PLSC 502 – Autumn 2016
Measures of Association
Ordinal Variables

November 3, 2016

- Key issue: *how to retain the information present in the ordering of the categories without giving the numerical values assigned to them cardinal content*.
- Key concept: **Concordance**

For a pair of values on two observations $i = \{1, 2\}$ and two variables X and Y, a *concordant pair* has:

$$\text{sign}(X_2 - X_1) = \text{sign}(Y_2 - Y_1)$$

and a *discordant* pair has:

$$\text{sign}(X_2 - X_1) = -\text{sign}(Y_2 - Y_1).$$

A(nother) Contingency Table

Consider:

		X 1	2	3	
Y	1	n_{11}	n_{12}	n_{13}	n_{1X}
	2	n_{21}	n_{22}	n_{23}	n_{2X}
	3	n_{31}	n_{32}	n_{33}	n_{3X}
		n_{Y1}	n_{Y2}	n_{Y3}	N

Concordance with $\{1, 1\}$ observations:

		X 1	2	3	
	1	n_{11}	n_{12}	n_{13}	n_{1X}
Y	2	n_{21}	n_{22}	n_{23}	n_{2X}
	3	n_{31}	n_{32}	n_{33}	n_{3X}
		n_{Y1}	n_{Y2}	n_{Y3}	N

Concordance with $\{1, 2\}$ observations:

			X		
		1	2	3	
	1	n_{11}	n_{12}	n_{13}	n_{1X}
Y	2	n_{21}	n_{22}	n_{23}	n_{2X}
	3	n_{31}	n_{32}	n_{33}	n_{3X}
		n_{Y1}	n_{Y2}	n_{Y3}	N

Concordant and Discordant Pairs

Discordance with $\{1, 2\}$ observations:

		X			
		1	2	3	
	1	n_{11}	n_{12}	n_{13}	n_{1X}
Y	2	n_{21}	n_{22}	n_{23}	n_{2X}
	3	n_{31}	n_{32}	n_{33}	n_{3X}
		n_{Y1}	n_{Y2}	n_{Y3}	N

Discordance with $\{1, 3\}$ observations:

		X 1	2	3	
	1	n_{11}	n_{12}	n_{13}	n_{1X}
Y	2	n_{21}	n_{22}	n_{23}	n_{2X}
	3	n_{31}	n_{32}	n_{33}	n_{3X}
		n_{Y1}	n_{Y2}	n_{Y3}	N

Concordant and Discordant Pairs

For a 3×3 table, the total number of *concordant pairs* is:

$$N_c = n_{11}(n_{22} + n_{23} + n_{32} + n_{33}) + n_{12}(n_{23} + n_{33}) + n_{21}(n_{32} + n_{33}) + n_{22}(n_{33})$$

and the total number of *discordant pairs* is:

$$N_d = n_{13}(n_{21} + n_{22} + n_{31} + n_{32}) + n_{12}(n_{21} + n_{31}) + n_{23}(n_{31} + n_{32}) + n_{22}(n_{31}).$$

This extends to a table of arbitrary size $M \times N$ straightforwardly...

Gamma (γ)

Gamma (γ) is the normed difference between the number of concordant and discordant pairs in the data:

$$\gamma = \frac{N_c - N_d}{N_c + N_d}$$

Equivalently:

$$\gamma = \frac{N_c}{N_c + N_d} - \frac{N_d}{N_c + N_d}$$

About γ

Gamma:

- does not count "ties."
- $\gamma \in [-1, 1]$.
- $\gamma = 0 \leftrightarrow$ no association between X and Y, though it can also happen whenever $N_c = N_d$. That is, $\gamma = 0$ is necessary but not sufficient for statistical independence.
- Higher absolute values of γ correspond to stronger associations between X and Y.
- $\gamma = \pm 1.0$ under conditions of (at least) *weak monotonicity* (e.g., γ will equal 1.0 whenever, as X increases, Y only increases or stays the same).

Inference on γ

Can be shown that:

$$\hat{\gamma} \sim \mathcal{N}(\gamma, \sigma_\gamma^2)$$

where

$$\sigma_\gamma^2 = \frac{N(1-\hat{\gamma}^2)}{N_c + N_d}$$

So

$$t \approx (\hat{\gamma} - \gamma)\sqrt{\frac{N_c + N_d}{N(1-\hat{\gamma}^2)}}.$$

τ(s)

(Goodman-Kruskal's) "Tau-a":

$$\tau_a = \frac{N_c - N_d}{\frac{1}{2}N(N-1)}$$

(Kendall's) "Tau-b":

$$\tau_b = \frac{N_c - N_d}{\sqrt{[(N_c + N_d + N_{Y^*})(N_c + N_d + N_{X^*})]}}$$

where N_{Y^*} and N_{X^*} are the number of pairs *not tied* on Y and X, respectively.

(Stuart's) "Tau-c":

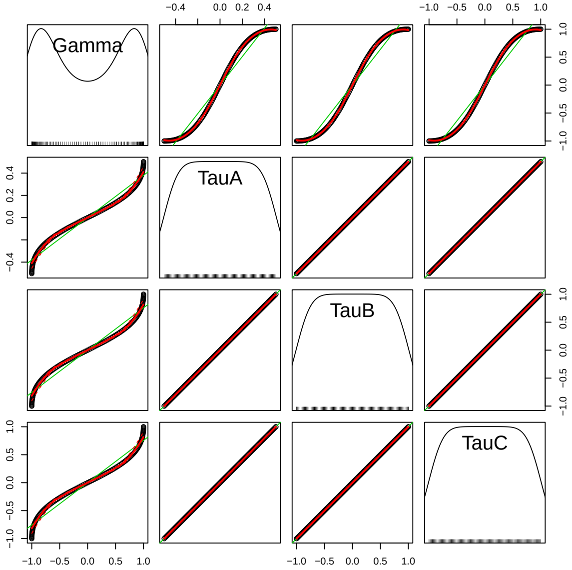
$$\tau_c = (N_c - N_d) \times \left\{ \frac{2m}{[N^2 2(m-1)]} \right\}$$

where m is the number of rows or columns, whichever is smaller.

τ Traits & Tips

- All except τ_a have $\tau_{(\cdot)} \in [-1, 1]$
- For all τs, the numerator signs the statistic.
- Like γ, τ_a doesn't do "ties" $\rightarrow$ attenuated range
- $|\tau_b| = 1.0$ only under *strict monotonicity*
- $\tau_b \rightarrow$ "square" tables
- $\tau_c \rightarrow$ "rectangular" (asymmetrical) tables
- $\gamma \geq \tau \ \forall \ \tau_{(\cdot)}$

γ and the τs Compared



Example: Sarah Palin Support...

September 2008 "Battleground" Poll in PA:

```
> summary(MamaGriz)
    caseid          female                            palin     
 Min.   :    2   Female:2370   Very Unfavorable    :1200  
 1st Qu.:30034   Male  :2221   Somewhat Unfavorable: 739  
 Median :31831                 Somewhat Favorable  :1132  
 Mean   :36776                 Very Favorable      :1520  
 3rd Qu.:60674                                             
 Max.   :62125                                             
          pid      
 Democrat   :1709  
 Independent:1391  
 GOP        :1491  
```

Palin Approval and Party ID

```
> palinpid<-with(MamaGriz, xtabs(~palin+pid))

> addmargins(palinpid)
                      pid
palin                   Democrat Independent  GOP  Sum
  Very Unfavorable           881         282   37 1200
  Somewhat Unfavorable       441         245   53  739
  Somewhat Favorable         291         412  429 1132
  Very Favorable              96         452  972 1520
  Sum                       1709        1391 1491 4591
```

Estimating γ and the τs

```
> # Gamma:
>
> GoodmanKruskalGamma(palinpid,conf.level=0.95)
  gamma  lwr.ci  ups.ci
0.73376 0.71529 0.75223

> #Tau-A:
>
> KendallTauA(palinpid,conf.level=0.95)
  tau_a  lwr.ci  ups.ci
0.38762 0.38639 0.38884

> # Tau-B:
>
> KendallTauB(palinpid,conf.level=0.95)
  tau_b  lwr.ci  ups.ci
0.55453 0.53784 0.57121

> # Tau-C:
>
> StuartTauC(palinpid,conf.level=0.95)
   tauc  lwr.ci  ups.ci
0.58130 0.56401 0.59859
```

Men vs. Women on Palin

```
> palinfemale<-with(MamaGriz, xtabs(~palin+female))

> addmargins(palinfemale)
                      female
palin                  Male Female  Sum
  Very Unfavorable      508    692 1200
  Somewhat Unfavorable  328    411  739
  Somewhat Favorable    575    557 1132
  Very Favorable        810    710 1520
  Sum                  2221   2370 4591
```

Men vs. Women on Palin

```
> GoodmanKruskalGamma(palinfemale,conf.level=0.95)
    gamma    lwr.ci    ups.ci
-0.136410 -0.179514 -0.093306

> KendallTauA(palinfemale,conf.level=0.95)
    tau_a    lwr.ci    ups.ci
-0.050259 -0.051137 -0.049382

> KendallTauB(palinfemale,conf.level=0.95)
    tau_b    lwr.ci    ups.ci
-0.082912 -0.109268 -0.056556

> StuartTauC(palinfemale,conf.level=0.95)
     tauc    lwr.ci    ups.ci
-0.100497 -0.132442 -0.068552
```