Module 2 – Section 7

Measures of Agreement

Variables

- n observations rated by two judges
- Variable 1 = Rating from Judge 1
  - J categories
- Variable 2 = Rating from Judge 2
  - J categories
- Categories can be ordered (usually) or unordered and are the same for each judge

Data

Judge 1	Judge 2
Level 1	Level 2
Level 2	Level 1
Level 3	Level 3
⋮	⋮
⋮	⋮
Level 4	Level 3
Level 4	Level 4

Contingency Table

- Form a contingency table based on the ratings of the two judges on the n observations (sample).

		Judge 2			
Judge 1	Level 1	Level 2	Level 3	Level 4	Total
Level 1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	$Y_{1.}$
Level 2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	$Y_{2.}$
Level 3	Y_{31}	Y_{32}	Y_{33}	Y_{34}	$Y_{3.}$
Level 4	Y_{41}	Y_{42}	Y_{43}	Y_{44}	$Y_{4.}$
Total	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	$Y_{.4}$	n

Agreement

- Agreement = both judges giving same rating to an observation.
- Study level of agreement between two judges

Population Proportions

| | Judge 2 | | | | |
Judge 1	Level 1	Level 2	Level 3	Level 4	Total
Level 1	p_{11}	p_{12}	p_{13}	p_{14}	$p_{1.}$
Level 2	p_{21}	p_{22}	p_{23}	p_{24}	$p_{2.}$
Level 3	p_{31}	p_{32}	p_{33}	p_{34}	$p_{3.}$
Level 4	p_{41}	p_{42}	p_{43}	p_{44}	$p_{4.}$
Total	$p_{.1}$	$p_{.2}$	$p_{.3}$	$p_{.4}$	1

Agreement between Judges

- Probability the two judges are in agreement:

$$\sum_{j=1}^{J} p_{jj}$$

- Perfect agreement = 1

Agreement between Judges

- If ratings are independent, the probability of agreement is

$$\sum_{j=1}^{J} p_{jj} = \sum_{j=1}^{J} p_{j.}p_{.j}$$

Cohen's Kappa

- Compare probability of agreement to what is expected if judges rate independently

$$\kappa = \frac{\sum_{j=1}^{J} p_{jj} - \sum_{j=1}^{J} p_{j.}p_{.j}}{1 - \sum_{j=1}^{J} p_{j.}p_{.j}}$$

Properties of κ

- If there is perfect agreement

  - $\kappa = \dfrac{\sum_{j=1}^{J} p_{jj} - \sum_{j=1}^{J} p_{j.}p_{.j}}{1 - \sum_{j=1}^{J} p_{j.}p_{.j}} = \dfrac{1 - \sum_{j=1}^{J} p_{j.}p_{.j}}{1 - \sum_{j=1}^{J} p_{j.}p_{.j}} = 1$

- If judges rate independently

  - $\kappa = \dfrac{\sum_{j=1}^{J} p_{jj} - \sum_{j=1}^{J} p_{j.}p_{.j}}{1 - \sum_{j=1}^{J} p_{j.}p_{.j}} = \dfrac{\sum_{j=1}^{J} p_{j.}p_{.j} - \sum_{j=1}^{J} p_{j.}p_{.j}}{1 - \sum_{j=1}^{J} p_{j.}p_{.j}} = 0$

Properties of κ

- If there is no agreement

$$\kappa = \frac{\sum_{j=1}^{J} p_{jj} - \sum_{j=1}^{J} p_{j.}p_{.j}}{1 - \sum_{j=1}^{J} p_{j.}p_{.j}} = \frac{0 - \sum_{j=1}^{J} p_{j.}p_{.j}}{1 - \sum_{j=1}^{J} p_{j.}p_{.j}} = \frac{- \sum_{j=1}^{J} p_{j.}p_{.j}}{1 - \sum_{j=1}^{J} p_{j.}p_{.j}} < 0$$

- Usually bounded away from -1
- Not usually concerned with negative values of κ.

Properties of κ

- Value depends on marginal distributions
- Small example:

| | Judge 2 | | |
Judge 1	Level 1	Level 2	Total
Level 1	0.40	0.09	0.49
Level 2	0.06	0.45	0.51
Total	0.46	0.54	1

| | Judge 2 | | |
Judge 1	Level 1	Level 2	Total
Level 1	0.80	0.10	0.90
Level 2	0.05	0.05	0.10
Total	0.85	0.15	1

$$p_{11} + p_{22} = 0.85$$

$$\kappa = 0.7$$

$$p_{11} + p_{22} = 0.85$$

$$\kappa = 0.32$$

Estimate of Cohen's Kappa

- Contingency Table

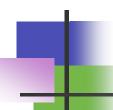| Judge 1	Judge 2				
	Level 1	Level 2	Level 3	Level 4	Total
Level 1	Y_{11}	Y_{12}	Y_{13}	Y_{14}	$Y_{1.}$
Level 2	Y_{21}	Y_{22}	Y_{23}	Y_{24}	$Y_{2.}$
Level 3	Y_{31}	Y_{32}	Y_{33}	Y_{34}	$Y_{3.}$
Level 4	Y_{41}	Y_{42}	Y_{43}	Y_{44}	$Y_{4.}$
Total	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	$Y_{.4}$	n

Estimate of Cohen's Kappa

$$\hat{\kappa} = \frac{\sum_{j=1}^{J} \frac{Y_{jj}}{n} - \sum_{j=1}^{J} \frac{Y_{j.}Y_{.j}}{n^2}}{1 - \sum_{j=1}^{J} \frac{Y_{j.}Y_{.j}}{n^2}} = \frac{n\sum_{j=1}^{J} Y_{jj} - \sum_{j=1}^{J} Y_{j.}Y_{.j}}{n^2 - \sum_{j=1}^{J} Y_{j.}Y_{.j}}$$

Example

- Rating of Student Teachers by 2 Supervisors
- Rating choices
  - Authoritarian (A)
  - Democratic (D)
  - Permissive (P)

Ex. Data

Judge 1	Judge 2
A	A
A	A
A	A
⋮	⋮
⋮	⋮
P	P
P	P

Ex. Judges Rating

		Judge 2 Rating		
Judge 1 Rating	A	D	P	Total
A	17	4	8	29
D	5	12	0	17
P	10	3	13	26
Total	32	19	21	72

Ex. Estimate of Cohen's Kappa

$$\hat{\kappa} = \frac{72(17 + 12 + 13) - (29(32) + 19(17) + 21(26))}{72^2 - (29(32) + 19(17) + 21(26))}$$

$$= 0.361$$

Interpretation: Estimate of Cohen's Kappa

Table 3.

Interpretation of Cohen's kappa.

Value of Kappa	Level of Agreement	% of Data that are Reliable
0–.20	None	0–4%
.21–.39	Minimal	4–15%
.40–.59	Weak	15–35%
.60–.79	Moderate	35–63%
.80–.90	Strong	64–81%
Above .90	Almost Perfect	82–100%

Ex. Judges Rating

- There is minimal agreement between the two judges on the ratings of the student teachers.

Weighted Cohen's Kappa

- κ only measures absolute agreement.

- Degree of agreement measured by observations in off diagonal cells is not captured.

Population Proportions

Judge 1	Judge 2				Total
	Level 1	Level 2	Level 3	Level 4	Total
Level 1	p_{11}	p_{12}	p_{13}	p_{14}	$p_{1.}$
Level 2	p_{21}	p_{22}	p_{23}	p_{24}	$p_{2.}$
Level 3	p_{31}	p_{32}	p_{33}	p_{34}	$p_{3.}$
Level 4	p_{41}	p_{42}	p_{43}	p_{44}	$p_{4.}$
Total	$p_{.1}$	$p_{.2}$	$p_{.3}$	$p_{.4}$	1

Population Proportions

- Darkest Cells – Agreement
- Medium Cells – Ratings are 1 cell off from agreement.
- Lightest Cells – Ratings are 2 cells off from agreement.
- White Cells – Ratings are 3 cells off from agreement.

Weighted Cohen's Kappa

- Takes into account "distance" from agreement in the table.

- Each cell is weighted according to this distance.

- Many different weights are possible.
  - We will use "squared" weights.

Weighted Cohen's Kappa

$$\kappa_w = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}p_{ij} - \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}p_{i.}p_{.j}}{1 - \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ij}p_{i.}p_{.j}}$$

$$w_{ij} = 1 - \frac{(i-j)^2}{(J-1)^2}$$

Estimated Weighted Cohen's Kappa

$$\hat{\kappa}_w = \frac{n \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ij} Y_{ij} - \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ij} Y_{i.} Y_{.j}}{n^2 - \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ij} Y_{i.} Y_{.j}}$$

$$w_{ij} = 1 - \frac{(i - j)^2}{(J - 1)^2}$$

Ex. Judges Rating

- $w_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0.25 & \text{if } |i - j| = 1 \\ 0 & \text{if } |i - j| = 2 \end{cases}$

- $\hat{\kappa}_w = 0.2156$

- There is minimal agreement between the two judges on the ratings of the student teachers.