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	
	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

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

		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 $\{1, 2\}$ observations:

$$
\begin{array}{cc|cccc}
 & & & X & & \\
 & & 1 & 2 & 3 & \\
\hline
 & 1 & n_{11} & n_{12} & n_{13} & n_{1X} \\
Y & 2 & n_{21} & n_{22} & n_{23} & n_{2X} \\
 & 3 & n_{31} & n_{32} & n_{33} & n_{3X} \\
\hline
 & & n_{Y1} & n_{Y2} & n_{Y3} & N
\end{array}
$$

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

Concordant and Discordant Pairs

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

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

So

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

Kendall's τ(s)

"Tau-a":

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

"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.

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

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  
```

```
Gamma2.f<-function(x, pr=0.95)
{
    # x is a matrix of counts.  You can use output of crosstabs or xtabs in R.
    # A matrix of counts can be formed from a data frame by using design.table.

    # Confidence interval calculation and output from Greg Rodd

    # Check for using S-PLUS and output is from crosstabs (needs >= S-PLUS 6.0)
    if(is.null(version$language) && inherits(x, "crosstabs")) { oldClass(x)<-NULL;
attr(x, "marginals")<-NULL}

    n <- nrow(x)
    m <- ncol(x)
    pi.c<-pi.d<-matrix(0,nr=n,nc=m)

    row.x<-row(x)
    col.x<-col(x)

    for(i in 1:(n)){
        for(j in 1:(m)){
            pi.c[i, j]<-sum(x[row.x<i & col.x<j]) + sum(x[row.x>i & col.x>j])
            pi.d[i, j]<-sum(x[row.x<i & col.x>j]) + sum(x[row.x>i & col.x<j])
        }
    }

    C <- sum(pi.c*x)/2
    D <- sum(pi.d*x)/2

    psi<-2*(D*pi.c-C*pi.d)/(C+D)^2
    sigma2<-sum(x*psi^2)-sum(x*psi)^2

    gamma <- (C - D)/(C + D)
    pr2 <- 1 - (1 - pr)/2
    CIa <- qnorm(pr2) * sqrt(sigma2) * c(-1, 1) + gamma

    list(gamma = gamma, C = C, D = D, sigma = sqrt(sigma2), Level = paste(
        100 * pr, "%", sep = ""), CI = paste(c("[", max(CIa[1], -1),
        ", ", min(CIa[2], 1), "]"), collapse = ""))
}
```

```
> Gamma2.f(palinpid)
$gamma
[1] 0.73376

$C
[1] 4824989

$D
[1] 740927

$sigma
[1] 0.0094232

$Level
[1] "95%"

$CI
[1] "[0.715293551681856, 0.752232009250073]"
```

Kendall's τ

```
> with(MamaGriz, cor.test(PID,Palin,method="kendall"))

Kendall's rank correlation tau

data:  PID and Palin
z = 43.5, p-value <2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
    tau
0.55453
```

Men vs. Women on Palin

```
> palinfemale<-with(MamaGriz, xtabs(~palin+female))
> addmargins(palinfemale)
                       female
palin                   Female Male  Sum
  Very Unfavorable         692  508 1200
  Somewhat Unfavorable     411  328  739
  Somewhat Favorable       557  575 1132
  Very Favorable           710  810 1520
  Sum                     2370 2221 4591
```

Men vs. Women on Palin

```
> Gamma2.f(palinfemale)
$gamma
[1] 0.13641

$sigma
[1] 0.021992

$Level
[1] "95%"

$CI
[1] "[0.0933060117469164, 0.17951420622549]"
```

```
> MamaGriz$Female <- with(MamaGriz, 1 - (as.integer(female)-1))
> with(MamaGriz, cor.test(Female,Palin,method="kendall"))

Kendall's rank correlation tau

data:  Female and Palin
z = -6.13, p-value = 8.9e-10
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
-0.082912
```