

경락경혈학회지에 게재된 논문의 통계적 오류에 관한 고찰(2007~2011년)

이민희* · 강경원* · 김정은 · 최선미 · 이상훈

한국한의학연구원

Analysis of the Statistical Errors in Articles of The Korean Journal of Meridian and Acupuncture

Minhee Lee*, Kyung-Won Kang*, Jung-Eun Kim, Sun-Mi Choi, Sanghun Lee

Korea Institute of Oriental Medicine

Abstract

Objectives : This study was to investigate statistical validities and trends of previously reported papers that used various statistical techniques such as t-test and analysis of variance.

Methods : To analyze the statistical procedures, 54 original articles using those statistical methods were selected from The Korean Journal of Acupuncture published from 2007 to 2011.

Results : T-test and analysis of variance were used in 23(25.27%), and 18 papers(19.78%) out of 54 papers, respectively. Seven articles(12.96%) did not report alpha values and 26(48.15%) out of 54 studies were not tested for normal distribution. One paper(1.85%) misused t-test and 7 papers(38.89%) did not carry out the multiple comparison.

Conclusions : To improve the quality of KJA, statistician involvement in research design would be necessary to reduce errors in statistical methods and interpretation of the results.

Key words : statistical error, statistical method, The Korean Journal of Acupuncture

I. 서 론

통계분석을 사용함에 있어서 먼저 연구의 목적을 바탕으로 타당한 연구방법 및 통계적 방법을 기술하고, 이를 바탕으로 적절한 기법의 선택과 응용을 해야 한다¹⁾. 그러나 잘못된 응용은 논문이나

학술지에 게재된 논문들에서 뿐만 아니라 엄격한 심사를 거쳤다고 보여지는 전문학술지에 게재된 논문들에서도 흔히 발생하고 있는 문제다²⁾. 이러한 문제는 외국의 영국정신의학지(British Journal of Psychiatry)에 게재된 논문의 83%에서 통계적 방법을 사용하고 있고, 그 중 45%에서 통계적 오류가 있다고 보고하였으며³⁾, 국내에서도 잘못된 통계 기법의 적용 및 결과해석의 오류 등 여러 통계 오류가 있었다고 보고하였다⁴⁻⁷⁾. 이와 같이 통계적 방법에서 객관적이지 못할 경우에 논문의 그릇된 결론을 도출할 수 있기 때문에 올바른 통계 기법의 적절한 선택과 적용여부가 바로 과학적 연구의 성패를 가늠한다고 할 수 있다⁸⁾. 또한 자료의 분포

· Corresponding author: Sanghun Lee, Korea Institute of Oriental Medicine, 1672, Yuseongdae-ro, Yuseong-gu, Daejeon 305-811, Korea

Tel: +82-42-868-9461, Fax: +82-42-863-9464

E-mail: ezhani@kiom.re.kr

· Received December 3, 2012. Revised December 18, 2012.

Accepted December 18, 2012

This study was supported by Korea Institute of Oriental Medicine(K12010, K12203).

*These authors contributed equally to this work.

나 자료에 대한 가정조차 확실하지 않은 상황에서 정확한 접근방법을 결정하기란 어려운 일이기 때문에 연구계획서를 작성하는 과정에서 어떻게 분석해야 할지를 통계전문가의 조언을 받아서 고민해야 하는데, 이런 일련의 과정 없이 수집된 자료에서는 분석조차 어려운 자료일 가능성이 높다고 볼 수 있다⁹⁾. 따라서 2007년도 제24권부터 2011년도 28권까지 5년간 발간된 한국경락경혈학회지(The Korean Journal of Acupuncture)를 대상으로 통계기법의 적용과 통계적 오류를 평가 분석하고자 한다.

II. 재료 및 방법

1. 연구기간 및 대상

본 연구는 2007년도 제24권부터 2011년도 제28권까지 5년간 발간된 한국경락경혈학회지를 대상으로 실시되었다. 학회지에 게재된 총 285편의 논문 중 종설, 증례, 동물실험 연구 등을 제외하고 임상 연구를 대상으로 확증적 통계기법을 사용한 논문 54편을 대상으로 하였으며, 탐색적 통계기법을 사용하거나 통계자료가 없는 논문은 제외하였다(Fig. 1).

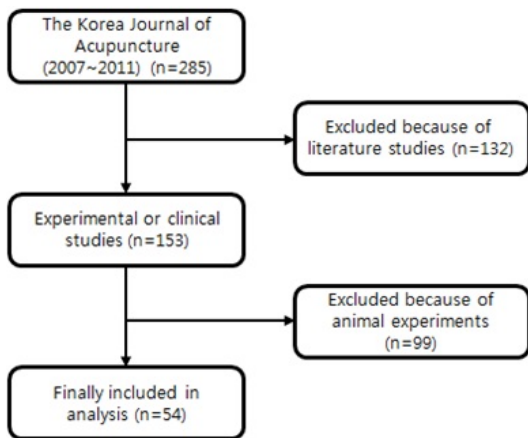


Fig. 1. Flow chart of selection process for analysing the statistical errors.

2. 방법

한국경락경혈학회지에 게재된 논문들을 대상으로 통계 사용 논문의 빈도, 통계 처리기법의 빈도, 통계적 오류의 빈도에 대하여 조사하였다. 통계적 오류의 빈도는 통계처리에 대한 서술의 오류, 유의수준(α)에 대한 오류, 중심경향(대표치) 또는 산포도에 대한 기술의 오류, 정규분포성에 대한 기술의 오류, 다중비교 시행에 대한 오류, 결과 제시에 대한 오류 등에 대하여 조사하였다.

III. 결 과

1. 통계 사용 논문의 빈도

한국경락경혈학회지 총 5권 285편의 논문 중 임상 연구를 대상으로 확증적 통계기법을 사용한 경우는 54편(18.95%)이었으며, 통계기법이 사용된 논문이 가장 많이 실린 횟수는 한 호에 5편으로 3회 있었으며, 통계기법이 사용된 논문이 전혀 실리지 않은 경우도 3번 있었다(Table 1).

2. 통계 처리기법의 빈도

통계분석을 이용한 54편의 논문에서 통계분석 기법은 총 91회 이용되었으며, 이 중 비모수적인 방법은 16회(17.58%) 이용되었다. t-검정(t-test)이 23회(25.27%)로 가장 많이 사용되었으며, 다음으로 분산분석(Analysis of Variance, ANOVA)이 18회

Table 1. Numbers of Statistical Articles in The Korean Journal of Acupuncture(n=285)

Year	Number				Total
	1	2	3	4	
07	14(2)	13(3)	16(5)	16(2)	59(12)
08	17(0)	16(4)	12(3)	15(3)	60(10)
09	10(0)	13(3)	11(0)	18(4)	52(7)
10	12(2)	16(3)	11(1)	13(5)	52(11)
11	16(2)	8(3)	20(5)	18(4)	62(14)
Total	69(6)	66(16)	70(14)	80(18)	285(54)

Table 2. Number of methods of analysis in The Korean Journal of Acupuncture(n=54)

Method of analysis	Year					Total
	07	08	09	10	11	
Student's t-test	3	3	2	3	3	14
Paired t-test	2	1	1	1	4	9
Analysis of variance	4	4	1	5	4	18
Repeated measures analysis of variance	2	0	1	1	2	6
Chi-square test	1	2	0	1	2	6
Fisher's exact test	1	0	0	0	0	1
Correlation	3	0	0	3	3	9
Regression	0	0	0	1	5	6
Mann-Whitney U test	1	0	2	1	0	4
Wilcoxon signed rank test	3	0	1	3	1	8
Kruskal-Wallis test	0	0	0	1	0	1
Etc.	3	4	2	0	0	9
Total	23	14	10	20	24	91

Table 3. Number of Description about Statistics Methods in The Korean Journal of Acupuncture (n=54)

Year	Description of statistical methods		
	No	Yes	Total
07	1	11	12
08	0	10	10
09	1	6	7
10	1	10	11
11	0	14	14
Total	3	51	54

(19.78%), 상관분석(correlation)이 9회(9.89%) 사용되었다. 총 사용된 통계기법 91개를 5년간 통계가 사용된 총 논문의 수로 나누어 보면 1.7로 대략 1편의 논문에 1.7개의 통계기법이 사용된 것을 알 수 있다(Table 2).

3. 통계적 오류의 빈도

1) 통계처리 방법에 대한 서술의 오류

사용한 통계처리 방법(t-test, ANOVA 등)에 대하여 서술하는 방법에 있어서, 54편의 논문 중에서

3편(5.56%)이 통계처리 방법을 밝히지 않았다. 26 권에서는 통계처리 방법을 밝히지 않은 비율이 14.29%였던 것에 비해 27권에서는 9.09%, 28권에서는 0.00%로 통계처리에 대하여 서술을 하는 비율이 높아졌다(Table 3).

2) 유의수준(α)에 대한 오류

유의수준은 0.05를 사용한 논문이 44편(81.48%)으로 가장 많았는데, 논문에서 유의수준으로 0.05/0.001을 함께 사용한 논문 2편은 0.05로 포함해서 합계를 구했다. 유의수준을 제시하지 않고 p value만 기록한 것이 4편(Unknown, 7.41%)이었으며, 유의수준과 p value를 모두 제시하지 않은 논문이 3편(None, 5.56%)이었다. 또한 유의수준 5%를 신뢰도 95%라고 기술한 논문이 1편(1.85%)이었다(Table 4).

3) 중심경향(대표치) 또는 산포도에 대한 기술의 오류

중심경향과 산포도에 대한 기술에 있어서 평균±표준편차(Mean±SD)로 나타낸 경우가 30편(57.69%)으로, 평균±표준오차(Mean±SE)로 나타낸 11편(21.15%)보다 많았다. 그리고 평균±표준편차인지

Table 4. Number of Articles Classified by Significant Level in The Korean Journal of Acupuncture(n=54)

Year	Significant level						Total
	0.005	0.01	0.05	0.1	Unknown	None	
07	0	1	9	1	1	0	12
08	0	0	7	0	2	1	10
09	0	0	5	0	1	1	7
10	0	0	11	0	0	0	11
11	1	0	12	0	0	1	14
Total	1	1	44	1	4	3	54

Table 5. Number of Articles Classified by Describing Central Tendency and Dispersion in The Korean Journal of Acupuncture(n=52)

Year	Central tendency and dispersion					
	Mean	Mean±SD	Mean±SE	Frequency	Unknown	Total
07	1	6	3	1	0	11
08	1	4	3	1	1	10
09	1	3	2	0	1	7
10	0	6	2	0	2	10
11	1	11	1	0	1	14
Total	4	30	11	2	5	52

평균±표준오차인지 명시하지 않는 경우가 5편(9.62%)이었으며, 평균만 제시한 경우가 4편(7.69%)으로 나타났다. 그리고 산포도를 기술할 때는 평균±표준편차로 제시하였지만 그래프에서는 평균±표준오차로 나타낸 경우가 1편(1.92%)이었다. 범주형 변수로만 이루어져 평균이 아닌 빈도만 제시한 경우는 2편(3.85%)이었으며, 비모수 검정을 시행한 논문 중 중심경향과 산포도를 중앙값과 사분위범위 등으로 나타낸 경우는 없었다(Table 5).

4) 정규분포성에 대한 기술의 오류

통계 분석을 위해서는 분석 전에 표본 집단의 정규분포성을 검정하여야 하며 정규성 여부에 따라 통계 분석 방법이 달라질 수 있다. 하지만 표본 집단의 정규분포성을 검정하여야 할 논문 54편 중 정규분포성에 대해 기술된 논문은 한 편도 없었다(Table 6).

Table 6. Number of Articles Classified by Describing Normal Distribution in The Korean Journal of Acupuncture(n=54)

Year	Normal		Total
	Yes	No	
07	0	12	12
08	0	10	10
09	0	7	7
10	0	11	11
11	0	14	14
Total	0	54	54

5) 다중비교 시행에 대한 오류

세 그룹 이상을 비교하기 위한 분석 방법이 쓰인 논문은 총 19편이었는데, 이 중 분산분석

Table 7. Number of Articles Classified by post hoc in The Korean Journal of Acupuncture(n=54)

Year	Post hoc							Total
	Scheffe	Duncan	Dunnett	LSD	Tukey	Unknown	None	
07	0	0	0	0	0	2	2	4
08	1	0	0	0	0	0	2	3
09	1	0	0	0	0	0	0	1
10	2	0	1	0	0	0	3	6
11	1	1	0	1	1	0	0	4
Total	5	1	1	1	1	2	8	18

(ANOVA)을 시행한 경우가 18편이었고, t-test를 이용한 경우가 1편이었다. 분산분석을 시행한 논문 18편 중 다중비교를 시행한 경우가 11편(61.11%)이었고, 세 그룹 이상을 비교한 결과 유의한 차이가 없다고 결론을 내려 다중비교를 시행하지 않은 경우가 7편(38.89%)이었다. 다중비교를 시행하였으나 어떤 방법을 이용하였는지 밝히지 않아 유의수준이 조정되었는지 여부를 확인할 수 없는 경우가 2편(11.11%)이었다(Table 7).

6) 결과 제시에 대한 오류

가설 검정을 위한 통계 분석 결과를 제시함에 있어 p value는 제시하지 않고 통계량만 제시한 경우가 1편이었다. 또한 종속변수의 형태가 연속형 변수이지만 Chi-square test를 수행하였다고 언급된 경우가 1편이었다.

IV. 고찰 및 결론

본 연구는 2007년도에서 2011년도까지 5년간 발간된 한국경락경혈학회지의 총 285편의 논문을 분석했으며, 그 중 임상연구를 대상으로 확증적 통계기법을 사용한 논문 54편을 대상으로 하였다.

5년간 임상연구에 확증적 통계분석을 사용한 논문의 개수는 총 54편이었으며, 사용된 총 통계기법의 수는 91개로 대략 1편의 논문에 1.7개의 통계기법이 사용된 것을 알 수 있었다. t-검정 및 분산분석으로 대표되는 모수적 방법으로 41건(45.05%), Mann-Whitney U test 및 Wilcoxon signed rank test로 대표되는 비모수적 방법으로는 16건(17.58%)

이었다. Park⁵⁾에 의하면 카이제곱검정이 31.7%로 가장 많이 적용되었고, 그 다음으로 t-검정과 분산분석 그리고 상관분석 순으로 많이 적용되는 통계분석이라고 보고하고 있다. Reed 등¹¹⁾의 연구에서는 의학 논문에서 주로 사용되는 추론통계 기법으로 t-검정, 카이제곱 검정 및 피셔의 정확 검정을 언급하고, 이들 통계 기법을 정확하게 이해하고 있으면 의학 논문의 약 70%를 이해하고 해석할 수 있다고 하였다.

통계 검정법에는 모수적 검정과 비모수적 검정으로 나누어지는데, 모수적 검정이란 정규분포(normal distribution)를 하는 자료의 평균을 비교하는 대부분의 검정, 즉 t-검정, 분산분석(ANOVA), 및 회귀분석을 일컫는다. 모수적 검정을 하기 위해서는 정규분포 외에도 각 집단의 분산이 서로 같아야 한다는 요구 조건이 만족되어야 한다. 반면 비모수적 검정은 자료가 정규분포를 하는 경우, 모수적 검정에 비해 낮은 검정력을 갖지만, 정규분포가 아닌 경우에는 모수적 검정보다 더 높은 검정력을 갖는다⁵⁾. 모수적 검정에서는 일반적으로 평균과 표준편차를 제시하고 비모수적 검정에서는 중앙값과 범위를 제시하는데, 검토된 논문 중 비모수 검정을 하였음에도 중앙값과 범위를 제시한 논문은 없었다.

적절한 통계기법의 선정은 첫째 분석하고자하는 변수의 수, 둘째 표본 집단의 수, 셋째 측정된 자료의 형태에 따라서 결정된다¹²⁾. 집단이 3개 이상이며, 정규성을 만족하면 분산분석(ANOVA)을 시행하여야 하는데, 본 연구에서 t 검정과 관련된 오류로는 보정 없이 t 검정을 반복 사용한 경우가 1

건이 있었는데, 이는 3군 이상간의 비교에서 보정 없이 반복적으로 t 검정을 사용함으로써 α 가 증가되어 잘못된 결과를 초래할 수 있다¹⁰⁾.

카이제곱 검정은 자료의 빈도 또는 비율을 비교하는 방법으로 전체 표본수가 20 이하이거나, 어느 한 셀의 기대치가 5 이하인 경우에는 카이제곱 검정을 그대로 적용하면 오차가 커지기 때문에 연속성을 수정하거나 Fisher의 정확 검정(Fisher's exact test)을 사용해야 한다. 카이제곱 검정을 사용한 6편의 논문 중 카이제곱 검정을 사용할 수 있는 전제조건을 간과하여 피셔의 정확검정을 사용하는 것이 적절하다고 권고 받은 논문이 2편이었다.

분산분석(ANOVA)을 한 결과 유의한 차이가 있을 때, 구체적으로 어떤 집단 간에 차이가 있는지를 추가로 검정할 필요가 있는데, 이 때 사용하는 검정방법이 다중비교(multiple comparison)이다. 모든 집단들간을 비교하는 경우에서 각 집단의 표본수가 같다면 Tukey, Duncan, Newman-Keuls 등의 검정방법을 사용하고, 각 집단의 표본수가 다르다면 Scheffe의 방법을 사용한다⁵⁾.

한 개체에 대하여 여러 번 반복적으로 측정된 자료를 분석하기 위한 분석법으로는 반복측정 분산분석법 등 반복 측정 자료를 위한 분석 방법을 이용해야 한다. 일원분산분석(ANOVA)을 이용하는 경우 개체 내 차이에 대한 고려가 되지 않고 각 관측치가 서로 독립이 아니기 때문에 결과가 왜곡될 수 있다. 검토된 논문 중 반복적으로 측정된 자료를 이용한 논문은 전체 54편 중 6편(11.11%)이었는데, 모두 반복측정 분산분석법을 이용하여 바르게 분석되었다. 또한 반복측정 분산분석법은 구형성 가정(Sphericity assumption)이 충족되었는지를 먼저 살펴보아야 하는데, 이에 대해 언급한 논문은 없었다.

회귀분석은 독립변수와 종속변수의 결합 관계를 알아보기 위하여 수행되는데, 오차항의 등분산성과 독립성, 정규성을 만족하여야 수행할 수 있다. 검토된 논문에서 회귀분석이 이용된 논문은 6편(11.11%)이었는데, 6편 모두 오차항 및 다중공선성에 대한 검정 여부를 명시하지 않았으며, 한 편은 회귀식의 R-square만 제시하고 선형 관계가 높다고 결론을 내렸으며 그 외의 p value는 제시하지

않았다. 또한 회귀모형을 설정할 때 모형에 중요 변수를 설명변수로 포함하지 못하거나 불필요한 설명변수가 추가되는 경우가 있다. 중요 변수를 포함하지 못하는 경우에는 변수 누락 편의(omitted variable bias)가 생길 수 있으며 이렇게 추정되는 회귀계수는 일치추정량이 되지 못한다. 부적합한 변수가 포함된 경우 이 변수가 기존 설명변수 간에 상관관계가 존재할 경우 회귀계수가 유의한데도 유의하지 않은 것으로 나타날 수 있으므로 모형 설정 시 적절한 모형을 선택하도록 충분한 연구가 필요하다.

연구 시작 전 통계방법론의 기술에서 반드시 포함해야 하는 항목은 양측/단측 검정, 사용한 통계 프로그램, 유의수준(α), 검정력 등이다. 이 연구에서 유의수준을 나타내지 않고 연구결과와 함께 p value를 제시한 논문이 4편(7.41%)이었다. 이러한 오류는 유의수준에 대한 이해가 부족한 것으로 생각되며, 연구방법에서 미리 설정을 해서 연구를 진행해야 할 것으로 생각된다. 또한 본 연구에서 오류로 포함하지는 않았지만, 상당수의 논문에서 p value의 값을 그대로 기재하지 않고 $p > 0.05$ 혹은 $p < 0.05$ 등과 같이 표시하였다. 예를 들어 p value가 0.06이라면, 실제로 통계적으로는 유의하지 않지만 해당 연구 분야에서 의미가 있는 차이일 수도 있지만 $p > 0.05$ 라고 기록한다면 그 값이 어느 정도 크기인지 알 수 없어 매우 제한된 정보만을 제시하는 셈이다. 그리고 p value는 직관적으로 이해하기 쉽지 않은 점이 있으므로 p value만 단독으로 제시하기보다는 보다 많은 정보를 제공해 주는 신뢰구간(confidence interval)을 사용하는 것이 바람직하다¹³⁾.

본 논문에서 언급한 통계적 오류를 방지하기 위하여 Nam과 Chung¹⁴⁾이 제안한 여러 가지 통계적 방법들을 정리한 순서도가 임상, 의학 연구자들이 자료의 특성 및 상황에 따라 어떤 통계학적 검정들을 사용하는 것이 좋을지 결정하는 데 도움이 될 것이다(Fig. 2).

본 연구의 제한점으로는 첫째 통계기법의 오류에 대한 설명을 포괄적으로 기술함으로써 각각의 통계기법에 대한 전제조건, 장점 및 단점에 대해서 설명이 부족하다. 둘째로 논문에 활용된 통계기법

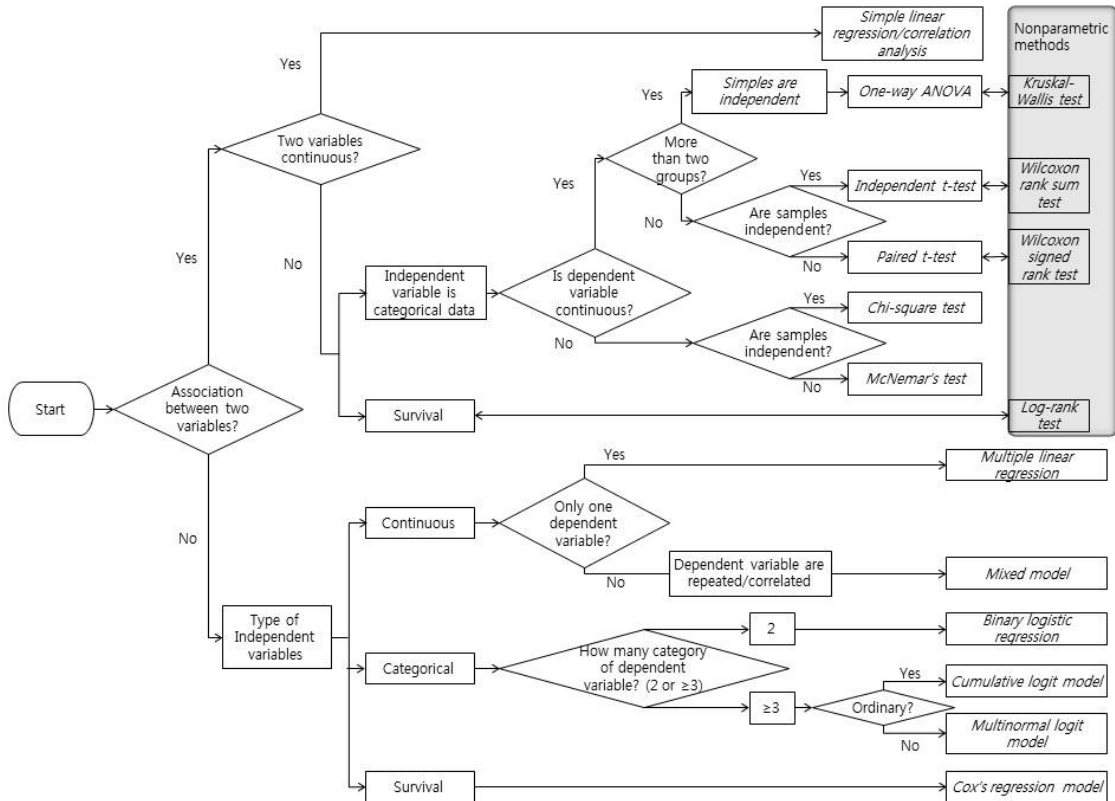


Fig. 2. A flowchart of appropriate tests of statistical inference.

보다 유용하게 이용될 수 있는 통계기법의 부연설명이 부족한데, 이는 연구시작부터 통계전문가의 의견을 참고하여 적절한 통계분석 방법을 선택할 수 있도록 노력할 필요가 있을 것으로 보인다.

이상의 결과를 통해서 양방연구에서 적용되고 있는 통계적 방법론이 아닌 한의학 연구에서 적용 가능한 연구설계의 개발이 필요해 보이며, 한의학 연구자들에게 통계기법 및 통계프로그램의 이용에 사전 교육이 강화되어야 할 것으로 보인다. 또한 연구 시작부터 종료까지 통계전문가의 참여로 통계적인 오류를 미연에 방지할 수 있는 체계를 갖추는 것이 필요하다고 생각된다.

감사의 글

This study was supported by Korea Institute of Oriental Medicine(K12010, K12203).

참고문헌

1. Korea Food and Drug Administration. Statistical Considerations of the Protocol and Report. 2009.
2. Kang HC. Analysis of using statistical methods for the Journal of Korean Academy of Nursing from 1999 to 2001. Journal of Korean Academy of Nursing. 2002 ; 32(12) : 929-35.
3. White SJ. Statistical errors in papers in the British Journal of Psychiatry. British Journal of Psychiatry. 1979 ; 135 : 336-342
4. Park YK. Statistical error of articles published in March 1993. Journal of the Korean Academy of Family Medicine. 1999 ; 20(4) : 412-3.

5. Park YK. A study on applying statistical methods in articles in Journal of the Korean Academy of Family Medicine. Journal of the Korean Academy of Family Medicine. 1998 ; 19(5) : 137-43.
6. Ahn YO, Ko UR. Statistical resumption from health articles in one series of medical journals. Journal of Preventive Medicine and Public Health. 1973 ; 6(1) : 81-6.
7. Lee HK, Huh BY, Ahn YO. An assessment of methodological and statistical validity of medical articles published in Korea, from 1980 to 1989. Journal of the Korean Academy of Family Medicine. 1991 ; 12(6) : 46-67.
8. Lee SM, Lee SW. An assessment of statistical validity of nursing researches published in Korea from 1986 to 1995. J of the Korean Society of Health Statistics. 1998 ; 3(1) : 42-64.
9. Korea Food and Drug Administration, National Institute of Toxicological Research. Basic Statistics for the design and analysis of clinical research.
10. Reed JF 3rd, Salen P, Bagher P. Methodological and statistical techniques: what do residents really need to know about statistics? J Med Syst. 2003 ; 27(3) : 233-8.
11. Park YK. A study on Applying Statistical Methods in Articles in Journals: for Manuscript-Review. Journal of the Korean Academy of Family Medicine. 1998 ; 19(11) : 1093-8.
12. Yu TY, Jin YH, Jeong TO, Lee JB. Statistical methods in the articles in the Journal of the Korean Society of Emergency Medicine Published from 1998 to 2002. Journal of the Korean Society of Emergency Medicine. 2003 ; 14(5) : 475-80.
13. Park ES. Recommendation of using statistical methods for study of bioscience. Korean Journal of Clinical Pharmacy. 2007 ; 17(2) : 96-101.
14. Nam CM, Chung SY. Statistical methods for medical studies. J Korean Med Assoc. 2012 ; 55(6) : 573-81.