

## 5

# 치의학 연구에서 반복 계측한 자료의 일치도 평가방법

연세대학교 치과대학 BK21 플러스 통합구강생명과학 사업단, 치의학 교육개발실  
김기열

## ABSTRACT

### Statistical methods for accessing agreement between repeated measurements in dental research

BK21 PLUS Project, Institute for Dental Education, Yonsei University College of Dentistry  
Ki-Yeol Kim, Ph.D.

The comparison of the repeated measurements is often needed to see whether they agree sufficiently, when a measurement is repeated under identical conditions by different raters. Such investigations are often analyzed inappropriately, by using correlation coefficient. The purpose of this study is to introduce statistical methods for accessing the agreement of the repeated measurements, which include Bland-Altman plot, intra class correlation, Passing-Bablok regression and Cohen's kappa coefficient, and to show how to execute them using examples.

Key words : agreement; repeated measurements; dental research

Corresponding Author

Ki-Yeol Kim, Ph.D.

BK21 PLUS Project, Institute for Dental Education, Yonsei University College of Dentistry

250 Seongsanno, Seodaemun-gu, Seoul 120-752, Korea

Tel : +82-2-2228-3039, Fax : +82-2-392-2959, E-mail : kky1004@yuhs.ac

본 연구는 연세대학교 치과대학 2016년도 교수연구비에 의하여 이루어졌음(6-2016-0045).

## I. 서론

환자로부터 어떤 값을 계측하는데 있어서 기존에 사용하던 계측기법과 새로운 계측기법을 교체하여 사용해도 될 지, 또는 두 가지 방법을 병용하여 사용해도 될 지에 관하여 알아보기 위해 종종 두 가지 방법의 비교가 필요할 때가 있다. 동일한 계측대상에 대해 여러 계측자(또는 평가자)들이 계측이나 평가한 결과가 일치하는 정도는 일치도(agreement)로 나타낼 수 있다.

일치도란 한 표본을 여러 번 반복하여 계측한 결과가 어느 정도 일치하는가를 알아보는 신뢰도 평가의 척도로, 한 명의 계측자가 동일한 표본을 반복하여 계측, 혹은 여러 명의 계측자가 동일한 표본을 계측할 때 일치하는 정도이다. 이것은 계측한 값들이 얼마나 일관적이고 동일한 결과를 보이는가 하는 개념이며, 임상연구나 실험연구에서 계측오차의 평가는 계측치의 타당성을 확인하기 위하여 필요한 과정이다.

동일한 환자에게서 다른 계측방법으로 계측한 값들이 완전히 일치하는 경우는 거의 없다. 우리는 단지 새로운 기법이 기존에 사용해 오고 있는 기법과 어느 정도 차이가 있는지, 또는 다른 계측자들이 계측한 값이 어느 정도 차이가 있는지 알기를 바란다. 만약 그 차이가 임상적인 해석에 문제가 될 만큼 크지 않다면 기존의 방법을 새로운 방법으로 교체하여 사용하거나 두 가지 방법을 병용하여 사용할 수 있다<sup>1)</sup>.

계측한 자료가 연속적인 값인 경우, 상관계수가 종종 사용되기도 하는데 이것은 옳은 방법이 아니며, 이러한 자료의 경우에는 간단하게 그래픽 방법인 Bland & Altman plot 을 이용하여 일치도를 탐색할 수 있다<sup>1)</sup>. 수치적인 방법으로는 급내상관계수(Intra-class correlation coefficient)와 Passing & Bablok regression 을 사용할 수 있다<sup>2)</sup>. 계측한 자료가 범주형 자료인 경우에는 카파 통계량을 사용할 수 있다. Cohen(1960)이 처음으로 두

명의 평가자 간의 일치도를 제시하여 카파통계량을 정의하였으며<sup>3)</sup>, 그 후 평가자가 여러 명인 경우를 고려한 다양한 카파 통계량이 제안되었다<sup>4)</sup>.

본 연구에서는 예제 데이터와 통계 프로그램을 사용하여 반복 계측된 자료의 일치도를 구하는 방법을 소개한다. 예제 데이터로는 연속형 값으로 계측된 데이터와 범주형 형태로 계측된 데이터 셋을 사용하였으며, 통계 프로그램으로는 연구자들이 주로 사용하고 있는 SPSS 와 최근 치의학 분야에서도 사용자가 증가하고 있는 R 을 이용하였다.

## II. 데이터 셋

계측한 값의 형태에 따라 일치도를 평가하는 방법이 다르므로 다른 형태의 데이터 셋을 사용하여 설명하고 분석하는데 활용하였다.

**[데이터 1]** 3D CBCT를 이용하여 BoNT-A를 한 번 injection한 그룹과 두 번 injection한 그룹에서의 masseter 및 bone volume change를 비교하여 second injection의 효과를 알아보고자 하는 연구를 위해 계측한 자료의 일부이다. 계측치의 재현성을 평가하기 위해 모든 계측치는 두 명의 계측자가 계측하였다(표 1).

**[데이터 2]** 두 명의 방사선 전문의가 85명 환자의 liver lesion을 다음의 네 단계로 평가하였다. 0=normal, 1=benign, 2=suspected, 3=cancer. 두 명이 평가한 조합은 16가지이며 다음과 같이 정리된다(표 2).

## III. 일치도 평가방법

표 1. Injection 전과 injection 후에 두 명의 계측자가 계측한 bone volume

ID	vol_pre	vol_pre2	vol_post	vol_post2
1	2333	2308	2423	2423
2	2575	2503	2597	2337
3	1924	2248	1866	1959
4	4506	4333	4422	4218
5	3661	3757	3665	3681
6	3660	3644	3621	3637
7	2092	2257	2053	2032
8	1809	2052	1758	1935
9	1819	1697	1878	1674
10	2976	2952	3010	2959
11	1996	1990	1811	1898
12	2051	1774	2026	1671
13	3172	3283	3019	3225
14	2208	2086	2205	2063
15	3033	2484	2827	2354
16	6687	6631	6575	6092
17	4531	4222	4507	4196
18	1553	1515	1518	1515
19	3403	3130	3388	3092
20	2667	2512	2592	2495

vol\_pre 와 vol\_pre2 는 injection 전에 두 명의 계측자가 계측한 bone volume 이며, vol\_post 와 vol\_post2 는 injection 후에 두 명의 계측자가 계측한 bone volume 이다.

표 2. 85명 환자의 liver lesion 단계를 두 방사선 의사가 평가한 자료

radiologistA	radiologistB	frequency
normal	normal	21
normal	benign	12
normal	suspected	0
normal	cancer	0
benign	normal	4
benign	benign	17
benign	suspected	1
benign	cancer	0
suspected	normal	3
suspected	benign	9
suspected	suspected	15
suspected	cancer	2
cancer	normal	0
cancer	benign	0
cancer	suspected	0
cancer	cancer	1

### 1. Bland-Altman plot<sup>1)</sup>

두 측정자가 측정한 값들을 그래픽 방법으로 탐색하는 가장 간단한 방법은 산점도를 그려보는 것이다(그림 1).

두 측정자가 측정한 값이 차이가 적을수록,  $Y=X$  직선 근처에 데이터가 몰려있을 것이다. 하지만, 이 그림보다 더 효과적인 방법은 두 측정자가 측정한 값의 차이를 이용하는 것이다. 측정치의 참값은 모르지만, 두 측정자가 측정한 값의 평균이 참값의 좋은 추정치가 된다. 따라서 두 측정치의 평균을 중심으로 차이 값들이 어떻게 분포되어 있는지 확인함으로써 일치도를 평가할 수 있다.

Bland & Altman plot은 두 측정치를 비교하는데 사용되는 그래픽 방법이다. 두 측정자가 측정한 값의 차이가 있는지 또는 추정된 값과 실제 측정한 값의 차이가 있는지에 대해 한눈에 살펴볼 수 있으며, 차이의 평균과 표준편차를 이용하여 그려진다(그림 2). 평균을 중심으로 차이 값들이 모여 있으면, 두 측정자의 측정값은 일치도가 크다고 할 수 있다.

### 2. Intra Class Correlation (ICC, 급내상관계수)<sup>5)</sup>

반복적으로 측정한 측정치의 일치도를 평가하는 적절한 방법으로 ICC가 사용될 수 있다. ICC를 계산하

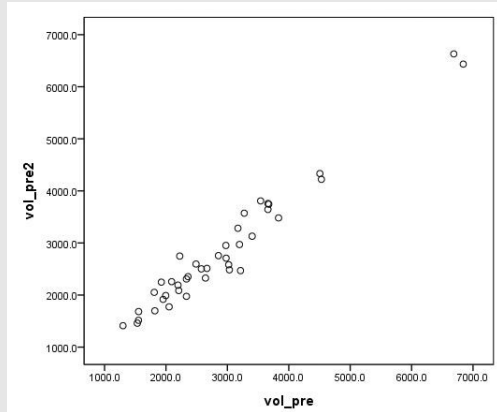


그림1. 두 측정자가 측정한 데이터의 산점도. X축은 첫번째 측정자, Y 축은 두번째 측정자가 측정한 값이다.

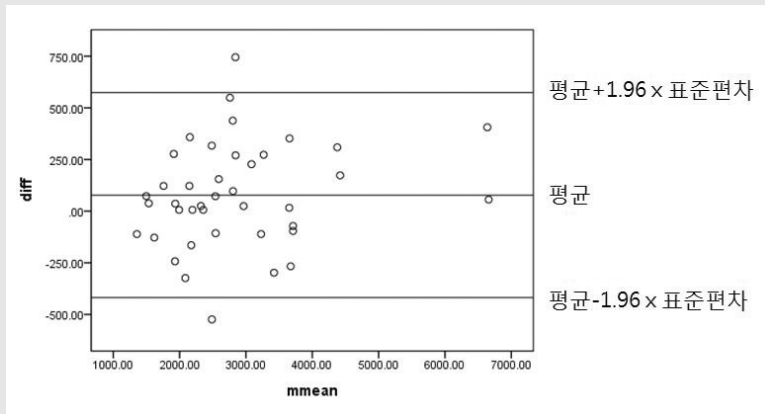


그림 2. 두 측정자가 측정한 값의 평균을 중심으로 차이의 분포를 그린 Bland & Altman plot.

기 위해서 계측한 자료의 변동은 다음 두 종류의 변동을 포함한다.

- 환자 간 변동(between patient variation)
- 환자 내 변동(within patient variation)

‘환자 내 변동’은 한 환자에서 두 번 계측한 값의 차이, ‘환자 간 변동’은 환자들 간 계측한 값의 차이이며, ICC는 이 두 요인의 비율에 의해 계산된다(그림 3).

두 계측자가 측정한 값이 정확하게 일치한다면, 환자 내 변동은 0이 되고 ICC는 1이 된다. Cicchetti (1994)는 ICC의 해석에 대해 표 3과 같이 가이드라인을 제시하였다<sup>6)</sup>.

### 3. Passing-Bablok regression<sup>2)</sup>

두 계측자가 계측한 값을 Y=X 형식의 선형회귀분석(linear regression analysis)을 적용하여 일치

도를 알아볼 수 있다(그림 4). 그러나, 선형회귀분석은 데이터와 오차에 관한 가정을 만족하는 경우에 사용할 수 있다. 데이터에 관한 가정은 두 변수가 선형관계이어야 한다는 것이고, 오차에 대한 가정은 오차값들이 정규분포이며 독립적이고 분산이 같아야 한다는 것이다. Passing & Bablok regression은 데이터의 분포나 오차에 관한 특별한 가정이 없는 비모수적인 선형회귀분석으로 기울기와 절편은 95% 신뢰구간으로 표현되며, 이러한 신뢰구간은 기울기가 1인지, 절편이 0인지를 결정하는데 사용된다.

Passing & Bablok regression은 다음의 가정을 전제로 사용할 수 있다.

- 두 계측자의 측정 값은 강한 양의 상관관계를 갖는다.
- 이 값들의 관계는 선형관계이다 (직선관계).
- 두 변수 각각의 분포에 대한 가정은 없다.

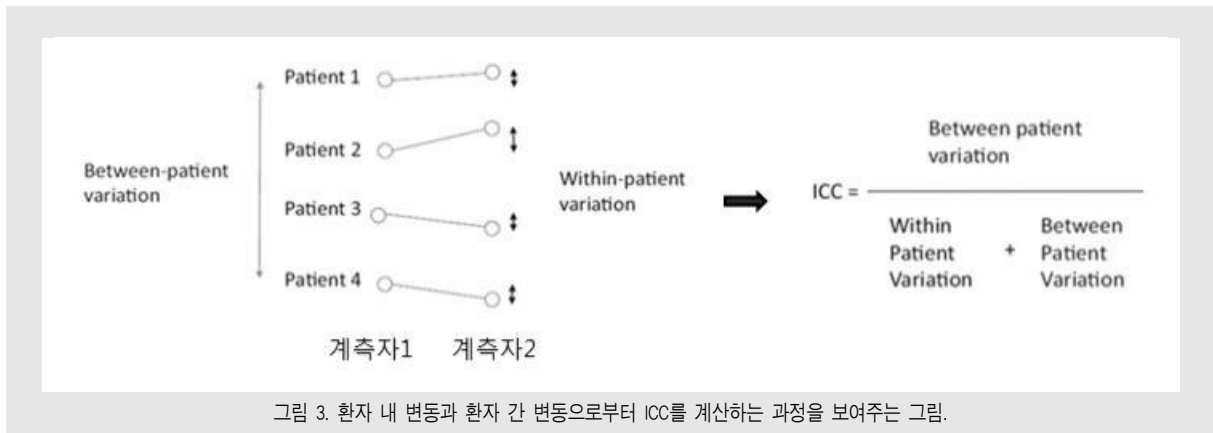


그림 3. 환자 내 변동과 환자 간 변동으로부터 ICC를 계산하는 과정을 보여주는 그림.

표 3. ICC의 크기에 따른 해석 (Cicchetti, 1994)

ICC	interpretation
Less than 0.40	Poor
Between 0.40 and 0.59	Fair
Between 0.60 and 0.74	Good
Between 0.75 and 1.00	Excellent

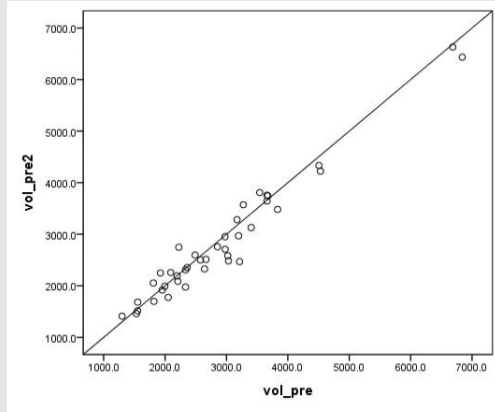


그림 4. 두 가지 방법에 의해 측정된 값들의 관계를 보여주는 그림으로 직선의 기울기는 1이며 45도 기울어져 있고, 절편은 0이다.

이 분석에서 귀무가설은 기울기=1, 절편=0 이며, 이 가설을 받아들여지게 되면 두 계측자의 계측치는 동일하다고 결론 내리게 된다. 이 방법은 두 측정값의 선형성을 전제로 하며, 비선형성인 데이터의 일치도를 평가하는 데는 적절하지 않다<sup>7)</sup>.

#### 4. Cohen's Kappa coefficient(k)<sup>3)</sup>

두 개의 질적 데이터간의 일치도를 평가하는 경우에는 Cohen's kappa (k, 카파통계량) 가 주로 사용된다<sup>3)</sup>. k 는 우연히 발생하는 일치도를 고려하기 때문에 일반적으로 안정적인 척도이다. k 를 계산하는 식은 다음과 같다.

$$k = \frac{\Pr(o) - \Pr(c)}{1 - \Pr(c)}$$

여기서, Pr(o)는 두 계측자가 평가한 값 중 일치한 확률이고, Pr(c)는 일치가 우연히 나타날 확률이다.

예를 들어, 두 사람의 계측자가 두 개의 범주(1 또는 2) 로 판단한 결과를 정리한 결과가 표4와 같다고 가정해 보자.

두 계측자가 평가한 값 중 일치한 확률은 0.7, 우연히 일치한 확률은 0.5이다. 이 값들을 이용하여 계산한 k는 0.4가 된다.

$$Po = \frac{a+d}{N} = \frac{40+30}{100} = 0.7$$

$$Pc = \frac{A_1}{N} \times \frac{B_1}{N} + \frac{A_2}{N} \times \frac{B_2}{N} = \frac{50}{100} \times \frac{60}{100} + \frac{50}{100} \times \frac{40}{100} = 0.5$$

$$k = \frac{Po - Pc}{1 - Pc} = \frac{0.7 - 0.5}{1 - 0.5} = 0.4$$

표 4의 경우를 확대하여 데이터의 형태가 표 5와 같이 복잡해지면 k는 다음과 같이 계산된다.

표 4. 2x2 의 형태로 정리된 데이터 형태(A) 및 예제 데이터(B)<sup>8)</sup>

(A)

	Rater B		Total	
	1	2		
Rater A	1	a	b	A <sub>1</sub>
	2	c	d	A <sub>2</sub>
Total	B <sub>1</sub>	B <sub>2</sub>	N	

(B)

	Rater B		Total	
	1	2		
Rater A	1	40	10	50
	2	20	30	50
Total	60	40	100	

$$\kappa = \frac{Po - Pc}{1 - Pc}$$

$$Po = \frac{1}{N} \sum_{i=1}^q n_{ii}$$

$$Pc = \frac{1}{N^2} \sum_{i=1}^q A_i B_i$$

카파통계량은 k값만을 제시하는 것이 아니라 신뢰 구간(confidence interval)을 함께 제시하는 것이 통계적 추론에 도움을 준다. 카파통계량의 표준오차 (standard error) 와 신뢰구간은 다음과 같이 구할 수 있다. 아래 공식에서  $\delta$ 는 표준오차를 나타낸다.

$$\delta\kappa = \sqrt{\frac{Po(1 - Po)}{N(1 - Pc)^2}}$$

95% Confidence intervals =  $\kappa \pm 1.96 \times \delta\kappa$

99% Confidence intervals =  $\kappa \pm 2.58 \times \delta\kappa$

k 의 크기에 대한 해석은 ICC 와 같은 가이드라인 을 적용한다<sup>6)</sup>.

#### IV. SPSS 와 R을 이용한 예

##### 1. Bland-Altman plot

[SPSS] 에서 실행

대부분의 데이터는 엑셀에 저장하는 경우가 많다. SPSS 에서 엑셀에 저장된 데이터를 불러온 후, 두 번

표 5. q x q 의 형태로 정리된 데이터

		Rater B				Total
		1	2	-	q	
Rater A	1	$n_{11}$	$n_{12}$	-	$n_{1q}$	$A_1$
	2	$n_{21}$	$n_{22}$	-	$n_{2q}$	$A_2$
	-			-		-
	q	$n_{q1}$	$n_{q2}$	-	$n_{qq}$	$A_q$
Total		$B_1$	$B_2$	-	$B_q$	$N$

	ID	vol_pre	vol_pre2	diff	mmean
1	1.0	2333.0	2308.0	25.00	2320.50
2	2.0	2575.0	2503.0	72.00	2539.00
3	3.0	1924.0	2248.0	-324.00	2086.00
4	4.0	4506.0	4333.0	173.00	4419.50
5	5.0	3661.0	3757.0	-96.00	3709.00
6	6.0	3660.0	3644.0	16.00	3652.00
7	7.0	2092.0	2257.0	-165.00	2174.50
8	8.0	1809.0	2052.0	-243.00	1930.50
9	9.0	1819.0	1697.0	122.00	1758.00
10	10.0	2976.0	2952.0	24.00	2964.00
11	11.0	1996.0	1990.0	6.00	1993.00
12	12.0	2051.0	1774.0	277.00	1912.50
13	13.0	3172.0	3283.0	-111.00	3227.50
14	14.0	2208.0	2086.0	122.00	2147.00
15	15.0	3033.0	2484.0	549.00	2758.50

그림5. SPSS 에서 엑셀로 정리된 데이터를 불러온 후, diff 와 mmean 을 계산한 그림

측정한 값의 차이(diff)와 평균(mmean)을 계산한다 (그림 5). 이것은 엑셀에서 미리 계산해 도 좋다.

먼저 두 계측자가 계측한 값(vol\_pre, vol\_pre2)을 산점도로 그려본다.

방법 : 그래프 → 레거시 상자 → 산점도(점도표)

X축 : mmean , Y축 : diff 으로 정하고 실행하면 다음과 같은 산점도가 그려진다(그림 6).

그려진 산점도에 평균과 평균±1.96 x 표준편차 를

표시하면 된다.

그림을 편집하는 단계는 ‘옵션 → Y축 참조 → 평균’ 으로 지정한다(평균=77.55).

그림 7를 실행하면 다음과 같이 평균선이 그려진다 (그림 8).

이 그림에 ‘평균±1.96 x 표준편차’ 를 표시하기 위해서는 차이(diff) 의 표준편차를 알아야 한다. 표준편차는 253.19 로 계산되며, 이 값을 이용하여 Y 축에  $77.55+1.96 \times 253.19$ ,  $77.55-1.96 \times$

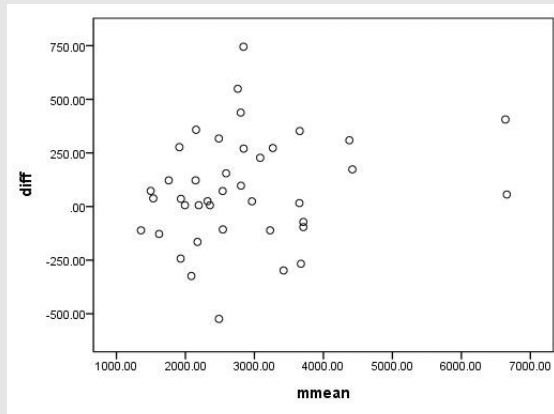


그림 6. 두 계측자가 측정한 값의 차이와 평균을 사용하여 그린 산점도.

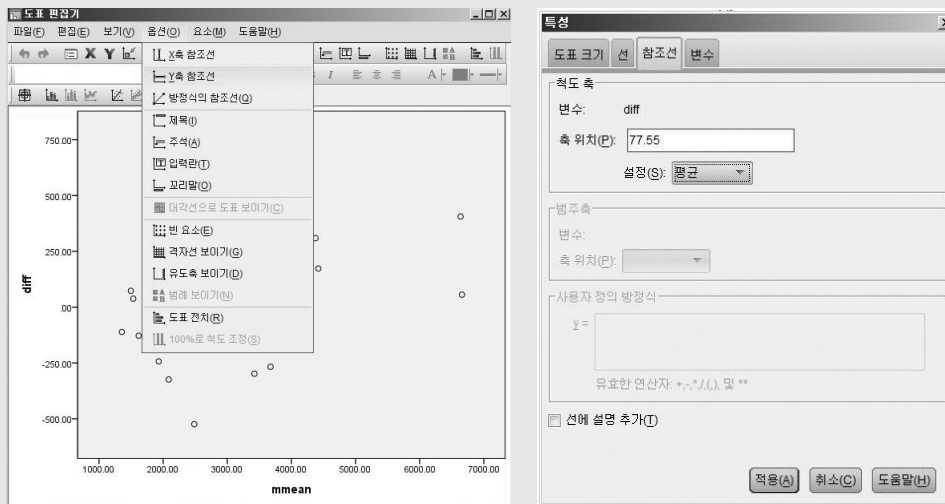


그림 7 산점도에 두 계측치 차이의 평균을 표시하는 방법



253.19 를 추가하면 Bland\_Altman plot 이 완성된다(그림 9).

[R] 에서 실행

R 을 사용하여 Bland\_Altman plot 을 그리기 위해서는 R 을 실행 시키고 데이터를 불러들인 후 다음의 명령문을 실행한다. 데이터는 엑셀이나 SPSS 형태 중 어느 것을 사용해도 된다.

R 을 사용하여 Bland\_Altman plot 을 작성할 때에는 평균과 표준편차를 계산할 필요가 없다(그림 10). 이 값을 계산하는 과정이 Bland\_Altman plot 을 그리기 위한 함수 (bland.altman.plot) 에 포함

되어 있다.

2. ICC(intra class correlation)

[SPSS] 에서 실행

방법 : 분석 → 척도 → 신뢰도 분석

오른쪽 <통계량>을 클릭 한 후, <급내상관계수>를 선택하고 모형을 <이차원변량>으로 바꾼 후, <계속>, <확인> 을 눌러서 분석을 진행한다(그림 11).

결과는 다음과 같이 나타난다(그림 12).

결과는 두 명의 측정자의 측정치는 98.8% 일치한다는 의미이고, 이는 통계적으로 유의하다고 해석한

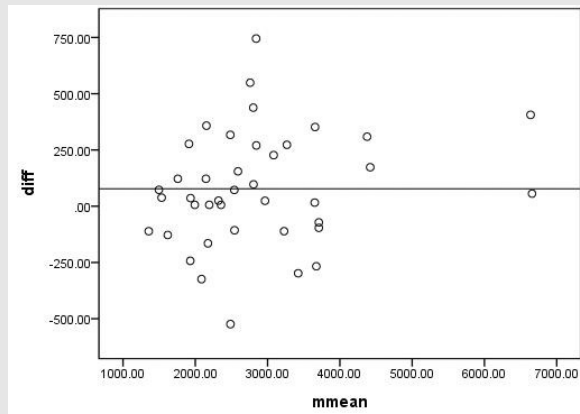


그림 8. 산점도에 차이의 평균을 표시한 그림

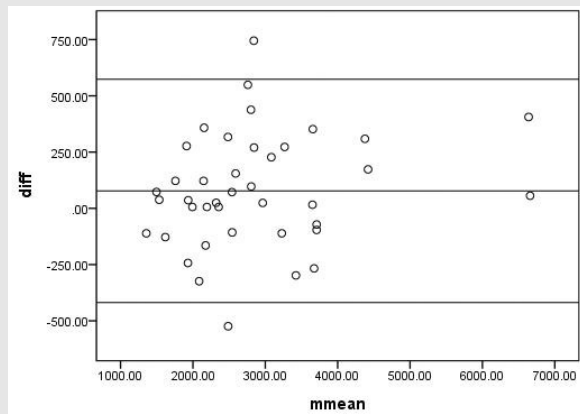


그림 9. Bland\_Altman plot. 산점도에 평균±1.96 x 표준편차를 표시한 그림.

```
install.packages("BlandAltmanLeh")
library(BlandAltmanLeh)
```

```
bland.altman.plot(vol_pre,vol_pre2, two=1.96, xlab="Mean value of vol_pre and vol_pre2",
ylab="vol_pre and vol_pre2")
```

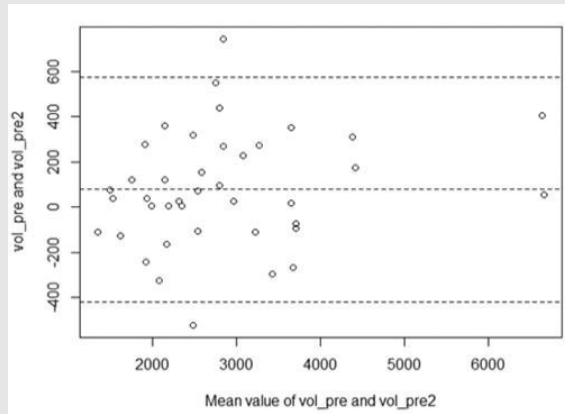


그림 10. R 로 작성한 Bland\_Altman plot

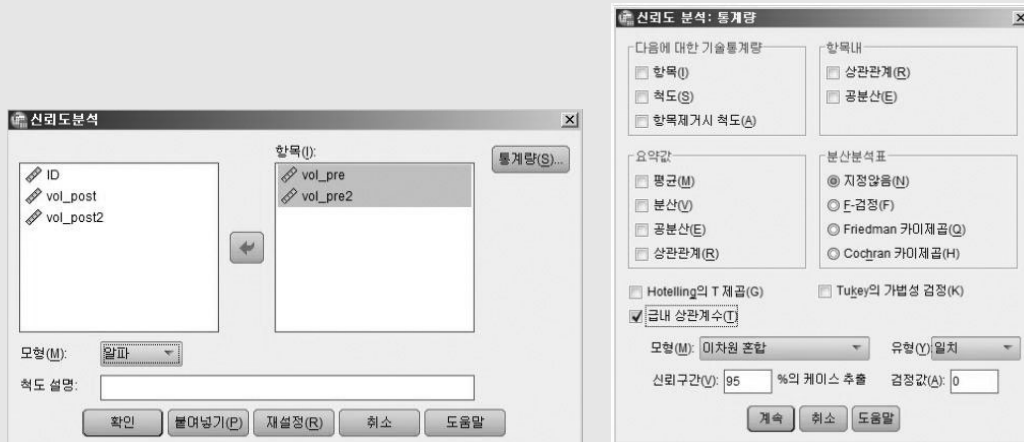


그림 11. SPSS 에서 ICC를 계산하는 과정

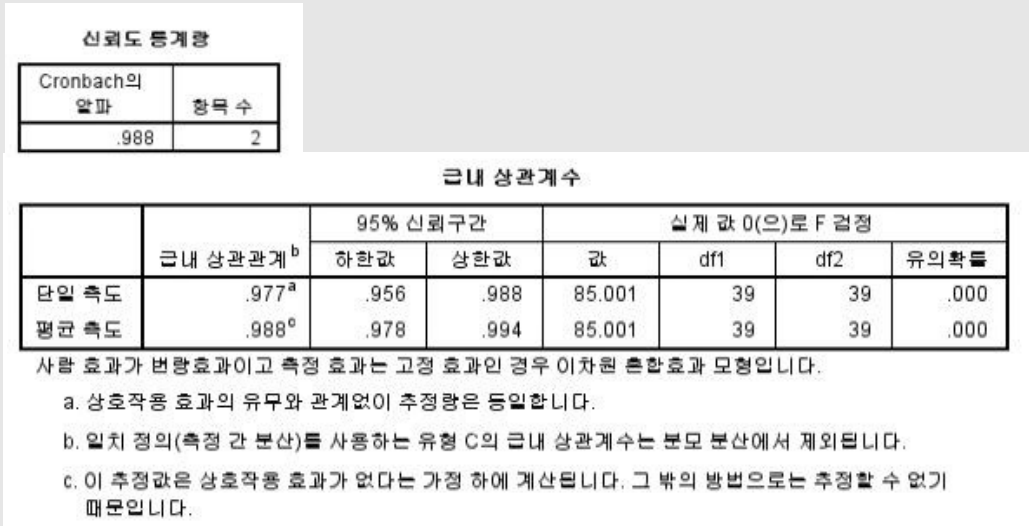


그림 12. SPSS 를 이용하여 ICC 를 계산한 결과

다(유의확률=0.000). 평균측도는 Cronbach's alpha 값과 동일하다.

[R] 에서 실행

R에서는 다음과 같이 ICC 를 계산한다.

R 실행 결과는 다양한 값을 제공하는데, ICC2k 는 two-way random effects (absolute

agreement)를 평가하는 값이므로 이 값을 해석한다 (그림 13).

### 3. Passing-Bablok regression

Passing-Bablok regression 은 기울기는 1, 절편은 0으로 하는 비모수적 회귀분석이지만, SPSS 에

```
install.packages("psych")
library(psych)

ICC(data.frame(vol_pre,vol_pre2))

> ICC(data.frame(vol_pre,vol_pre2))
Call: ICC(x = data.frame(vol_pre, vol_pre2))

Intraclass correlation coefficients
      type ICC  F df1 df2 p lower bound upper bound
Single_raters_absolute ICC1 0.98 80 39 40 0 0.95 0.99
Single_random_raters   ICC2 0.98 85 39 39 0 0.95 0.99
Single_fixed_raters     ICC3 0.98 85 39 39 0 0.96 0.99
Average_raters_absolute ICC1k 0.99 80 39 40 0 0.98 0.99
Average_random_raters  ICC2k 0.99 85 39 39 0 0.98 0.99
Average_fixed_raters   ICC3k 0.99 85 39 39 0 0.98 0.99

Number of subjects = 40  Number of Judges = 2
```

그림 13. R을 이용하여 ICC를 계산한 결과

는 Passing-Bablok regression 이 실행되지 않으므로 모수적인 선형회귀분석을 이용하여 설명하였다(그림 14).

방법 : 분석 → 회귀분석 → 선형

이때 귀무가설은 절편=0, 기울기=1 로 한다. 분석 결과에서 p 값으로 해석하기 보다는 각각에 대한 신뢰구간을 참고하면 된다. 기울기와 절편에 대한 분석결과, 각 추정값의 95% 신뢰구간이 0과 1을 포함하므로(-100.742~312.625, 0.870~1.003) 두 계측자가 측정한 값들의 일치도는 높다고 할 수 있다.

SPSS 에는 Passing-Bablok regression 분석이 포함되어 있지 않으나 우리가 쉽게 사용하는 엑셀에서 분석할 수 있는 방법을 소개한다. 엑셀에 'XLSTAT' 라는 모듈을 추가하면 200여가지의 분석방법을 사용할 수 있으며, XLSTAT 는 한달간 시험용 버전을 다운 받아 사용할 수 있다. 이를 다운받

아서 설치하면 다음과 같은 화면이 나타나고 엑셀에 새로운 메뉴 'XLSTAT' 이 추가된 것이 보인다(그림 15).

데이터가 있는 엑셀시트에서 다음과 같이 실행하면 된다.

방법 : Advanced features → Method validation → Passing and Bablok regression

분석결과, 추정된 절편과 기울기의 95% 신뢰구간이 0과 1을 포함하므로(-144.599~266.714, 0.874~1.038) 두 계측치의 일치도는 높다고 할 수 있다(그림 16).

엑셀에서 XLSTAT 을 사용한 결과가 SPSS 의 결과와 다르게 나타나는 이유는 SPSS 는 일반선형회귀 분석을 적용하였기 때문이다.

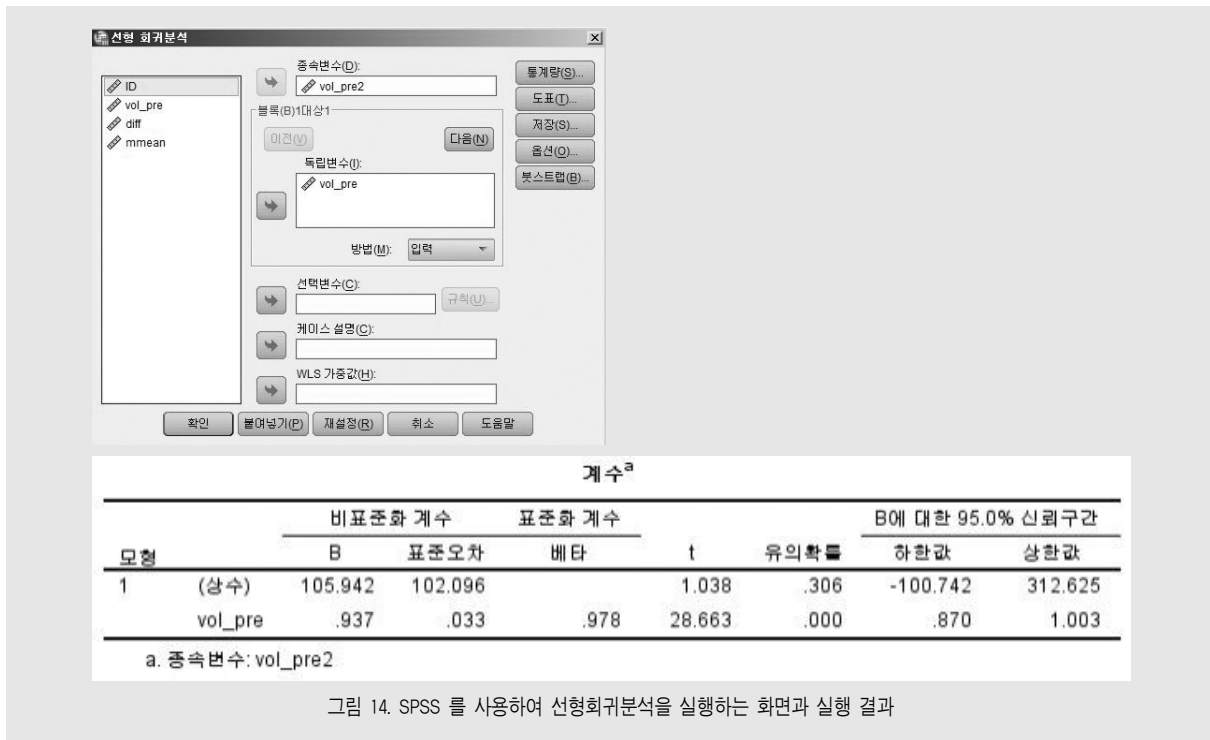


그림 14. SPSS 를 사용하여 선형회귀분석을 실행하는 화면과 실행 결과

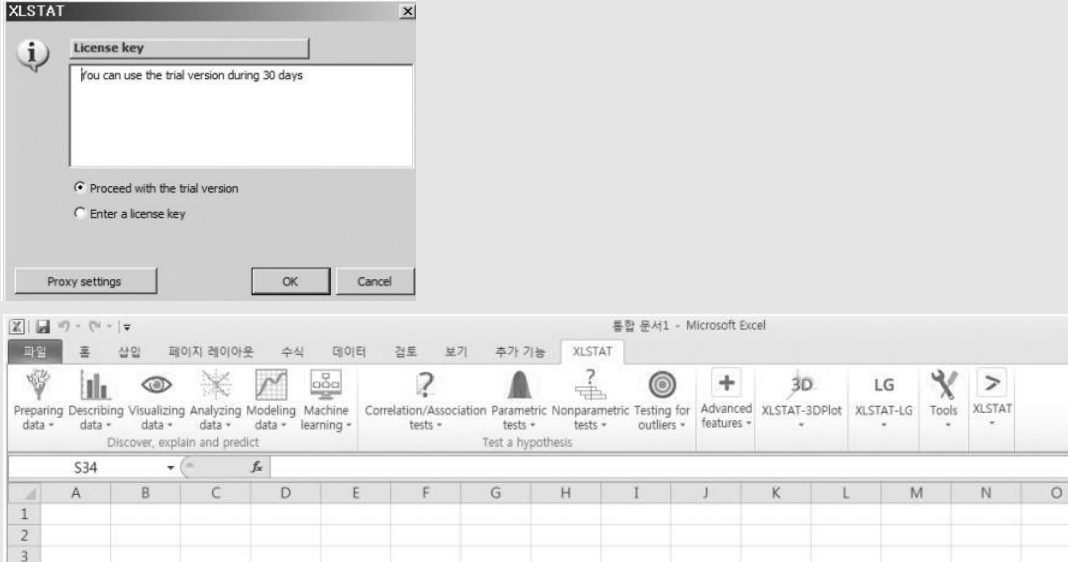


그림 15. 엑셀에 LXSTAT 을 설치한 화면

Model coefficients:			
	Value	Lower bound 95% (Mean)	Upper bound 95% (Mean)
Intercept	43.856	-144.599	266.714
Slope coef	0.957	0.874	1.038

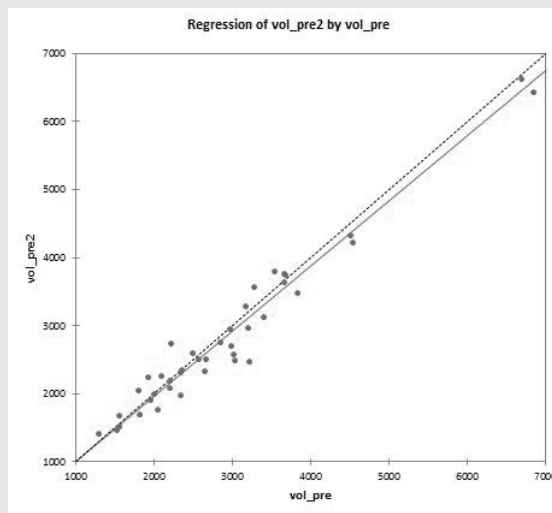


그림 16. 엑셀에서 Passing and Bablok regression 을 실행한 결과

[R] 에서 실행

R에서는 다음의 명령어를 실행하면 두 계측자가 계측한 값의 관계를 산점도로 표현하고 회귀직선이 그려진다. 회귀직선 식으로부터 절편과 회귀계수 추정치의 95% 신뢰구간이 XLSTAT 을 사용한 경우와 동일함을 알 수 있다(그림 17).

4. Cohen's Kappa coefficient(k)

[SPSS] 에서 실행

[데이터 2] 를 사용하여 Cohen's kappa 를 계산하기 위해서, 데이터 셋에서 frequency 는 빈도수를 나타내므로 이 변수를 가중케이스로 지정해 주어야 한다(그림 18).

방법 : 데이터 → 가중케이스 에서 'frequency' 를 빈도변수로 지정해준다.

방법 : 분석 → 기술통계량 → 교차분석 통계량 옵션에서 '카파' 를 선택하고 실행한다(그림 19).

```
install.packages("MethComp")
library(MethComp)

fit=PBreg(data.frame(vol_pre,vol_pre2), conf.level=0.05)
plot(fit, s=1)
```

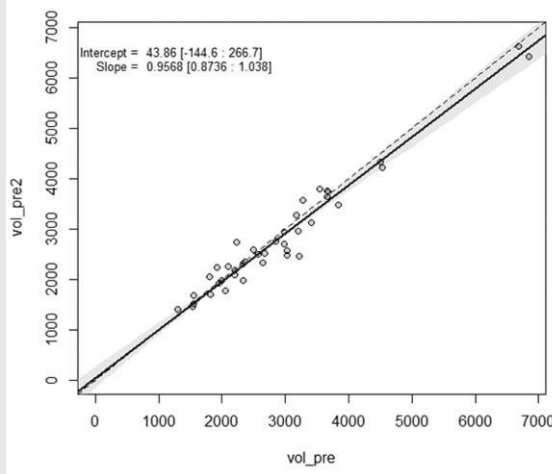


그림 17. 엑셀에서 Passing and Bablok regression 을 실행한 결과

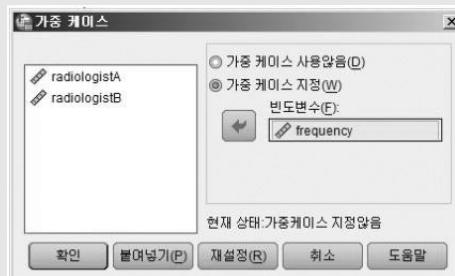


그림 18. 데이터에서 frequency를 가중케이스로 지정하는 방법



그림 19. frequency를 가중케이스로 지정한 후 교차분석을 실행하는 방법

SPSS 실행 결과이다. 두 방사선 학자의 판단결과를 정리한 표가 나타난다(그림 20).

카파를 계산한 결과이다(그림 21). 카파는 -1 에서 1의 값을 가지며<sup>9)</sup>, Cicchetti et al.(1994)의 기준에 의하면 0.473의 카파는 fair agreement로 해

석될 수 있다<sup>6)</sup>.

[R]에서 실행

R에서 다음의 명령을 실행하면 카파가 계산된다(그림 22).

**radiologistA ^ radiologistB 교차표**

빈도		radiologistB				전체
		normal	benign	suspected	cancer	
radiologistA	normal	21	12	0	0	33
	benign	4	17	1	0	22
	suspected	3	9	15	2	29
	cancer	0	0	0	1	1
전체		28	38	16	3	85

그림 20. SPSS 실행결과로 보여주는 데이터 요약

**대칭적 측도**

		값	점근 표준오차 <sup>a</sup>	근사 T 값 <sup>b</sup>	근사 유의확률
일치 측도	카파	.473	.073	6.815	.000
유효 케이스 수		85			

a. 영가설을 가정하지 않음.  
b. 영가설을 가정하는 점근 표준오차 사용

그림 21. 분석 결과를 보여주는 그림. 분석결과 카파는 0.473이다.

```
install.packages("fmsb")
library(fmsb)

KappaData=matrix(frequency, ncol=4)
Kappa.test(KappaData)

> Kappa.test(KappaData)
$Result

      Estimate Cohen's kappa statistics and test the null hypothesis that
      the extent of agreement is same as random (kappa=0)

data:  KappaData
Z = 6.53, p-value = 3.288e-11
95 percent confidence interval:
 0.3248650 0.6207132
sample estimates:
[1] 0.4727891

$Judgement
[1] "Moderate agreement"
```

그림 22. R 을 이용한 분석결과. 카파는 0.472, 95% 신뢰구간은 0.324~0.621 이다.

R 을 이용한 분석결과 계산된 카파에 대하여 'Moderate agreement' 로 판단하였으며, 이는 Landis et al.(1977) 의 기준에 의한 것이다<sup>10)</sup>. Cicchetti et al.(1994) 는 4 단계로 agreement 정도를 분류하였으며, Landis et al.(1977) 은 6 단계로 분류하였고 R 에서는 Landis et al.(1997) 을 참고하여 판단하였다<sup>6, 10)</sup>.

## V. 결론

계측치의 오류에 대한 검증은 임상연구나 실험연구에서 중요하다. 환자의 구강상태를 검진하는데, 데이터의 타당성을 보장하기 위해서는 한 사람의 계측자가 반복적으로 계측하거나 혹은 다른 두 계측자가 계측을 할 때, 계측한 총치의 개수나 치주낭 깊이(periodontal pocket depth) 등은 비슷해야 한다. 실험연구에서도 계측기계의 작은 계측오차를 확인하는 것은

연구의 중요한 과정을 시작하기 전에 선행되어야 하는 것이다.

최소한 두 명 이상의 연구자가 각각 독립적으로 데이터를 계측 또는 평가한 후 연구자들 사이에 얼마나 일치하는지 확인하기 위해서는 데이터의 형태에 따라 ICC 나 Cohen's 카파통계량을 사용하는 것이 일반적이며, 일치도를 평가할 때에는 p 값은 사용하지 않고 두 계측치 차이에 대한 95% 신뢰구간을 사용하는 게 좋다<sup>11)</sup>.

연구의 질은 데이터의 질에 따라 크게 좌우된다. 따라서, 연구자는 연구의 시작단계에서 계측한 값 또는 계측방법의 신뢰도를 확인하는 것이 좋다. 평가자의 역할이 중요한 계측방법에 대해서는 평가자 간의 신뢰도도 고려하여야 한다. 즉, 같은 대상을 다른 평가자가 평가하였을 때 어느 정도의 차이가 있고, 어느 정도 일치하였는지를 확인하여야 한다. 또한, 동일한 평가자가 계속 평가하지 않고 평가 대상에 따라 평가자가 다르다면, 연구결과에 심각한 뻘뻘림이 발생할 것이



다. 그러나, 간혹 대기관 연구에서는 여러 명의 평가자가 평가해야 하는 경우가 있다. 이런 경우에 1) 사전 훈련으로 평가자간 일치도를 향상시키고, 2) 같은 대

상자에 대한 전/후 측정을 할 때에는 동일한 평가자가 평가하도록 권고하고 있다<sup>2)</sup>.

## 참 고 문 헌

1. Bland JM AD. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1(8476): 307-10.
2. Passing H, Bablok. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *Journal of clinical chemistry and clinical biochemistry Zeitschrift fur klinische Chemie und klinische Biochemie* 1983; 21(11): 709-20.
3. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; 20(1): 37-46.
4. Fleiss JL, Gurland BJ, Cooper JE. Some contributions to the measurement of psychopathology. *The British journal of psychiatry : the journal of mental science* 1971; 119(553): 647-56.
5. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychological reports* 1966; 19(1): 3-11.
6. Cicchetti DV, Nelson LD. Re-examining threats to the reliability and validity of putative brain-behavior relationships: new guidelines for assessing the effect of patients lost to follow-up. *Journal of clinical and experimental neuropsychology* 1994; 16(3): 339-43.
7. Bilic-Zulle L. Comparison of methods: Passing and Bablok regression. *Biochemia medica* 2011; 21(1): 49-52.
8. Chang Un Park HJK. Measurement of Inter-Rater Reliability in Systematic Review. *Hanyang Med Rev* 2015; 35: 44-9.
9. Altman DG. *Practical statistics for medical research*. New York, NY: Chapman & Hall/CRC Press; 1999.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1): 159-74.
11. Shintani A. *Primer of statistics in dental research: Part II*. *Journal of prosthodontic research* 2014; 58(2): 85-91.
12. Pocock S. *Clinical Trials: A Practical Approach*. Chichester, UK: John Wiley and Sons; 1983.