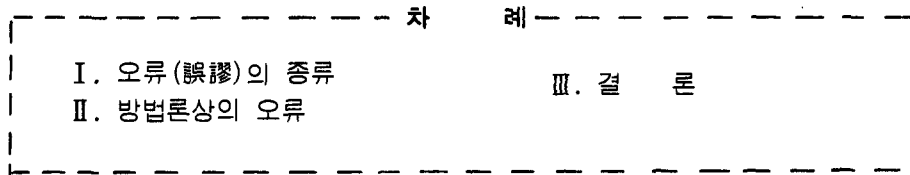


연구과정상의 오류 (誤謬)

— 방법론상의 오류를 중심으로 —

고 응 린*



I. 오류 (誤謬)의 종류

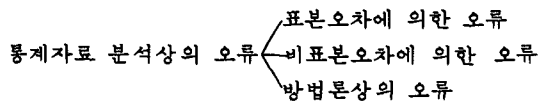
생명체를 대상으로한 조사나 연구결과의 특징은 그 변동의 폭이 크다는 데 있다.

특히 환자를 대상으로한 보건의학적 또는 행동과학적인 연구과제인 경우에는 그 변동의 크기는 더욱 심해지며 이와 같은 극심한 변동속에서도 어떤 결론이나 판정을 내려야하는 것이 연구자들이 당면하고 있는 고충의 하나라 볼 수도 있다.

이와 같은 고충을 덜기 위해서 또는 다양하게 변동하는 변동속에서도 그 무엇인가를 탐구해 보기 위해서 우리가 흔히 도움을 받게 되는 것이 바로 통계학적 방법론이라 할 수 있다.

그러나 이 통계학적방법론을 잘못 이용하거나 또는 잘못 해석함으로써 발생하는 오류 (誤謬)도 또한 연구자의 입장에서 볼 때 유의하지 않을 수 없는 고충의 요소라 할 수 있을 것이다.

일반적으로 연구결과 분석상 발생할 수 있는 오류의 종류는 크게 다음과 같이 분류해 볼 수 있다.



여기서 "표본오차에 의한 오류"라는 것은 우리의 관측대상이 주로 표본이기 때문에 표본에 따라 관측

결과가 달리 나와서 생기는 오류이다.

이때 표본에 따라 달리 나타나는 관측결과의 통계량(예를 들면 평균치)간의 변동의 크기를 표본오차 (Sampling error)라고 부르며 이론적으로 유도된 공식에 의하여 산출하게 된다.

이 표본오차는 통계학적이론에 의하여 어느 정도 조정 또는 관리할 수 있는 오류이다.

따라서 이 표본오차는 우리가 흔히 말하는 잘못 (Mistake)이라는 개념에 속하는 오류라고는 보기 힘들다고 표현하는 것이 타당할 것이다.

그러나 위에 제시된 "비표본오차"나 "방법론상의 오류"는 모두 잘못 (Mistake)이란 개념에 속한다고 볼 수 있는 오류들이다.

즉 "비표본오차 (Non-Sampling error)는 관측방법이나 응답내용 또는 기재나 계산 등의 잘못으로 인한 오류를 뜻하므로 이것은 "통계학적 이론"에 의하여 관리될 수 있는 오류라고는 볼 수 없다.

따라서 이와 같은 "비표본오차"는 관측방법 (측정기, 설문내용등)자체의 정확성을 높임으로써 관리할 수 있는 오차라 할 수 있다.

끝으로 "방법론상의 오류"란 조사 또는 실험계획의 잘못, 자료분석 방법의 잘못, 분석결과의 해석상의 잘못 등으로 인한 오류를 뜻한다.

여기서는 주로 이 "방법론상의 오류"를 중심으로 생각해 보기로 한다.

* 한양대학교의과대학

II. 방법론상의 오류

이런 오류로부터 우리 자신을 보호할 수 있는 방법이란 우선 우리 주변에서 많이 발생하고 있는 그런 오류의 내용을 알아보는 길이라 할 수 있을 것이다.

Schor박사는 1964년도 3개월 동안에 미국에서 발간된 의학분야의 학술잡지 중에서 이름난 것 10가지를 선택하여 그곳에 실려 있는 연구조사 논문의 내용을 통계학적으로 검토해 본 결과 다음과 같은 결과를 얻었다.

조사한 논문진수	우수하다고 판정된논문	수정이필요하다 판정된논문	완전실패작으로 판정된논문
149(100%)	41(27.5%)	101(67.8%)	7(4.7%)

통계자료의 분석과정과 결론유도에 있어서 타당하다고 볼 수 있는 논문이 28% 정도에 지나지 않았다는 데 우리의 관심을 끌게 한다.

이와 같은 사실은 연구조사를 기획하는 데 있어서나 그 결과를 분석하는 데 있어 통계학적 방법론을 올바르게 이용한다는 것이 얼마나 중요한가를 반영해 주고 있다.

이 잘못된 논문들이 가장 빈번하게 범한 오류의 내용을 알아본 결과는 다음과 같다.

- 1) 대조군(Control group)이 필요한데도 대조군을 설정하지 않고 연구결과를 분석하므로써 범한 오류.

예1) 어떤 약을 투약하기 전에 관측한 혈당치에 비해 투약후 5시간후에 관측된 혈당치가 높게 나타났다가 그 약에 혈당상승작용이 있다고 결론내리는 경우.

이런 때는 투약실험을 할 때 다른 쪽에는 약을 전혀 주지 않고 5시간 후에 혈당치에 어떤 변화가 왔는지를 알아보지 않고서는 아무런 결론을 내릴 수 없다.

예2) 80세 이상 장수한 노인들을 대상으로 조사해 본 결과 70%에서 새벽산책을 하는 것으로 응답했다고 "새벽산책"이 장수에 관련있는 것같이 결론내리는 경우

이때도 대조군, 즉 80세 이상 살지 못한 사람들의 "새벽산책"을 알지 못하고서는 아무런 결론을 내리지 못한다.

- 2) 어떤 (진단) 검사법의 적중률을 평가할 때 "민감도(sensitivity)"란 관측하고 "특이도(specificity)"는 계측하지 않거나 또는 제시하지 않고 결론을 내림으로써 범한 오류.

(여기서 "민감도"란 해당 질병에 걸려 있는 환자

들에게 이 진단검사를 적용시켰을 때 그중 몇%가 양성으로 나타나는가를 표시한 비율이고 "특이도"란 그 질병에 걸려 있지 않는 사람들에게 그 검사를 적용시킨다면 몇%에서 음성으로 나타나는가를 표시한 비율이다)

예1) 암환자 중 99%에서 어떤 피부반응검사 결과가 양성으로 나타났다고 해서(즉 이때 민감도는 99%)그 피부검사를 암진단에 이용할 수 있다고 결론내리는 경우.

이때는 암환자가 아닌 사람들에게 이 피부반응검사를 적용시켰을 때 몇%가 음성으로 나타나는가(즉, 특이도의 크기)를 알지 못하면 아무런 결론을 내릴 수 없다.

즉 이 피부반응검사의 민감도는 위에서와 같이 99%이었다 해도 그 특이도가 0%라면(즉 암환자 아닌 사람들 중에서 피부검사결과가 음성인 경우가 한 건도 없었다면)이 피부검사법은 암진단에 아무런 도움을 주지 못할것이다.

예2) 어떤 진단검사방법의 "민감도"만 가지고서는 그 검사법의 이용가치를 평가할 수 없다는 것은 다음과 같은 예를 생각해 보면 곧 알 수 있다.

즉 만일 자궁암을 진단하기 위해서 환자의 "눈이 돌인가?"란 검사를 하는 경우를 생각해 보자. 자궁암 환자 모두가 눈이 돌이 있으므로 이때 "눈이 돌인가?"란 검사법의 민감도는 100%이다.

그러나 "자궁암환자가 아닌 사람들의 경우도 모두 "눈이 돌"이므로 특이도는 0%이다. 따라서 이 "눈이 돌인가"를 검사해봄으로써 자궁암을 진단한다는 것은 불가능하다는 것을 알 수 있다.

- 3) 통계학적 방법론을 잘못적용함으로써 범하는 오류

원래 통계학적 방법론이란 각기 어떤 가정과 조건하에서 유도된 것들이다.

따라서 통계학적 방법론을 이용하려면 우선 그 가정과 조건들이 충족되어 있는가를 확인해 나가면서 적절한 방법론을 선택해야 한다.

예1) 이론적으로 산출된 빈도가 5보다 작게 나타난 경우에 χ^2 -test법을 적용시키는 경우.

원래 이 χ^2 -test는 이론적인 빈도가 5 이상은 되어 적용을 추천할 수 있다.

예2) 어떤 약제의 혈압에 미치는 영향을 실험하는 경우 투약전의 평균혈압치와 투약후의 평균혈압치를 비교하여 "통계학적으로 유의한 차이가 없다"라고 판정내는 경우.

이런 때는 투약전과 후의 혈압은 같은 환자로부터

연계 되므로(즉, 작을 지을 수 있는 때이므로)위와 같은 평균치의 비교방법보다는 환자 개개인의 투약전과 후의 혈압의 차이를 구하고 이 차들의 평균치를 가지고 0에 가깝다고 볼 수 있는가 없는가를 통계적으로 추정해 보는 것이 타당하다.

4) 통계학적인 검정결과를 잘못 해석하여 범하는 오류.

예1) 두 표본 평균치의 차이를 통계학적으로 검정하면 결과 $P < 0.01$ (즉 1%)로 나타났을 때 일반적으로 유의 (Significant)한 차가 있다고 판정내리고 있다.

그런데 이 결과를 마치 그 유의한 차가 있다는 판정이 맞을 가능성이 99%가 된다는 것으로 잘못 해석하는 수가 있다.

이 판정결과의 참뜻은 다음과 같다.

$P < 0.01$ 이라는 표시는 두 모집단의 "평균치가 같다"는 가정이 사실일 때 두 표본평균치의 차이가 실제로 관측된 것 만큼 이상으로 나타날 확률이 0.01 (즉 1%)보다 작다는 뜻이다.

이때 이 1%라는 확률을 작다고 (즉 드문일이 발생했다고) 보고 결국 "두 모집단의 평균치가 같다"는 가정을 버리고 그 대신 차이가 있다는 판정을 내리게 된 것이다.

따라서 $P < 0.01$ 에 대한 해석은 "실은 차이가 없는 데도 불구하고 차가 있다"고 잘못 결론 내리게 될 위험성이 1% 이하라는 뜻으로 해석하는 것이 타당한 것이다.

그렇다고 "유의한 차가 있다"고 내리게 된 경우 이 판정결과가 맞을 가능성이 99% 이상 된다는 뜻은 아니다. (이 "맞을 가능성"이란 영원히 알아낼 수 없다. 모집단의 진짜 평균치가 주어지지 않는 한...)

5) 극히 작은 수의 표본을 관측하여 얻은 자료로부터 통계학적으로 "Non-significant"하다는 판정이 내려졌을 때 두 표본간에는 차이가 없다는 것을 너무 강조함으로써 범하는 오류.

예를 들어 두 모집단의 평균치 사이에 차이가 있을 때 그 사실을 밝혀내는 데 필요한 표본의 크기는 그 모집단의 평균치의 차이의 크기에 반비례한다.

즉 그 차이가 크면 클수록 적은 수의 표본을 관측해 봄으로써 그 차이가 있다는 것을 증명할 수 있을 것이다.

만일 두 모집단의 평균치 사이에 차이는 있으나 그 크기가 작을 때에는 상당히 큰수의 표본을 관찰해야만 비로소 통계학적으로 유의한 차가 있다는 결론을 얻게 될 것이다.

거꾸로 말하면 표본의 크기를 크게 잡기만 하면 언

제나 두 평균치 사이에는 "Significant하다"라는 통계학적 판정을 내리도록 유도할 수도 있다는 해석도 가능한 것이다.

따라서 통계학적 검정결과가 "유의하다" 또는 "유의하지 않다"로 나왔을 때 반드시 점검해 보아야 할 것은 표본의 크기가 얼마나 되는가 하는 것이다.

즉 검정결과가 "유의하다"로 나타났을 때는 표본의 크기가 아무리 작아도 그 결론은 뜻이 있는 것은 사실이지만 "유의하지 않다"라는 판정이 내려졌을 때 표본의 크기가 작을 때는 그 결론을 별 뜻이 없다고 보아야 한다.

결국 "Non-significant"라는 판정이 나왔다면 이것은 두 모집단의 평균치간에 차이가 없다고 강조하는 것 보다는 "단지 이 정도의 작은 수의 표본가지고는 차이가 있다는 것을 주장할 수 없었다"라고 표현하는 것이 올바른 해석이라 할 수 있을 것이다.

예1) 몇명을 대상으로 관측했는지를 제시해 놓지 않고 통계학적으로 "Significant" 또는 "Non-significant" 등의 결론을 내리는 경우 이것은 아무런 의미가 없다.

반드시 관측대상이 된 표본의 크기를 제시해 놓아야 한다.

극단적인 예로서 1,000,000마리의 동물을 대상으로 실험해 본 결과 겨우 통계학적으로 Significant한 차이가 있다는 판정을 얻을 수 있었다면 그 차이가 임상적으로 실용의 가치가 있을지 의심이 갈 것이다.

실용성있는 차이가 있다고 인정할 수 있으려면 10여 마리 정도의 작은 수의 동물실험으로부터 "Significant"하다는 판정에 도달할 수 있어야 할 것이다.

예2) "5마리 정도의 동물실험결과를 가지고 결론 내린 논문도 가치가 있는가?"라는 질문은 그 결론의 내용에 따라서 다른 답을 얻게 된다.

즉 그 결론이 "Non-significant"인 경우는 답은 "가치가 없는 논문이다"가 될 것이고 만일 그와 반대로 "Significant"인 경우는 "가치가 있는 논문이다"라는 답을 얻게 될 것이다.

6) 통계학적인 검정을 거치지 않고 표본의 정보를 가지고 모집단에 대한 결론을 내림으로써 생기는 오류.

예1) 재래식 치료방법으로는 50%정도가 치료되었는데 이 새로운 치료방법으로서는 70%의 치료를 나타냈다고 해서 곧 후자가 전자에 비해 더욱 우수한 치료효과가 있다고 결론내리는 경우.

III. 결 론

- 1) 우리들의 관측대상은 거의 모두가 표본이다.
따라서 이 표본으로부터 얻은 정보를 가지고 모집단에 관한 어떤 결론을 내리려면 반드시 통계학적인 방법론의 힘을 빌려야 한다.
- 2) 통계학적 방법론이란 모두가 자기 특이한 가정과 전제조건하에서 유도된 것이므로 이용에 앞서 그 관측자료가 그런 가정과 조건을 어느 정도 충족시키고 있는가를 확인해 보아야 한다.
- 3) 통계전문가의 협조는 실험이나 조사의 기획 때 부터 받아야 한다.
기획에 참여하지 못한 통계전문가는 엄격히 말하면 자료분석도 도울 수 없다고 해도 과언이 아닌 것이다.
- 4) 연구목적과 알고자 하는 내용이 무엇인가를 확실히 파악하고 실험이나 조사에 임해야 한다. 내용 목적도 없이 연구에 참여하는 사람이 어디에 있겠는가 하고 반문할 사람도 있겠으나 실은 그런 목적이 구체화되어 있지 않는 상태에서 연구 조사가 진행되는 경우도 흔히 있다는 점을 반성해야 한다