Inter Rater Reliability

Computer Skills for Medical Students
The University of Jordan
KASIT /Computer Information Systems Department

Inter rater Reliability

Inter-rater reliability is the **level of agreement between raters or judges**.

If everyone agrees, IRR is 1 (or 100%) and if everyone disagrees, IRR is 0 (0%).

Inter rater Reliability Methods

1. Percent Agreement for Two Raters.
2. Cohen's Kappa .

1. Percent Agreement

The simple way to measure inter-rater reliability is to calculate the percentage of items that the judges agree on.

This is known as **percent agreement**, which always ranges between 0 and 1.

0 : indicating **no agreement** between raters.

1 : indicating **perfect agreement** between raters.

1. Percent Agreement (Example)

For example, suppose two judges are asked to rate the difficulty of 10 items on a test from a scale of 1 to 3. The results are shown below:

	Judge 1	Judge 2
Question 1	1	1
Question 2	1	1
Question 3	2	3
Question 4	2	2
Question 5	1	2
Question 6	2	3
Question 7	3	3
Question 8	2	2
Question 9	3	3
Question 10	3	3

1. Percent Agreement (Example)

For each question, we can write "1" if the two judges agree and "0" if they don't agree.

	Judge 1	Judge 2	Agree?
Question 1	1	1	1
Question 2	1	1	1
Question 3	2	3	0
Question 4	2	2	1
Question 5	1	2	0
Question 6	2	3	0
Question 7	3	3	1
Question 8	2	2	1
Question 9	3	3	1
Question 10	3	3	1

The percentage of questions the judges agreed on was 7/10 = **70%**

Percent Agreement Acceptable Level

The field you are working in will determine the acceptable agreement level. If it's a sports competition, you might accept a 60% rater agreement to decide a winner.

However, if you're looking at data from cancer specialists deciding on a course of treatment, you'll want a much higher agreement — above 90%.

In general, above 75% is considered acceptable for most fields.

2. Cohen's Kappa Statistic

Cohen's Kappa Statistic is used to measure the level of agreement between two raters or judges .

2. Cohen's Kappa Statistic

The formula for Cohen's kappa is calculated as:

$$k = (p_o - p_e) / (1 - p_e)$$

where:

p_o: Relative observed agreement among raters.

p_e: Hypothetical probability of chance agreement.

How to Interpret Cohen's Kappa ?

Cohen's Kappa always ranges between 0 and 1.

0 : indicating **no agreement** between the two raters

1 : indicating **perfect agreement** between the two raters.

How to Interpret Cohen's Kappa ?

The following table summarizes how to interpret different values for Cohen's Kappa:

Cohen's Kappa	Interpretation
0	No agreement
0.10 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Near perfect agreement
1	Perfect agreement

Cohen's Kappa Statistic (Example)

Suppose two doctors are asked to diagnose 70 patients on whether they're Covid-19 Patients or Not. The following 2×2 table shows the results of the diagnosis:

		Rater 2	
		Yes	**No**
Rater 1	**Yes**	25	10
	No	15	20

Cohen's Kappa Statistic (Example)

Step 1: Calculate relative agreement (p_o) between raters.

First, we'll calculate the relative agreement between the doctors(raters) . This is simply the proportion of total ratings that the raters both said "Yes" or both said "No" on.

We can calculate this as:

- p_o = (Both said Yes + Both said No) / (Total Ratings)
- p_o = (25 + 20) / (70) = 0.6429

Cohen's Kappa Statistic (Example)

Step 2: Calculate the hypothetical probability of chance agreement (p_e) between raters.

This is calculated as the total number of times that Rater 1 said "Yes" divided by the total number of responses, multiplied by the total number of times that Rater 2 said "Yes" divided by the total number of responses, added to the total number of times that Rater 1 said "No" multiplied by the total number of times that Rater 2 said "No."

Cohen's Kappa Statistic (Example)

For our example, this is calculated as:

P("Yes") = ((25+10)/70) * ((25+15)/70) = 0.285714

P("No") = ((15+20)/70) * ((10+20)/70) = 0.214285

p_e = 0.285714 + 0.214285 = **0.5**

Example

Step 3: Calculate Cohen's Kappa

Lastly, we'll use p_o and p_e to calculate Cohen's Kappa:

$k = (p_o - p_e) / (1 - p_e)$

$k = (0.6429 - 0.5) / (1 - 0.5)$

$k = 0.2857$

Cohen's Kappa turns out to be **0.2857**. Based on the table from earlier, we would say that the two Doctors only had a "fair" level of agreement.