로 오십시오.

LC타워(주)
첨단지구 LC타워
최대한 최고의 요지
최적의 조건으로 임대합니다.

일가공인중개사
상가건물 매매
*중장부 대지 75평 5층건물 24/1,700만 21억
*금강동 대지 150평 3층 388만 900만 19억

한일지도판매(주)
2007 최신광주지도
책 판매개시!!
*타사제품 미기득권 부분

호성공인중개사
대지매매(교회, 원, 투룸부지)
*북구 양산지구 대지 1,000평, 500평 3면도로 접 평당 135만

토우드공인중개사
토지
*신마동 800평 당 도로 접 평당 1만 1천

다우공인중개사
오봉리 전원주택지 (평당 30만)
*남양 현대자동차 출고장 건너편 오봉리(낙산) 바로 옆 210~360평 총 8필지 중 4필지 매각

법원경매
지지경매컨설팅(주)
(법원 경매임찰장)
☎ 062-226-0047
현장조사, 권리분석, 수익분석, 입찰신청, 대출안내, 명도까지 원스톱서비스로 확실한 채테크