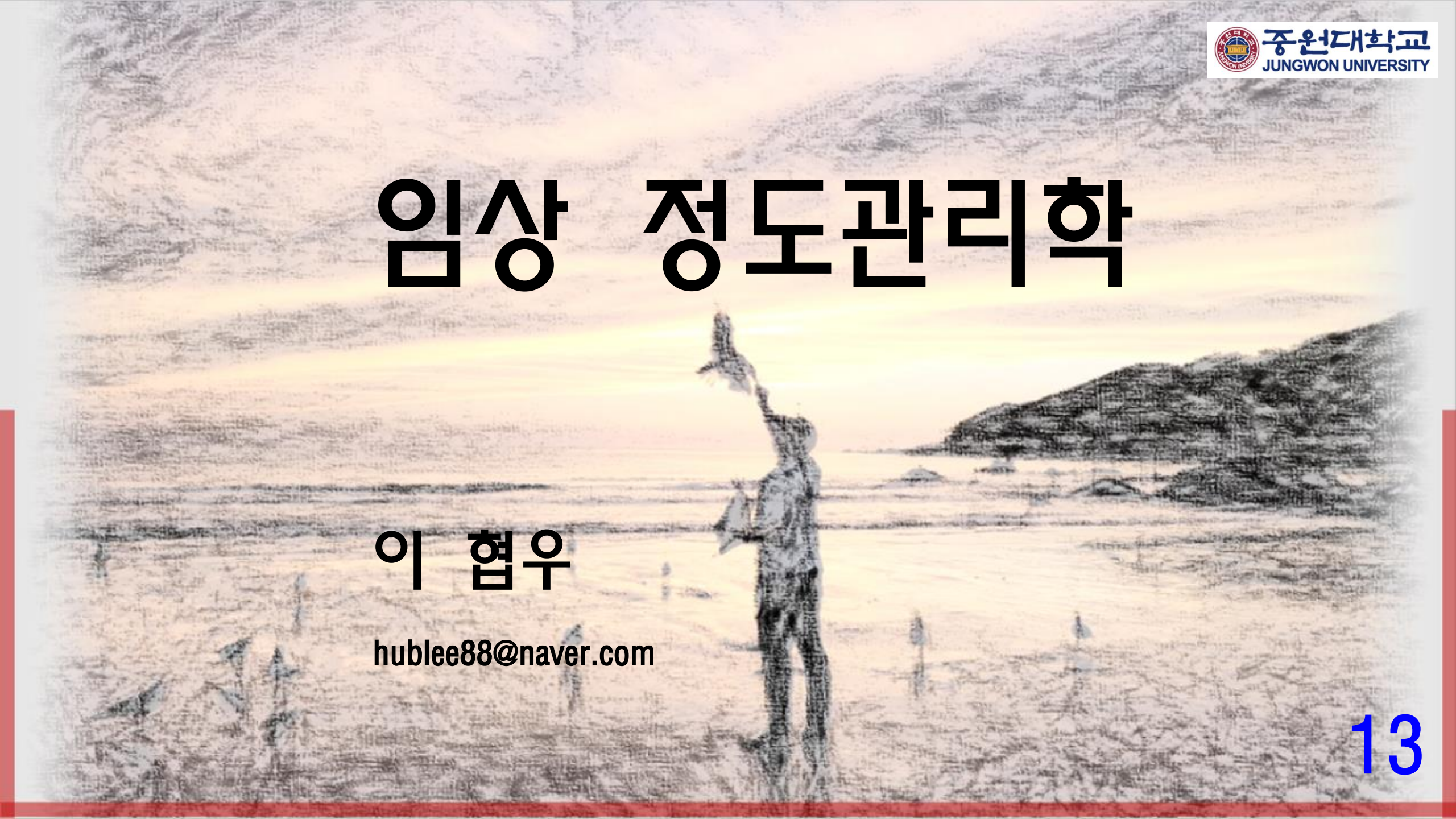


임상 정도관리학

이 협우

hublee88@naver.com

검사평가를 위한 통계 ☞ 상관성 분석

중원대학교 JUNGWON UNIVERSITY

● 상관성 분석

- 두 변수가 함께 변화하는 경향을 분석하는 것
- 상관계수를 분석하면 상관관계의 정도와 방향을 분석 가능
- 상관계수(correlation coefficient, r)는 두 변수 (x, y)의 관계를 알아보기 위해 n개의 데이터쌍 (x1, y1) ... (xn, yn)을 측정하여 두 변수의 산점도를 plot한 후 두 변수의 관련성을 나타내는 척도로 이용
- 상관계수(correlation coefficient, r)

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2]\,[n\sum y^2 - (\sum y)^2]}}$$

$-1 \le r \le +1$

- 진단검사에서 상관분석은 표준 또는 기준방법과 40개 이상의 검체를 비교

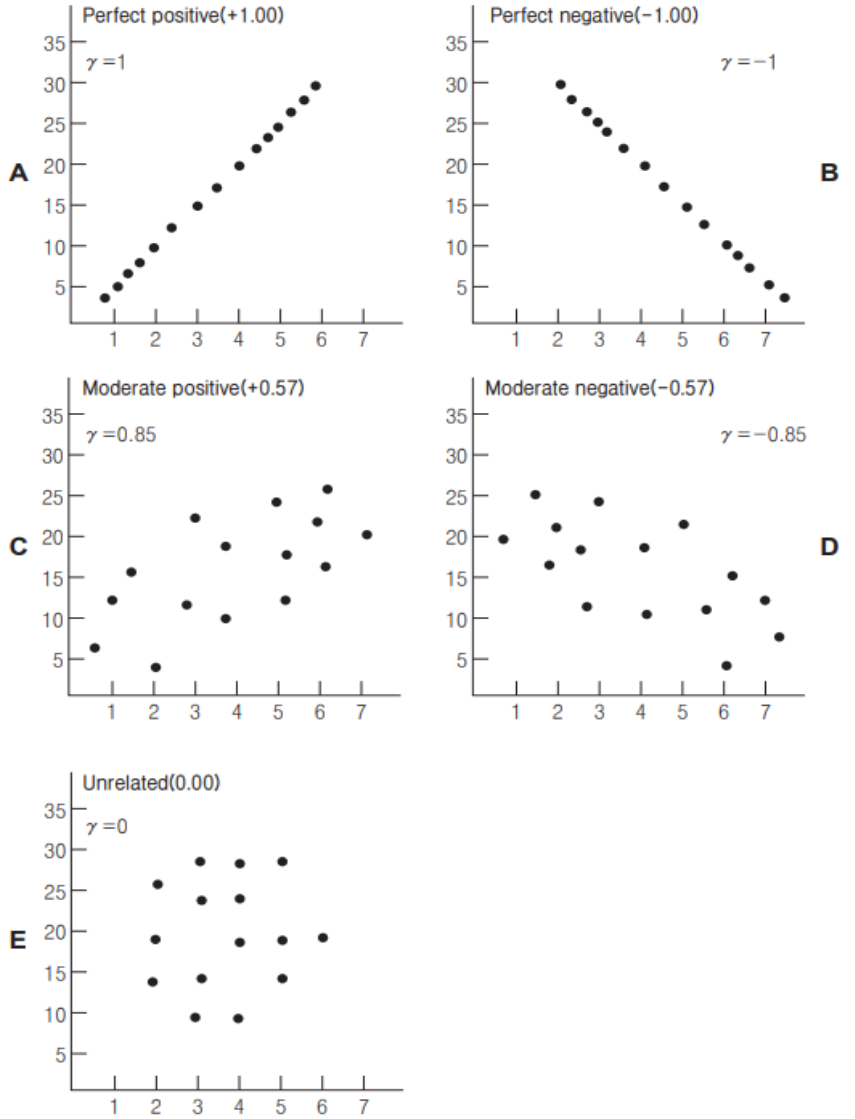


그림 12-6 상관계수 그래프와 상관계수 r값의 의미. r 값이 1이라는 것은 두 변수 간에 차이가 없다는 것을 의미하며, r 값이 0이라는 것은 두 변수 간에 상관성이 없음을 의미한다.

검사평가를 위한 통계 ☞ 상관성 분석

중원대학교 JUNGWON UNIVERSITY

● 상관성 분석

- 두 변수가 함께 변화하는 경향을 분석하는 것
- 상관계수를 분석하면 상관관계의 정도와 방향을 분석 가능
- 상관계수(correlation coefficient, r)는 두 변수 (x, y)의 관계를 알아보기 위해 n개의 데이터쌍 (x1, y1) ... (xn, yn)을 측정하여 두 변수의 산점도를 plot한 후 두 변수의 관련성을 나타내는 척도로 이용
- 상관계수(correlation coefficient, r)

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2]\,[n\sum y^2 - (\sum y)^2]}}$$

$-1 \le r \le +1$

- 진단검사에서 상관분석은 표준 또는 기준방법과 40개 이상의 검체를 비교

상관계수 계산 예

검체번호(n)	표준물질(x_i)	상용물질(y_i)	x_i^2	y_i^2	x_iy_i
1	50	61	2500	3721	3050
2	55	58	3025	3364	3190
3	60	59	3600	3481	3540
4	65	71	4225	5041	4615
5	70	80	4900	6400	5600
6	75	76	5625	5776	5700
7	80	90	6400	8100	7200
8	85	106	7225	11236	9010
9	90	98	8100	9604	8820
10	95	100	9025	10000	9500
11	100	114	10000	12996	11400
Sum	(Σx_i) 825	(Σy_i) 913	(Σx_i^2) 64625	(Σy_i^2) 79719	(Σx_iy_i) 71625

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2\,][n\sum y^2 - (\sum y)^2\,]}}$$

$$r = \frac{11(71625) - (825)(913)}{\sqrt{\sum [11(64625) - (825)^2][11(79719) - (913)^2]}}$$

$r = 0.96$ (포도당 상용물질을 분석하여 표준물질과 비교한 결과 상관성은 비교적 양호한 것으로 분석됨)

검사평가를 위한 통계 ☞ 회귀 분석

중원대학교 JUNGWON UNIVERSITY

● 회귀분석(regression analysis)

- 관찰된 연속형 변수들에 대해 독립변수와 종속변수 사이의 상관관계를 나타내는 선형 관계식을 구하는 기법
- 모형의 적합도를 측정하는 분석 방법
- 하나 이상의 예측 변수와 반응 변수 사이의 통계적인 관계를 설명
- 새 관측치를 예측하는 분석법으로도 유용하게 사용
- 진단검사에서는 표준방법과의 비교테스트, 또는 두 가지 검사법 간 혹은 시약 간의 측정값 차이의 경향과 오차를 비교할 때 사용
- 회귀분석을 실시하면 상승오차(proportional error)와 상가오차(constant error)를 파악할 수 있다(그림 12-7)
- 회귀방정식은 y=bx+a의 1차방정식으로 표현할 수 있는데 b는 상승오차를 a는 상가오차를 의미
- b는 slope(경사, 구배)값이고 a는 y절편(intercept)값 이다.

Y= bx + a (회귀방정식에서 b와 a 구하기)

$$b = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$a = \frac{\Sigma y - b\Sigma x}{n}$$

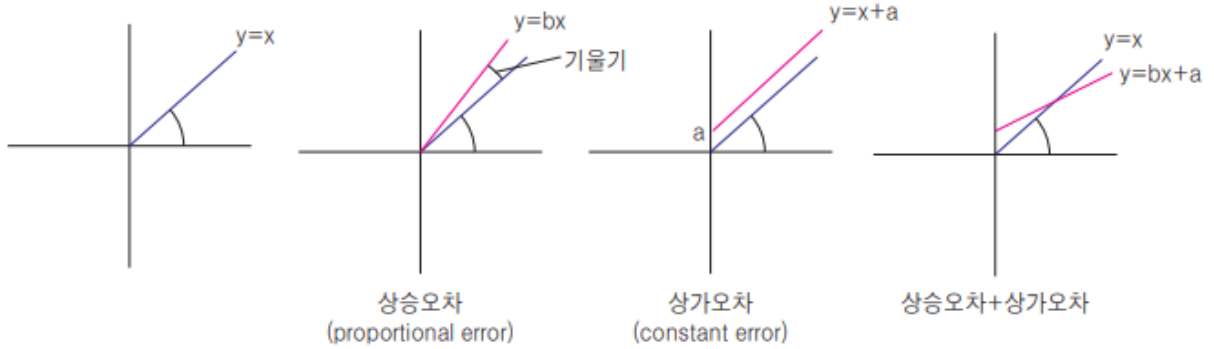


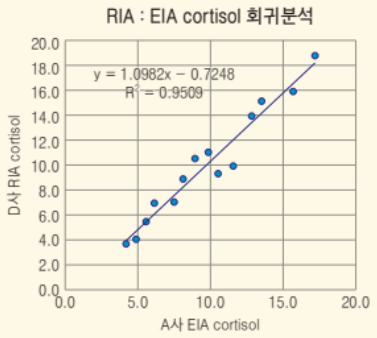
그림 12-7 상승오차와 상가오차의 비교.

검사평가를 위한 통계 ☞ 회귀 분석

중원대학교 JUNGWON UNIVERSITY

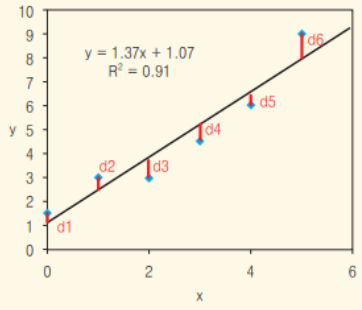
● 회귀분석(regression analysis)

검사방법 간 결과 차이 분석을 위한 회귀분석의 예



RIA로 분석하던 cortisol을 EIA법으로 바꾸어 분석을 할 때 새로 도입된 EIA법으로 측정한 결과가 과거 RIA법으로 측정하였을 때 예측되는 값이 얼마인지 알기 위하여 회귀분석을 실시한 결과이다. 즉, EIA 측정결과 x를 RIA측정결과 y로 변환하면 얼마인지를 계산하기 위해서 slope와 intercept를 구한 회귀방정식이 y=1.0982x + 0.7248이다.

회귀직선(최소자승직선) 그리기



산점도, 최소자승직선, 수직편차

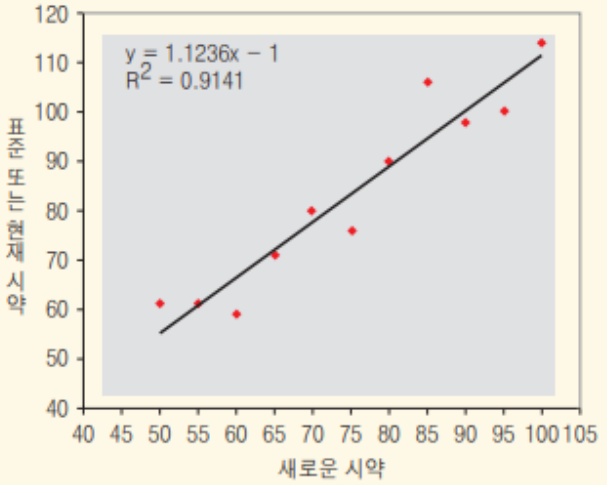
회귀직선 또는 최소자승직선이란?

회귀분석을 통하여 얻은 x와 y 두 변수의 관계를 나타내는 선을 최적합선이라 한다. 최소자승직선은 그래프에 나타난 모든 관찰점을 통과하지 않는다. 바꾸어 말하면 최소자승직선(least-squaros line)으로부터 다소간 편차가 있음을 알 수 있다.

관찰점들을 가로질러 그은 직선이 최적이란 다음과 같은 경우를 의미한다. '최소자승직선에서 관찰지점(yi)에 이르는 수직편차의 자승합이 다른 어떤 직선에서 그은 그 관찰치에 이르는 수직편차의 자승합보다 작다'. 즉, 각 관찰점(yi)로부터 최소자승직선에 이루는 수직편차의 자승을 구하여 그 합을 모두 더하면, 산점도 내에서 그을 수 있는 다른 어떤 직선에서의 수직편차의 자승의 합보다 작다는 것이다. 이런 이유로 그 직선을 최소자승직선이라 부른다.

회귀직선(최소자승직선) 그리기

최소자승직선 그리기 예



	새로운 시약(x)	표준 또는 현재 시약(y)
1	50	61
2	55	61
3	60	59
4	65	71
5	70	80
6	75	76
7	80	90
8	85	106
9	90	98
10	95	100
11	100	114

새로 도입예정인 시약과 표준 또는 현재 사용 중인 시 약간의 회귀분석 결과와 그래프

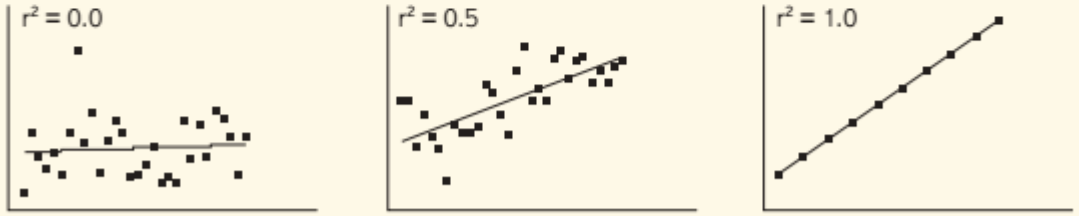
회귀분석을 통하여 b와 a를 구한 후 x값에 측정을 하여 얻은 결과값 중 2개를 넣어 각각 y값을 구한다. x와 y가 만나는 위치에 점을 plot하고 두 점을 연결하면 최소자승직선이 그려진다. 즉 y=1.1236x−1의 회귀식에 새로운 시약 측정값 중에서 50과 100을 넣어 각각 y값을 구한 후 만나는 점을 plot하고 이 점을 연결하면 최소자승직선이 된다. 이 회귀선을 이용하여 새로운 시약으로 측정한 결과가 표준 또는 현재사용 중인 시약으로 측정할 때 그 값이 얼마나 나오는지를 예측할 수 있게 된다.

검사평가를 위한 통계 ☞ 결정계수(r^2)

중원대학교 JUNGWON UNIVERSITY

결정계수 r^2, coefficient of determination

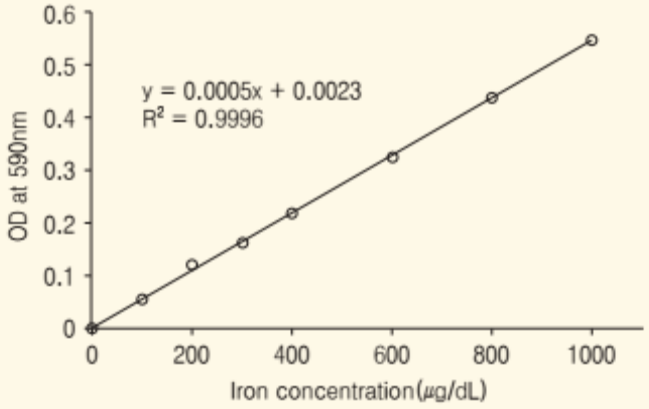
결정계수는 총 변동 중에서 회귀직선에 의해 설명되는 비율이고, 상관계수의 곱으로 계산한다. r^2은 x와의 관계에 의해 예측된 y의 변량 비율을 결정하기 때문에 r^2를 결정계수라 부르며, r^2의 범위는 $0 \le r^2 \le 1$ 이다. 만약 모든 관측점들이 회귀선상에 위치한다면 $r^2=1$이다. 이와 반대로 x와 y 사이에 회귀관계가 전혀 없으면 $r^2=0$이다. 결정계수가 1이라는 것은 독립변수(예: MCV)가 종속 변수(예: Hct)의 움직임을 완전하게 설명한다는 것을 의미한다.



결정계수 값에 의해 설명되는 회귀직선의 설명력

Hematocrit와 MCV의 상관관계를 분석하고 결정계수를 구하면 MCV 증가에 따른 hematocrit의 증가가 예측되므로 결정계수는 당연히 1에 가까운 결과를 보여줄 것이다. 그러나 예를 들어 몸무게와 키의 상관분석을 실시한 후 결정계수가 0.58이 나왔다고 하면 몸무게 변화의 58%는 키로 인한 증가로 설명되지만 나머지 42%는 성별, 나이, 체형 등에 의한 오차로 설명될 수 있다.

아래 그림은 혈청철을 측정하는데 가장 적합한 파장을 찾는 실험으로 590nm의 파장에서 흡광도 증가와 철농도 증가가 거의 1에 가까운 결정계수를 보여주고 있다. 즉, 590nm 파장을 사용하면 혈청철 농도 0~1,000μg/dL까지 농도변화에 따라 대응하는 흡광도의 변화가 선형적으로 일직선에 가까운 직선관계를 보여주어 590nm에서 가장 좋은 혈청철 농도를 측정할 수 있다는 연구결과이다.



검사평가를 위한 통계 ☞ Bland-Altman plot

- **Bland-Altman plot**
  - 2개의 검사방법 간의 비교 분석법
    *기존 방법(기준방법, reference method)과 새로운 방법(new method)
    *비교 후 결과가 문제가 될 정도의 차이가 있는지 없는지 평가(95% 유의 수준 내에서)
  - A 방법과 B 방법의 측정값의 평균과 측정치의 차이를 각각 x축과 y축으로 plotting
  - 2개의 측정값 간의 오차경향 파악하고 평가

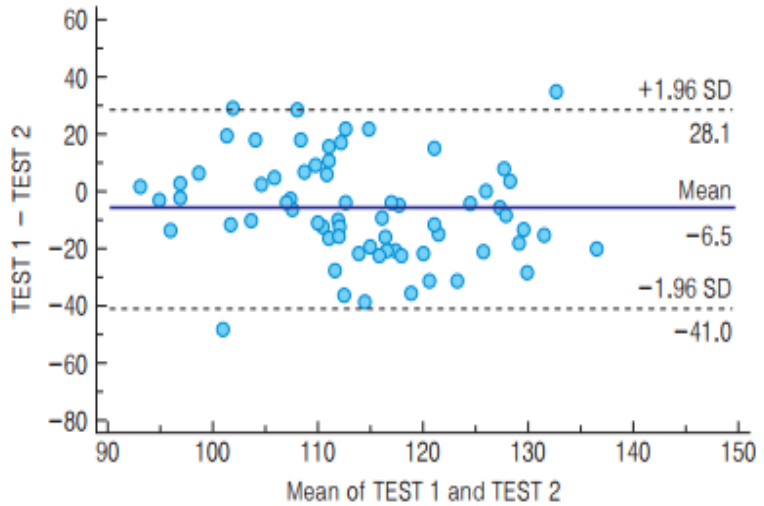
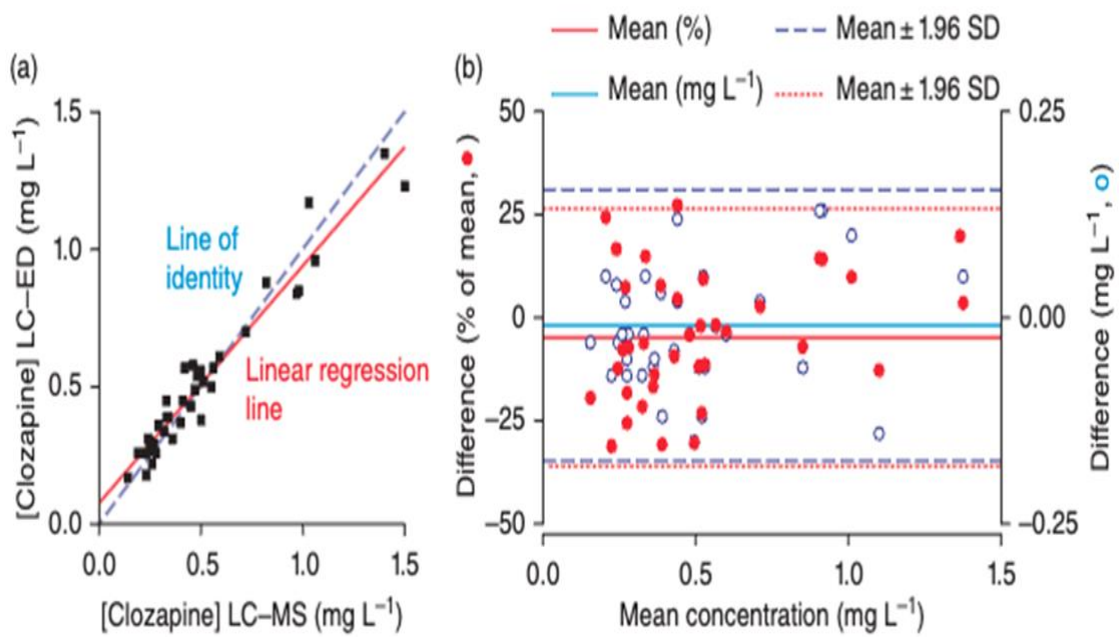
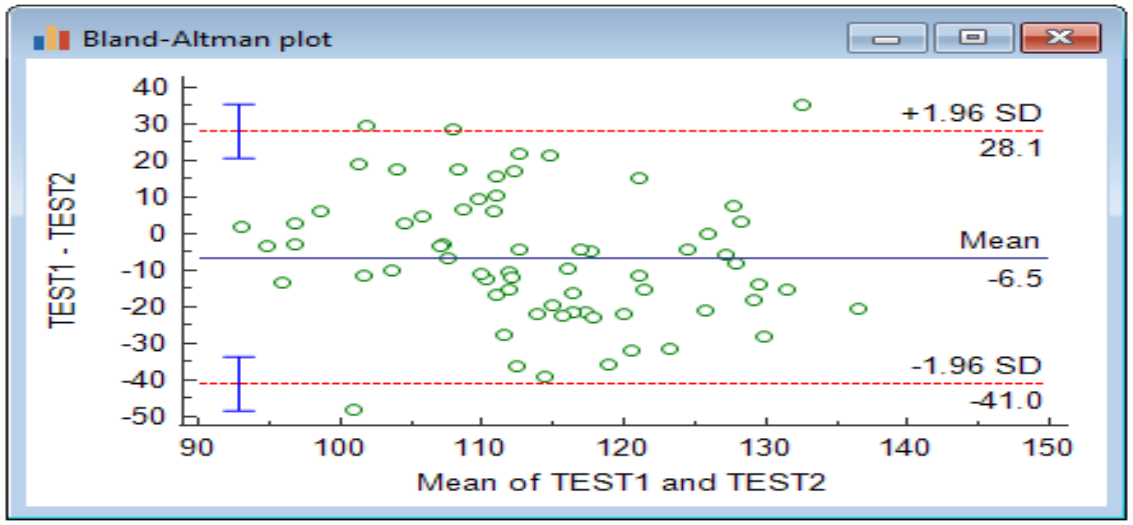


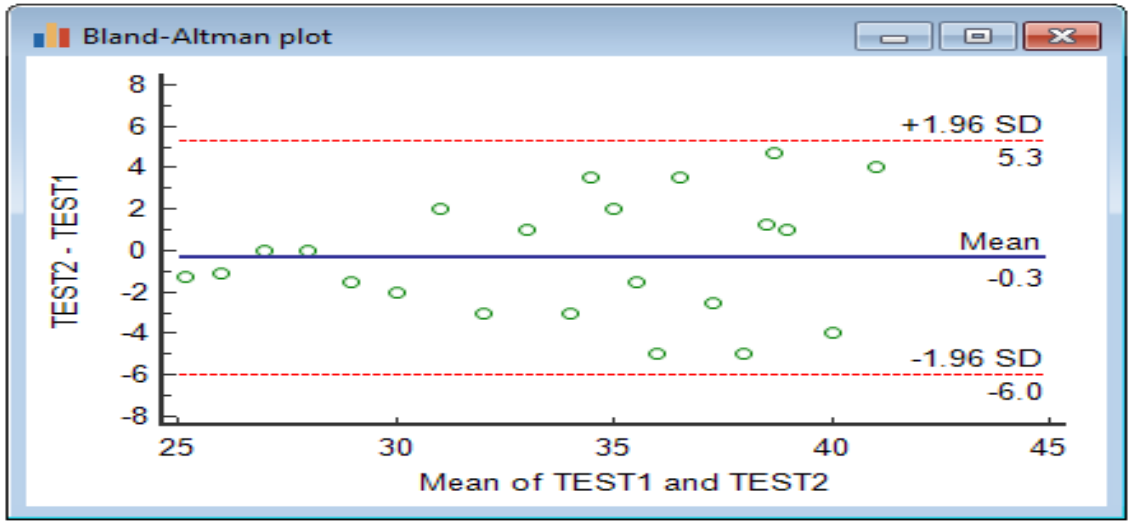
그림 12-8 Bland–Altman plot. X축은 두 측정치의 평균값이고, Y축은 측정방법1에서 측정방법 2 값을 뺀 값으로 두 측정값이 만나는 점을 plot한 것이다. Plot된 점들의 평균값과 표준편차를 그어 분산의 정도를 통계적으로 나타내었다.

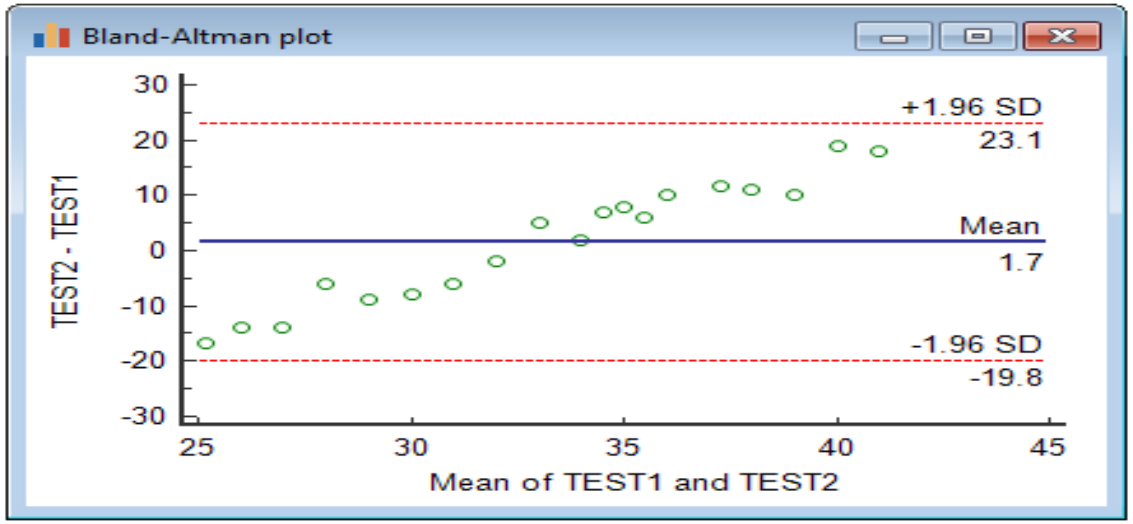


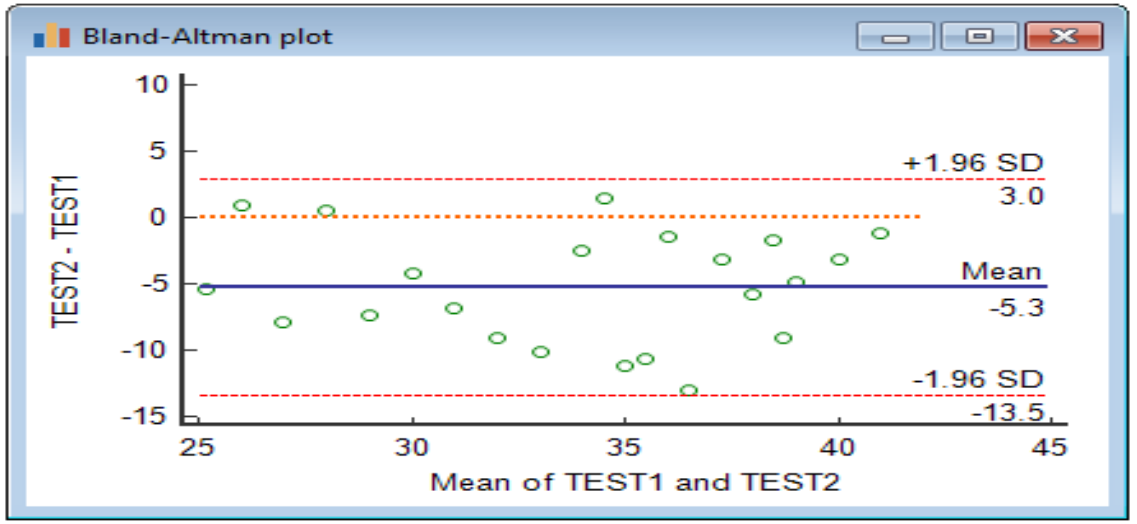
검사평가를 위한 통계 ☞ Bland-Altman plot

- Bland-Altman plot



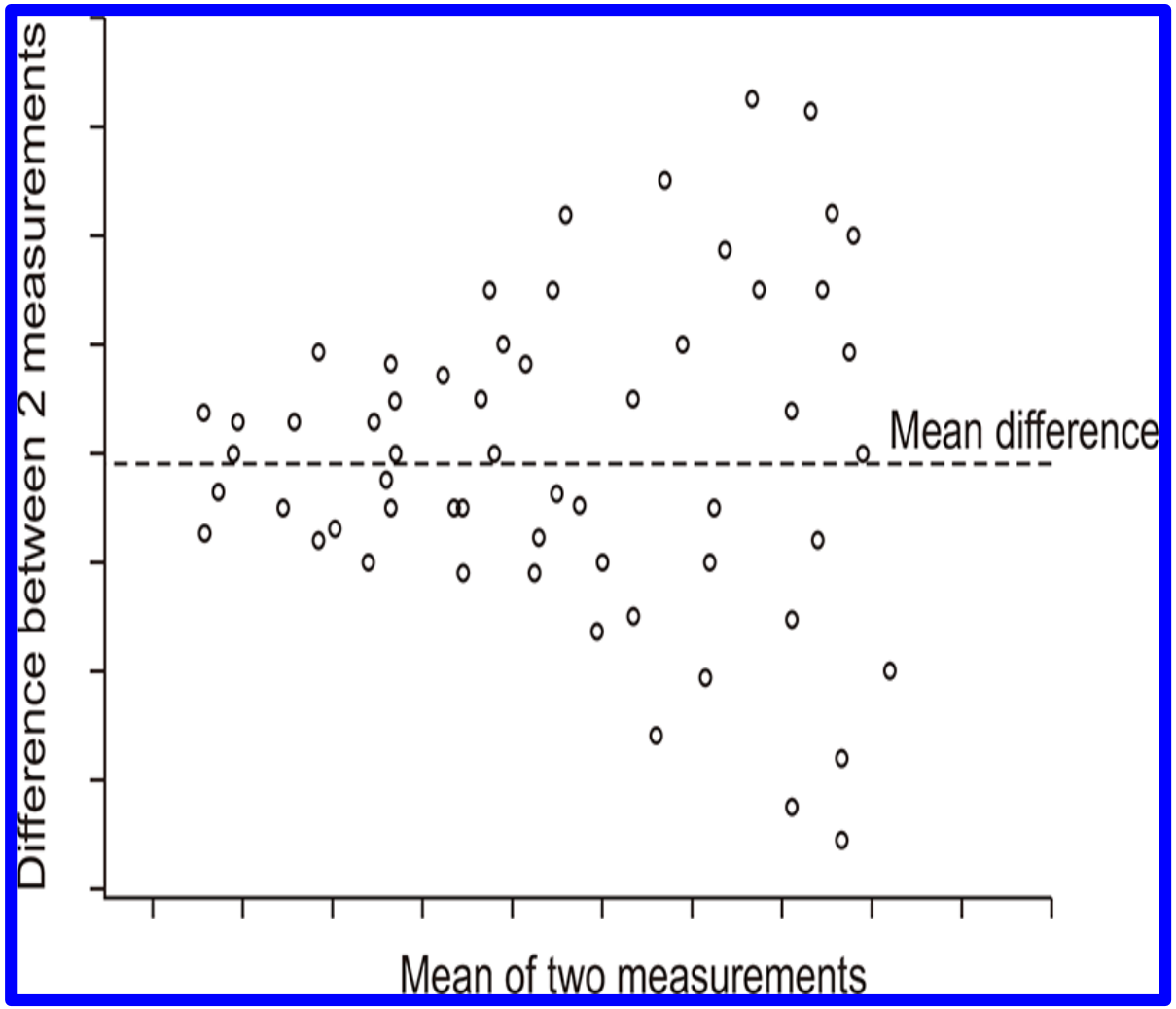


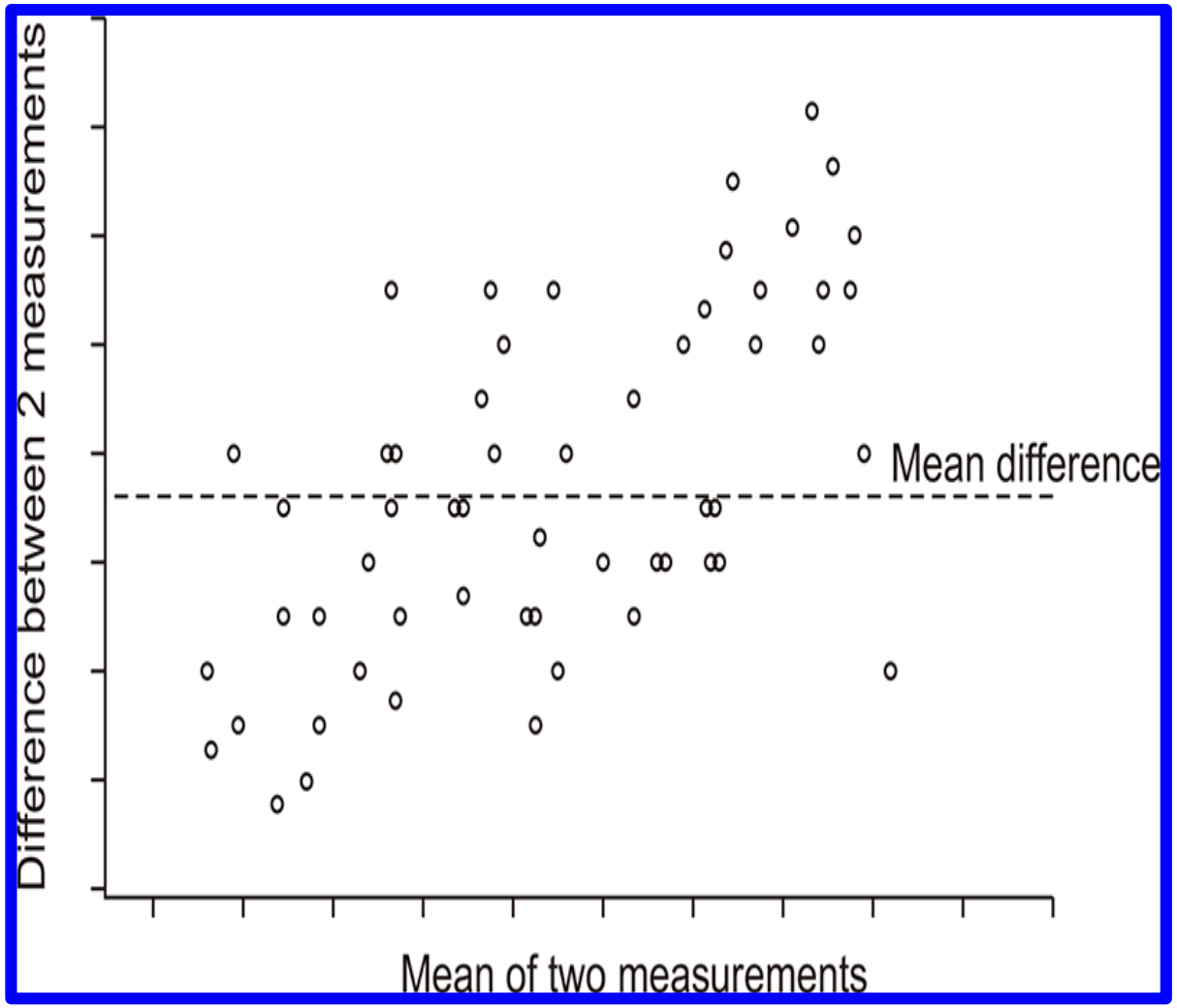




검사평가를 위한 통계 ☞ Bland-Altman plot

- Bland-Altman plot





검사평가를 위한 통계적 유의성 평가

중원대학교 JUNGWON UNIVERSITY

- t-test(t검정)
  - 표본수가 작은(보통 30개 이하)
  - 표본수(모집단)이 적어 정규분포가 되지 않을 때

- F-test(F검정)
  - 표본수(모집단)이 많아 정규분포 일 때

※ 평가: 두 집단 중 한 집단이 다른 집단의 95% 범위 내에 포함 여부

- 귀무가설(null hypothesis, H0: u1 = u2)
  - 현재의 가설이 의미 있는 차이가 없다고 주장

- 대립가설(alternative hypothesis, H1: u1 ≠ u2)
  - 새로운 것을 주장하는 가설로 기존 가설과 차이가 있다고 주장

민감도와 특이도의 평가

● 민감도(sensitivity)와 특이도(specificity) 평가

표 12-2. 분변잠혈검사의 진단민감도와 진단특이도 평가 예

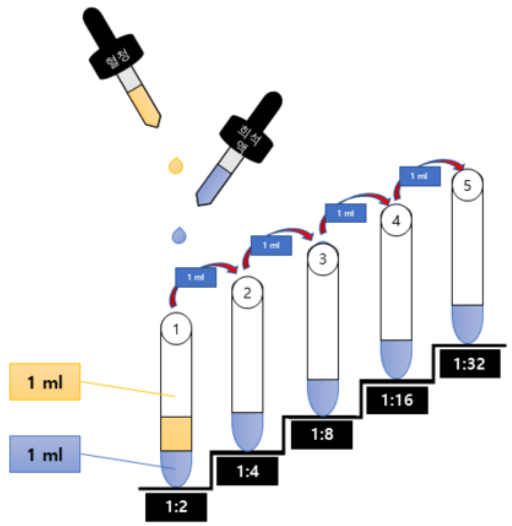
		대장암 환자(내시경검사로 확인)	
		양성 condition	음성 condition
분변 잠혈검사 결과	양성	진짜 양성(예: 20)	가짜 양성(예: 180)
	음성	가짜 음성(예: 10)	진짜 음성(예: 1820)
		진단민감도 $= \frac{TP}{(TP + FN)}$ = 20 / (20 + 10) = 67%	**진단특이도** $= \frac{TN}{(TN + FP)}$ = 1820 / (180 + 1820) = 91%

표 12-2는 내시경검사로 확인된 대장암환자의 분변을 이용하여 진단민감도와 진단특이도를 분석한 결과이다. 진단민감도의 경우 67%, 즉 대장암이 있는 환자 중 분변잠혈검사를 이용하여 67%를 찾아낼 수 있다는 것을 의미한다. 이것은 또한 33%의 위음성 오류를 범한다는 것이다. 진단특이도 91%는 대장암이 없는 환자를 대상으로 분변잠혈검사를 시행하였을 때 91%의 사람들을 대장암이 없는 환자로 판정한다는 것이다. 즉, 9%의 위양성 오류를 범한다는 것을 의미한다.

희석의 개념

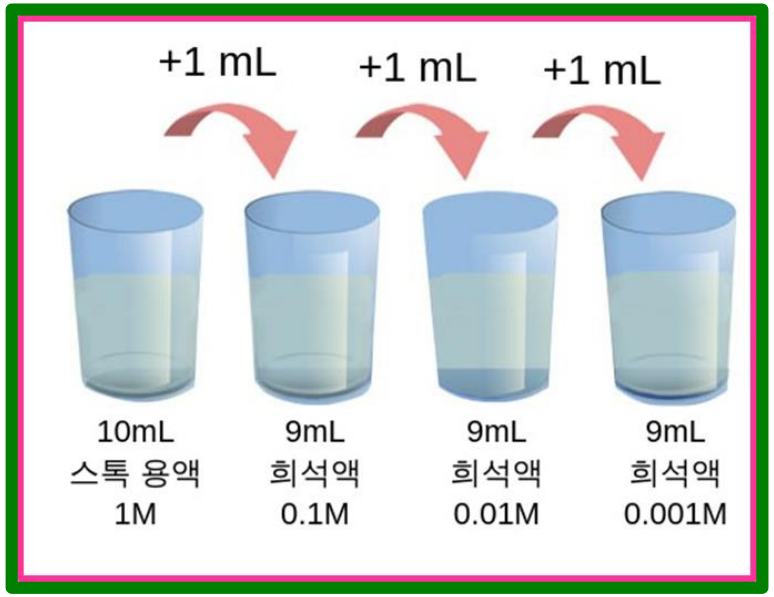
중원대학교 JUNGWON UNIVERSITY

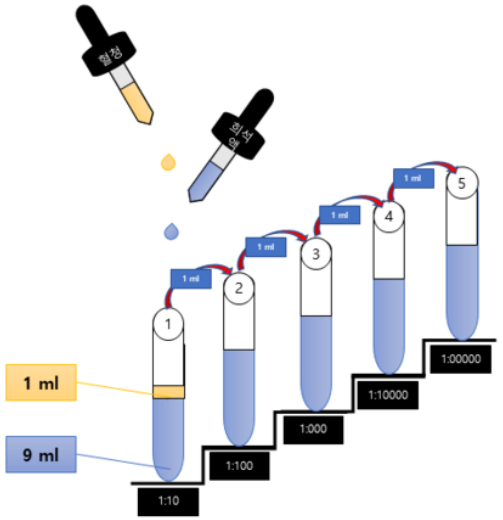
● 희석



By 희석맘몸

번호	검체	희석액	검체량/검체량 + 희석액	최종배수
1	1	1	$^{1}/_{1+1}$	1:2
2	0.5	1.5	$^{0.5}/_{0.5+1.5}$	1:4
3	0.25	1.75	$^{0.25}/_{0.25+1.75}$	1:8
4	0.125	1.875	$^{0.125}/_{0.125+1.875}$	1:16
5	0.0625	1.19375	$^{0.0625}/_{0.0625+1.19375}$	1:32





By 희석맘몸

번호	검체	희석액	검체량/검체량 + 희석액	최종배수
1	1	9	$^{1}/_{1+9}$	1:10
2	0.1	9.9	$^{0.1}/_{0.1+9.9}$	1:100
3	0.01	9.99	$^{0.01}/_{00.1+9.99}$	1:1000
4	0.001	9.999	$^{0.001}/_{0.001+9.999}$	1:10000
5	0.0001	9.9999	$^{0.0001}/_{0.0001+9.9999}$	1:100000

희석의 개념

중원대학교 JUNGWON UNIVERSITY

● 희석

$$\text{희석배수} = \frac{\text{희석후 용액의 양}}{\text{희석전 용액의 양}}$$

희석배수 조견표 (㎖=cc=g)

희석배수/물의양	1ℓ	2ℓ	3ℓ	4ℓ	5ℓ	10ℓ	20ℓ	100ℓ
50배 희석시	20㎖	40㎖	60㎖	80㎖	100㎖	200㎖	400㎖	2ℓ
100배 희석시	10㎖	20㎖	30㎖	40㎖	50㎖	100㎖	200㎖	1ℓ
200배 희석시	5㎖	10㎖	15㎖	20㎖	25㎖	50㎖	100㎖	500㎖
250배 희석시	4㎖	8㎖	12㎖	16㎖	20㎖	40㎖	80㎖	400㎖
500배 희석시	2㎖	4㎖	6㎖	8㎖	10㎖	20㎖	40㎖	200㎖
1,000배 희석시	1㎖	2㎖	3㎖	4㎖	5㎖	10㎖	20㎖	100㎖
2,000배 희석시	0.5㎖	1㎖	1.5㎖	2㎖	2.5㎖	5㎖	10㎖	50㎖
5,000배 희석시	0.2㎖	0.4㎖	0.6㎖	0.8㎖	1㎖	2㎖	4㎖	20㎖

희석배수 조견표(㎖=cc=g)

리터(ℓ) \ 희석배수	100배	200배	250배	500배	1000배	2000배	2500배	5000배
1ℓ	10㎖	5㎖	4㎖	2㎖	1㎖	0.5㎖	0.4㎖	0.2㎖
1.5ℓ	15㎖	7.5㎖	6㎖	3㎖	1.5㎖	0.75㎖	0.6㎖	0.3㎖
2ℓ	20㎖	10㎖	8㎖	4㎖	2㎖	1㎖	0.8㎖	0.4㎖
3ℓ	30㎖	15㎖	12㎖	6㎖	3㎖	1.5㎖	1.2㎖	0.6㎖
4ℓ	40㎖	20㎖	16㎖	8㎖	4㎖	2㎖	1.6㎖	0.8㎖
5ℓ	50㎖	25㎖	20㎖	10㎖	5㎖	2.5㎖	2㎖	1㎖
10ℓ	100㎖	50㎖	40㎖	20㎖	10㎖	5㎖	4㎖	2㎖
20ℓ	200㎖	100㎖	80㎖	40㎖	20㎖	10㎖	8㎖	4㎖
100ℓ	1000㎖	500㎖	400㎖	200㎖	100㎖	50㎖	40㎖	20㎖
500ℓ	5000㎖	2500㎖	2000㎖	1000㎖	500㎖	250㎖	200㎖	100㎖
1000ℓ	10000㎖	5000㎖	4000㎖	2000㎖	1000㎖	500㎖	400㎖	200㎖
2000ℓ	20000㎖	10000㎖	8000㎖	4000㎖	2000㎖	1000㎖	800㎖	400㎖
5000ℓ	50000㎖	25000㎖	20000㎖	10000㎖	5000㎖	2500㎖	2000㎖	1000㎖

Stay Hungry

Stay Foolish !!