

측정의 정확성과 신뢰성

김 동 현

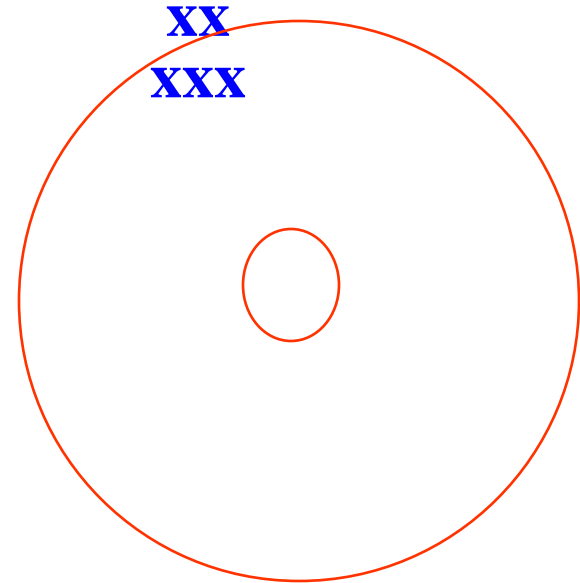
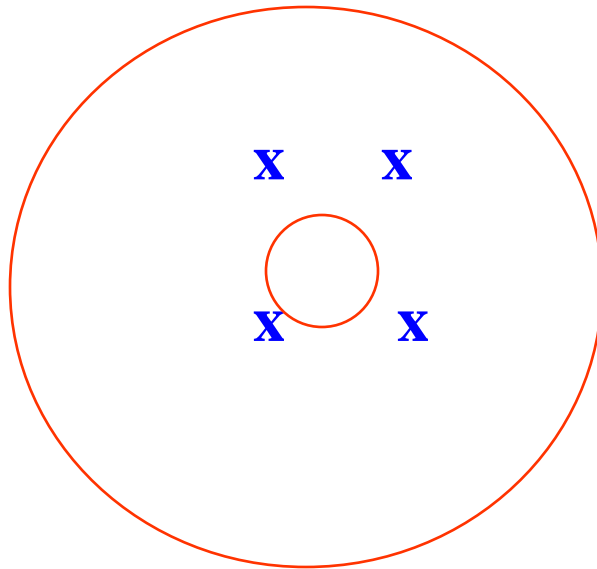
한림대학교 의과대학

학습 목표

1. 측정에서 타당도, 신뢰도의 개념과 차이를 설명한다.
2. 진단검사법의 타당도를 나타내는 기준으로 사용되는 다음 5가지 지표를 산출하고, 그 의미를 설명 한다.
 - (1) 민감도(sensitivity)
 - (2) 특이도(specificity)
 - (3) 예측도(predictability)
 - (4) 위양성(false positive)
 - (5) 위음성(false negative)
3. 신뢰도를 평가하는 kappa 통계량의 의미와 계산식을 이해한다.

Accuracy in Measurement

Validity and Reliability



Reliability, 신뢰도, 재현도

- Precision, reproducibility, repeatability
- Source of variability :
 - observer, test instrument -> measurement error
 - subject -> biological variation
- test-retest reliability, inter-rater reliability, internal consistency
- Especially important when true state could not be known
 - stress level, dietary intake, etc.

Evaluating Reliability of Study variables

- **Test-retest reliability**
 - Degree of correspondence bwt. answers at different points in time
 - Shorter period preferred (2-4 weeks)
 - Pearson correlation coefficient for interval-level data
 - Spearman rank order coefficient for ordinal-level data
- **Inter-rater reliability**
 - Equivalence of information among different gatherers or raters
 - Pearson correlation, Spearman rank order, kappa coefficients
- **Internal consistency reliability**
 - Constructing and evaluating summary scales
 - Inconsistency of different items intended to measure same concepts
 - Corrected item-total correlation > 0.4
 - Split-half reliability

Reliability

- Assessment of reliability

R2

		Ds+	Ds-	
R1	Ds+	a	b	a+b
	Ds -	c	d	c+d
		a+c	b+d	

1) proportion of overall agreement : $(a+d) / (a+b+c+d)$

- two measurements should be independent of each other
- percent exact agreement, most commonly used

Reliability

- Assessment of reliability

		R2		
	R1	Ds+	Ds-	
Ds+		a	b	a+b
Ds -		c	d	c+d
		a+c	b+d	

2) Kappa statistic : $(P_o - P_c) / (1 - P_c)$, agreement beyond chance

- P_o : Proportion of overall agreement, $(a+d)/(a+b+c+d)$

P_c : Proportion of chance agreement,

$$[(a+b)(a+c) + (b+d)(c+d)] / (a+b+c+d)^2$$

- > 0.8 excellent agreement

0.6 - 0.8 good agreement

Validity, 타당도

- should be reliable, gold standard
- Assessment of validity

		Gold standard		
		Ds+	Ds-	
Test	T+	a	b	a+b
	T-	c	d	c+d
		a+c	b+d	

a: True positive (TP)

b: False positive (FP)

c: False negative (FN)

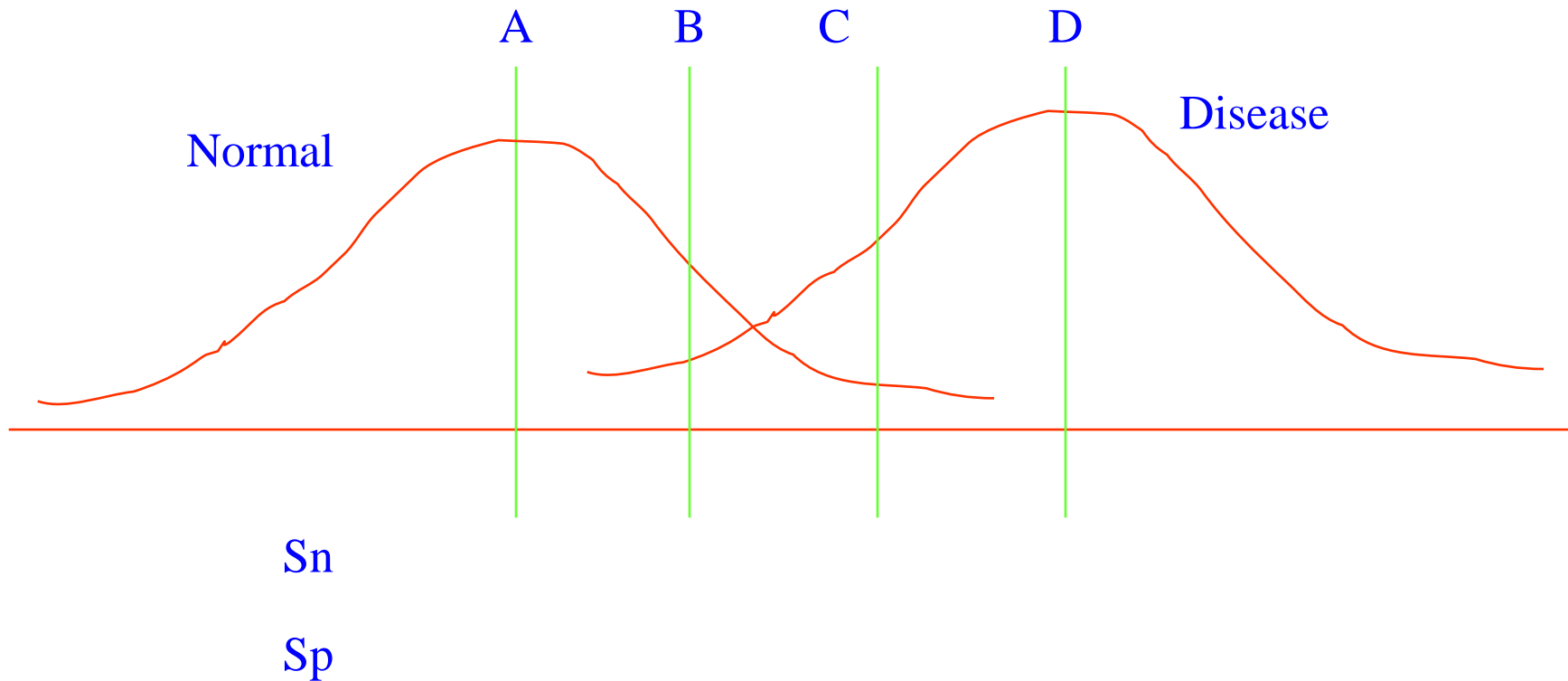
d: True negative(TN)

Validity

- **Sensitivity(Sn), 민감도 : $a/(a+c)$**
 - 참값(질병)을 옳게 측정(진단)해 줄 수 있는 검사(진단)도구의 능력
- **Specificity(Sp), 특이도 : $d/(b+d)$**
 - 질병이 없을 때 없다고 해 줄 수 있는 진단도구의 능력
- **질병의 선별검사에 적합한 검사 : higher Sn**
- **질병의 확진검사에 적합한 검사 : higher Sp**
- **조기진단의 치료 효과가 큰 질환 : higher Sn**
- **위양성에 의해 야기될 수 있는 문제가 큰 질환 : higher Sp**

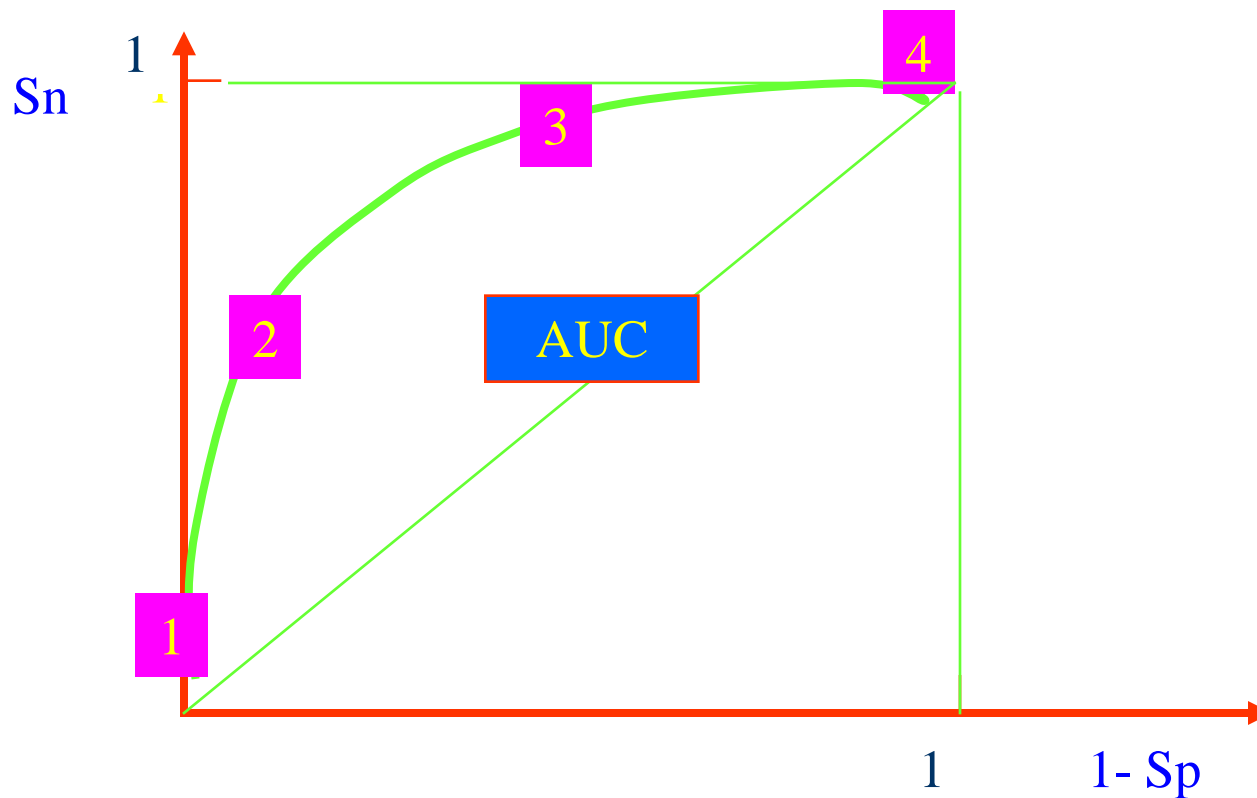
Validity

검사 성적이 연속변수로 제시될 때(혈당치, 혈압등),
진단을 위한 **cut-off point** 의 결정



Validity

- 각 검사치를 기준으로 S_n , $1-S_p$ 를 계산
- Receiver Operating Characteristic(ROC) Curve 를 plot
- Area Under the Curve (AUC) : 검사의 정확도를 종합, 평가하는 지표



Validity

- **Positive predictive value(+Pv), 양성 예측도 : $a/(a+b)$**
 - 검사에서 양성 나왔을 때, 실제로 질병이 있을 확률
- **Negative predictive value(-Pv), 음성 예측도 : $d/(c+d)$**
 - 질병이 없을 때 없다고 해 줄 수 있는 진단도구의 능력
- **+Pv, -Pv 는 대상 집단에서의 질병 유병률에 따라 변동**
 - $+Pv = (Sn * Pr) / [(Sn * Pr) + (1 - Sp)(1 - Pr)]$, Pr : Prevalence
 - 유병률이 높은 상황에서는 양성 예측도가 높다
 - 유병률이 낮은 상황에서는 음성 예측도가 높다
- **개별 환자의 진단에서 중요**

Validity

■ Serial measurements,

- 첫 검사에서 양성자만이 다음 검사, 연속된 모든 검사에서 양성이면 확진할 경우

⇒ **lower Sn, higher Sp, and higher +Pv 를 결과
test with higher Sn 를 먼저 적용**

■ Parrallel measurement,

- 여러 검사를 동시에 적용, 최소한 한 검사에서 양성이면 확진할 경우

⇒ **lower Sp, lower +Pv, and higher Sn 를 결과**

Evaluating Validity of Study Variables

- Are we measuring what we want to measure ?
- Content validity (내용타당도)
 - representativeness or sampling adequacy of the content
 - Content validation is basically judgmental
- Criterion validity (준거타당도)
 - Evaluated in comparison with an ‘expert’ judgment or a previously validated instrument
 - Concurrent, predictive validity
- Construct validity (구성타당도)
 - What factors or construct account for variance in test performance
 - Try to validate the theory behind the test, nature of reality