Measures of Association with Chi-square Tests

While chi-square helps us understand whether the studied relationship is statistically significant, we look to additional measures of association to tell us just how strong and consistent that relationship is. Here's are some tests to consider:

Test	Shape	Direction	Range	PRE	Variable type
Phi (φ)	Square	Symmetric	0 to 1	Non-PRE	Dich. x Dich.
Cramer's *V*	Either	Symmetric	0 to 1	Non-PRE	Nom. x Nom.
Lambda (λ)	Either	Asymmetric	0 to 1	PRE	Nom. x Nom
Somers' *d*	Either	Asymmetric	-1 to 1	PRE	Ord. x Ord.
Gamma (γ)	Either	Symmetric	-1 to 1	PRE	Ord. x Ord.
Tau-b (τ_b)	Square	Symmetric	-1 to 1	PRE	Ord. x Ord.
Tau-c (τ_c)	Rectangle	Symmetric	-1 to 1	PRE	Ord. x Ord.

Proportional Reduction in Error (PRE) measures of association tell us how much the independent variable helps us predict the outcome of the dependent variable compared to random chance/guessing. In other words, PRE measures allow us to infer the level of predictive accuracy in using the IV to explain the DV; non-PRE measures don't. Instead, non-PRE measures of association can be useful to help us gauge similarity in strength between categories (especially in controlled comparisons).

<u>Choosing the appropriate test:</u>
Phi (φ): dichotomous by dichotomous (square, i.e., 2x2 table)
Cramer's *V*: nominal by nominal (symmetrical)
Lambda (λ): nominal by nominal (asymmetrical)
Somers' *d*: ordinal by ordinal (asymmetrical)
Gamma (γ): ordinal by ordinal (symmetrical)

- **Tau-b (τ_b):** ordinal by ordinal (square)
- **Tau-c (τ_c):** ordinal by ordinal (rectangle)

The directionality, **Symmetric/Asymmetric**, refers to the importance the test places on causal direction between the variables of interest. If the test treats each variable equally, as though either variable could be the DV or IV then it is **symmetric**. If the result of the test will change depending on which variable is the DV and which variable is the IV, then the test is *directional* or **asymmetric**.

The shape, **Square/Rectangle**, refers to the structure of the crosstabulation table. When the dependent variable (DV) and the independent variable (IV) have the ***same*** number of categories, we consider that **square**. When the DV and IV have **different** numbers of categories, we consider that **rectangular**. See below for examples.

Square (e.g., 3x3)

	IV_{Cat1}	IV_{Cat2}	IV_{Cat3}
DV_{Cat1}	a	b	c
DV_{Cat2}	d	e	f
DV_{Cat3}	g	h	i

Rectangle (e.g., 3x4 or 3x2)

	IV_{Cat1}	IV_{Cat2}	IV_{Cat3}	IV_{Cat4}
DV_{Cat1}	a	b	c	d
DV_{Cat2}	e	f	g	h
DV_{Cat3}	i	j	k	l

	IV_{Cat1}	IV_{Cat2}
DV_{Cat1}	a	b
DV_{Cat2}	c	d
DV_{Cat3}	e	f

<u>Assessing the predictive value/strength of the test:</u>
Cramer's *V*, Lambda, and Phi all range from 0.0 to 1.0
Somers' *d*, Gamma, Tau-b, and Tau-c all range from -1.0 to 1.0

Regardless of the direction, values closer to the absolute value of 1.0 (so, 1.0 or -1.0) indicate a stronger relationship, and values closer to 0.0 indicate a weaker relationship. The ordinal measures of association simply include direction in addition to strength.

- From 0.0 to +/- 0.1: *very* weak relationship
- From +/- 0.1 to 0.2: weak relationship
- From +/- 0.2 to 0.3: moderate relationship
- From +/- 0.3 to 1.0: strong relationship

For Gamma (γ), we also care about significance:

- Gamma tells us how well the categories "match" each other.
- As the IV categories move from left to right, how consistent is the pattern among the DV categories?
- When gamma is **significant**, we can interpret the strength of the relationship just like above. When it's not, we can't.
- But gamma is also famous for being over-inflated (i.e., bigger than expected), so we regularly use Tau-b or Tau-c (depending on the shape of the crosstabulation) with it to help us gain additional insight.