

# 대한치과보철학회지에서 볼 수 있는 통계적 오류의 고찰(2006 - 2010)

강동완<sup>1</sup> · 서윤암<sup>2</sup> · 오남식<sup>3</sup> · 임희정<sup>4\*</sup>

<sup>1</sup>조선대학교 치의학전문대학원 보철학교실, <sup>2</sup>전남대학교 통계학과, <sup>3</sup>인하대학교 치의학교실,

<sup>4</sup>전남대학교 치의학전문대학원 교정학교실

**연구 목적:** 대한치과보철학회지의 최근 5년간 게재된 논문 중에서 통계 방법이 사용된 논문의 빈도를 조사하고 통계 방법의 종류 및 오류의 유형과 빈도를 살펴 보고 오류 별 적절한 통계 방법을 제시하고 통계분석 가이드라인을 개발하여 대한치과보철학회지의 학술적 발전에 기여하고자 한다.

**연구 재료 및 방법:** 2006년부터 2010년까지 게재된 논문 336편 중 통계방법이 사용된 255편을 연구대상으로 하여 연도별로 통계 방법 사용 여부, 주로 사용된 통계방법과 소프트웨어의 유형과 빈도를 조사하였다. 통계방법 별로 고안된 가이드라인에 따라 오류를 조사하였는데 주로 실험설계의 적절성, 분석법에 대한 가정검토, 표본의 독립성 및 변수의 적합성, 적절한 표본 수와 적절한 통계방법이 사용되었는지에 대한 평가를 하였고 이를 토대로 고안된 가이드라인을 제시하였다.

**결과:** 본 연구에서 조사된 논문 중 추론통계를 사용한 논문이 193편(75.9%)으로 대다수를 차지하였고 사용된 소프트웨어는 SPSS가 153편(59.77%)으로 가장 많았다. 추론통계를 사용한 논문들을 대상으로 통계방법을 분류한 결과 ANOVA (41.5%), t-test (20.0%), 비모수 방법(16.9%)순으로 많았고 최근에 가까울수록 다양한 분석법을 시도하였으나 유의한 변화는 관찰되지 않았다. 대부분의 논문들은 표본수의 산출근거를 제시하지 않았고, 분석법에 대한 가정(독립성, 정규성, 등분산성 등) 검토를 하지 않은 공통적인 오류를 범했다. 전체적으로는 61.2%의 통계적 오류를 범하였다.

**결론:** 대부분의 오류는 전체적인 분석 후 세부적인 분석으로 들어가야 하는데 두 요인에 대한 상호작용을 무시하고 단변수 분석을 여러 번 한 경우와 연구 계획 단계에서 적절한 표본 수를 산출하지 않아 나타난 경우 등이었다. 이러한 통계 오류를 최소화하기 위하여 분석 방법 별 검증 절차에 대한 통계분석 가이드라인을 제시하였다. (대한치과보철학회지 2012;50:258-70)

**주요단어:** 통계 오류; 통계 방법; 대한치과보철학회지

## 서론

연구결과를 과학적으로 입증하는데 필요한 객관적인 검증의 도구인 통계분석은 분석법에 대한 정확한 지식과 올바른 사용법이 수반되어야 한다. 그렇지 않은 경우 잘못된 결과를 낳아 많은 독자들에게 혼돈을 줄 수 있을 뿐 아니라 귀중한 자료의 낭비를 가져올 수 있다. 특히 치의학과 같이 사람을 대상으로 하는 임상연구에서는 올바른 결과를 도출하는 과정은 아무리 강조해도 지나치지 않을 것이다. 그러나 치의학 뿐 아니라 의학, 교육학, 생물학 등 많은 분야에서 객관적인 연구 결과를 검증하기 위해 통계 방법을 사용하고 있으나 분석 방법에 따른 가정을 무시하고 적절한 통계 방법을 사용하지 못함으로써 생기는 50% 이상의 오류가 각 저널에서 보고되고 있는 실정이다.<sup>1,2</sup>

국내의 치의학 분야에서의 통계 방법 관련 논문들을 살펴보면, 박 등<sup>3</sup>은 2000년부터 2006년까지의 대한치과보철학회지에 게재된 논문 중 통계적 방법이 사용된 292편의 영문판과 한글판의 논문을 조사하여 통계적 오류 항목을 크게 통계 프로그

램의 불명시, 통계 방법의 부적절한 기술, 통계 용어의 오용 등으로 나누어 오류의 유형과 빈도를 파악한 결과 논문의 통계적 타당성은 다소 부족한 것으로 나타났다고 보고하고 있다. 또한 김 등<sup>4</sup>은 2000년부터 2006년까지의 대한치주과학회지의 학술 논문 중 통계기법을 적용한 308편을 대상으로 통계 기법 오류를 분석한 결과 모수적 통계 기법과 비모수 통계 기법에 대한 명확한 이해가 선행되어야 함을 주장하였고 적절한 통계 방법을 제시하였다. 임<sup>5</sup>은 1999년부터 2003년까지의 대한치과 교정학회지(KJO)와 미국교정학회지(AJODO)의 사용된 통계 방법의 종류를 비교 분석한 결과 KJO에서는 5년간 통계방법의 종류에 유의한 변화를 관찰할 수 없었으나 AJODO에서는 통계방법의 종류에서 유의한 변화를 관찰할 수 있었다고 보고하였다.

외국의 치의학 분야에서의 통계 방법 관련 논문을 살펴보면, 김 등<sup>6</sup>은 1995년에서 2009년까지 10개의 외국의 치의학 저널 중 307개의 논문을 대상으로 적어도 한 개 이상의 통계 오류를 포함한 논문은 51%로 조사되었다. 이는 의학 저널보다는 낮은 것으로 나타난 것으로 보고 되었다. Bland와 Altman<sup>7</sup>은 대부분의 의

\*교신저자: 임희정

500-757 광주광역시 북구 용봉로 33 전남대학교 치의학전문대학원 교정학교실

062-530-5830; e-mail, hjlim@jnu.ac.kr

원고접수일: 2012년 8월 31일 / 원고최종수정일: 2012년 10월 4일 / 원고채택일: 2012년 10월 11일

\*2011년도 대한치과보철학회지 연구과제(치과보철생체공학연구회)

© 2012 대한치과보철학회

© 이 글은 크리에이티브 커먼즈 코리아 저작자표시-비영리 3.0 대한민국 라이선스에 따라 이용하실 수 있습니다.

학 저널에 출판되는 논문들이 통계학자의 도움 없이 분석이 수행되는데, 아무리 의학 연구자들이 기본적인 통계 지식을 가지고 있다 할지라도 중요한 통계 개념이나 원리에 대한 통찰력을 가진다는 것은 쉬운 일이 아니며 부분적으로는 이로 인해 논문에 사용된 통계 방법에 오류가 생겼다고 보고하고 있다. 통계 전문가의 견해는 연구 프로젝트를 계획할 때부터 얻어져야 한다고 보고하고 있고 통계 전문가의 어떤 통계방법을 사용할 것인지에 관한 의견을 듣는 것을 강력하게 권하고 있으며 각 통계 방법 별로 통계 가이드 라인을 제시하였다.

본 연구에서는 대한치과보철학회지의 최근 5년간의 게재 논문 중에서 통계 방법이 사용된 논문은 얼마나 차지하고 있으며, 이러한 논문에서 사용된 통계 방법의 종류 및 통계 오류의 유형과 빈도를 살펴 보고자 한다. 이에 근거하여 자주 범하는 오류와 자주 사용되는 통계 방법을 기반으로 분석 방법 별 검정 절차에 대한 통계분석 가이드라인을 제시하여 대한치과보철학회지의 학술적 발전에 기여하고자 한다.

## 연구대상 및 방법

### 1. 연구대상

2006년부터 2010년까지 5년간 대한치과보철학회지에 게재된 논문 336편이었고 그 중 통계 방법이 사용된 논문 258편과 사용되지 않은 논문 78편으로 분류되었고 통계방법이 사용된 258편을 연구대상으로 하였다.

### 2. 연구방법

#### 1) 연도별 통계 방법 사용 여부 분류 및 빈도조사

대한치과보철학회지에 게재된 논문 336편 중에서 통계 방법이 사용된 논문과 사용되지 않은 논문을 분류하였고 통계 방법이 사용된 논문 중에서 기술 통계를 사용한 논문과 추론 통계를 사용한 논문을 분류하여 빈도를 조사하였다.

#### 2) 연도별 사용된 통계 소프트웨어 분류 및 빈도조사

통계 방법 중 추론 통계를 사용한 논문 255편을 대상으로 사용된 통계 소프트웨어를 분류하고 빈도를 조사하였다.

#### 3) 연도별 사용된 통계 방법 분류 및 빈도 조사

연도별로 사용된 통계 방법을 분류하여 어떤 통계방법이 주로 사용되었는지 알아보았다. 본 연구에서 분류한 통계기법은  $\chi^2$ -test, Fisher's exact test, Independent t-test, Paired t-test, Analysis of Variance (ANOVA), Repeated-Measures (RM) ANOVA, Mann-Whitney test (or Wilcoxon rank-sum test), Wilcoxon signed-rank test, Kruskal-Wallis test, Pearson correlation analysis, Spearman correlation analysis, Linear regression, Logistic regression, Survival analysis, Others 등으로 분류하였으며 Others에는 위의 14가지 방법 이외에 군집분석,

Multivariate analysis of Variance (MANOVA), 요인분석, Analysis of Covariance (ANCOVA), Friedman test, Linear mixed model, two sample Kolmogorov-Smirnov test, Cronbach's alpha 등이 이에 속한다. 위의 통계방법 별로 빈도를 조사하였다.

#### 4) 통계기법의 오류 조사 및 적합한 통계 방법 제시

위의 통계방법 별로 오류를 조사하기 위하여 실험설계를 파악해서 적절한 분석법인지 판단했고 각 분석법에 대한 가정검토 및 정규성 검정, 표본 및 변수의 독립에 대한 적합성을 검토하였다. 분석법에 따른 적절한 표본을 확보하였는지에 대해 평가하고 적절한 통계 방법이 사용되지 않은 경우 이를 제시하고 결과가 올바르게 유도되었는지 확인하였다. 이를 토대로 고안된 통계 가이드라인을 완성하였다.

## 결과

### 1. 연도별 통계방법 적용 여부 조사

최근 5년간 대한치과보철학회지에 게재된 논문 336편 중 통계적 방법이 사용된 논문 258편과 사용되지 않은 논문 78편으로 기술통계를 사용한 논문은 오직 3편 뿐 이었고, 추론통계를 사용한 논문이 255편(75.9%)으로 대다수를 차지하였다(Table 1).

### 2. 연도별 사용된 통계 소프트웨어 조사

통계 방법을 사용한 논문에 대하여 사용된 통계 소프트웨어의 빈도를 조사한 결과 SPSS가 60.0%로 가장 많았고 그 다음이 SAS로 11.0%를 사용하였다. 하지만, 어떠한 소프트웨어를 사용하였는지 표기하지 않은 경우가 28.2%로 생각보다 많았다(Table 2). 소프트웨어마다 결과가 다소 다를 수 있기 때문에 반드시 사용된 소프트웨어를 기술해야 한다. 최근 들어 사용된 소프트웨어를 표기하지 않은 논문의 빈도는 줄어들었다.

**Table 1.** Frequency of descriptive and inferential statistics in the Journal of Korean Academy of Prosthodontics by year (%)

| Year  | Non-statistical Research | Statistical Research   |                        | Total       |
|-------|--------------------------|------------------------|------------------------|-------------|
|       |                          | Descriptive Statistics | Inferential Statistics |             |
| 2006  | 23 (35.9)                | 0 (0.0)                | 41 (64.1)              | 64 (100.0)  |
| 2007  | 12 (16.9)                | 0 (0.0)                | 59 (83.1)              | 71 (100.0)  |
| 2008  | 11 (17.7)                | 2 (3.2)                | 49 (79.0)              | 62 (100.0)  |
| 2009  | 13 (17.8)                | 0 (0.0)                | 60 (82.2)              | 73 (100.0)  |
| 2010  | 19 (28.8)                | 1 (1.5)                | 46 (69.7)              | 66 (100.0)  |
| Total | 78 (23.2)                | 3 (0.9)                | 255 (75.9)             | 336 (100.0) |

**Table 2.** Frequency of the statistical software in the Journal of Korean Academy of Prosthodontics by year (%)

| Statistical Software | 2006      | 2007      | 2008      | 2009      | 2010      | Total      |
|----------------------|-----------|-----------|-----------|-----------|-----------|------------|
| SPSS                 | 19 (46.3) | 34 (57.6) | 33 (67.3) | 39 (65.0) | 28 (60.9) | 153 (60.0) |
| SAS                  | 6 (14.6)  | 1 (1.7)   | 7 (14.3)  | 6 (10.0)  | 8 (17.4)  | 28 (11.0)  |
| Excel                | 0 (0.0)   | 0 (0.0)   | 0 (0.0)   | 1 (1.7)   | 0 (0.0)   | 1 (0.4)    |
| JMP                  | 0 (0.0)   | 0 (0.0)   | 0 (0.0)   | 1 (1.7)   | 0 (0.0)   | 1 (0.4)    |
| Missing              | 16 (39.0) | 24 (40.7) | 9 (18.4)  | 13 (21.7) | 10 (21.7) | 72 (28.2)  |
| Total                | 41        | 59        | 49        | 60        | 46        | 255        |

**Table 3.** Frequency of the use of statistical methods in the Journal of Korean Academy of Prosthodontics by year (%)

| Statistical Methods           | 2006 | 2007 | 2008 | 2009 | 2010 | Used Total            | Misused Total          |
|-------------------------------|------|------|------|------|------|-----------------------|------------------------|
| $\chi^2$ -test                | 1    | 1    | 2    | 9    | 4    | 17 (4.1) <sup>1</sup> | 12 (70.6) <sup>2</sup> |
| Fisher's exact test           | 0    | 1    | 1    | 1    | 1    | 4 (1.0)               | 3 (75.0)               |
| Independent t-test            | 13   | 7    | 15   | 9    | 14   | 58 (14.0)             | 43 (74.1)              |
| Paired t-test                 | 3    | 6    | 7    | 5    | 4    | 25 (6.0)              | 18 (72.0)              |
| ANOVA                         | 31   | 43   | 33   | 38   | 27   | 172 (41.5)            | 104 (60.5)             |
| Repeated-measures ANOVA       | 1    | 3    | 1    | 4    | 0    | 9 (2.2)               | 4 (44.4)               |
| Mann-Whitney U test           | 4    | 8    | 5    | 7    | 10   | 34 (8.2)              | 21 (61.8)              |
| Wilcoxon signed rank test     | 1    | 2    | 0    | 5    | 4    | 12 (2.9)              | 8 (66.7)               |
| Kruskal-Wallis test           | 4    | 7    | 3    | 6    | 4    | 24 (5.8)              | 21 (87.5)              |
| Pearson correlation analysis  | 3    | 3    | 6    | 4    | 4    | 20 (4.8)              | 3 (15.0)               |
| Spearman correlation analysis | 1    | 1    | 0    | 0    | 1    | 3 (0.7)               | 2 (66.7)               |
| Linear regression             | 3    | 2    | 2    | 3    | 2    | 12 (2.9)              | 4 (33.3)               |
| Logistic regression           | 0    | 0    | 1    | 1    | 0    | 2 (0.5)               | 1 (50.0)               |
| Survival analysis             | 0    | 1    | 1    | 3    | 1    | 6 (1.4)               | 3 (50.0)               |
| Others                        | 2    | 3    | 5    | 3    | 3    | 16 (3.9)              | 8 (50.0)               |
| Total                         | 67   | 88   | 82   | 98   | 79   | 414 (100.0)           | 255 (61.6)             |

<sup>1</sup>. Used total of each method/Used total of all methods, <sup>2</sup>. Misused total/Used total according to each method.

### 3. 연도별 통계 방법에 따른 분류 조사

추론 통계방법이 사용된 논문들을 대상으로 사용빈도를 조사한 결과 이중 가장 많이 사용한 통계분석 방법은 ANOVA (41.5%)였고, 그 다음은 독립표본 t-test (14.0%), Mann-Whitney test (8.2%)였다. 통계 방법을 연도별로 살펴 보았을 때 어떤 유의한 변화도 관찰할 수 없었으며 ANOVA 방법은 2010년에 들어 사용이 줄어들었으나 통계적으로 유의한 차이는 발견할 수 없었다 ( $\chi^2$ -test,  $P>.05$ ). 최근에 가까운 수록 다양한 분석을 시도하긴 했으나 여전히 ANOVA가 가장 많이 사용됨을 알 수 있었다 (Table 3).

### 4. 통계 방법의 오류 조사 및 적합한 통계 방법 제시

전체적인 통계 방법의 오류는 61.6%였고, 통계방법 별로 오류의 빈도를 살펴보면 독립표본 t-test, paired t-test,  $\chi^2$ -test, Fisher's exact test, kruskal-Wallis test 등에서는 70%가 넘는 오류가 발견되었다. 다음은 각 분석법에 대한 오류의 종류를 분류한 결과이다.

#### 1) $\chi^2$ -test(사용빈도: 17)

오류가 없었던 경우(5, 29.4%), 테이블을 적절치 않게 만들었

거나 결과 해석이 잘못된 경우(4, 23.5%), 순위 변수(나이 40대, 50대, 60대 등, 혹은 중학교/고등학교/대학교 등)인데  $\chi^2$ -test for trend를 하지 않은 경우(2, 11.8%), 표본의 독립을 위배(환자수는 26명, 임플란트 수는 62개)한 경우(3, 17.6%), 2×2 테이블이 아닌 테이블에서 기대빈도수가 5보다 작은데도  $\chi^2$ -test를 사용한 경우(2, 11.8%), 본페로니 보정을 하지 않은 경우(1, 5.9%)등의 오류가 있었다.

#### 2) Fisher's exact test (사용빈도: 4)

오류가 없었던 경우(1, 25.0%), 2×2 테이블 이외에서 Fisher's exact test를 사용한 경우(1, 25.0%), 표본이 독립적인지 않은 경우(2, 50.0%)등이 있었다.

#### 3) 독립표본 t-test (사용빈도: 58)

오류가 없었던 경우(15, 25.9%), two-way ANOVA를 사용하지 않고 one-way ANOVA와 독립표본 t-test를 혼합하여 사용한 경우(15, 25.9%), two-way 또는 three-way ANOVA를 사용하지 않고 one-way ANOVA와 독립표본 t-test를 여러 번 사용한 경우(3, 5.2%), 반복측정인 자료인데 t-test와 ANOVA를 사용한 경우(5, 8.6%), RM ANOVA를 사용하지 않고 대응표본 t-test와 독립표본 t-test를 혼합하여 사용한 경우(4, 6.9%), 변수가 많아 회귀분석이 적합한 경

**Table 4.** Frequency of the misused Independent t-test and suggested suitable techniques

| Error of statistical technique                   | Suitable technique   | Frequency  |
|--|----------------------|------------|
| No Error   |                      | 15 (25.9)  |
| t-test (multiple) + one-way ANOVA                | multi-way ANOVA      | 18 (31.1)  |
| t-test + ANOVA                                   | RM ANOVA             | 5 (8.6)    |
| t-test + paired T-test                           | RM ANOVA             | 4 (6.9)    |
| t-test (multiple)                                | Regression           | 1 (1.7)    |
| t-test   | nonparametric method | 1 (1.7)    |
| Others   |                      | 14 (24.1)  |
| Violation of Independence                        |                      | 6 (10.3)   |
| t-tests as post-hoc tests                        |                      | 4 (6.9)    |
| Incorrect description of statistical terminology |                      | 3 (5.2)    |
| Description of an unused method                  |                      | 1 (1.7)    |
| Total  |                      | 58 (100.0) |

**Table 5.** Frequency of the misused Paired t-test and suggested suitable technique

| Error of statistical technique          | Suitable technique   | Frequency  |
|---|----------------------|------------|
| No Error                                |                      | 7 (28.0)   |
| paired t-test + ind t-test or ANOVA     | RM ANOVA or ANCOVA   | 7 (28.0)   |
| paired t-test (multiple)                | 2-way RM ANOVA       | 2 (8.0)    |
| paired t-test + 1-way ANOVA or K-W test | 2-way or 3-way ANOVA | 2 (8.0)    |
| paired t-test                           | Bland- Altman Plot   | 2 (8.0)    |
| Others                                  |                      | 5 (20.0)   |
| Violation of Independence               |                      | 3 (12.0)   |
| Report of inappropriate results         |                      | 2 (8.0)    |
| Total                                   |                      | 25 (100.0) |

우(1, 1.7%), 두 집단의 표본수의 차이가 많은 경우에 비모수 방법을 사용하지 않았던 경우(1, 1.7%), 표본 또는 변수간이 독립 (한 토끼에 여러 개의 임플란트를 식립 등)이 아닌 경우(6, 10.3%), t-test를 분산 분석의 사후분석으로 이용한 경우(4, 6.9%), 결과를 제시하지 않았거나 통계용어를 잘못 기술(t-test of non-paired test 등)한 경우(3, 5.2%), t-test를 사용하지 않았는데 기술한 경우(1, 1.7%) 등의 오류가 있었다(Table 4).

#### 4) paired t-test (사용빈도: 25)

오류가 없었던 경우(7, 28.0%), RM ANOVA 또는 ANCOVA(사전 사후인 경우)를 사용하지 않고 대응표본 t-test와 독립표본 t-test 또는 ANOVA를 혼용 사용한 경우(7, 28.0%), two-way RM ANOVA를 사용하지 않고 대응표본 t-test를 여러 번 사용한 경우(2, 8.0%), two-way 또는 three-way ANOVA를 사용하지 않고 대응표본 t-test와 one-way ANOVA 또는 Kruskal-Wallis test를 혼용하여 사용한 경우(2, 8.0%), 두 시스템의 정밀성을 비교하는 실험에서는 Bland-Altman plot 등을 이용해야 하는데 그렇지 않은 경우(2, 8.0%), 표본 또는 변수간이 독립이 아닌 경우(3, 12.0%), 적절치 못과 결과를 제시하였거나 사용 목적을 알 수 없는 경우(2, 8.0%) 등의 오류가 발견되었다(Table 5).

#### 5) ANOVA (사용빈도: 172)

오류가 없었던 경우(68, 39.5%), two-way 또는 three-way ANOVA를 사용하지 않고 one-way ANOVA와 독립표본 t-test 또는 Kruskal-Wallis test를 혼용하여 사용한 경우(17, 9.9%)와 one-way ANOVA를 여러 번 사용한 경우(24, 14.0%), 반복측정 자료를 RM ANOVA를 사용하지 않고 one-way 또는 two-way ANOVA를 사용한 경우(14, 8.1%), two-way 또는 three-way RM ANOVA를 적용하지 않고 대응표본 t-test와 ANOVA를 혼용하여 사용한 경우(7, 4.1%), two-way 또는 three-way ANOVA를 사용하지 않고 one-way ANOVA와 two-way ANOVA를 혼용하거나 one-way ANOVA와 대응표본 t-test를 혼용하여 사용한 경우(2, 1.2%), one-way ANOVA 대신 ANCOVA를 사용해야 하는 경우(1, 0.6%), one-way ANOVA 대신 MANOVA를 사용해야 하는 경우(1, 0.6%), 변수가 많아 회귀분석을 해야 하는 경우(1, 0.6%), 주 효과 또는 상호작용에 대한 구체적인 결과표를 제시하지 않았거나 최종 모형에 대한 결과를 제시하지 않은 경우(14, 8.1%), 표본 또는 변수가 독립이 아닌 경우(7, 4.1%), 적절하지 못한 사후분석방법을 적용했거나 다른 분석의 사후분석으로 이용된 경우(7, 4.1%), 유의한 요인에 대하여 다중 비교를 하지 않은 경우(3, 1.7%), 통계용어를 잘못 기술(two-way ANOVA 인데 one-way ANOVA)한 경우(3, 1.7%) 등이 있었다

**Table 6.** Frequency of the misused ANOVA and suggested suitable technique

| Error of statistical technique                         | Suitable technique   | Frequency   |
|--|----------------------|-------------|
| No Error   |                      | 68 (39.5)   |
| 1-way ANOVA + ind t-test (multiple)                    | 2-way(3-way) ANOVA   | 16 (9.3)    |
| 1-way ANOVA + K-W test                                 | 2-way ANOVA          | 1 (0.6)     |
| 1-way ANOVA (multiple)                                 | 2-way or 3-way ANOVA | 24 (14)     |
| 1-way ANOVA + 2-way ANOVA                              | 3-way ANOVA          | 1 (0.6)     |
| 1-way ANOVA + Paired t-test                            | 2-way ANOVA          | 1 (0.6)     |
| Paired t-test + ANOVA                                  | RM ANOVA             | 7 (4.1)     |
| ANOVA  | RM ANOVA             | 14 (8.1)    |
| ANOVA  | ANCOVA               | 1 (0.6)     |
| ANOVA  | MANOVA               | 1 (0.6)     |
| ANOVA(factor $\geq 4$ )                                | Regression           | 1 (0.6)     |
| Others   |                      | 34 (19.8)   |
| No detailed results provided                           |                      | 14 (8.1)    |
| Violation of Independence                              |                      | 7 (4.1)     |
| Inappropriate Post-hoc test                            |                      | 7 (4.1)     |
| No post-hoc tests even with a significant ANOVA result |                      | 3 (1.7)     |
| Incorrect description of statistical terminology       |                      | 3 (1.7)     |
| Total  |                      | 172 (100.0) |

**Table 7.** Frequency of the misused Mann-Whitney test and suggested suitable technique

| Error of statistical technique                    | Proper technique                     | Frequencies |
|---|--------------------------------------|-------------|
| No Error  |                                      | 13 (38.2)   |
| M-W test + K-W test                               | (Nonparametric) ANCOVA or RM ANOVA   | 4 (11.8)    |
| M-W test  | $\chi^2$ -test                       | 1 (2.9)     |
| M-W test + K-W test or M-W test (multiple)        | (Nonparametric) 2-way or 3-way ANOVA | 4 (11.8)    |
| M-W test  | Kappa                                | 1 (2.9)     |
| Others  |                                      | 11 (32.4)   |
| Violation of Independence                         |                                      | 1 (2.9)     |
| No Bonferroni correction for multiple comparisons |                                      | 5 (14.7)    |
| M-W tests as post-hoc tests, following ANOVA      |                                      | 4 (11.8)    |
| Incorrect description of statistical terminology  |                                      | 1 (2.9)     |
| Total   |                                      | 34 (100.0)  |

(Table 6).

## 6) Repeated Measures ANOVA (사용빈도: 9)

오류가 없었던 경우(5, 55.6%), two-way RM ANOVA를 사용해야 하는데 one-way RM ANOVA를 사용한 경우(1, 11.1%), 사후검정을 하지 않은 경우(1, 11.1%), 통계용어를 잘못 기술(ANCOVA 또는 MANOVA로 표기함)한 경우(2, 22.2%) 등이 있었다.

## 7) Mann-Whitney test (사용빈도: 34)

오류가 없었던 경우(13, 38.2%), ANCOVA 또는 two-way RM ANOVA를 사용해야 하는데 Mann-Whitney test와 Kruskal-Wallis test를 혼용하여 사용한 경우(4, 11.8%), 범주형 자료이므로  $\chi^2$ -test를 해야 하는데 하지 않은 경우(1, 2.9%), two-way 또는 three-way ANOVA를 해야 하는데 Mann-Whitney test와 Kruskal-Wallis test를 혼용하거나 Mann-Whitney test를 여러 번 사용한 경우(4, 11.8%), 일치도 검정을 할 때 kappa를 사용하지 않은 경우(1,

2.9%), 표본의 독립이 아닌 경우(1, 2.9%), ANOVA의 사후검정으로 이용된 경우(4, 11.8%), 사후검정에서 본페로니 보정을 하지 않은 경우(5, 14.7%), 통계용어를 잘못 기술(Mann Whitney Rank Sum test라고 기술)한 경우(1, 2.9%) 등이 있었다(Table 7).

## 8) Kruskal-Wallis test (사용빈도: 24)

오류가 없었던 경우(3, 12.5%), two-way 또는 three-way ANOVA를 사용하지 않고 독립표본 t-test 또는 one-way ANOVA 또는 Mann-Whitney test와 Kruskal-Wallis test를 혼용하여 사용한 경우(6, 25.0%), three-way ANOVA를 사용해야 하는데 Kruskal-Wallis test를 여러 번 사용한 경우(1, 4.2%), Kruskal-Wallis test 대신에 Friedman test를 사용해야 하는 경우(2, 8.3%), 집단이 2개인데 Kruskal-Wallis test를 사용한 경우(1, 4.2%), 요인이 많아서 회귀분석을 해야 하는 경우(1, 4.2%), RM ANOVA를 사용해야 하는데 대응표본 t-test와 Kruskal-Wallis test 혼용하여 사용하거나 Kruskal-Wallis test를 여러 번 사용한 경우(2, 8.3%), 구체적 결과를 제시하지 않

**Table 8.** Frequency of the misused Kruskal-Wallis test and suggested suitable technique

| Error of statistical technique                               | Proper technique                     | Frequencies |
|--|--------------------------------------|-------------|
| No Error   |                                      | 3 (12.5)    |
| K-W test + (t-test or ANOVA or M-W test)                     | (Nonparametric) 2-way or 3-way ANOVA | 6 (25.0)    |
| K-W test (multiple)  | (Nonparametric) 3-way ANOVA          | 1 (4.2)     |
| K-W test   | M-W test                             | 1 (4.2)     |
| K-W test   | Friedman test                        | 2 (8.3)     |
| K-W test   | Regression                           | 1 (4.2)     |
| K-W test + paired t-test                                     | (Nonparametric) 2-way RM ANOVA       | 1 (4.2)     |
| K-W test (multiple)  | (Nonparametric) multi-way RM ANOVA   | 1 (4.2)     |
| Others   |                                      | 8 (33.3)    |
| No detail results provided                                   |                                      | 5 (20.8)    |
| No post-hoc tests even with a significant result of K-W test |                                      | 2 (8.3)     |
| Violation of Independence                                    |                                      | 1 (4.2)     |
| Total  |                                      | 24 (100.0)  |

**Table 9.** Frequency of the misused Wilcoxon signed rank test and suggested suitable technique

| Error of statistical technique                      | Proper technique               | Frequencies |
|---|--------------------------------|-------------|
| No Error  |                                | 4 (33.3)    |
| W-S test + M-W test                                 | (Nonparametric) 2-way RM ANOVA | 5 (41.7)    |
| Others  |                                | 3 (25.0)    |
| No Bonferroni correction after multiple comparisons |                                | 1 (8.3)     |
| Heterogeneity within a group                        |                                | 1 (8.3)     |
| No detail results provided                          |                                | 1 (8.3)     |
| Total   |                                | 12 (100.0)  |

았거나( $P$ -value 등) 사용하지 않은 통계방법인데 기술한 경우(5, 20.8%), 유의한 결과에 대하여 사후검정을 하지 않은 경우(2, 8.3%), 표본 또는 변수간이 독립이 아닌 경우(1, 4.2%) 등이 있었다(Table 8).

#### 9) Wilcoxon signed rank test (사용빈도: 12)

오류가 없었던 경우(4, 33.3%), RM ANOVA를 사용해야 하는데 Mann-Whitney test와 Wilcoxon signed rank test를 혼용하여 사용한 경우(5, 41.7%), 본페로니 보정을 하지 않은 경우(1, 8.3%), 군내가 이질적으로 구성 된 경우(1, 8.3%), 구체적인 결과( $P$ -value 등)를 제시하지 않은 경우(1, 8.3%)등이 있었다(Table 9).

#### 10) Pearson correlation analysis (사용빈도: 20)

오류가 없었던 경우(17, 85.0%), 시스템간의 일치도를 검증하는 분석에 잘못 사용된 경우(1, 5.0%), Spearman correlation analysis를 사용해야 하는 경우(1, 5.0%), 상관계수나  $P$ -value를 제시하지 않은 경우(1, 5.0%)등이 있었다.

#### 11) Spearman correlation analysis (사용빈도: 3)

오류가 없었던 경우(1, 33.3%), 통계용어를 잘못 기술한 경우(1, 33.3%), Pearson correlation analysis를 사용해야 하는 경우(1, 33.3%) 등이 있었다.

#### 12) Linear regression (사용빈도: 12)

오류가 없었던 경우(8, 66.7%), 유의하지 않은 변수를 제거하여 최종모형을 제시하지 않은 경우(1, 8.3%), 적절치 못한 표를 제시하거나 적절치 않은 통계량을 기술한 경우(2, 16.7%), 모형 비교 시 수정된 결정계수를 사용하지 않은 경우(1, 8.3) 등이 있었다.

#### 13) Logistic regression (사용빈도: 2)

표본이 독립적이지 않은 경우(1, 50.0%), 유의하지 않은 변수를 제거하여 최종 모형을 제시하지 않은 경우(1, 50.0%) 등이 있었다.

#### 14) Survival analysis (사용빈도: 6)

표본이 독립적이지 않은 경우(4, 66.7%), 적절한 결과를 제시하지 않은 경우(2, 33.3%) 등이 있었다.

#### 15) Others (사용빈도: 16)

오류가 없었던 경우(9, 56.3%), two sample kolmogrov-smimov test에서 구형성 검정으로 적절치 않게 사용된 경우(1, 6.3%), ANCOVA에서 결과의 수치를 잘못 표기한 경우(1, 6.3%), ANCOVA에서 공변량의 유의성을 제시하지 않은 경우(1, 6.3%), 신뢰도 분석에서 Cronbach's alpha의 결과를 잘못 표기한 경우(1, 6.3%), MANO-

VA를 사용하지 않았는데 기술한 경우(2, 12.6%), Mixed effect model에서 랜덤 효과와 고정효과에 대한 설명이 되어있지 않은 경우(1, 6.3%) 등이 있었다.

## 5. 공통 오류

### 1) 표본 수에 관한 오류

연구대상에 해당하는 어떤 논문에서도 표본 수에 대한 산출 근거가 명시된 논문은 단 한 건도 없었고 대부분의 논문에서는 각 그룹 당 매우 작은 표본으로 분석을 하였다. 적절한 표본 수를 정하는 것은 매우 중요한데, 만약 표본수가 매우 크다면 데이터의 분산이 작아져 작은 변동에도 유의한 결과를 나타내는 경향이 있고, 표본수가 매우 작다면 데이터의 분산이 커져 큰 변동에도 유의하지 않은 결과가 나올 수 있다. 즉, 표본수가 작다면 검정력이 낮아져 올바른 결론을 유도할 수 없다. 따라서 실험 데이터에 대한 분석은 검정력과 효과크기를 고려한 표본수를 산출해야 한다.

### 2) 가정 검정에 관한 오류

조사된 논문에서는 각 분석법에 따른 가정에 대한 검정 여부를 명시하지 않았다. 만약 가정에 대한 검토가 이루어져 있지 않고, 가정에 위배되었다면 분석결과를 신뢰할 수 없다. 가정 위배에 따라 통계분석 방법이 달라질 수 있기 때문에 각 분석법에 따른 가정 검정을 하고 이에 대한 사항을 명시하여야 하고 검정 결과에 따라 적합한 분석을 해야 한다.

## 고찰

결과에서 정리한 오류들 중 대표적인 오류에 대한 해결 방안 에 대하여 알아보려고 한다. 이는 공통 오류 및 가장 많은 오류를 가진 분석법 별 해결책을 제시 하고 이러한 오류를 최소화 할 수 있는 통계분석 가이드라인을 부록으로 제시하고자 한다.

## 1. 공통 오류

### 1) 표본 수에 대한 오류

무료 소프트웨어 G-POWER를 사용하여 연구자가 손쉽게 표본 수를 산출 할 수 있다. 적절한 표본 수와 산출근거를 실험 논문에서는 명시하는 것이 연구 결과에 대한 신뢰성을 높이는 방법 중 하나이다.

### 2) 관측된 표본의 독립성

표본이 독립성이 위배된 경우는 대부분 동물실험에서 나타난다. 예를 들어 하나의 토끼다리에 여러 개 또는 다른 종류의 임플란트를 식립하는 경우 표본의 독립성이 위배된다. 이는 각 토끼들의 특성이 다를 수 있기 때문에 실험에 의한 효과인지 토끼들의 차이에 대한 효과 인지 알 수 없는 교란이 일어나

기 때문이다. 이를 방지 할 수 있는 방법은 한 토끼에 하나의 임플란트를 식립하든지 혹은 한 토끼에 여러 개의 임플란트를 식립해야 하는 경우 Correlated Data Analysis를 해야 한다. 이에 대한 분석법은 관련 문헌을 참고하기 바란다.

### 3) 분석법에 대한 가정 검토

분석을 시행할 때 각 분석에 대한 가정 검토는 중요하다. 가정이 위배 될 경우 다른 분석법을 선택해야 하거나 변수를 적절히 변환하여 분석을 해야 한다. 일반적으로 정규성이 위배 될 경우 평균비교에 관한 분석을 비모수 검정을 하거나 표본 수를 늘려 모수적 방법을 써야 한다. 또는 변수 변환을 통하여 데이터를 정규성을 만족하게 하여 분석하면 된다.

일반적으로 분석법 별 검토해야 하는 주요 가정은 다음과 같다. t검정, 분산분석은 집단의 정규성 검정을 시행하고 이를 만족하지 않을 경우 비모수 방법을 사용해야 한다. 회귀분석일 경우 오차의 정규성, 등분산성을 만족해야 하고 이를 만족하지 않으면 변수변환을 통하여 만족하도록 만들어 분석한다. 공분산분석일 경우에는 종속변수와 독립변수의 선형성을 검정해야 하고 선형성을 만족하지 않으면 일반적인 분산분석을 시행한다. 반복측정 분산분석일 경우 구형성 검정을 하고 이를 만족하지 않으면 자유도를 보정을 한 결과(Greenhouse-Geisser, 혹은 Hyun-Felt)로 분석해야 한다. 표본의 독립성이 위배될 경우에는 Generalized Estimating Equation, Mixed effect model 등으로 분석해야 한다.

## 2. 통계 방법 별 오류

### 1) 모수적 분석법

Fig. 1과 같은 대표적인 실험을 예로 모수적 분석법에 대한 오류와 그에 대한 해결책에 대하여 설명하려 한다. Fig. 1A의  $2 \times 4$  실험을 했을 경우 일반적으로 two-way ANOVA를 하지 않고 식립법에 대하여 독립 t-test를 하고 스크류에 대하여 one-way ANOVA를 하는 경우가 많았다. 이는 식립법과 스크류에 대한 상호작용을 무시하고 분석을 하는 경우이다. 이와 비슷하게 one-way ANOVA를 여러 번 하는 경우도 있었다. 이러한 경우에 전체적인 분석을 위하여 two-way ANOVA를 이용하여 상호작용을 확인하고 주효과가 유의하면 사후검정을 시행하고 상호작용의 유의효과를 가지고 최적의 조건을 찾는다. Fig. 1B의 실험에서는 같은 개체가 2번 측정된 경우이다. 즉, 코팅 전과 코팅 후는 독립이 아니다. 이러한 경우에 전체적인 분석을 하지 않고 one-way ANOVA와 대응표본 t-test를 혼용하여 사용하는 경우가 많았다. 또한 대응표본 t-test 대신 독립표본 t-test를 쓰는 경우는 더욱 틀린 방법이다. Fig. 1B의 실험에서는 사전의 효과를 공변량으로 하는 공분산 분석을 시행해야 한다. 공분산 분석을 하더라도 공변량이 무엇인지 명시해야 하고 이에 대한 유의성 또한 기술하여야 한다. 공변량이 유의하지 않은 경우에는 공분산분석은 의미가 없기 때문이다. Fig. 1C의 실험은 요인

**A**

|       | 식립법 A | 식립법 B |
|-------|-------|-------|
| 스크류 A |       |       |
| 스크류 B |       |       |
| 스크류 C |       |       |
| 스크류 D |       |       |

**B**

|       | 코팅 전 | 코팅 후 |
|-------|------|------|
| 스크류 A |      |      |
| 스크류 B |      |      |
| 스크류 C |      |      |
| 스크류 D |      |      |

**C**

|       | 코팅 전  |       | 코팅 후  |       |
|-------|-------|-------|-------|-------|
|       | 식립법 A | 식립법 B | 식립법 A | 식립법 B |
| 스크류 A |       |       |       |       |
| 스크류 B |       |       |       |       |
| 스크류 C |       |       |       |       |
| 스크류 D |       |       |       |       |

Fig. 1. A: 2 × 4 experiment design, B: 2 × 4 dependent experiment design, C: 2 × 2 × 3 experiment design.

이 3개인 실험이다. 이 또한, 상호작용을 무시하고 각 요인에 대하여 one-way ANOVA, 독립표본 t-test, 대응표본 t-test를 혼합하여 사용한 경우가 많았다. 이러한 경우에는 코팅 전을 공변량으로 하는 two-way ANCOVA(요인이 2개인 공분산분석)를 해야 한다. 만약 코팅 여부에 대한 요인이 다른 독립 요인(재료 종류 등)이라면 이는 three-way ANOVA를 사용하여 분석해야 한다. 앞에서 살펴본 세 종류의 실험에서 자료가 반복측정 자료라면 반복측정 분산분석을 해야 한다. 반복측정 분산분석에서 가정 검토에 대한 오류 외에 범하는 오류는 실험에 사용된 요인을 모두 고려하지 않은 경우가 있었다. 이는 모든 요인을 고려하여 분석을 하고 유의한 요인과 그렇지 않은 요인을 판단하여 최종적인 분석을 해야 한다. 추가로 요인이 너무 많은 경우에는 회귀분석을 하여 결과를 확인하는 것이 해석에 용이하다. 이 외에도 one-way ANOVA와 비모수 검정(Mann-Whiney test, Kruskal Wallis test, Wilcoxon Singed Rank test 등)을 혼합하여 사용하는 경우에도 마찬가지로 전체적인 검증을 할 수 있는 분석을 시행하여야 한다.

## 2) 비모수 분석법

비모수 분석법에서 가장 많은 오류를 범하는 경우도 두 가지 분석을 혼합하여 하는 경우이다. 예를 들어 Fig. 1A의 실험의 경우에 Kruskal-Wallis test와 Mann-Whitney test 또는 독립 t-검정을 혼합하는 경우였다. 이는 표본 수를 늘려서 혹은 데이터 변환을 통해 정규성을 만족시켜 two-way ANOVA를 하여 상호작용과 주효과에 대한 분석을 하거나 정규성을 만족하지 못할 경우 R 통계소프트웨어에서 그에 대응하는 비모수 통계방법이 있는지 찾아보는 것도 한 방법이다. 무료 R 통계소프트웨어로 SPSS에서는 돌려볼 수 없는 거의 모든 통계방법의 결과를 얻을 수 있지만 다소 SPSS에 비해 다루기가 복잡한 단점이 있다. Kruskal-Wallis test를 하는데 자료가 반복측정 자료이면 반복측정 분석의 대응하는 비모수적 방법인 Friedman test를 해야 한다. 다른 오류로는 Kruskal-Wallis test 후 사후검정을 제대로 하지 않은 경우가 있다. 예를 들어 사후검정은 모수적인 방법을 사용하거나 Mann-Whitney test로 사후검정을 하고 Bonferroni Correction을 하지 않은 것이다.  $\chi^2$ -test 검정에서는 셀의 기대 빈도가 5이하인 셀이 있을 경우에는 2 × 2 분할표에 한하여 Fisher's exact test를 사용해야 한다.

## 3. 기존 연구와의 비교

박 등<sup>1)</sup>은 2000년부터 2006년까지의 대한치과보철학회지의 게재된 논문들의 통계기법의 사용빈도가 한글판과 영문판 통합하여 73.18%였고, 통계오류는 45.07%였다. 본 연구에서는 2006년부터 2010년 까지 게재된 논문들의 통계기법의 사용빈도는 76.79%로 3.61% 증가하였고, 통계오류는 61.59%이었다. 통계오류의 빈도의 비교는 박 등<sup>1)</sup>은 각 편 당 오류를 나타냈었고, 본 연구에서는 전체 통계사용기법 중 오류를 나타내었기 때문에 논문 간 오류의 차이가 있을 수 있다. Kuo<sup>6)</sup>가 조사한 바로는 해외 유명 의학 저널인 British Medical Journal, The Journal of the American Medical Association, The Lancet, The New England Journal of Medicine에서도 60%의 통계적 오류를 범하고 있다고 보고하였다. 박 등<sup>1)</sup>의 연구에서는 통계분석 기법을 7가지로 분류하여 부적절한 통계기법에 대하여 논한 반면, 본 연구에서는 통계분석 기법을 15가지로 분류하고 각 분석법에서 자주 범할 수 있는 오류를 세분화하여 이에 대한 해결책을 제시 하였다. 이로 인해 통계오류를 범하는 근본적인 문제를 해결 할 수 있을 것으로 기대해 본다.

## 4. 통계 가이드라인

앞서 언급한 공통적인 오류와 분석법에 대한 오류를 줄여주는 방법으로 통계분석 시 본 연구에서 제시한 통계 가이드라인(부록에 첨부)을 활용할 수 있다. 이에 대한 사용법에 대하여 간략히 설명하고자 한다.

본 연구에서 제시한 가이드 라인의 분석법은 크게 검정 변수



(종속 변수)가 연속형인 경우와 범주형인 경우로 나누어져 있다. 먼저 검정변수가 연속형인 경우를 보면, 검정할 집단의 수를 파악하여 질문의 지시대로 다음 질문으로 넘어가면 된다. 집단이 2개인 경우에는 독립표본 *t-test*와 대응표본 *t-test* 중 하나를 선택하여 각 분석법에 대한 가정을 검토하면 된다. 공통적인 가정인 정규성 검정을 시행하고 이를 만족하지 못할 경우 각각의 비모수 검정을 시행하고 가정을 만족할 경우 지시에 따라 검정을 하면 검정 집단이 2개인 경우가 해결된다.

다음으로 검정 집단이 3개 이상인 경우에는 처음의 질문에서 지시한 ANOVA 항목으로 넘어간다. 먼저 요인의 수를 파악하여 해당하는 항목으로 이동한다. 1요인인 경우에 그룹내의 표본이 독립인지 판단을 하고 독립일 경우에는 지시하는 항목으로 이동하고 그렇지 않을 경우에는 다른 방법을 모색해야 할 것이다. 표본이 독립일 경우에는 자료가 반복측정 자료인지 판단을 한다. 대부분 1요인에서 자료가 독립이 아닐 경우에 반복측정 자료인 경우가 많기 때문에 반복측정 항목으로 이동하면 된다. 반복측정이 아니라면 일반적인 분산분석을 하는 항목으로 이동하게 된다. 여기에서 가정 검토를 하게 되고 가정에 위배될 경우 이에 해당하는 항목으로 이동하여 분석을 시행한다. 분석 절차대로 분석을 하여 결과를 도출한다. 반복측정일 경우에는 반복측정 분산분석을 시행하는 항목으로 이동하게 된다. 반복측정 분산분석 항목의 절차는 가정에 대한 검정을 하고 각각의 지시에 따라 결과를 확인하면 된다.

이제 2요인 실험일 경우를 살펴보자. 2요인인 경우에는 대표적으로 세가지 two-way ANOVA, ANCOVA, two-way RM ANOVA로 분류된다. 예시와 비교하여 해당하는 항목을 선택하여 각 항목으로 이동한다. 선택한 항목에서 지시대로 분석하여 결과를 확인한다. 특히 ANCOVA에서 주의해야 할 점은 집단간 기울기가 동일한지 아닌지 판단해서 기울기가 다를 경우에 조정된 관측치로 분석해야 하는 점이다.

이제 요인이 3개 이상인 경우에는 multi-way ANOVA의 항목으로 이동하여 분석을 한다. 분석법은 two-way ANOVA와 유사하다. 하지만 요인이 많은 경우에는 회귀분석을 하는 것이 해석에 용이하다. 연속형 변수인 경우에 추가적으로 회귀분석과 상관분석이 있다. 회귀분석의 대략적인 분석절차를 기술하였으니 이에 따라 분석하여 결과를 얻을 수 있고, 상관분석인 경우에 데이터가 연속형인지 순위변수인지 판단하여 분석하면 된다. 검정변수(종속변수)가 범주형 변수인 경우에는 대표적

인 3가지 방법에 대하여 분석절차에 대하여 안내하였다. 범주형인 경우  $\chi^2$ -test와 Logistic Regression을 이용하여 분석 절차대로 시행하여 결과를 확인하고, GEE인 경우에는 관련 문헌을 참고하여 분석해야 할 것이다.

본 연구에서 제시한 가이드라인을 잘 활용하기 위해서는 각 분석법 별로 약간의 통계적 지식이 요구되며, 통계적 지식이 없는 상태라면 결과 해석에도 어려움이 따를 것으로 사료된다. 회귀분석이나 GEE 등은 관련 문헌을 참고하여 분석하기 바라며, 여기서 소개된 통계 가이드라인은 모든 통계기법에 대하여 안내한 것이 아니라 주로 대한치과 보철학회지에서 사용된 통계방법에 대하여 안내했음을 유념하기 바란다.

## 결론

대부분의 오류는 전체적인 분석 후 세부적인 분석으로 들어가야 하는데 두 요인에 대한 상호작용을 무시하고 단변수 분석을 여러 번 한 경우와 연구 계획 단계에서 적절한 표본 수를 산출하지 않아 나타난 경우 등이었다. 이러한 통계 오류를 최소화하기 위하여 분석 방법 별 검정 절차에 대한 통계분석 가이드라인을 제시하였다.

## 참고문헌

1. Park DG, Choi YG, Kim YS, Shin SW. An assessment of statistical errors of articles in the Journal of Korean Academy of Prosthodontics: Comparison between Korean version and English version. J Korean Acad Prosthodont 2009;47:273-85.
2. Kim YJ, Jang HS, Kim BO. Evaluation of statistical methods in the Journal of Korean Academy of Periodontology published from 2000 to 2006. J Periodont Implant Sci 2009;39:399-405.
3. HJ Lim. A Review of Statistical Methods in the Korean Journal of Orthodontics and the American Journal of Orthodontics and Dentofacial Orthopedics. Korean J Orthod 2004;34:371-9.
4. Kim JS, Kim DK, Hong SJ. Assessment of errors and misused statistics in dental research. Int Dent J 2011;61:163-7.
5. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
6. Kuo YH. Extrapolation of correlation between 2 variables in 4 general medical journals. JAMA 2002;287:2815-7.

# Guideline for Statistical Analysis

## Part I 종속(반응)변수가 연속형 변수인 경우

1. 집단의 수 파악
2. 집단의 수는 몇 개인가?

예)

집단이 2개인 경우

| 스크류 |  |  |  |  |  |
|-----|--|--|--|--|--|
| A   |  |  |  |  |  |
| B   |  |  |  |  |  |

집단이 3개인 경우

| 스크류 |  |  |  |  |  |
|-----|--|--|--|--|--|
| A   |  |  |  |  |  |
| B   |  |  |  |  |  |
| C   |  |  |  |  |  |

집단이 6(3 × 2)개인 경우

|       | 코팅 A | 코팅 B |
|-------|------|------|
| 스크류 A |      |      |
| 스크류 B |      |      |
| 스크류 C |      |      |

- 1) 집단이 2개이면 3번으로 가시오.
  - 2) 집단이 3개 이상이면 6번으로 가시오.
3. 집단 간 독립인가?
    - 1) 독립이면 4번으로 가시오.
    - 2) 독립이 아니면 5번으로 가시오.

### Independent t-test

#### 4. Independent t-test 시행(독립표본 t-test)

##### 4-1. 각 그룹 별 정규성 검정시행

- 1) 정규성을 만족하면 Independent t-test 시행
- 2) 정규성을 만족하지 않으면 비모수 검정인 Mann-Whitney U test 시행

##### 4-2. 등분산 검정에 따른 p-value 선택하여 검정

### Paired t-test

#### 5. Paired t-test (대응표본 t-test)

##### 5-1. 두 집단의 차이값(x-y)에 대한 정규성 검정 시행

- 1) 정규성을 만족하면 Paired t-test 시행.
- 2) 정규성을 만족하지 않으면 비모수 검정인 Wilcoxon Singed Rank test 시행.

#### 6. 요인의 수를 파악한다.

예)

##### 1요인

| 스크류 |  |  |  |  |  |
|-----|--|--|--|--|--|
| A   |  |  |  |  |  |
| B   |  |  |  |  |  |
| C   |  |  |  |  |  |
| D   |  |  |  |  |  |
| E   |  |  |  |  |  |

##### 2요인

|     | 코팅   |      |  |  |
|-----|------|------|--|--|
| 스크류 | 코팅 유 | 코팅 무 |  |  |
| A   |      |      |  |  |
| B   |      |      |  |  |
| C   |      |      |  |  |
| D   |      |      |  |  |

##### 3요인

|     | 처리 A |      | 처리 B |      |
|-----|------|------|------|------|
| 스크류 | 코팅 유 | 코팅 무 | 코팅 유 | 코팅 무 |
| A   |      |      |      |      |
| B   |      |      |      |      |
| C   |      |      |      |      |
| D   |      |      |      |      |

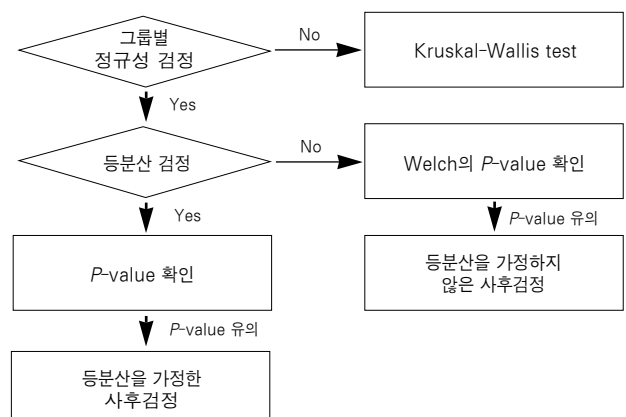
- 1) 1요인 이면 6-1번으로 가시오.
- 2) 2요인 이면 8번으로 가시오.
- 3) 3요인 이상이면 10번으로 가시오.

#### 6-1. 그룹 간 독립인가?

- 1) 각 그룹이 독립이고 반복측정이 아니라면 6-2로 가시오.
- 2) 각 그룹이 독립이 아닐 경우(예) 토끼 한 마리의 다리에 여러 종류의 임플란트 식립) 이거나 반복측정일 경우 (예) 한 개체에서 여러 번 반복측정 된 경우) 7번으로 가시오.

### One-way ANOVA

#### 6-2. One-way ANOVA (일원배치 분산분석)



## Repeated-Measures ANOVA

## 7. Repeated-Measures ANOVA (반복측정 분산분석)

예) 시간의 흐름에 대한 반복측정일 경우

시간: 반복요인

그룹: 처리요인 → 요인이 2개이므로 2-Way RM ANOVA

## 7-1. 차이에 대한 정규성 검정 (x-y, y-z)

1) 정규성을 만족하면 7-2로 가시오.

2) 정규성을 만족하지 못하면 7-5으로 가시오.

## 7-2. 구형성 검정

1) 구형성을 만족하면 7-3으로 가시오.

2) 구형성을 만족하지 않으면 7-4로 가시오.

## 7-3. P-value를 확인하고 유의하면 사후검정

7-4.  $\epsilon > 0.75$ 이면 Huynh-Feldt correction,  $\epsilon < 0.75$ 이면 Greenhouse-Geisser correction P-value를 확인 후 유의하면 사후검정

## 7-5. Friedman test를 시행하고 P-value가 유의하면 Wilcoxon Signed Rank test를 하고 Bonferroni Correction 된 유의 수준을 사용하여 사후검정

## 8.2요인에 대한 분석법 결정

예)

Two-way ANOVA

8-1번으로 가시오

|     | 처리(독립) |      |
|-----|--------|------|
| 스크류 | 처리 A   | 처리 B |
| A   |        |      |
| B   |        |      |
| C   |        |      |
| D   |        |      |

ANCOVA

9번으로 가시오

|     | 처리(독립 X) |    |
|-----|----------|----|
| 스크류 | 사전       | 사후 |
| A   |          |    |
| B   |          |    |
| C   |          |    |
| D   |          |    |

RM ANOVA

7번으로 가시오

|     | 반복측정(동인한 개체) |    |    |    |
|-----|--------------|----|----|----|
| 스크류 | 1주           | 2주 | 3주 | 4주 |
| A   |              |    |    |    |
| B   |              |    |    |    |
| C   |              |    |    |    |
| D   |              |    |    |    |

## Two-way ANOVA

## 8-1. Two-way ANOVA (이원배치분산분석)

\*자주 범하는 오류

전체적인 분석을 하지 않고

One-Way ANOVA + Independent t-test or Paired t-test

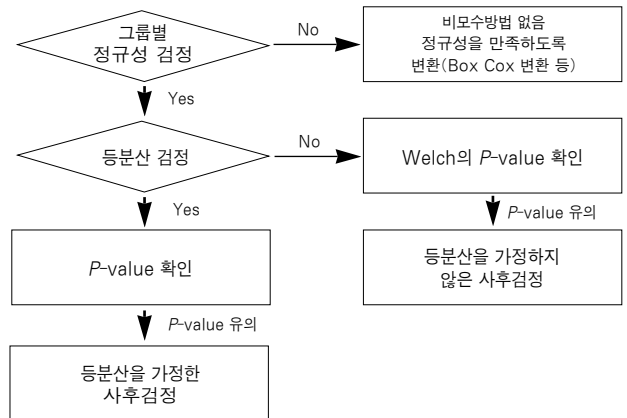
One-Way ANOVA + Mann-Whitney U test or Wilcoxon Signed

Rank test

One-Way ANOVA + One-Way ANOVA

등 혼합하여 사용하는 경우

이러한 경우에는 각 상황에 적합한 전체적인 분석을 해야한다.



## ANCOVA

## 9. ANCOVA (공분산분석)

정규분포와 등분산 가정도 검토한다.

## 9-1. 가정: 종속변수와 공변량은 선형성을 만족한다.

공변량의 유의성을 확인( $P < .05$ )

1) 공변량이 유의하면 9-2로 가시오.

2) 공변량이 유의하지 않으면 ANCOVA를 할 수 없다.  
(다른 방법 모색, RM ANOVA 등)

## 9-2. 집단간 기울기 동질성(Homogeneity of Regression Slopes) 가정 검정

확인 방법: 종속변수와 공변량이 집단에 따라 같은지 통계적으로 검정. 즉, 공변량과 집단간의 교호작용을 검정하여 유의하지 않으면( $P > .05$ ) 동질한 것임.

1) 기울기가 동일하면 9-3으로 가시오.

2) 기울기가 동일하지 않으면 9-4로 가시오.

## 9-3. 교호작용을 제거하고 주 효과의 유의성 확인 유의하면 사후검정

## 9-4. 각 집단에 따라 종속변수와 공변량의 관련성을 각각 추정해야 함.

각 집단의 기울기에 따라 조정된 종속변수로 집단 간 평균 차이가 검정

## Multi-way ANOVA

## 10. Multi-way ANOVA (다원배치 분산분석)

- 1) 반복요인이 있는 경우 7번으로 가시오.
- 2) 공변량이 있는 경우 9번으로 가시오.
- 3) 요인만 있는 경우 8-1번에서 요인만 하나 추가하면 된다.  
\*반복요인과 공변량 둘 다 있는 경우는 7번에서 공변량만 추가하여 분석하면 된다. (자세한 방법은 문헌을 참고)  
\*요인이 많은 경우는 회귀분석(ANOVA, ANCOVA 등), Mixed effect model (RM ANOVA)을 시행한다.

## Regression Analysis

## 11. Regression Analysis (회귀분석)

- 11-1. 종속변수와 독립변수의 산점도를 그려본다.  
선형성을 보이지 않으면 변수변환 또는 비선형 회귀분석
  - 11-2. 종속변수와 독립변수의 상관분석을 하여 종속변수에 유의한 독립변수를 선택.
  - 11-3. 11-2에서 얻은 독립변수들을 가지고 단계적 변수 선택 방법 등을 이용하여 최종 모델을 선택한다.
  - 11-4. 독립변수가 범주형이면 더미변수를 이용한다.
  - 11-5. 오차 가정: 독립성, 정규성, 등분산성
    - (1) 오차의 정규성: PP 도표와 QQ 도표 등으로 판단한다.  
정규성을 만족하지 않을 경우 변수변환 하여 분석
    - (2) 등분산성: 잔차와 예측치의 산점도가 이상이 없는지 확인  
이상이 있다면 변수변환을 하거나 비선형 회귀 또는 새로운 설명변수를 추가하여 분석
  - 11-6. 결정계수를 확인하여 모형의 설명력을 확인한다.
  - 11-7. 회귀계수에 대한 유의성을 확인하고 종속변수에 대한 영향력을 비교한다.  
(독립변수끼리 비교는 표준화된 회귀계수를 이용한다.)
- ps) 회귀분석 시 문헌을 참고하여 분석을 하는 것이 적합한 분석을 할 수 있다.

## Correlation Analysis

## 12. Correlation Analysis (상관분석)

- 1) Pearson 상관분석  
두 변수가 연속형일 경우 사용  
데이터의 분포가 이상치의 영향을 많이 받을 경우에는 이상치를 제거하여 분석하거나  
Spearman 상관분석을 하는 것이 바람직하다.
- 2) Spearman 상관분석  
둘 중 한의 변수라도 순위변수이면 시행 한다.

## Part II 종속(반응)변수가 범주형 변수인 경우

 $\chi^2$  test1.  $\chi^2$  test (카이제곱 검정)

|    | 성공 | 실패 |
|----|----|----|
| 남자 |    |    |
| 여자 |    |    |

범주형 자료로 구성된 분할표에서 변수간의 연관성을 검정할 때 사용

Cell의 기대빈도가 5이하인 셀이 있을 경우

→ Fisher's Exact test 시행(2×2에서만 가능)

검정결과가 유의하면 오즈비로 해석

\*Trend카이제곱검정: 범주가 순서형일 경우에 사용한다.

## Logistic Regression Analysis

## 2. Logistic Regression Analysis

종속변수는 이분형, 순서형, 명목형 일 때 모두 사용가능 하고, 독립변수는 연속형, 범주형일 경우 모두 사용 가능하다.

- 2-1. 각각의 독립변수와 종속변수 간에 Logistic Regression을 해보고  $P < .25$ 인 선택.
- 2-2. 2-1에서 선택된 독립변수로 전진선택법 등으로 Logistic Regression Analysis를 다시 돌려 최종 모델을 구축한다.
- 2-3. Hosmer-Lamshow test로 최종모형의 적합성을 검정한다 ( $P > .05$ )
- 2-4. 오즈비(Odds Ratio) 해석(유의한 변수에 한하여).

## Generalized Estimating Equations

## 3. Generalized Estimating Equations (GEE)

Logistic Regression을 하는 이유는 임플란트 성공/실패에 대해 어떤 요인이 유의한 영향을 주는지 보는 건데 한 사람에게 임플란트를 한 개씩 식립하였을 경우에는 독립성이 위배되지 않아가능하지만, 그렇지 않은 경우에는 독립성이 위배되므로 한 개체에 여러 개의 임플란트를 식립하여 얻어진 clustered data에 대하여 GEE를 사용해야한다.(분석법: 문헌참고)

## Assessment of statistical errors of articles published in the Journal of the Korean Academy of Prosthodontics: 2006 - 2010

Dong-Wan Kang<sup>1</sup>, DDS, MSD, PhD, Yunam Seo<sup>2</sup>, MS, Nam-Sik Oh<sup>3</sup>, DDS, DMS, PhD, Hoi-Jeong Lim<sup>4\*</sup>, PhD

<sup>1</sup>Department of Prosthodontics, School of Dentistry, Chosun University, Gwangju, Korea

<sup>2</sup>Department of Statistics, Chonnam National University, Gwangju, Korea

<sup>3</sup>Department of Dentistry, College of Medicine, Inha University, Incheon, Korea

<sup>4</sup>Department of Orthodontics, School of Dentistry, Chonnam National University, Gwangju, Korea

**Purpose:** Use of inappropriate statistical methods may lead to incorrect conclusions and a waste of valuable resources. The goal of this study was to assess the frequency and the types of several common statistical errors in the published articles of the Journal of the Korean Academy of Prosthodontics (JKAP) for a 5-year period. **Materials and methods:** Of 336 articles in the JKAP published from 2006 to 2010, 255 articles using statistics were reviewed and classified by statistical method and year. The frequency and types of the statistical methods were examined, and the statistical errors were evaluated by the appropriateness of the experimental design, assumption check, independent outcomes, proper sample size and suitable use of statistical method. Statistical guidelines were completed based on the appropriateness. **Results:** Of the 255 articles using statistics, 193 articles (75.9%) used inferential statistics and 153 articles used SPSS statistical software (60.0%). Of the articles using inferential statistics, the three most frequently used statistical methods were ANOVA (41.5%), t-test (20.0%), and the nonparametric method (16.9%). The average rate of statistical errors was 61.2 percent, similar to the rate reported by several studies completed for the medical journal. **Conclusion:** After the whole analysis of the difference among the groups, post-hoc tests for the pairwise comparisons are required. The optimal sample size calculation is an essential part of this study protocol. To minimize the occurrence of statistical errors, statistical guidelines were developed according to each statistical test procedure and will contribute to the academic improvement in the JKAP. (*J Korean Acad Prosthodont* 2012;50:258-70)

**Key words:** Statistical error; Statistical method; Journal of the Korean Academy of Prosthodontics

\*Corresponding Author: Hoi-Jeong Lim

Department of Orthodontics, School of Dentistry, Chonnam National University, 33 Yongbong-ro, Buk-gu, Gwangju, 500-757, Korea

+82 2 62 530 5830; e-mail, hylim@jnu.ac.kr

Article history

Received August 31, 2012 / Last Revision October 4, 2012 / Accepted October 11, 2012