가정의학회지 제23권 제3호 2002
J Korean Acad Fam Med Vol. 23, No. 3 March

■ 논 평

2002년 2월 게재 논문의 통계적 오류

박 용 규
가톨릭대학교 의과대학 의학통계학교실

서 론

2002년 2월 가정의학회지에 게재된 총 7편의 원저들 중, 이상래 등의 "성인 류마티스 관절염 환자의 민간요법실태-서울, 경기 지역 종합병원 외래 환자를 대상으로-"는 통계분석방법상의 특별한 문제점이 발견되지 않았다. 나머지 6편의 논문에서는 다음과 같은 통계적 오류가 발견되었다.

- 표본추출방법과 조사과정에 대한 서술
- 같은 내용의 중복표현
- 다중비교방법의 선택
- 표의 제목에 지나친 통계용어의 사용
- 논문의 주제에 합당한 결과의 제시

다음은 각 논문에 나타난 통계적인 문제점들을 제시한 것이다.

각 논문에서의 통계적 오류

1. 조홍준 등, 국민은 주치의제도에 대해 어떻게 생각하나

1) 조사대상자의 선정방법을 서술한 부분이 무작위 추출, 표본할당방법 등으로 되어있다. 연구방법에 있는 "가구주 연령별, 거주지별 모집단의 분포를 고려한"이라는 표현을 감안해 볼 때, 비례할당방법이 사용된 것 같다.[1)]

2) 이 경우 최종 조사대상자의 수가 657명으로 결정된 근거를 제시하는 것이 바람직하다. 예를 들어 표본수 결정과정에서 이용된 신뢰수준이나 표본오차(sampling error)가 제시되었으면 한다.

3) 표본의 선정은 가구주의 특성이 기준으로 되어있으나, 분석의 대상자들은 각 가구내의 응답한 사람의 특성이 기준이 되어있다. 그 결과 여성의 비율이, 특히 주로 주간에 집에 있는 전업주부들의 비율이 매우 높은 현상을 보이고 있다. 이 경우 가구주의 특성보다 기혼자들의 분포가 대상자선정의 기준이 되는 것이 더 적절하다고 본다.

4) 전화 또는 우편조사의 경우 응답률이 상당히 낮다는 점을 감안하여, 조사원이 직접 방문하여 일대일 면접조사하여 100%의 응답을 얻은 것으로 되어있다. 이렇게 높은 응답률은 처음부터 비례할당에 의해 산정된 각 특성별 가구수 보다 더 많은 가구를 선정한 후, 전체 대상자수가 657명이 될 때까지 조사하지 않으면 현실적으로 불가능한 일이 아닌가 생각된다.

5) 제한된 지역을 중심으로 조사한 결과이므로, 제목에 이러한 제약이 포함되었으면 한다.

6) 간행위원회에서 충분히 고려된 사안으로 생각되지만, 2001년 11월에 게재된 같은 저자의 "단골의사 보유와 연관되는 요인"이라는 논문과의 관련성을 한번쯤 짚어봐야 할 것으로 보인다.

2. 이인구 등, 초등학교 자녀의 비처방약물 중 진통제 사용에 관련된 부모의 행동-서울지역 일부 초등학교

를 대상으로-

1) 전반적으로 통계분석방법 자체에는 별다른 문제가 없었다고 생각된다. 다만 표본추출방법에 서술된 편의추출, 즉 조사의 편의에 따라 접근하기 쉬운 대상만 조사한 것이라면, 모집단에 대한 대표성이 결여되고, 비확률적인 표본이 되므로 통계적 검정에 문제가 생긴다.[2)]

2) 모든 분석이 단변량분석에 의존하고 있다. 적어도 표 6의 진통제 사용에 영향을 미친 요인들의 분석만큼은 다중 로지스틱분석을 사용했더라면 한다.

3. 한지혜 등, 젊은 여성의 체성분과 척추 골밀도의 상관관계

1) 표 1을 제외한 나머지 표 2에서 5까지의 제목이 모두 지나치게 통계적인 표현으로 되어있다.

2) 표 3의 상관계수들 중 L-BMD와 나머지 변수들간의 관련성과 표 4의 회귀분석의 결과는 결국 같은 내용을 제시한 것이다. 특히 본문의 결과서술에서도 각 표에 대한 해석이 모두 유의한 변수를 언급하는 것만으로 그치고 있기 때문에, 두 분석의 차이가 전혀 드러나지 않는다.

3) 젊은 여성을 대상으로 한 골밀도의 연구에서는, 특별히 영향력이 큰 요인이 발견되지 않는다는 점을 감안한다하더라도, 다중 회귀분석 결과 결정계수(R^2)가 0.065, 즉 제지방량으로 척추 골밀도를 설명할 수 있는 정도가 6.5% 밖에 되지 않는다는 것에 어느 정도 의미를 부여할 수 있을는지 의문이다.[3)]

4. 이형근 등, 교육정도와 우울성향과의 관계

1) 분산분석에서 유의한 결과를 얻었을 때, 다중비교를 하는 방법은 매우 다양하다. 이 논문에서는 Tukey의 방법을 사용하고 있다. 다중비교의 선택은 (1) 사전, 사후비교 (2) 각 집단의 동일 표본수 여부 (3) 비교의 내용에 따라 크게 구분된다. Tukey의 방법도 서로 다른 표본수에 대해 조화평균을 이용할 수 있다는 점에서 부적절한 적용이라고 할 수는 없지만, 이 경우에는 보다 일반적으로 사용되는 Scheffe의 방법을 권한다.[4)]

2) 제목이 말하는 것처럼 저자들의 관심은 교육정도가 우울성향에 미치는 영향을 파악하는데 있고, 논문의 서술과정도 이 목적에 맞추어 일관성 있게 진행된 것으로 보인다. 다만 단변량분석에서, 교육정도 외에 우울에 영향을 미치는 것으로 나타난 성별의 효과를 통제한 후에도, 교육정도가 우울과 관련이 있는지가 이 논문의 최종결론에 해당되리라 본다. 본문에서 다중회귀분석을 하여 의미있는 결과를 얻었다고 간단히 서술하고 있으나, 분석결과를 별도의 표로 나타내었으면 한다. 또한 이 경우 다중회귀분석보다 공분산분석으로 다른 변수들의 영향을 보정한 최소제곱 평균(least square means)을 제시하는 것이 더 적절할 것이다.[5)]

5. 강태환 등, 한국여성의 폐경 기간에 따른 요추의 골밀도의 변화

1) 표 4의 제목에서 regression analysis는 잘못된 표현이며, 첫머리에 있는 Pearson's correlation만으로도 표의 내용이 충분히 설명되고 있다.

2) 표 4의 BMD와 다른 변수들간의 상관계수는, 표 1의 값과 같아야 될 것으로 보이나 다른 값이 제시되어있다. 또한 만약 같은 값이라면 표 1에서 상관계수 값을 제시하지 않아야 한다.

3) 표 6의 다단계 회귀분석의 네 번째 모형에서, abortion의 회귀계수(0.003)과 표준오차(0.002)만으로 본다면 P값이 0.05보다 크게 되어야 한다. 혹시 반올림하는 과정 때문이 아닌지 저자들의 확인을 바란다.

6. 김현수, 비만여성의 신체구성 변화에 대한 임피던스법과 피지후법의 비교

1) 본문을 서술할 때 사용된 용어와 표 1에 제시된 용어가 서로 일치하지 않아 혼란을 준다. 표 1의 ST는 SF로, BF는 FW가 되어야 할 것이다.

2) 임피던스법과 피지후법을 비교하는 것이 저자의 주목적이라면, 각 방법에서 구한 평균, 상관계수값만 제시할 것이 아니라, 그 값들간의 유의한 차이까지 분석해 보는 것이 좋았으리라 생각된다. 참고로 이러한 관련된 두 상관계수간의 비교는 Hotelling의 T^2 검정으로 할 수 있다.[6)]

결 론

대규모 설문조사의 경우, 통계청의 census 자료를 이용하여 층화랜덤추출을 하여 얻은 표본을 자주 사용한다. 이 때 각 층별 표본수는 흔히 census자료의 특성별 분포와 동일한 비율로 결정하며 이를 비례할당이라고 한다. 이와 같이 표본을 선정하는 이유는, 연구자가 궁극적으로 알고자 하는 전체 집단과 실제로 조사한 대상자들의 성격을 가능한 비슷하게 하여, 모집단에 대한 치우쳐지지 않은 결론을 얻고자 하는 의도에서다. 그러나 이렇게 할당된 표본수에 따라 구체적인 대상자들을 선정했다해도, 실제 조사를 하는 과정에서는 여러 가지 이유, 특히 응답거부로 인해 처음 의도했던 표본의 대표성이 유지되지 못하는 것이 현실이다. 따라서 연구자들이 무응답률을 낮추기 위해 많은 고민과 노력을 하지만, 그 결과는 매번 만족스럽지 못한 것이 사실이다. 지난 호에 실린 조홍준 등의 논문에 대해 구체적인 논평을 한 이유는 논문의 주제나 내용의 문제라기보다, 이러한 표본조사의 방법을 보다 현실적으로 생각해 봄으로써, 다른 연구자들이 비슷한 문제에 부딪쳤을 때 해결책이 무엇인지 제시해주는 수준까지 끌어올리려는 생각에서 비롯된 것이다. 이를 위해서는 표본조사를 하는 각 연구자들이 실제 표본 선정과 설문조사과정에서의 현실적인 문제점들을 논문에 보다 상세히 소개하여, 연구자들간의 활발한 토의가 필요하다고 본다.

상관분석과 회귀분석의 차이는 연구내용에 따라 명확히 구분되어야 할 필요가 있다. 상관분석은 두 변수의 역할이 서로 대등할 때 사용된다. 예를 들어, 키와 몸무게가 관심이 있는 두 변수라면, 키를 이용하여 몸무게를 설명할 수도 있고, 반대로 몸무게를 이용하여 키를 설명할 수가 있다. 이처럼 연구자의 의도에 따라 목적변수의 역할이 서로 바뀔 수 있을 때(역할의 구분이 없을 때), 두 변수는 서로 대등하다고 한다. 이와는 달리, 혈압과 체질량지수라는 두 변수가 있다면, 혈압을 목적변수로 삼고, 체질량지수는 혈압을 설명하기 위한 변수로 두는 경우가 일반적이며, 혈압을 체질량지수를 설명하기 위한 변수로 생각하는 연구자는 극히 드물 것이다. 이와 같이 두 변수의 관계에 분명한 방향성이 있을 때에는 회귀분석을 해야 한다. 그러나 이처럼 두 변수의 역할에 따라 두 분석방법이 구분되지만, 두 변수간의 관련성에 대한 P값은 동일하게 나타난다. 따라서 단순히 유의한 변수를 선택하고자 하는 의도에서 두 가지 방법을 모두 적용한다면 중복표현에 해당된다고 할 수 있다.

참 고 문 헌

1. Cochran WG. Sampling techniques. 2nd ed. New York(NY) : John Wiley and Sons, Inc.;1963.
2. 박용규. 2001년 3월 게재논문의 통계적 오류. 가정의학회지 2001;22(4):584-7.
3. Draper NR, Smith H. Applied regression analysis. 3rd ed. New York (NY) : Wiley;1998.
4. Snedecor GW, Cochran WG. Statistical methods. 7th ed. Ames(IA) : Iowa state univ. press; 1980.
5. Neter J, Wasserman WW. Applied linear statistical models. Homewood(ILL) : Richard D. Irwin, Inc.;1974.
6. 박용규. 2000년 10월 게재논문의 통계적 오류. 가정의학회지 2000;21(11):1466-9.