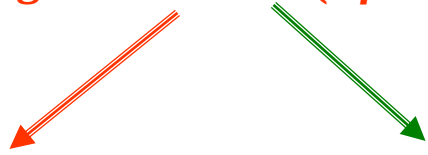
Cohen's kappa coefficient of agreement

Prof. Giuseppe Verlato

Unit of Epidemiology and Medical Statistics

University of Verona

Is there a gold standard (*riferimento aureo*)?



YES: SENSITIVITY and SPECIFICITY of the new diagnostic test are computed using the gold standard as reference

NO: Cohen's kappa coefficient of agreement is computed

Example:

The gold standard to diagnose colorectal cancer is colonoscopy + biopsy

***Inter-rater* o *intra-rater agreement*: concordance between the judgements expressed by different operators or by the same operator**

Study on tardive dyskinesia

Rater a		Rater b: *present*	Rater b: *absent*	Rater b: *total*
	pres.	**123**	**10**	**133**
	abs.	**6**	**29**	**35**
	total	**129**	**39**	**168**

$\chi^2 = 88.23 \longrightarrow P < 0.0000001$

Bergen et al (1992) Predictors of the Course of Tardive Dyskinesia in Patients Receiving Neuroleptics. *Biol Psychiatry* 1992; 32: 580-594

The chi-square test provides little information.

We simply falsified the null hypothesis that the diagnoses made by the first clinician were completely independent of the diagnoses made by the second clinician. However this result can also be achieved by inexperienced clinicians.

Proportion of agreement (Misura di concordanza osservata)

Rater b

Rater a	*present*	*absent*	*total*
pres.	**a**	**b**	**a+b**
abs.	**c**	**d**	**c+d**
total	**a+c**	**b+d**	**N**

$$P_0 = \frac{a + d}{N}$$

Proportion of agreement (Misura di concordanza osservata)

Rater b

Rater a	*present*	*absent*	*total*
pres.	**123**	**10**	**133**
abs.	**6**	**29**	**35**
total	**129**	**39**	**168**

$$P_0 = \frac{123 + 29}{168} = \frac{152}{168} = 0.905$$

Bergen et al (1992) Predictors of the Course of Tardive Dyskinesia in Patients Receiving Neuroleptics. *Biol Psychiatry* 1992; 32: 580-594

But even two people, completely unacquainted with medicine, manage to achieve a certain agreement, simply by chance.

We are interested in assessing how much the agreement between two experts exceeds the agreement expected by chance.

Agreement EXPECTED just by chance

$$P_{exp} = \frac{a_{exp} + d_{exp}}{N}$$

Under the null hypothesis of statistical independence

expected = (row total) * (column total) / (grand total)

		Rater b		
		present	*absent*	*total*
Rater a	*pres.*	a	b	a+b
	abs.	c	d	c+d
	total	a+c	b+d	N

$$a_{exp} = \frac{(a+b) * (a+c)}{N}$$

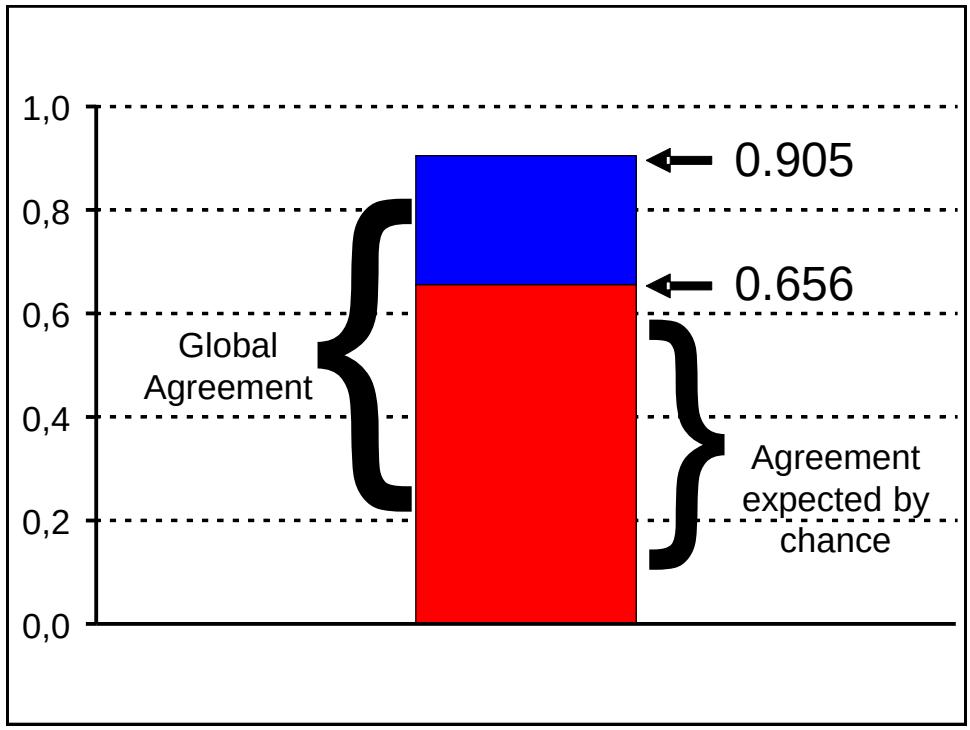
Expected agreement

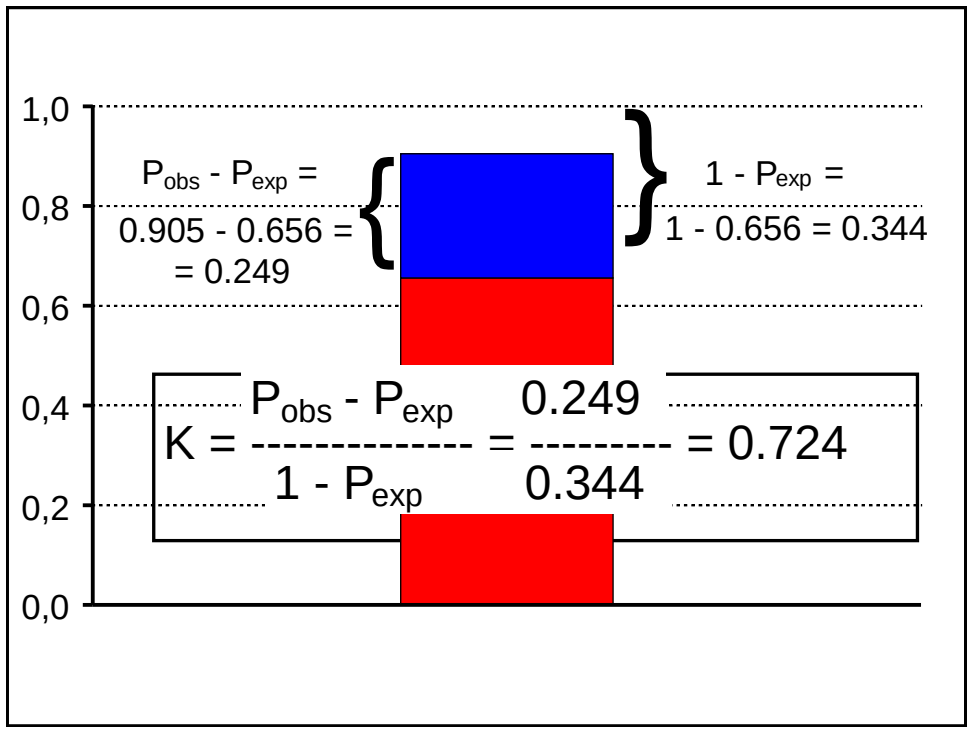
		Rater b		
		present	*absent*	*total*
Rater a	*pres.*	123	10	133
	abs.	6	29	35
	total	129	39	168

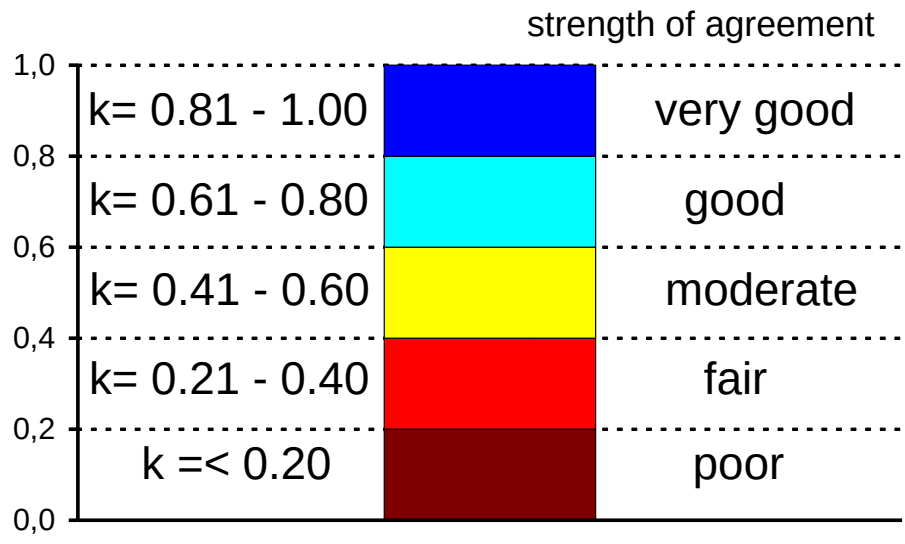
$a_{expected}$ = (133 * 129) / 168 = 102.125

$d_{expected}$ = (35*39) / 168 = 8.125

$$P_e = \frac{102.1+8.1}{168} = \frac{110.25}{168} = 0.656$$


1,0
0,8
0,6
0,4
0,2
0,0
0.905
0.656
Global
Agreement
Agreement
expected by
chance


1,0
0,8
0,6
0,4
0,2
0,0
Pobs - Pexp =
0.905 - 0.656 =
= 0.249
1 - Pexp =
1 - 0.656 = 0.344
K = Pobs - Pexp / 1 - Pexp = 0.249 / 0.344 = 0.724



Altman DG (1991) Practical statistics for medical research. Chapman & Hall: London.

Given a certain level of agreement, the k coefficient is maximum when the prevalence of positives (sick people) is close to 0.50.

Rater b

Rater a	*present*	*absent*	*total*
pres.	**70**	**10**	**80**
abs.	**10**	**10**	**20**
total	**80**	**20**	**100**

P_{obs} = 80 / 100 = 0.80
a_{exp} = (80*80) / 100 = 64
d_{exp} = (20*20) / 100 = 4
P_{exp} = (64+4) / 100 = 0.68
k= (0.80-0.68)/(1-0.68) =0.38

Rather b

Rather a	*present*	*absent*	*total*
pres.	**40**	**10**	**50**
abs.	**10**	**40**	**50**
total	**50**	**50**	**100**

P_{obs} = 80 / 100 = 0.80
a_{exp} = (50*50) / 100 = 25
d_{exp} = (50*50) / 100 = 25
P_{exp} = (25+25) / 100 = 0.50
k= (0.80-0.50)/(1-0.50) =0.60

It is possible to separately address the agreement between positive and negative responses through the following indices:

P_{pos} = proportion of positive agreement

P_{neg} = proportion of negative agreement

Proportion of negative agreement

2nd survey

First survey		*No*	*Yes*	*Total*
	No	**a**	**b**	**a+b**
	Yes	**c**	**d**	**c+d**
	Total	**a+c**	**b+d**	**N**

$$P_{neg} = \frac{a}{(a+b+a+c)/2}$$

Proportion of positive agreement

2nd survey

First survey		*No*	*Yes*	*Total*
	No	**a**	**b**	**a+b**
	Yes	**c**	**d**	**c+d**
	Total	**a+c**	**b+d**	**N**

$$P_{pos} = \frac{d}{(b+d+c+d)/2}$$

Cicchetti & Feinstein, J Clin Epidemiol, 1990