

# 이화의대지에 게재된 논문의 통계분석방법에 관한 조사연구

- 1978년 3월(창간호)부터 1995년 6월호를 중심으로 -

이화여자대학교 의과대학 예방의학교실 및 이화의과학 연구소 역학부  
목동병원 안과학교실 및 이화의과학 연구소 세포조직학부\*

하 은 희 · 최 규 룡\*

= Abstract =

## Statistical Techniques and Misuse of Statistical Methods in The Ewha Medical Journal

- From 1978 to 1995 -

Eun Hee Ha · Kyu Ryong Choi\*

*Department of Preventive Medicine, Ophthalmology,\* and Ewha Medical Research Center  
College of Medicine, Ewha Womans University*

Over the past 17 years, the number of medical research papers in the Ewha Medical Journal has increased. And the use of statistics to evaluate experimental and clinical research data has increased. We observed statistical methods used in 730 articles from the Ewha Medical Journal for a 17 year period(from 1978 to 1995). The most commonly used method was measurement of central tendency with t-test,  $\chi^2$ -test, ANOVA, correlation analysis and regression analysis(74.8%). The second most common method was  $\chi^2$ -test with t-test, ANOVA, pearson correlation(9.2%), followed by epidemiological statistics(7.6%).

The observed errors of statistical method in 314 original articles were 585 spells(repeat No./paper).

The most common error was that of no significant value and confidence interval(18.4%), followed by misuse of SE instead of SD(15.5%).

Misuses of applying statistical method included the use of the t-test instead of ANOVA in comparing three groups(0.2%) and the use of the student t-test instead of paired t-test and vice versa.

KEY WORDS : Statistics · Statistical method · Misuse.

## 서 론

일반적으로 의과학(medical science)에서 추구하는 내

용은 연구 대상(인간의 생명, 건강, 질병 등)의 본질(essence), 기원(foundation), 원인성(causes), 그리고 그 궁극성(finality)에 관하여 시간과 공간을 초월하는 지적인 이해(intellectual understanding)를 얻고자 하는

것이라 하였는데 이와 같은 과학적 추구의 가장 보편적 과정은 인간의 지혜와 능력으로 가능한 범위 내에서 대상 현상을 관측하고 그 관측된 결과로부터 실제의 본질이나 기원, 원인성 또는 궁극성 등에 관한 이헤나 지식을 쌓아가는 방법이다<sup>1)</sup>.

한편 연구 대상 및 연구 결과에 대한 연구자의 견해는 일반적으로 주관적이라 할 수 있는데 따라서 동일한 주제에 대한 많은 연구에도 불구하고 전문가들이 주장하는 바가 다양하게 된다.

그러므로 의학 연구의 결말은 무엇을 증명해 내거나 결론을 내리는 것이 아니고 그 양이나 관련성에 관한 기존의 지식에 대하여 변화를 주는 것이며 그 변화는 연구에서 얻어지는 정보나 증거들을 바탕으로 하여 연구자에 의해 만들어진다. 이때 증거라 함은 지금까지의 각종 정보나 가정, 그리고 자료의 요약 혹은 분석 처리를 통하여 얻게 된다. 자료 분석이 끝 연구 대상에 대한 결론에 실제로 도달하는 것이 아님은 확실하다. 자료 분석은 연구자의 지식에 변화를 일으키게 하는 증거를 요약하는 데에 그 목적이 있는 것이다<sup>2)</sup>.

따라서 지금 현재까지의 개발된 많은 방법 중 연구 결과 분석시 연구자가 도달하고자 하는 연구의 가설을 유도하고 연구 결과를 계량화하기 위하여 거의 모두 통계적(statistical) 또는 확률적(probability)기법에 의존하고 있다. 이러한 노력들의 이론적 근거에는 연구 자료에서 관측한 내용들이 확률적 모형을 따르고 있다는 것이 전제되어 있기 때문이다.

이러한 통계적 가설은 경험적 사실에 해당하는 표본에서부터 통계적 추론의 특성을 그대로 지니게 된다. 따라서 통계적 가설의 타당성 여부를 추론하는 가설검정은 표본분포를 바탕으로 한다. 그러므로 의학 연구에서도 연구 대상의 분포와 이들 연구 결과들을 추론해 나갈 때 적절한 연구 방법론과 통계 기법을 활용하여 연구자가 하고자 하는 연구를 의미 있게 마무리하는 작업은 연구 논문을 쓸 때에 가장 핵심적인 작업이라 할 수 있다<sup>3)</sup>. 최근 의학 논문 및 학회지에 있어서도 통계방법의 적용이 필수 요건이 되고 있다<sup>3)</sup>.

그러나 어떤 현상을 찾아내고자 하는 통계적 방법을 이용함에 있어 통계학을 전공하지 않은 연구자들이 이를 사용하는데는 통계 방법 적용의 오류를 범할 수 있는 여러 가지 어려움이 있다. 통계적 방법론을 평가한 외국문헌으로는 영국 정신과학회지(British Journal of Psychi-

atry)에 1977년 7월부터 1978년 6월까지 게재된 168편의 논문에 대해 자료의 표본추출 방법, 모집단 특성 규모, 기술 여부, 표본 자료 기술, Student's t-test, Chi-square test등을 분석<sup>4)</sup>한 바 이 중 140편(83%)이 통계적 기법을 통하여 결론을 도출했는데 이 가운데 45%에 해당되는 63편의 논문이 통계적 오류를 범했다는 것을 보고했으며 이 중 표본추출 방법, 자료의 측정에 대한 오류가 가장 많았다고 하였다. 그 외 다중 비교(multi-compairson)분석에서의 통계적 오류를 평가한 논문과<sup>5)</sup> Chi-square test 방법에서의 적정성 문제로 야기되는 통계적 오류 등을 관찰<sup>6)</sup>한 바 있다. 안윤옥 등도<sup>7)</sup>의학연구 논문의 통계처리 기법의 타당성 평가를 위한 점검표를 개발한 바 있다. 이들은 통계자료의 측정 방법, 기법의 적정성 및 적용 절차, 분석 상의 오류, 이에 따른 결론 유출 등을 공통적으로 지적했다.

국내에서는 1980년대에 발표된 대한의학협회지에 수록된 논문 379편을 21항목으로 구성된 점검 표로 평가하였는데, 통계 처리 기법의 타당성 평가에 있어서는 297편중 290편인 97.6%가 하나 이상의 점검항목에 대해 통계적 오류가 있음<sup>8)</sup>을 지적하였던 바 있다. 또한 1989년도 예방의학회지에 게재된 논문에 대한 분석에서도 통계분석 방법 적용의 오류가 73.0%라고 하였던 것<sup>9)</sup>과 김윤연 등<sup>10)</sup>이 1990년 안과학 회지를 분석한 결과 각 점검 항목별로 9.8~100%의 오류를 지적하였다. 그러나 기존의 이러한 논문들의 연구 대상이 일개 학회지를 대상으로<sup>10)</sup>하였거나 일년간의 기간을 대상으로 하였거나 또는 수년간을 분석했더라도 일개 학과<sup>11)</sup>를 중심으로 분석하였다.

따라서 본 연구에서는 통계분석 방법을 통해 결론을 도출하는 연구 논문에 대해 엄격한 논문 심사 과정이 제도화된 외국 학술 논문집과 국내 의학계의 여러 학회지 연구 논문에서도 위와 같이 통계적 오류가 지적되었던 바, 이회대의지에 논문들의 과별 특성과 연도별 통계분석의 동향에 관한 적절성을 평가함으로써 향후 통계분석 방법의 오류를 최소화하기 위한 기초자료를 마련하고자 하였다.

평가 지표의 기준이 비교적 명확하게 평가될 수 있는 항목을 선정하여 최대한 주관적인 관점을 배제한 상태에서 객관적으로 연구 논문을 고찰했으나 여러가지 제한점들이 있을 수도 있다.

따라서 본 연구에서는 대부분의 연구 논문이 통계적 분석을 통해 결론을 도출하고 있는 것을 고려하여 이회대의

지 논문에 활용된 통계분석 방법의 사용 현황을 알아보고, 통계적 기법을 활용한 연구 논문에서 나타난 통계분석 방법과 이에 대하여 통계분석방법의 효율적인 이용 및 통계적 오류를 최소화 할 수 있는 해결 방안을 모색하고자 시도되었다.

## 연구방법

### 1. 연구 대상

본 연구의 대상은 1978년 3월 이화의대지 창간호부터 1995년 6월호까지의 기간동안 총 53권에 게재된 논문 730편 전수를 연구 자료로 하였으나 통계분석방법의 분석 시에는 증례 및 증설, 분류가 불가능한 연구를 제외한 총 314편의 논문에 사용된 통계분석 방법을 분석 대상으로 하였다.

### 2. 연구 내용 및 방법

1995년 7월 1일에서 9월 30일까지의 기간동안 이화여 대학교 의과대학 도서관을 방문하여 1978년부터 1995년까지의 이화의대지를 수집하였다. 논문에 사용된 통계 적용의 적절성을 분석하기 위하여 안윤옥<sup>7)</sup>, 안윤옥<sup>8)</sup>, 이관형<sup>9)</sup>, 김용연<sup>10)</sup>의 점검 표와 외국 문헌의 점검표<sup>12)15)</sup>를 참고로 하여 연구자가 고안한 점검표를 예비 조사를 통하여 구조화하였으며 기초학 교수와 임상 교수 각각 한 명씩 의견을 교환한 후 수정 및 보완하여 통계분석 방법 적용의 적절성에 대한 점검표(Table 1)를 완성하였다.

이렇게 완성된 점검 표를 가지고 통계분석 방법론을 분석하였으며 연구의 틀(Fig. 1)은 다음과 같다.

연구의 내용은 연구의 틀(Fig. 1)에서 보는 바와 같이 연구 논문에 사용된 통계방법론에 대하여 연도별, 과별로 분석하고 통계분석 방법의 적정성 여부에 대하여 평가하였으며 통계분석시 컴퓨터 통계 패키지 사용에 대하여 과별 및 통계 패키지 종류별로 분석하였다.

통계분석방법의 분류는 통계분석을 하지 않은 경우나 명확하게 구분할 수 없는 경우는 not-define, 평균이나 이에 준하는 중심 경향만 분석한 경우는 central tendency only, 민감도, 특이도, 상대위험도, 교차비와 같은 역학적 지표가 사용되었을 경우 epidemiologic statistics, t-test,  $\chi^2$ -test, ANOVA, Pearson correlation, non-parametric test, simple linear regression, 기타

**Table 1.** Check list for statistical methods for Ewha Medical Journal

- 
- I. Was there a complete description of basic datas
1. Were there frequencies and percent distribution?
  2. Were there means and standard deviations?
- II. Were the presentation of statistical material used appropriate?
1. T-Test :
    - 1) Was there a adequate use of dispersion?
    - 2) Were the "t-value" and "degree of freedom" given for the main results?
    - 3) Was there the use of paired t-test for paired data?
    - 4) Was there description of power level and confidence interval?
  2.  $\chi^2$ -test :
    - 1) Were the " $\chi^2$ -value" and "degree of freedom" given for the main results?
    - 2) Was there description of power level and confidence interval?
    - 3) Was there correction for continuity?
  3. ANOVA :
    - 1) Were the "F-value" and "degree of freedom" given for the main results?
    - 2) Was there description of power level and confidence interval?
  4. Correlation Analysis :
    - 1) Was there description of correlation coefficient (r)?
    - 2) Was there description of power level and confidence interval?
  5. Regression Analysis :
    - 1) Was there description of regression coefficient ( $r^2$ )?
    - 2) Was there presentation of regression equation?
- III. Were the statistical methods used appropriate?
1. Wasn't there multiplicity on hypothesis testing?
  2. Wasn't there  $r^2$  instead of r and vice versa?
  3. Wasn't there misuse of  $\chi^2$ -test instead of ratio difference analysis?
- IV. Was the inscription of statistical methods appropriate?
- 

(survival analysis) 등으로 구분하였다.

### 3. 통계 분석 방법

본 연구는 현황 분석을 목적으로 하였으므로 모든 자료

를 전산 입력한 후 PC-SAS package(6.04)를 이용하여 자료처리 하였으며 기술 통계와  $\chi^2$ -test를 분석에 활용하였고 유의 수준은 5%를 선정하였다.

## 연구결과

1978년에서 1995년 6월호까지 게재된 총 730편의 논문

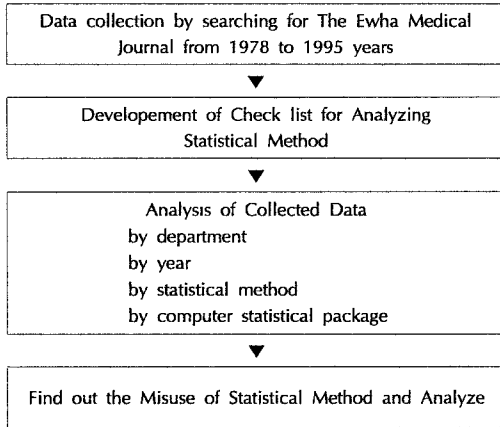


Fig. 1. Outline of research design.

중 외과 계열 학과의 논문이 332편(45.5%), 내과계열학과 논문이 232편(31.8%), 기초학 논문이 166편(22.7%)이었다(Table 2).

연도 별로는 1989년도부터 증편되기 시작하여 1991년도가 58편으로 가장 많았으며 90년대에 와서는 계속 50편 이상을 유지하고 있음을 알 수 있다.

통계 분석 방법의 적용유무를 살펴 본 결과 통계분석 방법을 적용한 논문이 180편으로(57.3%) 더 많았으나 과별로는 내과계열 학과에서 비교적 통계분석을 많이 하는 것(65.3%)으로 나타났으며 기초학, 외과계열 순으로 통계분석 방법을 적용하고 있는 것을 알 수 있으나 이의 차이는 유의하지 않았다(Table 3).

Table 3. Application of statistical analysis by department Unit : No(%)

Department	Application of statistical analysis			$\chi^2$ -test <df>
	Yes	No	Total	
Basic science	59(55.7)	47(44.3)	106(100.0)	3.8 <sup>NS</sup>
Medical part	62(65.3)	33(34.7)	95(100.0)	<2>
Surgical part	59(52.2)	54(47.8)	113(100.0)	
Total	180(57.3)	134(42.7)	314(100.0)	

NS : Non Significant

Table 2. Research papers in the Ewha Medical Journal by years

Unit : No(%)

Year	Department			Total
	Basic science	Medical part	Surgical part	
78	13(36.1)	11(30.5)	12(33.3)	36(100.0)
79	14(43.8)	7(21.9)	11(34.4)	32(100.0)
80	8(21.6)	7(18.9)	22(59.5)	37(100.0)
81	8(26.7)	5(16.7)	17(56.7)	30(100.0)
82	11(36.7)	6(20.0)	13(43.3)	30(100.0)
83	12(22.6)	11(20.8)	30(56.6)	53(100.0)
84	13(31.7)	8(19.5)	20(48.8)	41(100.0)
85	9(24.3)	14(37.8)	14(37.8)	37(100.0)
86	8(25.0)	5(15.6)	19(59.4)	32(100.0)
87	6(19.4)	9(29.0)	16(51.6)	31(100.0)
88	11(30.6)	14(38.9)	11(30.6)	36(100.0)
89	5(10.6)	19(40.4)	23(48.9)	47(100.0)
90	6(11.3)	23(43.4)	24(45.3)	53(100.0)
91	9(15.5)	22(37.9)	27(46.6)	58(100.0)
92	6(12.0)	25(50.0)	19(38.0)	50(100.0)
93	11(20.8)	21(39.6)	21(39.6)	53(100.0)
94	13(25.5)	15(29.4)	23(45.1)	51(100.0)
95*	3(13.0)	10(43.5)	10(43.5)	23(100.0)
Total	166(22.7)	232(31.8)	332(45.5)	730(100.0)

\*include 1995. 6.

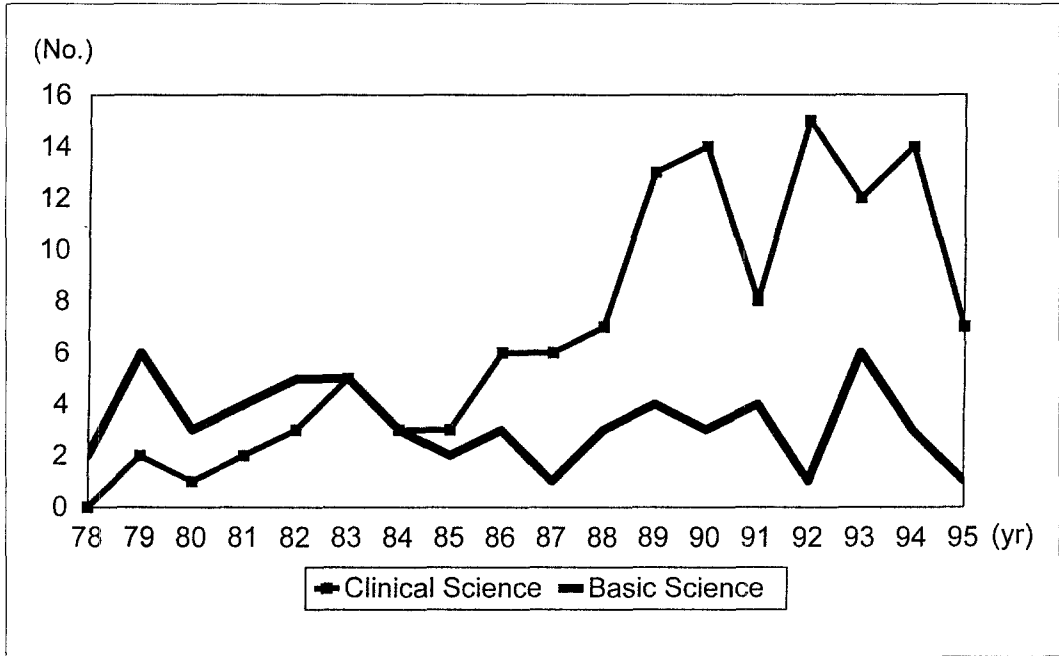


Fig. 2. The use of statistical analysis by year.

이를 연도별로 기초 과목과 임상 과목으로 구분하여 살펴보면 다음과 같다(Fig. 2).

78년부터 83년까지는 기초 과목에서 통계분석을 더 많이 활용하였으나 그 이후에는 임상 과목에서 통계분석을 많이 시행하고 있으며 특히 89년부터 급격하게 증가하고 있음을 알 수 있다.

한편 컴퓨터 통계 패키지 사용 유무 현황을 살펴보면 컴퓨터 통계 패키지를 사용하지 않은 경우가 244편(77.7%)으로 더 많았으나 과별로는 내과 계열, 외과 계열, 기초학 순이었으며 과별로 유의한 차이를( $p < 0.05$ ) 나타내었다(Table 4).

이를 연도별로 기초 과목과 임상 과목으로 구분하여 살펴보면 80년대 초에는 기초 과목에서 더 많이 시행하였으나 80년대 중반부터 임상 과목에서 훨씬 증가하였으며 역

시 89년부터 급격히 증가하고 있음을 알 수 있다(Fig. 3).

이는 연구논문의 전체수 증편과 통계분석, 통계 패키지의 사용이 일치되고 있음을 알 수 있다(Table 1, Fig. 2, 3).

기초학 연구 논문에 있어서 대부분이(83.0%) 통계 패키지를 사용하지 않았으나 컴퓨터 통계 패키지를 사용한 경우에 대하여 살펴보면 명확히 명시된 경우에는 SAS를 이용하였다고(6.6%) 하였으나 다른 경우에는 컴퓨터 패키지명을 제시하지 않고 있었다(Table 5-1).

임상 연구 논문에 있어서도 대부분이(75.0%) 통계 패키지를 사용하지 않았으나 사용한 경우에는 내과, 마취과, 흉부외과 등에서 SPSS(3.4%)를, 정신과, 피부과, 일반 외과에서 SAS를(1.9%) 사용하였으나 41편의 논문에서는 사용한 통계 패키지에 대한 기술이 불충분하여 컴퓨터 통계 패키지를 이용하였다 등으로만 언급하고 있었다(Table 5-2).

주로 사용하는 통계분석방법으로는 산포도를 기본으로 하여 이에 t-test나  $x^2$ -test, ANOVA, correlation analysis, regression analysis, 비모수 검정 등을 같이 시행하고(74.8%)있었으며 다음은  $x^2$ -test를 기본으로 하고 t-test, ANOVA, Pearson correlation을 실시하는 경우가(9.2%) 많았다. t-test나 z-test 단독 시행(4.0%)한

Table 4. Statistical analysis with computer statistical package Unit : No(%)

Department	Computer statistical package			$x^2$ -test <df>
	Yes	No	Total	
Basic science	18(17.0)	88(83.0)	106(100.0)	8.5*
Medical part	31(32.6)	64(67.4)	95(100.0)	<2>
Surgical part	21(18.6)	92(81.4)	113(100.0)	
Total	70(22.3)	244(77.7)	314(100.0)	

\* $p < 0.05$

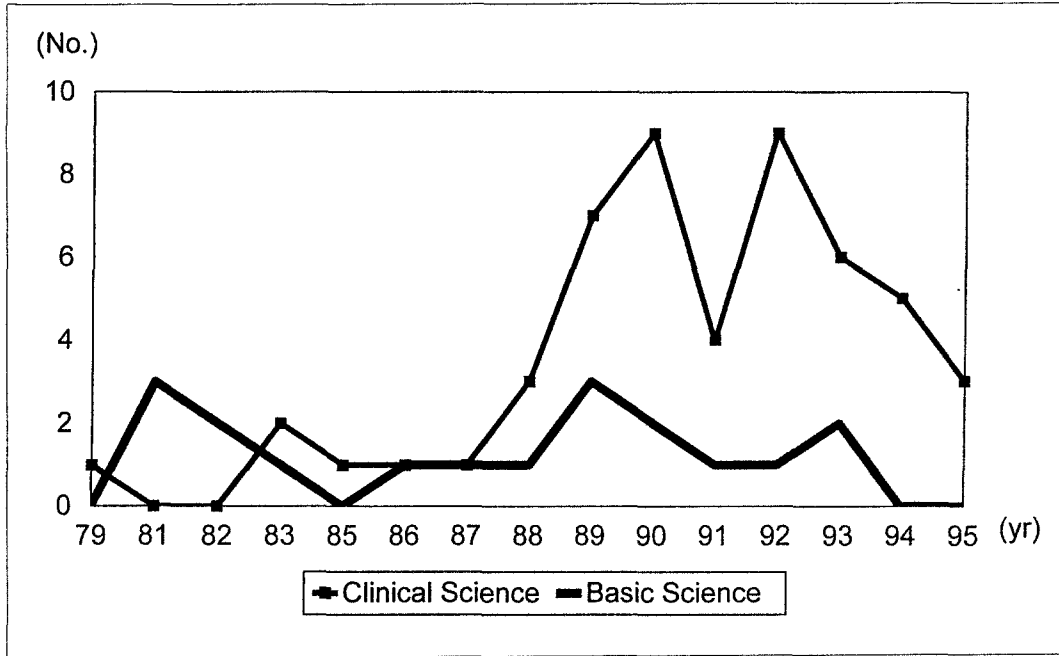


Fig. 3. Use of computer statistical package by year.

Table 5-1. The use of computer statistical package in basic science

Unit : No(%)

Department	Computer statistical package			Total
	Non	SAS*	Others®	
Anatomy	7(100.0)	0(0.0)	0(0.0)	7(100.0)
Parasitology	14(87.4)	1(6.3)	1(6.3)	16(100.0)
Physiology	3(75.0)	0(0.0)	1(25.0)	4(100.0)
Biochemistry	25(96.2)	0(0.0)	1(3.8)	26(100.0)
Preventive medicine	7(46.7)	5(33.3)	3(20.0)	15(100.0)
Microbiology	1(50.0)	1(50.0)	0(0.0)	2(100.0)
Pharmacology	11(100.0)	0(0.0)	0(0.0)	11(100.0)
Pathology	13(76.5)	0(0.0)	4(23.5)	17(100.0)
Biology	7(87.5)	0(0.0)	1(12.5)	8(100.0)
Total	88(83.0)	7(6.6)	11(10.4)	106(100.0)

# : Statistical Analysis System @ : Not defined computer package

경우와 Pearson correlation analysis를 사용한 경우도 1편(0.41) 있었다. 한편 epidemiologic statistics도 19편(7.6%)있었으며 linear regression과 correlation(3.2%), survival analysis(2.4%), multivariate analysis(0.8%)도 시행되고 있음을 알 수 있다(Table 6).

통계분석방법의 숫자가 통계분석 사용 논문 편수 보다 많은 것은 통계분석을 중복 시행한 경우를 모두 건수로 계산하였기 때문이다.

기초학 연구 논문에 있어서 주로 사용된 통계 분석 방법을 과별로 살펴보면 대부분이 central tendency를 중심으로 하여 t-test & z-test 또는  $\chi^2$ -test를 이용하였으며 비슷한 경향이었으나 예방 의학과 미생물학에서 epidemiologic statistics를 사용하였던 것이 특기할 만하다 (Table 7-1).

임상 연구 논문에서도 주로 central tendency와 t-test,  $\chi^2$ -test등을 사용하는 경향이 있었으나 내과 치료방사선과에서는 주로 survival analysis를 이용하였으며 소

**Table 5-2.** The use of computer statistical package in clinical science.

Unit : No(%)

Department	Computer statistical package				Total
	None	SAS <sup>#</sup>	SPSS <sup>§</sup>	Others <sup>@</sup>	
Internal medicine	14(60.9)	0(0.0)	2(8.7)	7(30.4)	23(100.0)
Pediatrics	5(33.3)	0(0.0)	0(0.0)	10(66.7)	15(100.0)
Psychiatrics	10(71.4)	1(7.1)	1(7.1)	2(14.3)	14(100.0)
Neurology	3(50.0)	0(0.0)	0(0.0)	3(50.0)	6(100.0)
Dermatology	15(78.9)	1(5.3)	0(0.0)	3(15.8)	19(100.0)
Radiology	16(94.1)	0(0.0)	0(0.0)	1(5.9)	17(100.0)
OB & Gyn	12(70.6)	0(0.0)	0(0.0)	5(29.4)	17(100.0)
General surgery	8(72.7)	1(9.1)	0(0.0)	2(18.2)	11(100.0)
Neuro surgery	19(90.5)	0(0.0)	0(0.0)	2(9.5)	21(100.0)
Orthopadics	7(87.5)	0(0.0)	0(0.0)	1(12.5)	8(100.0)
Plastic surgery	17(94.4)	0(0.0)	0(0.0)	1(5.6)	18(100.0)
Chest surgery	3(60.0)	0(0.0)	2(40.0)	0(0.0)	5(100.0)
ENT	6(100.0)	0(0.0)	0(0.0)	0(0.0)	6(100.0)
Urology	5(83.3)	0(0.0)	0(0.0)	1(16.7)	6(100.0)
Anesthesiology	8(57.2)	1(7.1)	2(14.3)	3(21.4)	14(100.0)
Ophthalmology	7(100.0)	0(0.0)	0(0.0)	0(0.0)	7(100.0)
Others	1(100.0)	0(0.0)	0(0.0)	0(0.0)	1(100.0)
Total	156(75.0)	4(1.9)	7(3.4)	41(19.7)	208(100.0)

# : Statistical analysis system    § : Statistical package of social science

@ : Not defined computer package(including QBSTAT, DBSTAT, VIDS IV)

**Table 6.** The kind of statistical analysis method in the Ewha Medical Journal

Statistical method	No(%)
Central tendency or dispersion only	61(24.6)
Central tendency or dispersion only and t-test & z-test	77(31.0)
Central tendency or dispersion only and $\chi^2$ -test	8(3.2)
Central tendency or dispersion only and ANOVA	7(2.8)
Central tendency or dispersion only and Pearson correlation	7(2.8)
Central tendency or dispersion only and simple linear regression	2(0.8)
Central tendency or dispersion only and non-parametric	6(2.4)
t-test & z-test and $\chi^2$ -test	10(4.0)
$\chi^2$ -test	18(7.2)
t-test & z-test	1(0.4)
t-test & z-test and Pearson Correlation	1(0.4)
Epidemiologic statistics	19(7.6)
$\chi^2$ -test and ANOVA	3(1.2)
$\chi^2$ -test and Pearson correlation	1(0.4)
ANOVA and Pearson Correlation	4(1.6)
ANOVA and non-parametric	3(1.2)
Simple linear regression and Pearson correlation	8(3.2)
Pearson correlation and non-parametric	1(0.4)
Time series	3(1.2)
Factor analysis	1(0.4)
Multivariate cluster analysis	2(0.8)
Survival analysis	6(2.4)
Total	249(100.0)

**Table 7-1.** Major statistical method in basic science

Department	Major statistical Method	
	First	Second
Anatomy	t-test & z-test	Central tendency or dispersion only
Parasitology	Central tendency	$\chi^2$ -test
Physiologys	Central tendency	t-test & z-test
Biochemistry	t-test & z-test	Central tendency
Preventive medicine	Epidemiologic statistics	$\chi^2$ -test
Microbiology	Central tendency	Epidemiologic statistic
Pharmacology	t-test & z-test	Central tendency
Pathology	Central tendency	t-test & z-test
Biology	Central tendency	t-test & z-test

**Table 7-2.** Major statistical method in clinical science

Department	Major statistical Method	
	First	Second
Internal medicine	t-test/ $\chi^2$ -test	Survival
Pediatrics	t-test/central tendency	Simple linear regression/correlation
Psychiatrics	$\chi^2$ -test/ANOVA	Central tendency/multivariate analysis
Neurology	t-test/central tendency	ANOVA/non-parametric
Dermatology	Central tendency	ANOVA/correlation
Radiology	Central tendency	Correlation
OB & Gyn	t-test	Central tendency
General surgery	Central tendency	t-test
Neuro surgery	t-test/central tendency	$\chi^2$ -test
Orthopadics	Central tendency t-test	ANOVA/correlation
Plastic surgery	Central tendency t-test	ANOVA
Chest surgery	Central tendency/non-parametric	
ENT	Central tendency	Epidemiologic statistics
Urology	t-test/central tendency	
Anesthethology	t-test/central tendency	Non-parametric
Ophthalmology	Central	Tendency t-test

아과에서 correlation analysis나 regression analysis를 사용하였고, 정신과에서는 multivariate analysis를, 신경과, 흉부외과, 마취과 등에서 비모수 통계 분석에도 이용한 것으로 나타났다(Table 7-2).

각 통계분석 방법의 사용 여부가 기초학, 내과 계열 학과, 외과 계열 학과에 따라 차이가 있는지를 살펴보기 위하여  $\chi^2$ -test를 시행한 결과 통계분석을 사용 안하는 경우는 기초학이 내과, 외과 계열보다 많은 편으로 유의한 차이( $p < 0.01$ )를 나타내었으며 기타 다른 통계분석방법에 대하여 과별로 유의한 차이를 나타내지는 않았다. 통계분석치가 생략된 것(linear regression analysis 이하)은  $\chi^2$ -test에서 칸의(cell) 기대숫수가 5이하이기 때문이다

(Table 8).

창간 이후 1978년에서 1995년 6월호사이의 총 730편중 통계분석 방법의 적절성에 대하여 살펴 본 결과 총 314편의 원저중 180편의 논문에서 통계분석을 시행하였고 총 585건의 오류가 관찰되었다(Table 9).

이는 논문에서 나타난 통계분석 방법이 한가지 이상이 있을 때를 중복하여 건수를 계산하였기 때문에 585건이 되었다. 많은 논문에서 t-test를 사용하였으나 t-test의 유의 수준을 선정하지 않거나(18.4%), t-value를 제시하지 않았고(3.1%), SD대신 SE를 사용하고 있었다(15.5%).  $\chi^2$ -test에서는  $\chi^2$ -value와 자유도를 제시하지 않았고(5.0%) 유의 수준을 선정하지 않는 경우도(2.4%) 있었



**Table 8.** The difference of statistical method among department

	Basic science	Medical Part	Surgical Part	Total	Unit : No(%) $\chi^2$ -test <df>
<u>Not defined</u>					16.0**
Yes	95(89.6)	63(66.3)	87(77.0)	245(78.0)	<2>
No	11(10.4)	32(33.7)	26(23.0)	69(22.0)	
<u>t-test</u>					3.1 <sup>NS</sup>
Yes	31(29.3)	39(41.1)	41(36.3)	111(35.4)	<2>
No	75(70.7)	56(58.9)	72(63.7)	203(83.8)	
<u>Central tendency</u>					6.2 <sup>NS</sup>
Yes	47(44.3)	57(60.0)	51(45.1)	155(49.4)	<2>
No	59(55.7)	38(40.0)	62(54.9)	159(50.6)	
<u><math>\chi^2</math>-test</u>					2.2 <sup>NS</sup>
Yes	5(4.7)	9(0.4)	6(0.3)	20(6.4)	<2>
No	101(95.3)	86(90.6)	107(94.7)	294(93.6)	
<u>Simple linear regression<sup>#</sup></u>					
Yes	1(0.9)	6(6.3)	3(1.2)	10(3.2)	
No	105(99.1)	89(93.7)	110(97.4)	304(96.8)	
<u>Pearson correlation<sup>#</sup></u>					
Yes	1(0.9)	14(14.7)	3(2.7)	18(5.7)	
No	105(99.1)	81(85.3)	110(97.3)	296(94.3)	
<u>ANOVA<sup>#</sup></u>					
Yes	1(0.9)	7(7.4)	5(4.4)	13(4.1)	
No	105(99.1)	88(92.6)	108(95.6)	301(95.9)	
<u>Epidemiologic statistics<sup>#</sup></u>					
Yes	0(0.0)	3(3.2)	6(5.3)	9(2.9)	
No	106(100.0)	92(96.8)	107(94.7)	305(97.1)	
<u>Non-Parametric<sup>#</sup></u>					
Yes	1(0.9)	14(14.7)	3(2.7)	18(5.7)	
No	105(99.1)	81(85.3)	110(97.3)	296(94.3)	
Total	106(100.0)	95(100.0)	113(100.0)	314(100.0)	

\*\*p < 0.01 \*p < 0.05 NS : Non-significant  
# : The statistics were omitted because of cell number(under 5).

으며 표본에 대한 적합도 검정을 실시하지 않았거나 연속성 보정을 실시하지 않은 경우도 있었다(0.8%). ANOVA test에서도 F-value나 자유도를 제시하지 않고(2.1%), 유의수준을 선정하지 않은 경우도(1.7%) 있었으며 특히 다중 비교를 하지 않고 t-test를 여러 번 시행한 경우도(2.7%) 있었다. 상관 분석에서는 상관 계수의 유의수준을 제시하지 않거나 상관 계수를 제시하지 않은 경우도(4.2%)있었으며 회귀 분석에서는 회귀식을 제시하지 않은 경우가 1편 있었다. 다음은 통계 분석 방법 적용의 오류에 대하여 살펴보면 비록 편수는 많지 않아 각각 1편씩이었으나 세 집단 비교시에는 ANOVA를 실시해야 하는데 t-test를 여러번 실시하는 경우, t-test를 실시해야

함에도 ANOVA를 실시한 경우도 있었으며(0.6%), paired t-test를 시행하여야 하는 경우에 student t-test를 시행하였고(11.1%), student t-test를 시행하여야 하나 paired t-test를 시행하고 있었다(0.9%). 또한 상관계수(r)를 회귀계수( $r^2$ )로 제시하였으며 비율차에 대한 검정을 실시해야 하는데  $\chi^2$ -test를 실시하고 있었다. 한편 분석 방법 표기 오류로는 평균 ± 표준편차의 형태로 제시하여야 하는데 평균 + 표준편차로 표기하였으며 paired t-test에서 p-value를 2개씩 표기한다던가 t-test를 실시하고  $\chi^2$ -test를 실시한 것으로 잘못 표기한 경우도 있었다.

**Table 9.** Misuse of statistical methods in the Ewha Medical Journal

Misuse of statistical methods	No*(%)
<u>Incomplete description of basic data</u>	
No verification of normal distribution	79(13.4)
No frequency	45(7.7)
No percent	18(3.1)
Mean without SD	23(3.9)
Without Mean $\pm$ SD	2(0.3)
<u>t-test</u>	
No verification of control	11(1.9)
Misuse of SE(instead of SD)	91(15.5)
No degree of freedom or t-value	18(3.1)
No significant value and level	109(18.4)
Student t-test instead of paired t-test	65(11.1)
Paired t-test instead of student t-test	5(0.9)
<u><math>\chi^2</math>-test</u>	
No degree of freedom or $\chi^2$ -value	29(5.0)
No significant value and levels	14(2.4)
No verification of fitness test about sample	3(0.5)
No adjustment of continuity	2(0.3)
<u>ANOVA</u>	
No degree of freedom or F-value	12(2.1)
No significant value and level	10(1.7)
Repeated t-test instead of Manova	16(2.7)
<u>Correlation Analysis</u>	
No significant value and level about correlation coefficient(r)	12(2.1)
No correlation coefficient(r)	12(2.1)
<u>Regression Analysis</u>	
No regression equation	1(0.2)
<u>Misuses of applying statistical method</u>	
z-test instead of t-test(2 groups)	1(0.2)
t-test instead of ANOVA(3 groups)	1(0.2)
ANOVA instead of T-test(2 groups)	1(0.2)
Regression coefficient( $r^2$ ) instead of correlation coefficient(r)	1(0.2)
$\chi^2$ -test instead of difference of ratio	1(0.2)
<u>Misuses of inscripting statistical Method</u>	
Mean + SD instead of Mean $\pm$ SD	1(0.2)
Two p-values in paired t-test	1(0.2)
$\chi^2$ -test instead of t-test	1(0.2)
Total	585(100.0)

# : Include the repeated event

## 고찰

본 연구는 이화의대지 논문의 창간 이후 지금까지의 통

계분석방법을 대상으로 하여 어떤 종류의 통계 방법이 사  
용되었는지 통계분석 방법은 적절하였는지를 살펴보고자  
시도되었다. 본 연구가 이화의대지 논문의 질을 평가하고  
자 시도된 것이 아니며 본 연구자들이 점검하고자 개발한

점검표가 완벽한 평가서이라고 장담할 수도 없다. 다만, 통계분석 의뢰 또는 상담을 통하여 얻어진 경험을 바탕으로 이화의대지에 사용된 통계분석 방법의 종류와 이의 사용이 적절한지의 현황을 파악하여 향후 이러한 오류가 최소화되고자 하는데 기본 자료를 제공하고자 하였을 뿐이다.

의학분야에서의 통계분석 방법의 적용에 대하여는 많은 통계학 교과서에<sup>119)</sup> 그 이론적 근거가 제시되어 있다. 그러나 통계 전문가가 아닌 많은 연구자들이 쉽게 접근하기에는 너무 난해하게 만들어져 있어 어려움이 많다. 그러므로 초기에는 가능하면 통계학자의 도움을 통해 많은 부분이 이해되어야만 점차적으로 독자적인 해결이 가능할 수 있을 것이다. 최근 수년 사이에 컴퓨터 통계 패키지는 눈부시게 발전하여 어떠한 자료든지 그 자료로부터 기술 통계치에서부터 추정통계치까지 산출 가능하기 때문에 컴퓨터를 사용할 줄 안다면 원하는 통계치를 얻기란 어려운 일이 아니다. 컴퓨터는 단순하기 때문에 구체적인 통계적 방법이나 결과에 대한 해석 방법의 기본 원리를 몰라도 컴퓨터가 스스로 입력된 자료에 대해 연구자가 강요한 모든 결과를 내준다. 그러므로 논문이 통계 처리가 되어 통계적으로 유의한 결과를 유도하였다는 것보다는 연구 가설 또는 목적에 맞는 적절한 자료 수집 과정을 거쳤는지, 통계분석 방법의 적용은 적절하였는지, 통계 패키지를 사용하였다면 컴퓨터에서 나온 결과를 제대로 해석하였는지, 논문에서의 표기 방법은 적절하였는지가 더욱 중요하다고 할 수 있다. 또한 누구나 가장 익숙하고 다루기 쉬운 통계 패키지를 사용하면 되지만 반드시 무슨 통계 패키지를 사용하였는지를 제시하여야 한다.

본 연구에서 사용된 통계분석 방법에서는 창간 이후 1978년에서 1995년 6월호사이의 총 730편중 통계분석 방법의 적절성에 대하여 살펴 본 결과 총 314편의 원저중 180편의 논문에서 통계분석을 시행하였고 가장 많이 사용된 통계분석방법은 산포도를 제외하면 t-test와  $\chi^2$ -test였으며 총 585건의 오류가 관찰되었다. 이는 일년간의 안과 학회지의 통계처리방법에서 총 204편의 논문중 통계처리가 없는 경우가 113편, 통계처리된 89예중 통계방법을 밝히지 않은 예가 29예로 가장 많았고 사용된 통계기법중에는 t-test와 분할표본이 많이 사용되었다고 한 것<sup>120)</sup>과 Juzych등<sup>14)</sup>이 Ophthalmology 1990년도 232편중에서 통계처리가 없는 경우가 77편이었으며 통계처리된 155예중 통계처리 방법을 밝히지 않은 경우가 4.7%, 가장 많이 사

용된 방법은 t-test였던 것과 비슷한 연구결과라고 할 수 있다.

통계적 오류에 있어서는 많은 연구에서 SD대신 SE를 사용하고 있었고 t-test,  $\chi^2$ -test에서는 t-value나  $\chi^2$ -value 그리고 자유도를 제시하지 않았으며 유의 수준을 선정하지 않는 경우도(2.4%) 있었다. ANOVA test에서도 F-value나 자유도를 제시하지 않고 유의수준을 선정하지 않은 경우도 있었으며 특히 다중 비교를 하지 않고 t-test를 여러 번 시행한 경우도 있었다. 상관 분석에서는 상관 계수의 유의 수준을 제시하지 않거나 상관 계수를 제시하지 않은 경우도 있었으며 회귀 분석에서는 회귀식을 제시하지 않은 경우가 1편 있었다. 또한 통계 분석 방법 적용의 오류에 대하여 살펴보면 비록 편수는 많지 않아 각각 1편씩이었으나 t-test를 실시해야 하는데 ANOVA를 실시한 경우도 있었으며, paired t-test를 시행하여야 하는 경우에 student t-test를 시행하였고, student t-test를 시행하여야 하나 paired t-test를 시행하고 있었다.

자료의 표시에서 대표값(mean, median, mode)과 산포측도중 어느 하나가 생략되면 기초자료를 정확히 기술하였다고 할 수 없으며  $A \pm B$ 로 표시되는 기초자료도 여러 대표값과 산포측도중 어느것인지 명확히 표시하여야 한다<sup>11)</sup>.

또 다른 측면은 표준 오차와 표준 편차의 혼동이다. 표준편차는 어떤 표본에서 각 측정치의 산포도를 표시하는 것이며 표준오차는 가능한 모든 표본들의 평균치의 표준편차 즉 미래의 표본들의 표본평균들의 산포도를 추정한 추정 통계량이다. 표준오차는 표준편차보다 작아 자료의 정밀도 또는 신뢰도를 좋아보이게 하기 때문에 표준오차를 사용하려는 유혹을 받게 된다고 한다<sup>120)</sup>. 따라서 기초자료의 표시에는 표준오차를 사용하지 않아야 하며 또한 표준편차와 표준오차는 성격이 다르므로 혼용하면 안된다<sup>120)</sup>.

현대통계의 큰 특징중 하나인 추측통계(statistical inference)에 의한 자료분석은 p-값을 구하는 가설검정(hypothesis testing)과 신뢰구간(confidence interval) 추정을 포함한 모수추정(parameter estimation)으로 나누어진다<sup>120)</sup>. 통계적 가설검정에서는 대립가설(alternative hypothesis : 가설추정자가 맞다고 주장하고자 하는 가설)이나 귀무가설(영가설, null hypothesis : 대립가설과 반대되는 입장의 가설)을 검정하게 되는데 논리적으로, 통계학적으로 귀무가설을 기각함으로써 상대적으로

대립가설을 지지하는 방법이 편리하여 사용된다. 통상 0.05를 생정통계학(biostatistics)에서는 기준으로 삼는데, 이는 만일 0.05를 유의수준으로 하였다면 5%의 잘못을 감수하고 귀무가설을 부정하여 상대적으로 연구자의 가설을 지지하겠다는 것이다. 연구자들은 때때로 p-값에 대해 어떤 환상을 갖는데, 안윤옥<sup>1)</sup>은 p-값이 어떤 통계적 모형 하에서 나온 하나의 통계이나 연구가설의 옳고 그름을 판가름해 주는 결정인자는 아니라고 하였으며 p-값이 유의수준이하의 값이 얻어지야만 그 연구 또는 자료수집이 잘된 것이라고 믿는 것은 잘못이라고 하였다. 그러나 통계분석에서 유의수준 등을 표시하지 않거나 명확히 표현되지 않고 p-값을 남발한 경우도 잘못된 통계분석 결과라고 할 수 있다.

세계 각국의 308개 Journal에서 참여하고 있는 International Committee Medical Journal Editors (ICMJE, 1988)<sup>3)</sup>에서 결의한 biomedical journal의 uniform requirement에 제시된 통계분석 방법 적용에 대한 지침에 관한 15가지 설명서<sup>17)</sup>의 일부를 소개하면 다음과 같다.

- 그 분야에 관한 지식이 있는 독자가 논문에 제시된 결과를 가지고 원 자료(original data)를 검증할 수 있을 만큼 자세하게 통계 방법을 기술하라.
- 가능하면 신뢰 구간과 같은 측정 오차 또는 불확실성에 대한 적절한 지표표를 사용하여 결과치를 기술하라.
- 통계적 가설 검정에 대한 p-value에만 의존하는 것을 피하고 신뢰 구간과 함께 표시하여 더 많은 정보성을 추구하는 것이 바람직하다.
- 무작위 추출방법에 대하여 보다 상세하게 기술하라.
- 실험대상의 적격성에 대하여 토의하라.
- 어떤 개입 연구나 치료 효과에 관한 연구 등에서는 치료의 함병증에 관하여 보고하라.
- 기본적인 관찰 단위를 명확하게 하고 관찰 수를 제시하라.
- 관찰이 누락된 수에 대하여 보고하라(특히 임상시험에서의 탈락과 같은).
- 어떤 통계 프로그램을 사용하였는지를 자세히 명시하라.
- 연구 방법에서는 통계분석 방법을 명시하고 연구 결과에서는 결과를 분석하기 위하여 사용된 통계 방법을 자세히 기술하라.
- 표와 그림은 필요한 만큼만 제시하도록 하고 표와 그

림은 중복하지 말라.

- 통계에서의 기술적인(technical) 용어에 대하여 비기술적인(non-technical)용어의 사용을 금하라.

- 통계 용어와 약어, 모든 부호에 대하여 명시하라.

놀라운 사실은 이웃나라 중국, 일본, 인도 등도 참여하고 있는 실정임에도 불구하고 ICMJE의 uniform requirement에 동참하고 있는 한국 잡지는 하나도 없다는 현실이다.

그 외에도 임상 연구 논문에 대한 통계학적 내용을 정리하거나 점검표를 제시하고 있으며 국내에서도 안윤옥<sup>7)</sup>, 안윤옥<sup>8)</sup>, 이관형<sup>9)</sup>, 김용연<sup>10)</sup>의 연구에서도 이미 지적된 바 있다.

본 연구의 제한점으로는 비록 수년간의 연구 자료를 수집하기는 하였으나 일개 논문 잡지에 국한하였고 보다 많은 임상가와 통계학자의 의견을 수렴하지 않고 문헌 등에 근거하여 점검표를 개발하여 논문을 분석하였다는 것이다.

그럼에도 불구하고 많은 논문에서 연구 결과를 종합 정리하고 결론을 유도함에 있어서 통계적 검정 과정을 거치지 않고 있는 점과 적절한 통계 방법을 적용하고 있지 못한 점 등의 단순화된 평가를 통해 살펴보고자 하였기에 비교적 객관성을 가지고 있을 것으로 판단된다. 또한 어떤 논문의 한 부분이 잘못되었다는 것을 꼬집어서 지적하기 보다는 향후 그러한 오류들이 최소화되기 위한 노력의 필요성을 제시하고자 하였다는데 본 논문의 의의가 있다고 할 수 있다.

## 결론 및 요약

### 연구 목적 :

본 연구는 이화 의대지 논문에 활용된 통계적 분석 방법의 사용 현황을 알아보고, 통계분석을 한 연구 논문에서 나타난 통계분석 방법의 내용과 이에 대하여 통계분석 방법의 효율적인 이용 및 통계적 오류를 최소화 할 수 있는 해결 방안을 모색하고자 시도되었다.

### 연구 내용 및 방법 :

1978년 3월 이화의대지 창간호부터 1995년 6월호까지의 기간동안 총 53권에 게재된 논문 730편 전수를 연구 자료로 하였으며 통계분석방법의 분석시에는 증례 및 종설, 분류가 불가능한 연구를 제외한 총 314편의 논문에 사용된 통계분석 방법을 분석 대상으로 하였다. 1995년 7월 1일에서 9월 30일까지의 기간동안 이화여자대학교

의과대학 도서관을 방문하여 자료를 수집하였다. 기초학 교수와 임상 교수 각각 한 명씩 의견을 교환한 후 수정 및 보완하여 통계분석 방법 적용의 적절성에 대한 점검표를 완성하였다.

연구 결과 :

1978년에서 1995년 6월호까지 게재된 총 730편의 논문 중 의과 계열 학과의 논문이 332편(45.5%), 내과계열학과 논문이 232편(31.8%), 기초학 논문이 166편(22.7%)이었다. 연도 별로는 1989년도부터 증편되기 시작하여 1991년도가 58편으로 가장 많았으며 90년대에 와서는 계속 50편 이상을 유지하고 있다.

주로 사용하는 통계분석방법으로는 central tendency or dispersion, t-test나  $\chi^2$ -test, ANOVA, correlation analysis, regression analysis이었으며 이들을 중복 시행하고 있었다(78.4%). 한편 epidemiologic statistics도 19편(7.6%) 있었으며 survival analysis(2.4%), multivariate analysis(0.8%)도 시행되고 있었다.

314편의 원저중 180편의 논문에서 통계분석을 시행하였고 총 585건의 오류가 관찰되었으며 통계적 오류의 내용은 많은 논문에서 t-test, ANOVA,  $\chi^2$ -test 시행후 t-value, F-Value,  $\chi^2$ -value와 자유도를 제시하지 않았으며 SD대신 SE를 사용하고 있었다. 상관 분석에서는 상관 계수를 제시하지 않은 경우도 있었으며 회귀 분석에서는 회귀식을 제시하지 않은 경우도 있었다. 통계 분석 방법 적용에 있어서는 비록 편수는 많지 않았으나 ANOVA를 실시해야 하는데 t-test를 여러번 실시하는 경우, t-test를 실시해야 하는데 ANOVA를 실시한 경우도 있었으며, paired t-test를 시행하여야 하는 경우에 student t-test를 시행하였고, student t-test를 시행하여야 하나 paired t-test를 시행하고 있었다.

결 론 :

본 연구에서는 비교적 객관적인 점검표를 사용하여 이 회의대지 논문에 나타난 통계적 오류에 대하여 고찰하였으며 향후 그러한 오류들이 최소화되기 위한 노력의 필요성을 제시하고자 하였는데 본 논문의 의의가 있다고 할 수 있다.

References

1) 안윤옥 : 실용의학통계론. 서울대학교 출판부 1990 : 1-63

2) 안윤옥 : 보건통계학이해. 정문각 1993 : 47-93  
 3) International committee of medical journal editors : Uniform Requirements for Manuscripts Submitted to Biomedical Journals. Ann. Inter. Med 1988 : 108 : 258-265  
 4) White SJ : Statistical errors in papers in the British Journal of Psychiatry. Br J Psychiatry 1979 : 135 : 336-42  
 5) Dawkins HC : Mutiple comparisons misuses : Why so frequently in response-curve studies. Biometrics 1983 : 39 : 789-90  
 6) Delucci KL : The use and misuse of chi-square : Lewis Burke Revisited. Psychological Bulletin 1983 : 94(1) : 166-76  
 7) 안윤옥 · 이형기 : 의학연구논문의 방법론 및 통계 처리기법의 타당성 평가를 위한 점검표 개발. 한국 의학교육 1991 : 3(1) : 19-28  
 8) 안윤옥 : 한국의학연구논문에 대한 방법론적 검토 - 의협산하 분과학회지에 게재된 논문예를 중심으로 -. 대한보건협회지 1980 : 6(1) : 101-107  
 9) 이관형 · 정호근 : 산업보건과 통계적 방법론 - 에 방의학회지에 발표된 연구논문의 통계적 방법론적 타당성 평가 -. 근로복지공사 중앙병원 직업병연구소 1990 : p1-27  
 10) 김용연 · 이은일 · 정해륜 : 국내 안과학 논문에 사용된 연구방법 및 통계처리기법. 대한안과학 회지 1994 : 35(4) : 104-117  
 11) 서홍관 · 황인홍 · 허봉렬 : 가정의학회지에 게재된 논문 분석. 가정의학회지 1990 : 11(2) : 8-16  
 12) 서울대학교 의과대학 예방의학교실 의학연구자료 분석상당실 편 : PC-SAS 공개강좌, 1992 : p25-119  
 13) Gore SM, Jones IG, Rytter EC : Misuse of statistical methods : critical assessment of articles in BMJ from January to March 1976. BMJ 1977 : 1 : 85-87  
 14) Juzych MS, Shin DH, Mahmoud Seyedsadr, Siegner SW, Juzych LA : Statistical Techniques in ophthalmic Journals. Arch Ophthalmol 1992 : 110 : 1225-1229  
 15) Fred Ederer : Refereing Clinical Research Papers for Statistical Content. Am J Ophthalmol 1985 : 100 : 735-737  
 16) Gardner MJ, Machin D, Campbell MJ : Use of check lists in assessing the statistical content of medical studies. BMJ 1986 : 292 : 810-812.  
 17) John C, Bailar III, Mosteller F : Guidelines for Sta-

*istical Reporting in Articles for Medical Journals.*  
*Ann Inter Med 1988 : 108 : 266-273*  
18) Glantz SA : *Biostatistics : How to detect, correct*  
*and prevent error in the medical literature.* Cir-

*culatation 1980 : 61 : p1-7*  
19) 이동우 : *보건통계학방법*, 초판, 서울, 신광출판사  
1986 : p13-32