

COLUMN percentages

- I used that phrase 33 times Tue 4/28, and 22 more on Tues 5/5
- I included an *entire slide* about this on Tues 5/5, including the quiz question that so many missed on 4/28: Crosstab cells should have the cell count (observed frequency) and...

The row percentage	1 respondent	7%	
All of these	10 respondents	71%	
The column percentage	2 respondents	14%	✓
The total percentage	1 respondent	0%	
The expected count	1 respondent	7%	

- Even fewer of the *same students* got it right the second time:

The row percentage	2 respondents	75%	
All of these	1 respondent	8%	
The column percentage	1 respondent	8%	✓
The total percentage	1 respondent	8%	
The expected count	1 respondent	8%	

- I'm worried about confusion about something I said 55 times ☹️

Outline for Today...

- Crosstabs
 - Basic Questions Review
 - Chi-square-based measures
- Measures of Association
 - Overview
 - Concordant & Discordant Pairs
 - Gamma, Tau-b
 - Skipping Tau-b math ☺️
- PRE Exercise

SOC424 w/ Dr. Ellis Godard -- Slide

Golden Quiz Bananas

- Highest Quiz Average in Each Section
- Tues Labs...
 - Shadie Arnold
- Thurs Labs...
 - Job Guiwa

SOC424 w/ Dr. Ellis Godard

Summary of Cross-tab Points

- First question: Is there a relationship?
 - See green/red/blue slide for basics
 - Compare across, Modal %s, Chi-square's p
- Three key questions for intervals (corr/reg)
 - Strength? – weak or strong?
 - Form? – linear? Curvilinear?
 - Direction? – positive? Negative?
 - Remember valence!!
- Three key questions for crosstabs, too:
 - Relationship? (chi-square)
 - Size/Strength? (gamma or lambda)
 - Direction? (gamma or percentage comparisons)

SOC424 w/ Dr. Ellis Godard -- Slide

SOC424 – Statistics
w/ Dr. Ellis Godard

Measures of (the strength of) Association

Chi-square-based measures

- χ^2 , varies with df of $(r-1)*(c-1)$:
$$\sum \frac{(f_o - f_e)^2}{f_e}$$
- χ^2 also varies with sample size, so Phi =
$$\sqrt{\frac{\chi^2}{n}}$$
- Pearson's contingency coefficient C =
$$\sqrt{\frac{\chi^2}{\chi^2 + n}}$$
- And, it varies w/ table size, so Cramer's V =
$$\sqrt{\frac{\chi^2}{\min(r, c) - 1}}$$

SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

PRE Measures

- **Gamma** (8.5, p.223) –
(Concordant–Discordant) / (Concordant+Discordant)
- Somer’s D
(concordant – discordant) / all ties
- Goodman & Kruskal’s tau b (8.4, p.224)
 - based on row marginals
- G&K’s tau c
 - based on column marginals
- **Lambda**
 - nominal only, based on modal category

11 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

Paired Cases in a Crosstab

- Who’s more likely to be on camera

	Male	Female
Never on camera	Paul	Patricia
Always on Camera	Peter	Penelope

- 1 pair of cases suggests women (Paul vs Penelope)
- 1 pair suggests men (Patricia & Peter)
- Other pairs are “ties” (males Paul & Peter, on camera Paul & Patricia, etc.)

- But now, more pairs suggest women than men:

	Male	Female
Never on camera	Paul	Patricia & Susan
Always on Camera	Peter & Mike	Penelope

14 SOC424 w/ Dr. Ellis Godard -- Slide

Review Pairs Gamma Interpretations Tau B PRE

A comparison, using 2 pairs

- If John got a higher degree than Mary and gets paid more, that pair of cases is **concordant** with the claim that educational and financial attainment go together

Note: these are NOT crosstabs!!

	John	Mary
Educ	Higher	Lower
Pay	Higher	Lower

- But if John got a higher degree than Mary and was paid less than she was, that pair of cases is **discordant**

	John	Mary
Educ	Higher	Lower
Pay	Lower	Higher

12 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

The Idea Behind “Gamma” () γ

- Compares most pairs of cases in the sample
- **Focuses on those that are concordant with a relationship** (that is, those that show that higher values on one variable are associated with higher values on the other one)...
- ...and those that are **discordant with a relationship** (that is, the reverse of or at discord with a positive relationship; a higher value on one variable but lower on the other)
- $\text{Gamma} = (C - D) / (C + D)$, that is, the difference as a fraction of the combination

15 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

C vs. D vs. “ties”

Note: these are NOT crosstabs!!

- Concordant pair:

	John	Mary
Educ	Higher	Lower
Pay	Higher	Lower

- Discordant pair:

	John	Mary
Educ	Higher	Lower
Pay	Lower	Higher

- Tied pairs:

	J	M
Educ	H	H
Pay	?	?

	J	M
Educ	?	?
Pay	L	L

13 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

Example with Larger Crosstab

Social Class	College Education	
	Less than college (Low)	College or more (High)
Low	15	5
Med	10	10
High	5	15
Total	30	30

16 SOC424 @ CSUN - Ellis Godard

Concordance vs. Discordance

Social Class	College Education	
	Less than college (Low)	College or more (High)
Low	15 L,L	5 H,L
Med	10 L,M	10 H,M
High	5 L,H	15 H,H
Total	30	30

For Gamma to be positive (above zero), we want more pairs of cases that are higher on both than are higher on one but lower on the other (for example, more pairs from blue to blue than from green to green)

17 SOC424 @ CSUN - Ellis Godard

Gamma: Discordant Cells

- Start w/ cell lowest on one, highest on other
- Which cells are concordant with it (lower on both vars)?
- Move down the cells, one at a time, going lower in each column then each row.
- Stop when there's nothing lower on both

	< College	College+
L Class	L,L	H,L
M Class	L,M	H,M
H Class	L,H	H,H

	< College	College+
L Class	L,L	H,L
M Class	L,M	H,M
H Class	L,H	H,H

	< College	College+
L Class	L,L	H,L
M Class	L,M	H,M
H Class	L,H	H,H

20 SOC424 @ CSUN - Ellis

Gamma: Concordant Cells

- Start w/ cell that's highest on each value
- Which cells are concordant with it (lower on both vars)?
- Move down the cells, one at a time, going lower in each column then each row.
- Stop when there's nothing lower on both

	< College	College+
L Class	L,L	H,L
M Class	L,M	H,M
H Class	L,H	H,H

	< College	College+
L Class	L,L	H,L
M Class	L,M	H,M
H Class	L,H	H,H

	< College	College+
L Class	L,L	H,L
M Class	L,M	H,M
H Class	L,H	H,H

18 SOC424 @ CSUN - Ellis

Gamma: Discordant Pairs

Multiply the pairs of observed cell counts, for each of set of discordant cells:

$$5 \times 10 = 50$$

$$5 \times 5 = 25$$

$$10 \times 5 = 50$$

$D = 50 + 25 + 50 = 125$ pairs of cases in which each case is higher on one variable than the other one

	< College	College+
L Class	15	5
M Class	10	10
H Class	5	15

	< College	College+
L Class	15	5
M Class	10	10
H Class	5	15

	< College	College+
L Class	15	5
M Class	10	10
H Class	5	15

21 SOC424 @ CSUN - Ellis Godard

Gamma: Concordant Pairs

Multiply the observed cell counts, for each of set of concordant cells (3 here):

$$15 \times 10 = 150$$

$$15 \times 15 = 225$$

$$10 \times 15 = 150$$

$C = 150 + 225 + 150 = 525$ pairs of cases in which one R is higher on both vars than the other

	< College	College+
L Class	15	5
M Class	10	10
H Class	5	15

	< College	College+
L Class	15	5
M Class	10	10
H Class	5	15

	< College	College+
L Class	15	5
M Class	10	10
H Class	5	15

19 SOC424 @ CSUN - Ellis Godard

Calculating Gamma

$$\gamma = \frac{C - D}{C + D}$$

$$= \frac{525 - 125}{525 + 125}$$

$$= \frac{400}{650} = 0.615$$

22 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

Interpreting Gamma: Overview

- Asks which pair type is dominant, Cs or Ds?
 - Does more of the data concord or discord w/ H_a ?
- If no pairs are at discord, $\text{Gamma} = C/C = 1$
 - strongest possible *positive* relationship
- If all pairs are at discord, $\text{Gamma} = -D/D = -1$
 - strongest possible *negative* relationship
- If all pairs are ties (no Cs or Ds) $\text{Gamma} = 0/0$
 - closer to zero, fewer pairs of cases to tell us either direction, and the relationship is called *weak*

23 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

Interlude: Fun w/ C & D

- Gamma = difference in proportions between concordant % & discordant %

$$\hat{\gamma} = \frac{C - D}{C + D} = \frac{C}{C + D} - \frac{D}{C + D}$$

$$= \frac{525}{650} - \frac{125}{650} = .808 - .192 = .616 \approx .615$$
- While 19.2% of the non-tied pairs are discordant, 80.8% are concordant — *four times* as many pairs represent a *positive* relationship as represent a *negative* one

26 SOC424 @ CSUN - Ellis C

Review Pairs Gamma Interpretations Tau B PRE

Interpreting Gamma 1: Direction

- If $\text{gamma} > 0$, relationship is positive
 - predict that a higher value on one probably *does* mean a higher value on the other
- If $\text{gamma} < 0$, relationship is negative
 - predict that a higher value on one probably means a *lower* value on the other
- In above example:
 - 0.615 -> positive relationship

24 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

What's a PRE?

- Proportional Reduction in Error
 - By what percentage do we reduce our error in predicting values of the *dependent* variable by knowing values of the *independent* variable?
 - For example, imagine that I want to guess which students hadn't finished an exam. By how much is my guesswork improved if I know which of you aren't on time for class the day that exam's due?

27 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

Interpreting Gamma 2: Strength

- Weak relationship if 0 to .3 (or 0 to -.3)
- Moderate if 0.3 to .6 (or -.3 to -.6)
- Strong if 0.6 to 1 (or -.6 to -1)

25 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau B PRE

PRE Formula

- All PREs follow same basic formula

$$(E1 - E2) / E1$$
- where E1 is the errors made (in predicting values of the dependent variable) when nothing is known about the independent,
- and E2 is the number of errors when the value of the independent variable is known

28 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PRE

- Errors under Rule 1 (E1):
 - All we know is that there are 650 pairs, so we predict they're all the same, C or D
 - Our estimated error rate here is 50% (0.5), like a flip of a coin – C or D on each side
 - $E1 = (C+D) * .5 = 650 * .5 = 325$
- Errors under Rule 2 (E):
 - Once we know the IV, we can predict that any random pair is of the dominant type (here, concordant). The pairs that aren't actually of the dominant type are errors.
 - Here, C is dominant so the 125 discordant pairs would be incorrectly predicted as concordant
 - $E2 = \min(C, D)$

29 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PRE

A Cautionary Note...

- Gamma assumes a higher *value* (i.e. number) means a higher *value label*
- If your variable is higher numbers for lower values (e.g. 1= Upper class, 2= middle, 3= lower) you'll need to *either* recode the variable to reverse the values *or* interpret gamma upside-down

32 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PRE

Gamma as PRE (Interp. Part 3)

- $(E1-E2) / E1 = (325 - 125) / 325 = .615$
- There are 61.5% fewer errors in predicting whether a pair is C or D if we know which type of pair happens more frequently, than there are if we *don't* know which type of pair is dominant

30 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PRE

Goodman & Kruskal's tau b

- Like gamma, tau b is asymmetrical
 - Infers a "direction" to the relationship
 - χ^2 is symmetrical – does not distinguish between DV and IV (though column percents *do*)
 - (plain) tau also symmetrical – use for 2 nominals
 - Use tau b for 2 ordinals
- Not absolute; relative only
 - below .5ish weaker; above .5ish is stronger

33 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PRE

Disadvantage of Gamma

- Does not take into account tied pairs
- Therefore overstates strength
 - Possible for gamma to suggest a perfect positive association (+1) with only one concordant pair: $(C-D)/(C+D) = (1-0)/(1+0) = 1$
- **Somer's D** takes ties into account $= (C-D) / (\text{all tied pairs})$

31 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PRE

Consider earlier dataset

- Political orientation = IV (columns)
- Child-rearing = DV (rows)
- E1 = errors if assign to categories of child-bearing w/o knowledge of polit. Affiliation
- E2 = errors once known polit. affiliation

	Conservative	Moderate	Liberal	Total
Permissive	7 (21.9%)	9 (30.0%)	14 (51.9%)	30
Moderate	10 (31.3%)	10 (33.3%)	8 (29.6%)	28
Authoritarian	15 (46.9%)	11 (36.7%)	5 (18.5%)	31
Col. Marginal	32	30	27	89

35 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PREs

G&K's Tau b: Errors w/ Rule 1

- E1 = error rate when we randomly assign all cases into categories of the DV while maintaining the marginal (row) distribution
- Here, we would randomly assign the 89 cases to the 3 categories of the DV, by row
 - We know the row marginals (univariate frequency of the DV) are 30, 28, and 31
- Two major steps:
 - Error rate = [row marginal * (N - RM)]/N
 - Errors = row marginal * error rate

36 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PREs

Measures to use...

- Chi-square always, to establish dependence
- If both variables are ordinal, use Gamma
 - Also okay for a binary (dichotomous) variable (only 2 values)
 - Interpret using guides above (slides 24, 25, & 30?)
 - Also interpret the p-value – can you reject the null (that the gamma is simply 0)?
- Lambda if *either* variable is nominal *and* has more than 2 categories
 - Eh... Below 0.5 is weak-ish, above 0.5 is strong-ish
 - Always positive, can't be negative
 - Also interpret the p-value – can you reject the null (that the lambda is simply 0)?

44 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PREs

G&K's tau b, calculated

- **Calculation:**
 - $\text{Tau b} = (59.3 - 56.6) / 59.3 = .046$
- **Interpretation:**
 - In the long run, knowing a subject's political orientation reduces our expected number of errors in predicting their child-rearing tendency by .046 or 4.6%
- **Comparison:**
 - when tau b = 0, chi-square = 0
 - Would be table of statistical independence

42 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PREs

All measures have a p value

- Some have a distribution; others are involved in a test statistic
- Ps interpreted basically the same always:
 - probability of getting a difference at least as large as the one observed, merely by chance
 - risk that we'd be wrong if we reject the null that the statistic (such as gamma) is zero in the population
 - We want that risk to be lower than 5% (below .05)
- Remember:
 - p never equals zero, though might be close

45 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PREs

PRE Measures

- **Gamma** (8.5, p.223) –
(Concordant–Discordant) / (Concordant+Discordant)
- Somer's D
(concordant – discordant) / all ties
- Goodman & Kruskal's tau b (8.4, p.224)
 - based on row marginals
- G&K's tau c
 - based on column marginals
- **Lambda**
 - nominal only, based on modal category

43 SOC424 @ CSUN - Ellis Godard

Review Pairs Gamma Interpretations Tau b PREs

One final lab – optional / extra credit ☺

- **Pick any 2 ordinal variables from any data set**
 - Gamma & tau b take into account order of values
- **Analyze relationship with χ^2 & gamma** (& tau-b, if you want?)
 - Examine univariate distributions & analysis 1st
 - Any problems? e.g. no variation or missing values?
 - Needn't submit this – but *always* do it, and do it *first*
 - Then look at cross-tabs & statistics
 - Chi-square 1st (is there a dependent relationship?)
 - Then assess **gamma** (direction, strength, & PRE; & p)
 - Submit output w/ a short paragraph summary

46 SOC424 @ CSUN - Ellis Godard