

2002년 4월 게재논문의 통계적 오류

가톨릭대학교 의과대학 의학통계학교실
박 용 규

서 론

2002년 4월 가정의학회지에 게재된 총 8편의 원자들 중, 안미홍 등의 '말기 암환자에서 진정제 사용 실태 조사'는, 한 병원에 입원한 156명의 말기암 환자들을 대상으로 연구한 내용이지만 단순히 집계만을 사용한 논문이었고, 나머지 7편의 논문에서는 다음과 같은 통계적 오류가 발견되었다.

- 편의적 표본추출
- 부정확한 통계용어 사용
- 다중비교의 표현방법
- 표에서 대상자 수에 대한 표현 누락

다음은 각 논문에 나타난 통계적인 문제점들을 제시한 것이다.

각 논문에서의 통계적 오류

1. 이병환 등, 두통과 우울성향 및 가족기능과의 관계-서울 모 여자 고등학생을 대상으로

1) 방법에 제시된 것처럼 설문지를 편의적으로 배포하였다면, 확률표본으로 볼 수 없기 때문에 통계적 검정의 결과, 특히 P값의 해석에 문제가 생긴다. 또한 한 학교의 한 학년만을 대상으로 조사한 결과를 전체 여고생에게 확장할 수 없다고, 고찰부분에서 연구의 제한점으로 저자들이 충분히 기술하고 있기는 하지만, 대표성의 문제는 편의표본에서도 발생하게 된다.¹⁾

2) 통계적 표현 중에서 다중검증은 다중비교로, 카이제곱은 카이제곱으로 고쳐야 한다.²⁾

3) 본문의 결과 3에 서술된 내용으로 볼 때, 세 집

단 간 가족기능 평가점수의 차이를 비교하기 위해 분산분석과 카이제곱 검정을 모두 한 것으로 짐작된다. 물론 그중 카이제곱 검정에 해당되는 내용만 Table 6으로 제시하였지만, 동일한 내용을 변수의 척도를 달리하여 반복적으로 분석하지 않아야 한다는 관점에서 볼 때, 분산분석에 관한 서술은 불필요한 것이다.

4) Table 5에는 빈도수만 제시되어 있다. 백분율이 함께 제시되어야 한다.

2. 오한진 등, 폐경 후 골다공증 여성의 알렌드 로네이트와 여성 호르몬 2년 병합치료 결과

1) 치료의 효과에 대해서만 기술하지 않고, 치료의 부작용적인 측면인 탈락자들의 흔한 이유에 대해서도 종류별로 언급(Table 1)한 것은 바람직하다고 본다. 그러나 이러한 이유들이 치료와 직접적인 관련이 있는 것인지에 대해서도 밝혔더라면, 부작용 측면이 보다 명확하게 부각되었을 것이다.

2) Table 3과 4에서는 세 시점 간을 비교한 것이므로 paired t-test가 아닌, 반복측정 분산분석법과 다중비교를 해야 한다.³⁾

3) Fig. 1은 Table 4의 내용이 반복적으로 표현된 것이며, Fig. 2는 Table 4의 값으로부터 계산된 증가율을 나타낸 것이다. Table 4의 값으로 증가량을 계산하면 본문에 기술된 값들과 약간의 차이를 보인다.

4) Table 1, 2와 같이 Table 3, 4에서도 각 집단의 대상수를 제시해야 한다.

3. 신호철 등, 한국형 노인 건강 상태 평가 도구의 개발

1) 개발된 도구의 신뢰도와 타당도는 추후의 연구를 통해 보고할 것이라고 했으므로, 엄밀히 말해 논

문에 제시된 도구는 아직 완성되지 않은 것이라 할 수 있다. 그리고 내용 중 마지막 단계인 예비조사의 결과를 통해 도구의 신뢰도에 대해 어느 정도 언급한 점을 감안할 때, 구인타당도 중 요인분석을 통해 구인타당도까지 검토해 보았으면 한다.

4. 조금주 등, 일개 대학병원 환자를 대상으로 시행한 혈장 Dehydroepiandrosterone sulfate (DHEA-S)와 심혈관계질환 위험인자와의 연관성

1) 방법에서는 나이집단 간의 비교 시 다중비교를 했다고 하였고, 본문에도 다중비교의 결과가 서술되어 있으나, Table 2와 3에는 이에 대한 표현이 없다.

2) Table 4는 남녀별로 DHEA-S와 cardiovascular risk factor들 간의 단순상관계수를 제시한 것이다. 이에 대한 본문의 결과부분을 보면, 다중회귀분석 결과 나이 외에는 DHEA-S와 관련된 요인이 없었다고 되어 있으며, 고찰부분에서도 이러한 내용이 정확히 기술되어 있지만, 만약 Table 4의 내용만 보는 독자가 있다면, 혈압이나 지단백 등이 관련된 요인이라고 오해할 수도 있다.

5. 서상연 등, B형 간염 예방 접종을 받지 않은 건강증진센터 수진자에서 3년 후 B형 간염 표지자의 변동

1) 통계분석방법에서 반복측정값의 변화를 검정하기 위해 Cochran-Mantel-Haenszel 통계량을 사용했다고 하였고, 실제 이 방법은 Table 2에 적용되었다. 일반적으로 Cochran-Mantel-Haenszel 방법은 3차원 분할표(3-dimensional contingency table), 예를 들어 성별 양성반응률의 차이 정도가 여러 나이집단에서 동일한지 아니면 다른지를 검정하는, 세 변수 간의 관련성을 검정하는 데 사용된다. 표에 제시된 검정결과인 P값이, 반복측정의 경우에 사용되는 Cochran의 Q 검정의 P값과 같은 점으로 미루어, 잘못 표기한 것으로 짐작된다.⁴⁾

참고로 Table 2의 내용 중 Anti-HBs의 total만을 보면, Table 2와 Table 6의 값으로부터 다음 표를 얻을 수 있다.

Baseline	Follow-up		Total
	+	-	
+	61	2	63
-	7	33	40
Total	68	35	103

여기서 Baseline의 양성률 61.2% (63/103)와 Follow-up의 양성률 66.0% (68/103)의 차이는, 대응표본(matched sample)에서 +에 해당하는 각 행과 열의 주변합(marginal)의 비율 간의 차이를 검정하는, 즉 marginal homogeneity를 검정하는 Cochran의 Q 검정이 된다. 이 때 만약 반응값의 범주수가 위와 같이 둘뿐인 경우(+와 -)에는 McNemar 검정과 동일하다.

2) Table 2의 결과에 대해 P=0.09, P=0.08을 모두 유의하다고 결론내린 것으로 보아 유의수준을 10%로 하였거나, 아니면 유의수준을 5%로 하되 단측검정(one-sided test)을 한 것으로 보인다. 이에 대한 기술이 필요하다.

6. 정유석 등, 외래 미방문 환자에 대한 전화가 순응도에 미치는 영향

1) 통계분석방법에서 유의수준을 0.05로 하였다면, 본문의 결과를 서술할 때 그 기준을 따라야 한다. P=0.06에 대해 “다소 통계적 유의성이 있었다”라는 표현은 적절하지 않다.

2) 분할표에 대해 Mantel-Haenszel chi-square test는 보통 linear trend (또는 linear by linear association) test를 할 때 사용된다. 이 분석방법은 Table 2와 같이 2x2 분할표에서는 Pearson의 카이제곱 검정과 동일한 분석결과를 주지만, Table 3에서와 같이 카이제곱 검정이라고 하는 것이 무난한 표현이다.

3) Table 5를 분석한 다중 로지스틱 회귀분석이 본문 중에는 다중회귀분석으로 잘못 표기되어 있다.

7. 박지영 등, 일개 3차 병원에서 시행한 3일 청소년 금연교실 프로그램

1) Table 2와 3은 대응표본이 3개의 범주를 가지므로 generalized McNemar 또는 Bowker의 검정이 정확한 표현이다.⁵⁾

2) Table 3, 4에는 각 집단의 대상자 수에 대한 표현이 누락되었다.

3) Table로 제시되지 않았지만, 본문 결과의 마지막 부분에 언급된 분석은 다중 회귀분석이 아닌 로지스틱 회귀분석이며, odds ratio와 함께 95% 신뢰구간을 제시하는 것이 바람직하다.⁶⁾

결 론

논문을 구성하는 한 부분인 방법(materials and methods)에는 통계적 분석방법에 대해서도 기술하게 된다. 그 내용과 양은 저널에 따라 상당한 차이를 보이고 있지만, 다음은 대부분의 저널에서 요구하고 있는 공통적인 내용이라 할 수 있다.⁷⁾

1. 가설과 통계검정방법

연구를 통해 밝히고자 하는 가설들을 적고, 그것을 분석하기 위해 어떤 통계적 검정법을 사용했는지 기술한다.

(예) “DHEA-S값이 세 집단 간에 차이가 있는지를 비교하기 위해 분산분석과 Tukey의 다중비교를 하였고, ...”.

2. 유의수준

차이가 있다는 연구자의 주장이 잘못될 확률인 유의수준 α 를 제시하여, 연구자가 내린 결론에 대한 신뢰의 정도를 제시하고, 이후 모든 분석결과에 대한

의사결정의 기준을 정한다. 그리고 특별한 언급이 없는 경우에는 양측검정을 한 것으로 간주되지만, 만약 단측검정을 했다면 반드시 기술해야 한다.

3. 통계패키지

널리 사용되는 프로그램인 경우에는 그 정확성을 인정받고 있지만, 개별적인 패키지 가운데에는 가끔 잘못된 결과를 제시하는 경우가 있으므로, 분석에 사용된 프로그램을 밝히는 것이 최근의 추세이다.

참 고 문 헌

1. 박용규. 2001년 10월 게재논문의 통계적 오류. 가정 의학회지 2001;22(11):1670-2.
2. 한국통계학회. 통계용어사전. 서울: 자유아카데미, 1987.
3. 박용규, 송혜향. 반복측정과 교차계획자료의 분석법. 서울: 자유아카데미; 1998.
4. Fleiss JL. Statistical methods for rates and proportions. 2nd ed. New York (NY): John Wiley and Sons, Inc.; 1981.
5. SAS Institute Inc. SAS guide for personal computers, Version 6.03. Cary (NC): SAS Institute Inc.; 1988.
6. Hosmer DW, Lemeshow S. Applied logistic regression. New York (NY): John Wiley and Sons, Inc.; 1989.
7. Lang TA, Secic M. How to report statistics in medicine - Annotated guidelines for authors, editors, and reviewers. Philadelphia (PA): American College of Physicians; 1997.