

Lab Class NLP:III

To solve the following task, please make yourself familiar with Cohen's κ , Fleiss' κ , Davies' κ , Scott's π , and Krippendorff's α first. [NLP:III-174]

Exercise 1 : Validity vs. Reliability

- (a) Define *validity*.
- (b) Define *reliability*.
- (c) Which implications hold for annotations? Mark the correct answer(s).
 - validity \Rightarrow reliability
 - reliability \Rightarrow validity

Exercise 2 : Inter-Annotator Agreement

Let S be a sample of $n = 20$ objects, C a set of $k = 2$ classes, and a group of $m = 2$ independent annotators γ_1 and γ_2 . Assume you have collected sentiment assessments for a set of documents. Each document was judged by two annotators A and B as either "positive" or "negative".

		γ_1		total
		positive	negative	
γ_2	positive	3	2	
	negative	1	14	
total				

Table 1: Contingency table of sentiment annotations assigned by annotators γ_1 and γ_2 .

- (a) Which inter-annotator agreement measures can be used in this scenario?
- (b) Compute the inter-annotator agreement for all nominal disagreements measures from (a).
- (c) How would you interpret the obtained agreement values?

Exercise 3 : Inter-annotator agreement measures

Complete the table using the terms in the box.

Fleiss' κ • Scott's π • Cohen's κ • Krippendorff's α • Davies' κ

	joint distribution	per-annotator distribution
$m > 2$ annotators		
$m = 2$ annotators		