Chapter 10 measures of strength of association in crosstab

Two aspects in statistics of bivariate situation 1) the significant of the relationship (p value), 2) strength of the association (how strongly the two variables are correlated). In the context of crosstab, we discuss lambda and gamma.

1) Lambda (λ)

$$\lambda = \frac{\#errors\ without\ the\ independent\ variable - \#\ errors\ with\ the\ independent\ variable}{\#errors\ without\ the\ independent\ variable}$$

	wm	ww	bm	bw	RM
Joe Biden	495	577	67	92	1,231
Donald Trump	687	597	19	14	1,317
CM	1,182	1,174	86	106	N = 2,548

$$\lambda = \frac{\#errors\ without\ the\ independent\ variable - \#\ errors\ with\ the\ independent\ variable}{\#errors\ without\ the\ independent\ variable}$$

$$\#errors\ without\ the\ independent\ variable = 1{,}231$$

$$\#\ errors\ with\ the\ independent\ variable = 495 + 577 + 19 + 14 = 1{,}105$$

$$\lambda = \frac{1{,}231 - 1{,}105}{1{,}231} = 10.2\%$$

Lambda is PRE (Proportion Reduction in Error)

PRE indicates that knowing the independent variable reduces x% of errors in estimating the value of the dependent variable.

Knowing race/gender of the voter reduces errors in estimating their votes by 10.2%.

$$0 \leq \lambda \leq 1\ or\ 100\%$$

Which one is better between lambda = 20%, and the other lambda = 80%. 80% is better

2) Exercise: calculating and explaining lambda (λ) for the following crosstab

	men	women	RM
Pickup trucks	86	63	149
SUVs	19	5	24
Sedans	2	10	12
CM			N = 185

$$\lambda = \frac{\#errors\ without\ the\ independent\ variable - \#\ errors\ with\ the\ independent\ variable}{\#errors\ without\ the\ independent\ variable}$$

$$\#errors\ without\ the\ independent\ variable = 36$$

$$\#\ errors\ with\ the\ independent\ variable = 21 + 15 = 36$$

$$\lambda = \frac{36 - 36}{36} = 0\%$$

3) Gamma (G)

Gamma can be used to measure the strength of association in crosstab between two ordinal variables.

$$G = \frac{\#same - \#opposite}{\#same + \#\ opposite}$$

$\#same = same\ rank\ order$; $\#opposite = opposite\ rank\ order$

	No treatment	Moderate treatment	Great deals of treatment
Great improvement	52	34	96
Moderate improvement	58	27	49
No improvement	112	38	10

$$\#same = 58 \times 34 + 58 \times 96 + 27 \times 96 + 112 \times 27 + 112 \times 49 + 112 \times 34 + 112 \times 96 + 38 \times 49 + 38 \times 96 = 38{,}714$$

$$\#opposite = 52 \times 27 + 52 \times 49 + 52 \times 38 + 52 \times 10 + 34 \times 49 + 34 \times 10 + 58 \times 38 + 58 \times 10 + 27 \times 10 = 11{,}508$$

$$G = \frac{\#same - \#opposite}{\#same + \#\ opposite} = \frac{38{,}714 - 11{,}508}{38{,}714 + 11{,}508} = 54\%$$

4) Range and interpretation of Gamma (G)

$$-1 \leq G \leq 1; Gamma\ is\ PRE$$

PRE indicates that knowing the independent variable reduces x% of errors in estimating the value of the dependent variable.

Knowing drug treatment reduces errors in estimating the health improvement by 54%, and this relationship is positive.

5) Exercise: calculating and interpreting Gamma

	<HS	HS	College
better	26	75	38
Worse	49	77	16

$$\#same = 49 \times 75 + 49 \times 38 + 77 \times 38 = 8{,}463$$

$$\#opposite = 26 \times 77 + 26 \times 16 + 75 \times 16 = 3{,}618$$

$$G = \frac{\#same - \#opposite}{\#same + \#\ opposite} = \frac{8463 - 3618}{8463 + 3618} = 40\%$$

Knowing the level of education reduces errors in estimating the financial situation by 40%, and the relationship is positive.