

Module 8

David Bishop

September 2024

1 Task 1

1.1 Data

Table 1: Comparison of Day 1 and Day 2 Annotations for Within-Rater Reliability

Ontology	Day 1	Day 2
Albums	9	9
Artists	39	41
Bands	3	3
Genres	12	19
Instruments	18	17
Locations	17	15
Release Dates	7	8
Songs	2	2
Sub Genres	22	44
Lyrics	0	0

1.2 ICC Calculation

Mean of Day 1

$$\bar{X} = \frac{\sum X}{n} = \frac{9 + 39 + 3 + 12 + 18 + 17 + 7 + 2 + 22 + 0}{10} = \frac{129}{10} = 12.9$$

Mean of Day 2

$$\bar{Y} = \frac{\sum Y}{n} = \frac{9 + 41 + 3 + 19 + 17 + 15 + 8 + 2 + 44 + 0}{10} = \frac{158}{10} = 15.8$$

Overall Mean

$$OM = \frac{\bar{X} + \bar{Y}}{2} = \frac{12.9 + 15.8}{2} = 14.35$$

Sum of Squares Between Targets

$$SSB = \sum (\text{Row Mean} - OM)^2 \times k$$

Category	Row Mean	$(\text{Row Mean} - GM)^2$	$(\text{Row Mean} - GM)^2 \times k$
Albums	9	$(9 - 14.35)^2 = 28.62$	$28.62 \times 2 = 57.24$
Artists	40	$(40 - 14.35)^2 = 658.82$	$658.82 \times 2 = 1317.64$
Bands	3	$(3 - 14.35)^2 = 128.82$	$128.82 \times 2 = 257.64$
Genres	15.5	$(15.5 - 14.35)^2 = 1.32$	$1.32 \times 2 = 2.64$
Instruments	17.5	$(17.5 - 14.35)^2 = 9.92$	$9.92 \times 2 = 19.84$
Locations	16	$(16 - 14.35)^2 = 2.72$	$2.72 \times 2 = 5.44$
Release Dates	7.5	$(7.5 - 14.35)^2 = 46.92$	$46.92 \times 2 = 93.84$
Songs	2	$(2 - 14.35)^2 = 152.42$	$152.42 \times 2 = 304.84$
Sub Genres	33	$(33 - 14.35)^2 = 347.42$	$347.42 \times 2 = 694.84$
Lyrics	0	$(0 - 14.35)^2 = 205.82$	$205.82 \times 2 = 411.64$

Table 2: Calculation of SSB for Each Category

$$\sum (\text{Row Mean} - OM)^2 \times k = 57.24 + 1317.64 + 257.64 + 2.64 + 19.84 + 5.44 + 93.84 + 304.84 + 694.84 + 411.64 = 2966.36$$

$$SSB = 2966.36$$

Sum of Squares within Targets

$$SSW = \sum [(X_i - \text{Row Mean})^2 + (Y_i - \text{Row Mean})^2]$$

Category	Day 1 (X)	Day 2 (Y)	Row Mean	$(X - \text{Row Mean})^2$	$(Y - \text{Row Mean})^2$	Total
Albums	9	9	9	0	0	0
Artists	39	41	40	1	1	2
Bands	3	3	3	0	0	0
Genres	12	19	15.5	12.25	12.25	24.5
Instruments	18	17	17.5	0.25	0.25	0.5
Locations	17	15	16	1	1	2
Release Dates	7	8	7.5	0.25	0.25	0.5
Songs	2	2	2	0	0	0
Sub Genres	22	44	33	121	121	242
Lyrics	0	0	0	0	0	0

Table 3: Calculation of SSW for Each Category

$$SSW = 0 + 2 + 0 + 24.5 + 0.5 + 2 + 0.5 + 0 + 242 + 0 = 271.5$$

Mean Square Between

$$MSB = \frac{SSB}{n - 1} = \frac{2966.36}{10 - 1} = \frac{2966.36}{9} \approx 329.596$$

Mean Square Within

$$MSW = \frac{SSW}{n \times (k - 1)} = \frac{271.5}{10 \times (2 - 1)} = \frac{271.5}{10} = 27.15$$

ICC

$$ICC = \frac{MSB - MSW}{MSB + MSW} = \frac{329.596 - 27.15}{329.596 + 27.15} = \frac{302.446}{356.746} \approx 0.848$$

2 Task 2

1. Confusion Matrix

Table 4: Confusion Matrix for Task 2.1

	Yes	No	Don't Know
Yes	5	7	4
No	3	10	2
Don't Know	5	2	8

(a) IAA using percent agreement

$$\text{Yes/Yes} = 5$$

$$\text{No/No} = 10$$

$$\text{Dont know/Dont know} = 8$$

$$\text{Total Agreements} = 23$$

$$\text{Number of comparisons} = 5 + 7 + 4 + 3 + 10 + 2 + 5 + 2 + 8 = 46$$

$$\text{Percent Agreement} = 23/46 * 100 = 50\%$$

(b) Calculate IAA using Bennett et al's S

$$Q = 3$$

$$P_a = .5$$

$$S = \frac{3 * .5 - 1}{3 - 1} = 0.25$$

(c) IAA using Scott's pi

$$P_o = .5$$

$$P(\text{Yes}) = \frac{16}{46}$$

$$P(\text{No}) = \frac{15}{46}$$

$$P(\text{don know}) = \frac{15}{46}$$

$$P_e = 0.1209 + 0.1063 + 0.1063 = 0.3335$$

$$\pi = \frac{.5 - .3335}{1 - .3335} = .6665$$

(d) IAA using Cohens Kappa

$$P_o = \frac{15}{25} = .6$$

$$N = 25$$

$$n_{yes,A1} = 12$$

$$n_{yes,A2} = 8$$

$$n_{no,A1} = 13$$

$$n_{no,A2} = 17$$

$$P_e = \frac{1}{25^2} * ((12 * 8) + (13 * 17)) = 0.5072$$

$$\kappa = \frac{0.6 - .5072}{1 - .5072} = 0.1882$$

2. Confusion Matrix

Table 5: Confusion Matrix for Task 2.2

	Yes	No	Don't Know
Yes	12	3	1
No	3	10	4
Don't Know	2	5	11

(a) IAA using percent agreement

$$\text{Yes/Yes} = 12$$

$$\text{No/No} = 10$$

$$\text{Dont know/Dont know} = 11$$

$$\text{Total Agreements} = 33$$

$$\text{Number of comparisons} = 12 + 3 + 1 + 3 + 10 + 4 + 2 + 5 + 11 = 51$$

$$\text{Percent Agreement} = 33/51 * 100 \approx 64.71\%$$

(b) Calculate IAA using Bennett et al's S

$$Q = 3$$

$$P_a = .6471$$

$$S = \frac{3 * .6471 - 1}{3 - 1} = 0.4704$$

(c) IAA using Scott's pi

$$P_o = .5$$

$$P(\text{Yes}) = \frac{16}{51}$$

$$P(\text{No}) = \frac{17}{51}$$

$$P(\text{don know}) = \frac{18}{51}$$

$$P_e = 0.1048 + 0.1167 + 0.1114 = 0.333$$

$$\pi = \frac{.6471 - .333}{1 - .333} = .471$$

(d) IAA using Cohens Kappa

$$P_o = \frac{22}{28} = 0.7857$$

$$N = 28$$

$$n_{yes,A1} = 15$$

$$n_{yes,A2} = 15$$

$$n_{no,A1} = 13$$

$$n_{no,A2} = 13$$

$$P_e = \frac{1}{28^2} * ((15 * 15) + (13 * 13)) = 0.5026$$

$$\kappa = \frac{0.7857 - .5026}{1 - .5026} = 0.569$$

3. Confusion Matrix

Table 6: Confusion Matrix for Task 2.3

	Yes	No
Yes	12	3
No	3	10

(a) IAA using percent agreement

$$\text{Yes/Yes} = 12$$

$$\text{No/No} = 10$$

$$\text{Total Agreements} = 22$$

$$\text{Number of comparisons} = 12 + 3 + 3 + 10 = 28$$

$$\text{Percent Agreement} = 22/28 * 100 \approx 78.57\%$$

(b) Calculate IAA using Bennett et al's S

$$Q = 2$$

$$P_a = .7857$$

$$S = \frac{2 * .7857 - 1}{2 - 1} = 0.5714$$

(c) IAA using Scott's pi

$$P_o = .7857$$

$$P_e = 0.5026$$

$$\pi = \frac{.7857 - .5026}{1 - .5026} = .569$$

4. Qualitatively describe how percent agreement, Bennett et al's S, Scott's pi, and Cohen's kappa differ in their results based on questions 1-3.

Percent Agreement: Results 1: 50%, Result 2: 64.71%, Result 3: 78.51%.

Bennett et al.'s S: Results 1: .25, Result 2: 0.4704, Result 3: 0.5714.

Scott's Pi: Results 1: 0.1665, Result 2: 0.471, Result 3: 0.569.

Cohen's Kappa: Results 1: 0.1882, Result 2: 0.569, Result 3: 0.569.

For each questions we see that everything except Cohen's Kappa increases. This makes sense for a few reasons. Between questions 1 and 2 the data is much more similar in question 2 than in task one meaning that pretty much every single value we calculate should also be higher in question 2. This is because each of these are calculating some form of similarity or error between the two raters. The difference for Task 3 is that we removed the raters Don't Know category. This made the similarity between the raters go up even more as the data was fairly symmetrical for just yes and no, Cohen's Kappa did not increase as we already were doing this calculation without the Don't Know category.

5. Calculate IAA using Fleiss kappa.

Calculate P_o :

$$P_o(1) = \frac{1}{8(8-1)} [0(0-1) + 6(6-1) + 2(2-1) + 0(0-1)] = \frac{32}{56} = 0.5714$$

$$P_o(2) = \frac{1}{8(8-1)} [1(1-1) + 2(2-1) + 4(4-1) + 1(1-1)] = \frac{14}{56} = 0.25$$

$$P_o(3) = \frac{1}{8(8-1)} [2(2-1) + 0(0-1) + 2(2-1) + 4(4-1)] = \frac{16}{56} = 0.2857$$

$$P_o(4) = \frac{1}{8(8-1)} [0(0-1) + 1(1-1) + 6(6-1) + 1(1-1)] = \frac{30}{56} = 0.5357$$

$$P_o(5) = \frac{1}{8(8-1)} [0(0-1) + 1(1-1) + 5(5-1) + 2(2-1)] = \frac{22}{56} = 0.3929$$

$$P_o(6) = \frac{1}{8(8-1)} [1(1-1) + 6(6-1) + 1(1-1) + 0(0-1)] = \frac{30}{56} = 0.5357$$

$$P_o(7) = \frac{1}{8(8-1)} [1(1-1) + 6(6-1) + 0(0-1) + 1(1-1)] = \frac{30}{56} = 0.5357$$

$$\hat{P}_o = \frac{0.5714+0.25+0.2857+0.5357+0.3929+0.5357+0.5357}{7} = 0.4444$$

$$p_1 = \frac{1+1+2+0+0+1+1}{8 \times 7} = \frac{6}{56} = 0.1071$$

$$p_2 = \frac{6+2+0+1+1+6+6}{8 \times 7} = \frac{22}{56} = 0.3929$$

$$p_3 = \frac{2+4+2+6+5+1+0}{8 \times 7} = \frac{20}{56} = 0.3571$$

$$p_4 = \frac{0+1+4+1+2+0+1}{8 \times 7} = \frac{9}{56} = 0.1607$$

$$\hat{P}_e = (0.1071^2) + (0.3929^2) + (0.3571^2) + (0.1607^2)$$

$$\hat{P}_e = 0.0115 + 0.1544 + 0.1275 + 0.0258 = 0.3192$$

$$\kappa = \frac{\hat{P}_o - \hat{P}_e}{1 - \hat{P}_e} = \frac{0.4444 - 0.3192}{1 - 0.3192} = \frac{0.1252}{0.6808} = 0.1839$$

6. Different Annotators and columns represent items to be annotated

(a) Krippendorff's Alpha

Obseverd Disagreement

$$A \text{ vs } B = \frac{2}{9} = 0.2222$$

$$A \text{ vs } C = \frac{1}{9} = 0.1111$$

$$A \text{ vs } D = \frac{2}{9} = 0.2222$$

$$B \text{ vs } C = \frac{2}{9} = 0.2222$$

$$B \text{ vs } D = \frac{2}{9} = 0.2222$$

$$C \text{ vs } D = \frac{2}{9} = 0.2222$$

$$D_o = \frac{0.2222+0.1111+0.2222+0.2222+0.2222+0.2222}{6} = 0.2037$$

Expected Disagreement

Label 1 Frequency: 10

Label 2 Frequency: 10

Label 3 Frequency: 9

$$p_1 = 10/29 = .3448$$

$$p_2 = 10/29 = .3448$$

$$p_3 = 9/29 = .3103$$

$$D_e = 1 - \sum_{k=1}^3 P_k^2 = 1 - ((0.3448)^2 + (0.3448)^2 + (0.3103)^2) = 0.6659$$

Krippendorff's Alpha

$$\alpha = 1 - \frac{0.3254}{0.6659} = 0.6941$$

(b) recognition error

Ignore non annotated items KA

Obseverd Disagreement

$$A \text{ vs } B = \frac{2}{6} = 0.3333$$

$$A \text{ vs } C = \frac{1}{5} = 0.2$$

$$A \text{ vs } D = \frac{2}{7} = 0.2957$$

$$B \text{ vs } C = \frac{2}{5} = 0.4$$

$$B \text{ vs } D = \frac{2}{6} = 0.3333$$

$$C \text{ vs } D = \frac{2}{5} = 0.4$$

$$D_o = \frac{0.3333+0.2+0.2857+0.4+0.3333+0.4}{6} = 0.3254$$

Expected Disagreement

Label 1 Frequency: 10

Label 2 Frequency: 10

Label 3 Frequency: 9

$$p_1 = 10/29 = .3448$$

$$p_2 = 10/29 = .3448$$

$$p_3 = 9/29 = .3103$$

$$D_e = 1 - \sum_{k=1}^3 P_k^2 = 1 - ((0.3448)^2 + (0.3448)^2 + (0.3103)^2) = 0.6659$$

Krippendorff's Alpha

$$\alpha = 1 - \frac{0.3254}{0.6659} = 0.5114$$

recognition error

$$\text{Recognition error} = \alpha_{all} - \alpha_{labeled} = 0.6941 - 0.5114 = 0.1827$$

(c) K-vitality

K alpha no A

Obseverd Disagreement

$$B \text{ vs } C = \frac{2}{9} = 0.2222$$

$$B \text{ vs } D = \frac{2}{9} = 0.2222$$

$$C \text{ vs } D = \frac{2}{9} = 0.2222$$

$$D_o = \frac{0.2222+0.2222+0.2222}{3} = 0.2222$$

Expected Disagreement

Label 1 Frequency: 8

Label 2 Frequency: 7

Label 3 Frequency: 6

$$p_1 = 8/21 = .3809$$

$$p_2 = 7/21 = .3333$$

$$p_3 = 6/21 = .2857$$

$$D_e = 0.663$$

Krippendorff's Alpha

$$\alpha = 1 - \frac{0.2222}{0.663} = 0.6647$$

K alpha no B

Obseverd Disagreement

$$A \text{ vs } C = \frac{1}{9} = 0.1111$$

$$A \text{ vs } D = \frac{2}{9} = 0.2222$$

$$C \text{ vs } D = \frac{2}{9} = 0.2222$$

$$D_o = \frac{0.1111+0.2222+0.2222}{3} = 0.1852$$

Expected Disagreement

Label 1 Frequency: 6

Label 2 Frequency: 9

Label 3 Frequency: 7

$$p_1 = 6/21 = .2857$$

$$p_2 = 9/21 = .4286$$

$$p_3 = 7/21 = .3333$$

$$D_e = 0.6236$$

Krippendorff's Alpha

$$\alpha = 1 - \frac{0.1852}{0.6236} = 0.7030$$

K alpha no C

Obseverd Disagreement

$$A \text{ vs } B = \frac{2}{9} = 0.2222$$

$$A \text{ vs } D = \frac{2}{9} = 0.2222$$

$$B \text{ vs } D = \frac{2}{9} = 0.2222$$

$$D_o = \frac{0.2222+0.2222+0.2222}{3} = 0.2222$$

Expected Disagreement

Label 1 Frequency: 10

Label 2 Frequency: 7

Label 3 Frequency: 6

$$p_1 = 10/23 = .4348$$

$$p_2 = 7/23 = .3043$$

$$p_3 = 6/23 = .2609$$

$$D_e = 0.6502$$

Krippendorff's Alpha

$$\alpha = 1 - \frac{0.2222}{0.6502} = 0.6581$$

K alpha no D

Observed Disagreement

$$A \text{ vs } B = \frac{2}{9} = 0.2222$$

$$A \text{ vs } C = \frac{1}{9} = 0.1111$$

$$B \text{ vs } C = \frac{2}{9} = 0.2222$$

$$D_o = \frac{0.2222+0.1111+0.2222}{3} = 0.1852$$

Expected Disagreement

Label 1 Frequency: 9

Label 2 Frequency: 7

Label 3 Frequency: 5

$$p_1 = 9/21 = .4286$$

$$p_2 = 7/21 = .3333$$

$$p_3 = 5/21 = .2381$$

$$D_e = 0.6485$$

Krippendorff's Alpha

$$\alpha = 1 - \frac{0.1852}{0.6485} = 0.7144$$

K Vitality

$$\text{A K Vitality} = 0.6941 - 0.6633 = 0.0308$$

$$\text{B K Vitality} = 0.6941 - 0.7030 = -0.0089$$

$$\text{C K Vitality} = 0.6941 - 0.6633 = 0.0360$$

$$\text{D K Vitality} = 0.6941 - 0.6633 = -0.0203$$

Based on the above calculations Rater C has the most positive impact while rater D has the most negative.