

원저

대한침구학회지 논문의 통계적 오류에 관한 연구

이승덕

동국대학교 한의과대학 침구학교실

Abstract

An Assessment of Statistical Validity of Articles Published in the Journal of Korean Acupuncture & Moxibustion Society - from 1984 to 2002 -

Lee Seung-deok

Department of Acupuncture & Moxibustion,
College of Oriental Medicine Dong-Guk University

This study was carried out to investigate statistical validity of medical articles that used various statistical techniques such as t-test, analysis of variance, correlation analysis, regression analysis and chi-square test.

For study 429 original articles using those statistical methods were selected from Journal of Korean Acupuncture & Moxibustion Society published from 1984 to 2002. 429 original articles were reviewed to analyzed the statistical procedures.

Results are summarized as follows :

1. In this study 93 articles(21.68%) of 429 ones didn't report statement of statistical method in detail.
2. 53 articles(12.53%) didn't report p-value in correctly, and 245 articles(57.11%) used mean±standard error (Mean±SEM.) and 109 articles used mean±standard deviation(Mean±SD.). All of 23 articles using nonparametric statistical techniques made an error to central tendency or dispersion.
3. 175 articles(59.93%) and 14 articles(4.79%) of 292 ones made an error to description of equal variances and normal distribution.
4. 99 articles(50%) of 185 ones misused t-test and 4 articles of 5 ones misused chi-square test.

* 이 연구는 동국대학교 논문게재 연구비 지원으로 이루어졌음.

· 접수 : 2004년 1월 8일 · 수정 : 2004년 1월 16일 · 채택 : 2004년 1월 17일
· 교신저자 : 이승덕, 서울특별시 강남구 논현동 37-21 동국대학교 강남한방병원 침구과
Tel. 02-3416-9738 E-mail : chuckman@dongguk.edu

5. 28 articles(73.68%) of 38 ones using discrete variable misused parametric technique such as t-test or ANOVA.
2 articles and 1 article of 125 ones choosing paired samples misused independent t-test and Mann-Whitney U test.
6. 20 articles using analysis of variance didn't use multiple comparison.

Key Words : Statistics, Validity Assessment, Acupuncture, Articles.

I. 서론

의학 논문의 결론은 객관적이고도 과학적으로 설명되어지고, 뒷받침할만한 충분한 근거의 제시되어야만 비로소 하나의 사실로 받아들여지게 되는데¹⁾, 과학 연구분야에서 연구자들이 결론을 뒷받침할 근거를 제시하기 위하여 가장 많이 사용하고 있는 방법이 통계적 방법이다²⁾.

그러므로 연구 주제가 아무리 참신하고 연구 결과가 대단한 발견으로 이어진다고 해도, 연구 주제를 구체화하거나 연구의 결론을 유도하기 위해서 사용된 연구방법론 및 통계분석 기법의 타당성이 의심된다면 연구 전체가 무효화 될 수도 있다³⁾.

현재 과학, 특히 의학분야 연구에 있어서의 통계 기법 적용 및 자료처리 방법에 관하여 국내외로 많은 저서들이 나와 있으며⁴⁾, 실제 그 적용이 얼마나 정확히, 적절하게 이루어지고 있는가에 대한 연구가 여러 분야에서 이루어지고 있다. 그런데 서양의학의 분야에서는 국내 의학잡지 게재 논문에 대한 안 등²⁾의 연구, 보건학 석사학위에 대한 하⁵⁾의 연구, 보건의학잡지에 대한 한⁶⁾의 연구, 대한의학협회지에 게재된 논문에 대한 이 등⁷⁾의 연구, 가정의학협회지에 게재된 논문에 대한 박⁸⁾의 연구 등이 있으나, 한의학 분야에서는 한의학협회지에 게재된 논문이 여러가지 통계적

오류를 범하였다는 이 등⁹⁾의 보고가 있을 뿐이며, 가장 많은 연구논문이 발표되고 있는 대한침구학회지의 게재 논문들이 올바르게 통계를 적용하고 있는지에 대한 연구는 없었다.

이에 저자는 침구학의 과학적 연구에 필수적인 통계학의 적절한 이용과 오류의 방지를 위한 지침을 마련하기 위하여 연구를 시작하였으며, 침구학회지 창간호부터 19권 6호까지 40권에 게재된 논문 921편 중 분석적 통계기법을 사용한 429편에 대하여 통계학적 견지에서 분석 검토하여 대한침구학회지에 게재된 논문들의 통계적 오류에 관하여 약간의 지견을 얻었기에 보고하는 바이다.

II. 연구

1. 연구대상 및 방법

1) 대상

연구대상은 1984년도 1권부터 2002년도 19권 6호까지 발간된 대한침구학회지(The Journal of Korean Acupuncture & Moxibustion Society : 이하 JKAMS)의 40권에 게재된 논문 921편 중 분석적 통계기법을 사용한 429편을 대상으로 하였다.

2) 방법

대한침구학회지에 게재된 논문들의 일반적인 통계적 오류와 척도 및 집단에 따른 통계방식 선택의 오류 그리고 통계방식 운용에 따른 오류 등에 대하여 조사하였다.

III. 결 과

1. 통계처리에 대한 서술의 오류

사용한 통계처리에 대하여 서술하는 방법에 있어서, 10권까지는 통계처리 방법을 밝히지 않은 경우(69.86%)가 통계처리 방법을 밝힌 경우보다 많았으나, 11권 이후부터는 통계처리 방법을 밝혀 놓은 비율이 점차 높아져 전체 논문의 58%가 넘는 논문에서 통계처리 방법에 대하여 밝혀 놓았으며, 19권에서는 통계기법을 사용한 71편의 모든 논문에서 통계처리 방법에 대하여 자세하게 밝혀 놓았다<Table 1>.

2. 유의수준(α)에 대한 오류

유의수준은 352편(82.05%)에서 0.05를 사용하였는데, 통계처리 방법을 서술하는 부분에서 유의수준을 정확하게 밝힌 경우가 371편(86.48%)이었고, 유의수준을 나타내지 않고 p-value 자체를 기록한 것이 26편(6.06%)이었으며, 유의수준이나 p-value를 나타내지 않은 논문이 27편(6.29%)이었다. 그리고 5편의 논문에서는 통계처리 방법을 밝히는 부분에서 유의수준을 나타내지 않고, 연구 결과를 나타내는 부분에서 통계 결과와 함께 표시하였다<Table 2>.

3. 중심경향(대표치) 또는 산포도에 대한 기술의 오류

중심경향과 산포도에 대한 기술에 있어서 평균±표준오차(Mean±SEM. or Mean±SE.)로 나타낸

경우가 245편(57.11%)으로 평균±표준편차(Mean±SD.)로 나타낸 경우 109편(25.41%)보다 많았으나, 18권 이후부터는 평균±표준편차로 중심경향과 산포도를 나타낸 경우가 평균±표준오차보다 많아지고 있다<Table 3>.

그리고 비모수 검정을 시행한 논문 중 중심경향과 산포도를 중앙값과 범위로 나타낸 경우는 없었고, 평균±표준편차로 나타낸 경우가 15편(65.22%)이었으며, 평균±표준오차로 나타낸 경우가 2편(8.7%)이었다. 그리고 중심경향 또는 산포도를 나타내지 않은 경우가 6편(26.08%)이었다.

Table 1. Number of Description about Statistics Methods in JKAMS

Vol	Description of Statistical methods		
	None	Described	Total
1	6(100.00)	0	6
2	2(50.00)	2	4
3	2(50.00)	2	4
4	4(80.00)	1	5
5	3(100.00)	0	3
6	6(85.71)	1	7
7	4(50.00)	4	8
8	8(66.66)	4	12
9	8(66.66)	4	12
10	8(66.66)	4	12
11	10(41.66)	14	24
12	8(38.09)	13	21
13	4(21.05)	15	19
14	2(7.40)	25	27
15	6(15.00)	34	40
16	7(12.96)	54	61
17	3(9.67)	28	31
18	2(3.33)	60	62
19	0(0.00)	71	71
Total	93	336	429

Table 2. Number of p-value in JKAMS

Vol	p-value						Value	Total
	0.1	0.05	0.01	0.001	0.0001	None		
1	0	5	1	0	0	0	0	6
2	0	3	1	0	0	0	0	4
3	0	4	0	0	0	0	0	4
4	0	4	0	0	0	0	1	5
5	0	2	0	0	0	0	1	3
6	0	4	0	0	0	0	3	7
7	0	5	0	0	0	0	3	8
8	0	10	1	0	0	0	1	12
9	0	9	1	0	0	1	1	12
10	0	9	0	0	0	2	1	12
11	0	17	1	0	0	2	4	24
12	0	14	0	0	0	3	4	21
13	0	16	0	0	0	0	3	19
14	0	25	1	0	0	0	1	27
15	0	36	1	0	0	1	2	40
16	1	57	0	0	0	3	0	61
17	0	21	3	1	2	4	0	31
18	0	53	1	0	0	7	1	62
19	0	58	8	1	0	4	0	71
Total	1	352	19	2	2	27	26	429

Table 3. Central Tendency and Dispersion of Statistical Articles in JKAMS

Vol	Central tendency and Dispersion					Total
	None	Mean	Mean±SD.	Mean±SEM.	etc	
1	0	0	0	6	0	6
2	0	0	0	4	0	4
3	0	0	0	4	0	4
4	1	0	0	4	0	5
5	0	0	0	3	0	3
6	0	0	1	6	0	7
7	0	0	0	8	0	8
8	1	0	1	10	0	12
9	0	0	2	10	0	12
10	1	0	0	11	0	12
11	0	3	6	14	1	24
12	2	2	2	15	0	21
13	1	1	5	12	0	19
14	7	0	5	15	0	27
15	5	0	6	28	1	40
16	8	0	18	35	0	61
17	6	1	9	15	0	31
18	10	2	25	24	1	62
19	19	2	29	21	0	71
Total	61	11	109	245	3	429

SD : Standard deviation : 표준편차
SEM(SE) : Standard error mean : 표준오차

4. 정규 분포성에 대한 기술의 오류

표본집단의 정규 분포성을 검정하여야 할 논문 292편 중 정규 분포성에 대해서는 18권과 19권에서 각각 1편씩 언급을 하고 있는데, 그 방법으로는 통계 패키지를 사용하여 Kolmogorov-Smirnov test를 시행하여 정규 분포성을 검정하였다. 정규 분포성을 검정하지 않은 논문 중 표본의 크기가 25에 이상인 경우 정규 분포성을 가정할 수 있는 경우가 115편이었다<Table 4>.

5. 등분산성에 대한 기술에 대한 오류

등분산성의 검정이 필요한 모수 검정을 시행한 292

편의 논문 중 등분산성을 밝힌 논문은 없었으나, 통계 패키지를 사용하여 등분산성을 검정해야 하는 문제를 해결한 경우가 116편(37.73%)이었고, 표본수의 차이가 2배 이하여서 등분산성을 인정할 수 있는 경우가 162편(55.48%)이었다. 그러나 14편(4.7%)에서는 표본수가 2배 이상의 차이가 있는데도 등분산성을 검정하지 않았다<Table 5>.

6. 표본집단의 수에 따른 통계기법의 오류

연구대상 429편의 논문 중 표본집단이 셋 이하인 경우가 185편이었고, 넷 이상인 경우가 242편이었으며, 표본집단의 수를 정확하게 밝히지 않은 경우가

Table 4. Normal Distribution of Statistical Articles in JKAMS

Vol	Normal distribution			Total
	None	Normal distribution	Above 25case	
1	0	0	0	0
2	2	0	0	2
3	2	0	1	2
4	1	0	0	1
5	0	0	0	0
6	1	0	3	1
7	4	0	4	4
8	4	0	0	4
9	3	0	0	3
10	4	0	0	4
11	9	0	1	9
12	13	0	15	13
13	13	0	26	13
14	22	0	38	22
15	31	0	9	31
16	49	0	7	49
17	29	0	7	29
18	53	1	1	54
19	50	1	3	51
Total	290	2	115	292

Table 5. Equal Variances of Statistical Articles in JKAMS

Vol	Equal variances			Total
	None	Package	Equal case	
1	0	0	0	0
2	1	0	1	2
3	2	0	0	2
4	1	0	0	1
5	0	0	0	0
6	0	0	1	1
7	2	0	2	4
8	1	1	2	4
9	0	1	2	3
10	0	0	4	4
11	0	5	4	9
12	1	3	9	13
13	0	3	10	13
14	0	4	18	22
15	0	13	18	31
16	3	14	32	49
17	1	13	15	29
18	1	30	23	54
19	1	29	21	51
Total	14	116	162	292

2편이었다. 이중 198편의 논문에서 t-test를 시행하였는데, t-test를 시행한 논문 중 표본집단이 셋 이하인 경우가 99편(50%)이었고, 넷 이상인 경우가 99편(50%)이었다<Table 6>.

7. χ^2 -검정과 Fisher exact test의 선택 오류
 연구대상 429편의 논문 중 χ^2 -검정을 시행한 경우가 14편이었는데, 이중 빈도수가 5보다 작은 칸의 비율이 전체 칸수의 25% 이상이어서 작은 표본수에 해당되는 경우가 5편, 25% 이하인 경우가 4편이었으며, 분할표를 나타내지 않아 빈도수를 알 수 없는 경우가 5편이었다. 작은 표본수를 가진 논문 5편 중

Table 6. Number of Groups of Statistical Articles in JKAMS

Vol	Samples			t-test
	None	Under 3 group	Above 4 group	
1	0	1	5	0
2	0	2	2	2
3	0	2	2	2
4	0	1	4	1
5	0	0	3	0
6	0	1	6	1
7	0	4	4	4
8	0	4	8	3
9	0	3	9	0
10	0	4	8	3
11	0	7	17	2
12	1	6	14	9
13	0	4	15	10
14	0	6	21	13
15	0	13	27	22
16	1	22	38	32
17	0	18	13	16
18	0	38	24	41
19	0	49	22	37
Total	2	185	242	198

Table 7. χ^2 -test of Statistical Articles in JKAMS

Vol	Number of 5 cases			Total
	Unknown	Under 25%	Above 25%	
1~11	0	0	0	0
12	0	0	1	1
13	0	1	0	1
14	0	0	0	0
15	0	0	0	0
16	0	0	0	0
17	1	1	0	2
18	1	2	3	6
19	3	0	1	4
Total	5	4	5	14

18편의 1편의 논문에서만 Fisher exact test를 시행하였고, 나머지 4편에서는 χ^2 -검정을 시행하였다<Table 7>.

8. 반응변수의 척도에 따른 통계기법의 오류
 연구대상 429편의 논문 중 반응변수가 연속형 척도인 경우가 391편(91.14%)이었고 이산형 척도인 경우가 38편(8.86%)이었는데, 이산형 척도인 경우 중 모수 검정방법인 t-test와 분산분석을 시행한 경우가 각각 14편(36.84%)과 4편(10.53%)이었다.

9. 쌍체표본과 독립표본에 따른 통계기법의 오류

연구대상 429편의 논문 중 표본의 형태가 독립표본인 경우가 304편(70.86%)이었고, 쌍체표본인 경우가 125편(29.13%)이었다. 쌍체표본인 논문 중 t-test를 시행한 경우가 62편이었는데, 이중 paired t-test를 시행하였다고 밝힌 경우가 3편이었으며, independent t-test를 시행하였다고 밝힌 경우가 2편이었다. 그리고 어떤 기법의 t-test를 시행하였

는지 정확하게 밝히지 않아 paired t-test를 시행하였는지 independent t-test를 시행하였는지 알 수 없는 경우가 57편이었다. 그리고 비모수 검정방법인 Wilcoxon signed rank test를 시행한 경우가 4편, Mann-Whitney U test를 시행한 경우가 1편이었다.

10. 다중비교의 시행오류

분산분석을 시행한 논문은 총 119편이었는데, 이중 일원 분산분석을 시행한 경우가 111편이었고, 분산분석을 시행하였다고 서술을 하지 않았으나 결과의 형태가 분산분석인 경우가 5편이었으며, 비모수적 검정방법을 시행한 경우가 4편이었다.

분산분석을 시행한 논문 119편 중 다중비교를 시행한 경우가 99편(83.19%)이었고, 다중비교를 시행하였는지 자체를 언급하지 않은 경우가 17편(14.29%)이었으며, 다중비교를 시행하지 않은 경우가 3편이었다.

IV. 고 찰

통계는 관찰 가능한 현상을 한정된 관찰 결과로서 일반성을 추론할 수 있는 방법이므로¹⁰⁾, 확률표본을 대상으로한 과학적 연구논문에서 통계기법의 적용은 제한된 연구대상으로부터 추론된 결과를 객관화, 일반화하는데 그 목적이 있다¹¹⁾.

이러한 통계적 방법은 과학논문에서 연구 결과를 발표할 때 그것을 설명하고 뒷받침할만한 충분한 근거의 제시를 위하여 가장 많이 사용되고 있는 방법이라고 할 수 있는데²⁾, 적절한 통계기법의 적용은 의학 연구논문의 타당성을 보증하는 최소한의 조건이며 그 구성체계의 주요한 밑받침이 된다¹²⁾.

그러나 많은 과학적 연구논문에서는 그 나름의 의미있는 연구 목적이나 도출된 결과의 중요성에도 불

구하고 연구방법상의 문제, 즉 부적절한 표본의 선택, 통계기법 적용상의 오류와 잘못된 결과해석 등과 같은 통계적 오류로 인하여 논문 자체의 객관적 가치를 오히려 떨어뜨리는 경우가 있다¹¹⁾.

과학적 연구에서 많은 연구자들은 통계기법을 적용하는데 있어서 오류들을 범하고 있으며, 이에 따라 과학연구 논문에 적용된 통계기법의 오류들에 대한 연구가 여러 분야에서 진행되고 있는데, 국내의 의학 연구 논문에 적용된 통계적 오류에 대한 연구로는 안 등²⁾이 의학 종합잡지에 게재된 논문 중 54.3%, 하⁵⁾가 서울대학교 보건대학원 석사학위 논문 중 76.0%, 현⁶⁾이 보건의학 잡지에 게재된 논문 중 69.3%, 이 등¹³⁾은 대한의학협회지에 게재된 논문 중 97.6%에서 하나 이상 통계적 오류가 있었다고 보고하였다.

그리고 서양의학 분야에서는 분석기법의 타당성을 평가하고 통계적 오류를 파악하기 위하여 이¹⁴⁾, 최 등¹⁵⁾ 및 안 등³⁾이 점검표(checklist)를 개발하는 등 활발한 연구가 진행되고 있으나, 한의학의 연구논문에서의 통계적인 오류에 대한 연구로는 이 등⁹⁾의 보고가 있을 뿐 대한침구학회지에 게재된 논문에 대한 통계적 오류를 검토한 연구는 없었다.

이에 저자는 대한침구학회지 창간호부터 2002년 19권 6호까지의 논문 중 분석적 통계기법을 사용한 429편의 논문에 대하여 그 자료처리 과정에서의 통계적 방법의 적용에 대하여 주로 통계학적 견지에서 검토한 바 약간의 결과를 얻었으며 이 결과가 앞으로의 과학논문, 특히 통계적 방법 응용이 필요한 논문에서 보다 정확하고 적절한 통계적 방법을 적용하는데 다소의 보탬이 되고자 보고하는 바이다.

논문의 연구 방법 및 대상 부분에서는 통계처리 방법에 대하여 밝혀놓아야 하는데, 연구 대상으로부터 관측된 변수의 특성, 분포에 대한 언급과 변수변환의 타당한 이유를 설명하여야 하며, 분석에 사용된 검정방법을 빠짐없이 적고, 분석에 이용한 통계프로그램과 유의수준을 명시하여야 한다¹⁶⁾.

대한침구학회지에 게재된 분석적 통계기법을 사용한 429편의 논문 중 93편(21.68%)에서 통계처리 방법에 대하여 밝히지 않았는데, 이러한 결과는 가정 의학회지에서 25.3%에서 통계처리 방법을 밝히지 않았다는 서 등¹⁷⁾의 연구 결과와 유사하였다. 그러나 이렇게 통계처리 방법을 밝히지 않는 오류를 범한 논문의 비율이 10권까지는 올바른 경우보다 많았으나, 11권 이후부터는 올바른 논문의 비율이 더 많아지다가 19권에서는 모든 논문에서 통계기법을 올바르게 밝혀놓고 있어서 향후에는 이런 오류를 범하는 경우가 적을 것으로 생각된다.

검정 통계량은 그 특성에 따라 특이한 분포를 하게 되는데, 그 특이한 분포의 발생 정도에 있어서 가장 일어나기 힘든 경우부터 산출된 검정 통계량에 이르기까지의 발생확률의 누적이 p-value가 된다. 즉 p-value는 probability value로서 그러한 검정 통계량이 일어날 확률을 의미한다¹⁸⁾.

그런데 유의수준(α)이 사전에 설정되지 않으면 임상적 의미를 발견하기 위한 적당한 검정력을 제공할 표본의 크기를 계산할 수 없다. 따라서 연구의 설계단계에서 유의수준을 정하지 않고 결과를 제시했을 때는 통계분석 후에 유의수준을 정하는 결과가 되며 결과적으로 몇가지 오류가 발생하게 된다¹⁹⁾.

이 연구에서는 유의수준을 통계처리 방법을 서술한 부분에서 올바르게 밝힌 경우가 371편(86.48%)이었고, 통계처리 방법을 밝히는 부분에서 나타내지 않고 연구결과에서 유의수준을 나타낸 경우가 5편이었다. 그러나 유의수준을 나타내지 않거나 p-value로 나타낸 경우가 53편(12.35%)이었다. 이와 같이 약 12.35%에서는 유의수준의 설정에 오류를 범했으며, 또한 5편에서는 유의수준을 나타내는데 방법에 잘못이 있는 것으로 나타났다. 이러한 오류들은 유의수준에 대한 올바른 이해의 부족에서 발생된 것으로 생각되는데, 이러한 오류를 범하지 않고 보다 높은 타당성을 얻기 위해서는 연구설계 부분에서 타당성

을 입증할 유의수준을 설정하고, 이 유의수준에 따라 적당한 표본의 크기를 정하여 연구를 진행하여야 할 것으로 생각된다.

중심경향과 산포도는 측정 수준과 분포의 특성을 고려하여 적절하게 선택되어야 하고, 아울러 중심경향과 산포도가 함께 제시되어야 하며, 사용된 종류에 대한 설명을 하여야 하는데¹⁴⁾, 모수적 방법을 사용한 경우에는 일반적으로 평균±표준편차를 제시하며, 비모수적 방법을 사용한 경우에는 중앙값과 범위(최소값-최대값)를 가장 자주 사용한다⁸⁾.

그런데 중심경향과 산포도를 나타낼 때 평균±표준편차대신 평균±표준오차를 사용하는 경우가 흔히 있는데, 표준편차는 원자료의 퍼져 있는 정도를 나타내는 값이고, 표준오차는 계산된 평균의 정밀성(precision)을 나타내는 추론 통계량이므로 표준오차를 표준편차 대신 사용하여서는 안된다^{3),8)}.

이 연구결과에서도 중심경향과 산포도를 평균±표준오차로 나타낸 경우가 245편(57.11%)으로 평균±표준편차로 나타낸 경우 109편(25.41%)보다 많았다. 이와 같이 표준편차 대신 표준오차를 사용하는 이유에 대하여 Brown G.W.²⁰⁾는 표준오차가 표준편차나 기타 산포를 나타내는 측도들에 비하여 작은 값을 취함으로써 자료의 정밀도 또는 신뢰 정보를 좋아 보이게 하기 때문이라고 하였다. 그런데 대한침구학회지에서 18권 이후부터는 평균±표준편차로 나타낸 경우가 잘못된 방법인 평균±표준오차보다 많아지고 있어서, 향후에 표준편차대신 표준오차를 사용하는 오류가 감소될 것으로 생각된다.

그리고 순위 척도에 대한 분석법인 비모수 검정(nonparametric test : 예를 들어, Wilcoxon signed Rank test, Wilcoxon rank sum test, Mann-Whitney U test, Kruskal-Wallis test 등)을 하였다면 표의 내용도 평균과 표준편차가 아닌 중앙값과 범위로 표현되어야 하는데⁸⁾, 이 연구의 대상 논문들에서는 비모수 검정을 한 논문에서도 모두 평균±표준편차도

는 평균±표준오차로 중심경향 또는 산포도를 나타내었으며, 중위값이나 범위를 사용한 경우는 없었다. 이와 같이 비모수 검정에서 중심경향과 산포도의 표기에 오류가 많은 이유는 연구자들이 아직까지 비모수 검정방법에 익숙하지 않기 때문인 것으로 생각된다.

자료의 평균을 비교하는 검정방법인 t-test, 분산분석과 회귀분석과 같은 모수적 검정을 시행하기 위해서는 표본집단의 정규 분포성과 등분산성의 요구조건이 만족되어야 하는데, 만약 두가지 조건이 충족되지 않을 경우에는 각 분석에 대응되는 비모수적 검정방법을 선택하거나 변수변환을 하여야 한다^{8),16)}.

그런데 자료의 정규 분포성 여부는 히스토그램, 줄기그림, 또는 상자 그림 등을 통한 분포모양을 파악하거나, 정규 확률지를 이용하여 정규성에서 벗어나는 정도를 관찰하거나, 정규성 검정을 통해 판단할 수 있으며, 등분산성은 F검정을 통해 진단할 수 있다⁸⁾.

이 등¹⁴⁾은 정규 분포성을 밝힌 경우와 정규 분포성을 가정할 수 있는 25예 이상인 논문의 편수²¹⁾에 대하여 113편의 t-test 시행 논문 중 각각 4편과 54편으로 전체 49%이었다고 하였고, 91편의 일원분석 시행 논문 중에는 각각 1편과 13편으로 전체 14%이었다고 하였는데, 이 연구에서는 모수적 검정을 사용한 292편의 논문 중 정규성을 밝힌 논문이 2편이고, 25예 이상인 경우가 115편이어서 39.38%를 차지하였다. 이는 이 등¹⁴⁾의 논문과 비슷한 양상을 보였으나 오류의 비율은 상대적으로 높았다. 그리고 정규 분포성을 밝힌 2편의 논문은 통계 패키지를 사용하여 Kolmogorov-Smirnov test를 시행하여 정규 분포성을 검정하였는데, 통계 패키지의 사용이 증가하는 추세를 보아 앞으로는 정규 분포성 검정의 오류가 감소될 것으로 생각된다.

SAS나 SPSS 같은 통계 패키지를 사용하면 등분산성을 검정해야 하는 문제가 저절로 해결해 되며, 두 표본의 크기가 비슷한 경우에는 등분산성을 가정

할 수 있다¹⁴⁾.

이와 같이 통계 패키지를 사용했거나, 표본의 크기가 2배 이하인 경우와 등분산성을 밝힌 논문 편수에 대하여, 이 등¹⁴⁾은 t-test에서 69편과 16편으로 75%가 적절하였고, 28편(25%)이 부적절하였다고 하였고, 일원분산 분석에서는 91편 중 10편만이 표본의 크기가 비슷(2배 이하)하였으나, 23편에서는 표본의 크기를 제시하지 않았으므로 파악할 수 없었다고 하였다.

그런데 이 연구에서는 모수적 검정을 시행한 292편의 논문 중 등분산성을 밝힌 논문은 없었고, 표본수가 2배 이하로 비슷하거나 통계 패키지를 사용하여 등분산 문제를 해결한 경우가 278편(95.2%)이었으며, 표본수가 2배 이상의 차이가 있으면서도 통계 패키지를 사용하지 않은 경우가 14편(4.7%)이었다. 이러한 결과는 등분산성을 검정해야 하는 문제에 있어서 이 등¹⁴⁾의 결과보다 적게 오류를 범한 것인데, 이것은 표본의 크기가 2배 이하로 비슷한 경우가 많았으며, 통계 패키지의 이용이 많아지는 경향 때문이라고 생각된다.

통계기법의 분류는 기준을 분석하는 집단의 개수에 의해 결정되는데¹⁸⁾, 척도가 연속척도인 구간척도 및 비척도인 경우에는 변수가 1개면서 집단이 2개이면 t-test, paired t-test 등을 시행하여야 하고, 집단이 3개 이상이면 분산분석을 시행하여야 한다¹⁸⁾.

대한침구학회지에 게재된 논문들의 연구에서는 Normal, Control, Sample 1, Sample 2와 같이 표본집단을 만들어 통계 처리할 때는 Normal을 제외하고 통계 처리한 것이 대부분이었다. 따라서 표본집단이 셋 이상인 경우는 실제로는 두 집단간의 평균비교에 해당한다고 볼 수 있으며, 이에 따라 이 연구에서는 표본집단이 넷 이상인 경우를 세 집단간의 비교에 해당한다고 가정하였다. 이 연구에서는 t-test를 시행한 198편의 논문 중 표본집단이 넷 이상인 경우가 99편(50%)이었는데, 이러한 결과는 t-test를 시행

한 50%의 논문들이 표본집단이 셋 이상인 경우에 2집단간의 평균비교 방법인 t-test를 사용하는 오류를 범했다는 것을 나타내는 것이다. 물론 여러 집단을 통계 처리하더라도 2집단간의 평균만을 비교한다면 t-test를 시행하여도 중복 가설검정의 오류에 해당되지는 않는다. 그러나 대한침구학회 게재 논문 중에 표본집단을 여럿으로 한 연구들은 약침종류나 농도에 따라 표본집단을 구분한 것 같이 다중비교나 상관성 비교를 시행하였다면 연구 결과에 대하여 보다 나은 해석을 얻을 수 있었던 경우가 많았는데, 2집단간의 평균비교인 t-test를 주로 사용하여 연구 결과에 대한 만족할만한 해석을 하지 못한 것으로 생각된다.

χ^2 -검정은 t-test나 분산분석과 같이 복수집단간 차이에 대한 검정이기는 하지만 종속변수가 명목 척도나 서열척도로 측정되었을 때 활용되는데¹⁴⁾, χ^2 -검정은 비모수 통계에 속하므로 모수통계에서 요구하는 까다로운 요건들을 충족시킬 수 없는 상황에서도 쉽게 활용할 수 있는 기법이다. 그러나 2×2보다 큰 분할표인 경우에는 기대도수가 5보다 작은 칸의 비율이 25%보다 적고 모든 칸의 기대도수가 1 이상이어야 χ^2 -검정을 할 수 있다¹¹⁾.

이 연구에서는 χ^2 -검정을 시행한 논문 14편 중 5편의 경우에는 분할표를 나타내지 않았으며, 빈도수가 5보다 작은 칸의 개수가 전체 칸수의 25% 이상인 작은 표본수의 경우가 5편, 25% 이상인 경우가 4편이었는데, 작은 표본수를 가진 5편 중 1편의 논문에서만 Fisher exact test를 시행하였고, 나머지 4편에서는 χ^2 -검정을 시행하는 오류를 범했다.

이 등¹⁴⁾은 χ^2 -검정을 사용된 논문 43편 중 6편에서 분할표를 제시하지 않았고, 표본의 크기가 적절한 것과 부적절 하게 작은 것이 각각 18편이었다고 보고하였는데, 표본수의 크기에 대한 적절과 부적절의 정도가 각각 50%를 차지한 이 연구의 결과와 유사하였다. 그러나 분할표를 나타내지 않은 경우에서

는 이 등¹⁴⁾의 결과에 비하여 이 연구의 논문들에서 상대적으로 높았는데, 이러한 이유는 아직까지 서양 의학 논문에 비하여 침구학 논문에서 χ^2 -검정의 사용비율이 낮아서 침구학 연구자들의 χ^2 -검정에 대한 이해부족에서 발생된 것으로 생각되므로, 향후 χ^2 -검정의 적용빈도가 높아지면 χ^2 -검정에 대한 이해도 높아져 오류의 발생비율이 감소될 것으로 생각된다.

모수적 검정과 비모수적 검정의 선택과정에서 반응변수가 연속형이고 정규 분포성과 등분산이 검정되었을 때는 t-test, 분산분석, 회귀분석과 같은 모수적 통계기법을 적용하며²²⁾, 정규 분포성이나 등분산성이 입증되지 않거나, 이산형 척도인 경우는 비모수적 검정을 적용하여야 한다^{4),23)}.

이 연구에서는 연구대상 429편의 논문 중 반응변수가 이산형 척도인 경우가 38편(8.86%)이었는데, 이중 모수적 검정방법인 t-test와 분산분석을 시행한 경우가 각각 14편(36.84%)과 4편(10.53%)이었다. 이러한 결과는 t-test를 시행한 논문 중 반응변수가 연속형이었던 경우가 11편(98%)이었고, 이산형이었던 경우가 2편(2%)이었다는 이 등¹⁴⁾의 보고에 비하여 높았다. 이와 같이 반응변수의 척도에 따른 비모수와 모수적의 검정방법의 선택에서 오류가 많았던 이유는 15권까지 비모수적 통계기법을 사용하지 않았기 때문으로 생각되며, 점차 비모수적 통계기법의 적용빈도가 증가하는 추세로 보아 이러한 오류도 점차 감소되리라 생각된다.

독립표본이란 각 집단을 구성하는 개체들이 서로 관련이 없는 표본을 말하고, 쌍체표본이란 주요 관심의 대상인 처리 이외의 요인들을 가능한 같게 하여 비교 실험한 표본자료를 말하는 것인데, 만약 쌍체표본임에도 불구하고 독립표본의 검정방법을 적용한다면 잘못된 추정량의 분산이 얻어지므로 검정의 결과를 믿을 수 없게 된다⁸⁾.

이 연구의 대상 논문 중에는 쌍체표본인 경우가

125편이었는데, 두 집단의 쌍체표본의 평균비교에 사용되는 모수적 검정방법인 paired t-test를 사용한 경우와 비모수적 검정방법인 Wilcoxon signed rank test를 사용한 경우가 각각 3편, 4편씩 총 7편이었으며, 독립표본에 사용하는 independent t-test와 Mann-Whitney U test를 잘못 사용한 경우가 각각 2편, 1편이었다. 이와 같은 결과는 가정의학회지 110편 중 3편(3%)에서 쌍체표본의 평균비교에 unpaired t-test를 사용한 오류가 있었다는 박²²⁾의 연구와 비슷하였다.

그러나 이 연구에서는 t-test를 시행하였다고만 하고 paired t-test인지 independent t-test인지를 밝히지 않은 경우가 57편으로 대다수를 차지하여 정확한 오류의 정도를 알기 어려웠으며, 이 또한 사용한 통계기법을 정확히 밝히지 않은 오류에 해당된다고 생각된다.

이와 같이 쌍체표본에 대하여 독립표본에 사용하는 통계기법을 사용한 이유에 대하여 박²²⁾은 대부분의 연구자들이 독립자료의 분석방법에 익숙해져 있기 때문이라고 하였는데, 저자는 이러한 오류가 표본의 형태와 표본에 따른 적용 가능한 통계기법에 대한 연구자들의 이해부족 때문에 발생된 것으로 생각한다.

분산분석은 셋 이상의 집단에 있어서 모든 집단의 평균이 같은가 아니면 적어도 하나라도 평균이 다른 집단이 있는 것인가 하는 평균을 비교하는 방법으로 유의한 차이가 있을 때, 구체적으로 어떤 집단간에 차이가 있는지를 추가로 검정할 필요성이 생긴다. 이때 사용하는 검정방법이 다중비교(multiple comparison)인데, 모든 집단들간을 비교하는 경우에서 각 집단의 표본수가 같다면 Tukey, Duncan, Newman-Keuls 등의 검정방법을 사용하고, 각 집단의 표본수가 다르다면 Scheff의 방법을 사용한다⁸⁾.

이 연구에서는 분산분석을 시행한 119편의 논문 중 다중비교를 시행한 경우가 99편(83.19%)이었고,

다중비교를 시행하였는지 자체를 언급하지 않은 경우가 17편(14.29%)이었으며, 사후비교를 시행하지 않은 경우가 3편이었다. 그런데 분산분석을 시행한 119편의 모든 논문에서 집단간에 유의성 있는 차이가 있었다고 하였으므로, 어떤 집단간의 차이가 있는지를 추가적으로 검정하여야 하는데, 20편(16.61%)의 논문에서는 다중비교를 시행하지 않았으므로 오류를 범한 것에 해당된다.

일반적으로 표본집단을 셋 이상으로 할당하여 연구결과를 나타낼 때는 분산분석을 시행하고 그 결과에 따라 적절한 다중비교를 시행하는 것이 표본집단을 여럿으로 선택하여 연구하는 목적에 부합되리라 생각되므로 앞으로 분산분석과 표본수와 검정의 신뢰도에 따른 다중비교를 시행하여야 할 것으로 생각된다.

Feinstein은 의사가 표준편차와 표준오차, t-test, χ^2 -검정을 이해하면 의학논문의 3/4를 이해할 준비가 되어 있는 셈이라고 하였으나²⁴⁾, 대부분의 연구자들은 통계에 대해 충분한 교육을 받지 못하고 있으며, 이로 인해 부적절한 표본의 선택, 잘못된 통계적 기법의 적용, 결과해석에 있어서 많은 오류를 범하고 있다. 이에 따라 박⁸⁾은 논문작성 과정에서의 통계적 오류를 최소화하기 위하여 통계교육의 필요하다고 하였고, 또 논문 심사과정에서의 레프리 제도의 활성화와 통계 상담서비스의 활용 등을 제안하였는데, 저자도 박의 이같은 의견에 동의하며, 향후에 침구학 연구자들은 연구의 결과를 좀더 객관적으로 입증하기 위하여 통계에 대한 기본적인 이해를 가지고 연구계획을 설정하고, 통계학자들과의 공동 연구를 통하여 통계적 오류를 줄여나가야겠다. 또한 대한침구학회지의 편집위원회에는 침구학 연구자 이외에 통계를 전공한 편집위원을 두어 침구학회지에 투고되는 논문들의 통계처리 방법에 대하여 검토하여 통계적 오류를 줄여갈 수 있도록 지도하여야 할 것이다.

V. 결 론

저자는 침구학회지의 논문들의 통계적용에 있어서 오류의 여부와 정도를 알아보고자, 침구학회지 1권부터 19권 6호까지 총 40권 921편의 논문들 중 분석적 통계기법을 사용한 429편에 대하여 통계적 처리과정을 분석하여 통계적 오류에 대해 검토한 결과 다음과 같은 결론을 얻었다.

1. 대상 논문 429편 중 93편(21.68%)에서 통계 처리방법을 밝히지 않는 오류를 범했던 것으로 나타났다. 서술의 빈도는 점차 증가하였으며, 19권에서는 통계 방법을 적용한 모든 논문들이 통계 처리방법을 정확히 밝혀놓았다.

2. 유의수준을 설정하거나 밝히는데 있어서 53편(12.35%)에서 오류를 범했고, 중심경향과 산포도의 기술에서는 109편(25.41%)에서만 올바른 방법인 평균±표준편차(Mean±SD.)로 나타내었고, 245편(57.11%)에서는 평균±표준오차(Mean±SEM. or Mean±SE)로 나타내는 오류를 범했으며, 비모수 통계기법을 사용한 논문 23편에서는 중앙값이나 범위로 나타낸 경우없이 모두 오류를 범하였다.

3. 정규분포성과 등분산성의 검정에 있어서는 292편의 논문 중 175편(59.93%)이 정규 분포성을 검정에 있어서 오류를 범했고, 14편(4.79%)이 등분산성의 검정에서 오류를 일으켰다.

4. t-test를 시행한 185편의 논문 중 표본집단이 셋 이상이어서 분산분석을 시행하여야 했는데 t-test를 시행하는 오류를 범했던 경우가 99편(50%)이었으며, χ^2 -검정을 시행한 14편의 논문 중 Fisher

exact test를 시행하여야 할 작은 표본수를 가진 논문이 5편이었는데, 이중 1편에서만 Fisher exact test를 시행하였고, 나머지 4편에서는 χ^2 -검정을 시행하는 오류를 범하였다.

5. 반응변수가 이산형 척도였던 38편의 논문 중에서 t-test나 분산분석을 시행하는 오류를 범한 논문이 각각 14편(36.84%)과 4편(10.53%)이었다. 또한 125편의 논문에서는 쌍체표본에 대하여 independent t-test와 Mann-Whitney U test를 시행하는 오류를 범한 경우가 각각 2편과 1편이었다. 또한 단순히 t-test를 시행하였다고 서술하여 paired t-test인지 independent t-test인지 알 수 없었던 경우가 57편이었다.

6. 분산분석을 시행한 119편의 논문에서 집단간의 유의한 차이가 있는 경우에도 다중비교를 시행하지 않거나, 시행하였는지 자체를 언급하지 않은 경우가 각각 3편과 17편으로, 다중비교의 시행하는 문제에 있어서 20편(16.81%)의 논문이 오류를 일으킨 것으로 나타났다.

VI. 참고문헌

1. 안윤옥. 한국의학 연구논문에 대한 방법론적 검토. 대한보건협회지. 1980 ; 6(1) : 101-107.
2. 안윤옥, 고웅린. 자료 처리과정에 대한 통계학적 검토. 예방의학회지. 1973 ; 6(1) : 81-85.
3. 안윤옥, 이형기. 의학 연구논문의 방법론 및 통계처리 기법의 타당성 평가를 위한 점검표 개발. 한국의학교육. 1991 ; 3(1) : 52-69.
4. 안윤옥. 실용의학통계론. 서울. 서울대학교 출판부. 1990.

5. 하헌선. 보건학 석사학위 논문에 대한 통계적 평가-서울대학교 보건대학원 석사학위 논문을 중심으로. 서울대학교 보건대학원 석사학위 논문. 1984.
6. 현혜진. 보건학 관련 연구논문에 대한 통계기법 적용과 방법론 검토. 서울대학교 보건대학원 석사학위 논문. 1990.
7. 이형기, 안윤옥. 1980년대에 발표된 국내 의학 연구논문의 방법론 및 통계처리 기법의 타당성에 관한 평가연구. 한국의학교육. 1991; 3(1) : 52-69.
8. 박용규. 가정의학회지 논문의 통계 적용. 가정의학회지. 1998; 19(5) : s137-43.
9. 이정열, 이선동. 한국 한의학 논문의 연구 설계와 통계 분석에 관한 연구 경향. 한의학 연구소 논문집. 1998; 1(1) : 105-118.
10. Dixon, Massey, Introduction to Statistical Analysis, 3rd Ed. McGraw Hill, 1969.
11. 박용규. 학회지 게재논문의 통계기법 적용 : 논문심사의 관점에서. 가정의학회지. 1998; 19(11) : 1093-1098.
12. Anonymous : A pillar of medicine[editorial]. J Am Med Assoc. 1966; 195 : 1145.
13. 이형기, 허봉렬, 안윤옥. 1980년대에 발표된 국내의학 연구논문의 방법론 및 통계처리 기법의 타당성에 관한 평가연구. 가정의학회지. 1991; 12(6) : 46-67.
14. 이선미, 이승옥. 국내 간호학 연구논문에 활용된 통계기법의 타당성 평가연구. 한국보건통계학회지. 1998; 23(1) : 42-64.
15. 최중후, 이재창. 학술논문과 통계적기법. 서울. 자유아카데미. 1990.
16. 박용규. 1998년 대한가정의학회지 게재논문의 분석단계별 오류 : 1998년 4월에서 12월까지. 가정의학회지. 1999; 20(5) : 457-461.
17. 서홍관, 황인홍, 허봉렬. 가정의학회지에 게재된 논문분석. 가정의학회지. 1990; 11(2) : 8-16.
18. 이승옥. 통계학의 이해. 자유아카데미. 1991.
19. 이은옥. 통계적 유의성 검정상의 오류와 임상적 유의성의 중요성. 간호학논문집. 1999; 13(1) : 1-6.
20. Brown G.W. Standard deviation, standard error. Which 'standard' should we use? Am J Dis child. 1982; 136 : 937-41.
21. Sawilowsky, S.S. & Blair, R.C., A more Realistic look at the Robustness and Type II Error, Properties of the t-test to Departures from Population Normality. Psychological Bulletin. 1992; 11 : 352-360.
22. 박용규. 1999년 6월 게재논문의 통계적 오류. 가정의학회지. 1999; 20(7) : 940-942.
23. 윤기중, 안윤기, 김범수. 통계의 오용과 효율적 이용에 관한 연구. 산업과 경영. 1987; 24(3) : 37.
24. Feinstein AR. Clinical biostatistics IV : A survey of the statistical procedures in general medical journals, Clin Pharmacol Ther. 1974; 15 : 97-107.