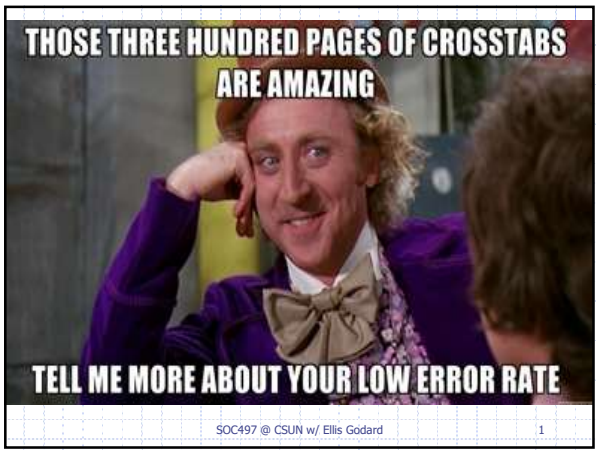
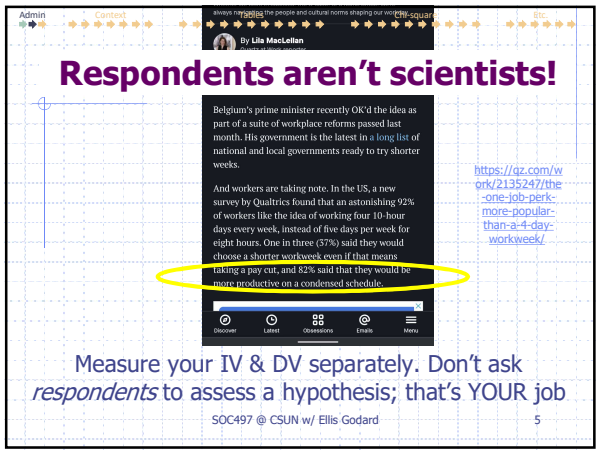
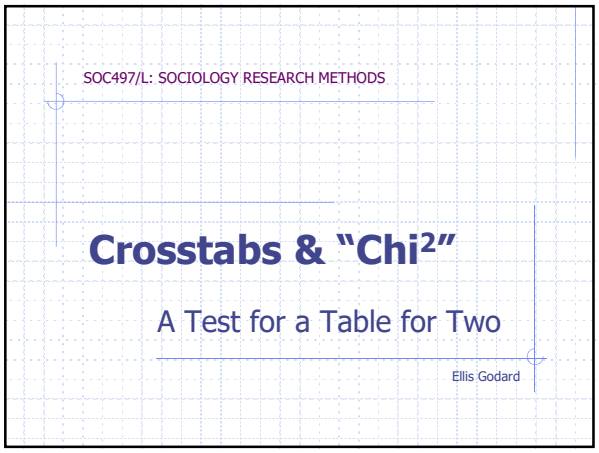
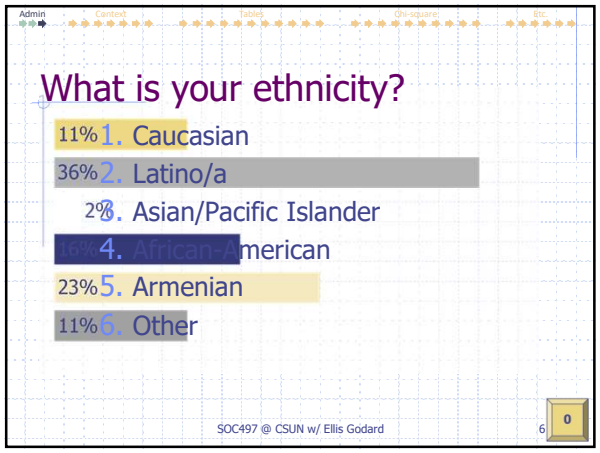
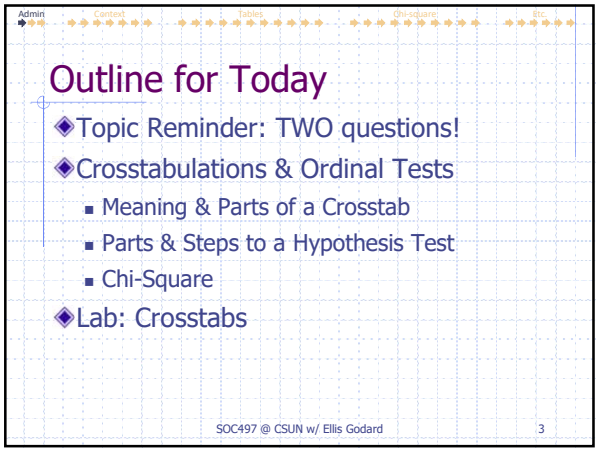
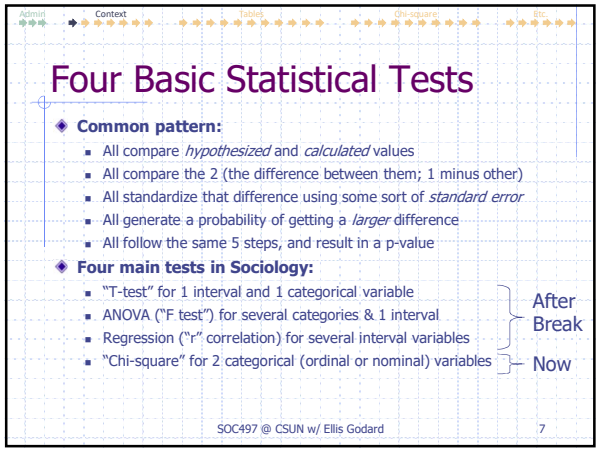

THOSE THREE HUNDRED PAGES OF CROSSTABS ARE AMAZING
TELL ME MORE ABOUT YOUR LOW ERROR RATE


Admin
Context
Tables
Chi-square
Etc.
Respondents aren't scientists!
By Lila MacLellan
Belgium's prime minister recently OK'd the idea as part of a suite of workplace reforms passed last month. His government is the latest in a long list of national and local governments ready to try shorter weeks.
And workers are taking note. In the US, a new survey by Qualtrics found that an astonishing 92% of workers like the idea of working four 10-hour days every week, instead of five days per week for eight hours. One in three (37%) said they would choose a shorter workweek even if that means taking a pay cut, and 82% said that they would be more productive on a condensed schedule.
https://qz.com/work/2135247/the-one-job-perk-more-popular-than-a-4-day-workweek/
Measure your IV & DV separately. Don't ask *respondents* to assess a hypothesis; that's YOUR job
SOC497 @ CSUN w/ Ellis Godard 5


SOC497/L: SOCIOLOGY RESEARCH METHODS
Crosstabs & "Chi²"
A Test for a Table for Two
Ellis Godard


Admin
Context
Tables
Chi-square
Etc.
What is your ethnicity?
11% 1. Caucasian
36% 2. Latino/a
2% 3. Asian/Pacific Islander
16% 4. African-American
23% 5. Armenian
11% 6. Other
SOC497 @ CSUN w/ Ellis Godard 6
0


Admin
Context
Tables
Chi-square
Etc.
Outline for Today
Topic Reminder: TWO questions!
Crosstabulations & Ordinal Tests
Meaning & Parts of a Crosstab
Parts & Steps to a Hypothesis Test
Chi-Square
Lab: Crosstabs
SOC497 @ CSUN w/ Ellis Godard 3


Admin
Context
Tables
Chi-square
Etc.
Four Basic Statistical Tests
Common pattern:
All compare *hypothesized* and *calculated* values
All compare the 2 (the difference between them; 1 minus other)
All standardize that difference using some sort of *standard error*
All generate a probability of getting a *larger* difference
All follow the same 5 steps, and result in a p-value
Four main tests in Sociology:
"T-test" for 1 interval and 1 categorical variable
ANOVA ("F test") for several categories & 1 interval
Regression ("r" correlation) for several interval variables
"Chi-square" for 2 categorical (ordinal or nominal) variables
After Break
Now
SOC497 @ CSUN w/ Ellis Godard 7

Admin | Context | Tables | Chi-square | Etc.

Q1. How many basic statistical tests are used by Sociologists?

1. 1 0%
2. 2 0%
3. 3 2%
4. 4 84%
5. 5 14%

SOC497 @ CSUN w/ Ellis Godard 8

Admin | Context | Tables | Chi-square | Etc.

Relationships

- Definition
  - Distribution of DV varies by IV, e.g. …
  - i.e. conditional distributions differ
- Question
  - Is the observed sample difference large enough to infer that there is probably difference in the population?
- Pieces
  - Observed difference
  - Sample size
  - # of rows & columns

SOC497 @ CSUN w/ Ellis Godard 11

Admin | Context | Tables | Chi-square | Etc.

Q2. Which is required if you have two interval variables?

A. ANOVA/"F" test 5%
B. Chi-square 2%
C. Regression/"r" 86%
D. "t"-test 7%
E. None of these can be used w/ interval variables 0%

SOC497 @ CSUN w/ Ellis Godard

Admin | Context | Tables | Chi-square | Etc.

Comparing Relationships

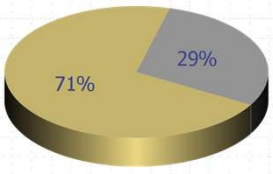
- Can compare one difference to another
- Could look at different IVs:
  - What accounts for partyid best (views, educ, relig, inc.)?
- Could look at different DVs:
  - What are the consequences of partyid (for voting, social life, etc.)

SOC497 @ CSUN w/ Ellis Godard 12

Admin | Context | Tables | Chi-square | Etc.

Should gay marriage be legal?

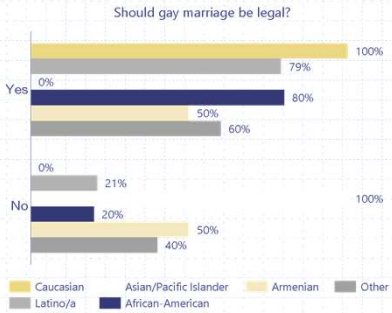
1. Yes
2. No



SOC497 @ CSUN w/ Ellis Godard 10

Admin | Context | Tables | Chi-square | Etc.

Out of each ethnicity (i.e. each value of the IV), what proportion said yes (i.e. what was the distribution of the DV)?



Spring 2025

	Camera	No Cam.	Total
Yes			
No			
Total			

Spring 2023 lecture example

	on camara	no on camera	
did volunteer	2 (33%)	3 (7.5%)	5 (12.4%)
didnt	4 (66%)	37 (92.5)	41 (87.6%)
			46

Fall 2024 Lecture Example

	Camera	No Camera	Total
Yes	0 0%	2 5.9%	2 5.6%
No	2 100%	32 94.1%	37 94.4%
Total	2 5.6%	34 94.4%	36 100%

Summer 2021 lecture example

	on camara	no on camera	
did volunteer	1 (25%)	3 (23%)	(24%)
didnt	3 (75%)	10 (77%)	(76%)
	4	13	

Fall 2023 Lecture Example

	Camera	No Camera	Total
Yes	1 50%	1 3%	2 5%
No	1 50%	34 97%	37 95%
Total	2	35	37

Spring 2021 lecture example

	on camara	no on camera	
did volunteer	2 (66%)	5 (12%)	7 (15%)
didnt	1 (33%)	38 (88%)	39 (85%)
	3	43	

(Re-)Introduction to Crosstabs

- Tables summarize the intersections of variables
  - Each row is a value of the DV ("effect")
  - Each column is a value of the IV ("cause")
    - Each col = frequency distrib of the IV, for 1 value of the DV
    - Row totals ("marginals") is freq distrib of IV (all cases)
- Each intersection (cell) is the combination of values of that row and that column (DV & IV)
  - Each cell has its *frequency* & *column percentage*

Column percents show reverse

The Nonequivalence of High School Graduates:
Educational Attainment and Employment Status at Age 25

	High School Dropout	GED	High School Graduate	
Jobless	17	8	31	56
	6.0%	7.0%	2.1%	3.0%
	E=8.53			
Employed	267	106	1435	1808
	94.0%	93.0%	97.9%	97.0%
	E=275.46			
	284	114	1466	1864
	15.3%	6.1%	78.6%	

Source: NLSY, analyzed by Cameron and Heckman 1991.

Q3. In a conventional Xtab, each row is a value of the…

1. Control variable 0%
2. Dependent variable 82%
3. Independent variable 18%
4. Variable view 0%
5. Variation Ratio 0%

Q4. To read Xtabs, focus on…

0% 1. The cell percentages
95% 2. The column percentages
0% 3. The raw frequencies
5% 4. The row percentages
0% 5. Everything at once so you get confused

Here, 17 is more than 8…

The Nonequivalence of High School Graduates:
Educational Attainment and Employment Status at Age 25

	High School Dropout	GED	High School Graduate	
Jobless	17	8	31	56
Employed	267	106	1435	1808
	284	114	1466	1864

Source: NLSY, analyzed by Cameron and Heckman 1991.

Examining Crosstabs

- Compare *percentages*, not raw frequencies
  - Compare *across,* not up & down (!)
    - Comparing up & down uses only one column, always just adds to 100%, and doesn't involve variation in the IV
  - Compare each column to the marginals (?)
  - Compare each column to a reference group (eh…)
    - Could be one particular column, or another sample, or even another variable *(e.g. those absent)*
- Look for "modal percentages"
  - Highest % in each row (like the mode – biggest)
  - Do they fall on the "main diagonal"
    - the line of cells predicted by your research hypothesis
    - Be careful to know whether slants up or down!

Hypothesis Testing Steps

1. Assumptions
   - Sample size, level of measurement, random sample
2. Hypotheses
   - Null (the one you want to reject) and alternate (research)
3. Calculated Value
   - Compares observed differences from null hypothesized value(s)
   - Chi-square, Student's "t", F ratio (ANOVA), r (corr.), r^2 (regression)
4. Compare to critical Value and/or get a "P" value
   - Alpha level - 1 minus confidence coefficient (95%, 99%, etc.) / 2
   - Assesses probability that we would be wrong if we reject the null
5. Make conclusion (about Ho and Ha)
   - Either reject the null or fail to reject the null
   - Are the sample observations different enough from null to reject it and instead find support for research hypothesis?

Chi-square & Dependence

- In a 2x2 table, two variables are *statistically independent* if the population conditional distributions equal the marginal distribution.
- They are statistically *dependent* if the population conditional distributions differ.

Steps in Chi-Square Test of Independence

1. Assumptions:

a. Two nominal variables

b. Random Sample or stratified random sample

c. for 2x2 table, the *expected* number of observations in *each* must be 5 or greater; for larger tables (with more cells), at least 3/4 of the cells should have an expected count of 5 or greater and all cells must have an expected count greater than 1.

*2. **Hypotheses***

H_0: The variables are statistically independent in population

H_a: Statistical dependence of the variables

Significance Testing w Chi-square

- Looking for a relationship – i.e. the distribution of the DV varies across values of the IV
- Here, they don't - the 2 are **independent**:

	LibProt	ConsProt	Catholic	None
IRM ok	750 30%	1200 30%	900 30%	150 30%
Not ok	1750 70%	2800 70%	2100 70%	350 70%
Total	2500	4000	3000	500

- Distribution of the DV is the *same* for every IV

Steps in Chi-Square Test of Independence

3. Test Statistic: $\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$

where $f_e = \frac{(\text{Row Total})(\text{Column Total})}{\text{Total Sample Size}}$

4. P-value:

Chance you would be wrong if you reject the null

SPSS gives value - but note p is NEVER ZERO.

Call .00000 "close to zero", etc.

*5. **Conclusion:** **Reject H_0 & support H_a at α-level (a conventional level of significance) if $P < \alpha$ (typically .05)***

Chi^2: Another Example

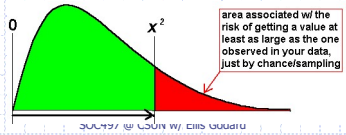
- Consider political affiliation crosstabulated with parental style:

	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

- The row marginals (highest % in each row) fall on the "main diagonal" (the corner-to-corner pattern predicted by our H_A)

The Chi-Square Distribution

- As with all statistical tests, some observed distance is associated w/ a p-value (probability level) of getting *more* than that distance away from the null hypothesized condition
- Here, the "distance" is between observed frequencies (your data) & expected frequencies (what the data would look like if there were *no* relationship, i.e. if the null were "true")
  - The more percentages across rows differ from the null, the bigger chi-square is
- As always, the bigger the test statistic, the smaller the p
  - If & only if p is smaller than 0.05 (5%), we reject the null..



Trump Ratings, NBC/WP, 10/29/17

	Dem	Ind	Rep	All
Approve	7%	34%	81%	38%
Diapprove	89%	57%	17%	58%

Sample Crosstabs: Strong

Chi-square's p is < .00001

Recycle	Attend church... Weekly	Monthly	Less than that	
Never	54 54.5%	55 29.6%	14 14.9%	123
Some-times	39 39.4%	87 46.8%	16 17.0%	142
Always	6 6.1%	44 23.7%	64 68.1%	114
	99	186	94	379

Reminder: Crosstabs & x^2

Here, **females are more than twice as likely to say they can't still remain friends (22% vs. 9%);**

males are more likely to say they can, and modal percentages fall along the "main diagonal";

& the difference is statistically significant (p<.05).

If they cross the line, can you still remain friends? * What is your gender? Crosstabulation

			What is your gender? Male	Female	Total
If they cross the line, can you still remain friends?	Yes	Count	71	167	238
		% within What is your gender?	91.0%	78.0%	81.5%
	No	Count	7	47	54
		% within What is your gender?	9.0%	22.0%	18.5%
Total		Count	78	214	292
		% within What is your gender?	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	6.398[b]	1	.011		
Continuity Correction[a]	5.565	1	.018		
Likelihood Ratio	7.196	1	.007		
Fisher's Exact Test				.011	.007
Linear-by-Linear Association	6.376	1	.012		
N of Valid Cases	292				

Sample Crosstabs: None

Chi-square's p is .99959

Smoke Crack	Listen to Rap Never	Sometimes	Often	
Never	94 80.3%	55 80.9%	110 80.9%	259
Once	19 16.2%	11 16.2%	22 16.2%	52
Regularly	4 3.4%	2 2.9%	4 2.9%	10
Totals:	117	68	136	321

Chi-square and Correlations

- For parts 3&4 of HW5
  - Just need to find the statistic
  - Don't need to calculate, interpret tables, or get other measures
- What you *do* need – TWO TESTS!!
  - P value for chi-square
    - **Analyze > Descriptives > Crosstabs; stats > chisqr**
    - "probability of being wrong if you reject the null, that there is not an dependent relationship between the two variables"
    - If it's less than 0.05, you can reject the null & find support for the alternative (that there is a dependent relationship)
  - Correlation coefficient from matrix…

Correlation Matrices in SPSS

- Analyze – Correlate – Bivariate
  - Options – Means & Standard deviations
- Matrix gives values for each pairing:
  - Correlation coefficient *(higher = stronger)*
    - Note asterisked notes
    - Easily identify which are statistically significant
  - P-values *(probability of Type I error in inferring correlation coefficient to population)*
  - N *(number of cases with values on both variables)*

SOC497 @ CSUN w/ Ellis Godard 38

Team Scores

Points	Team	Points	Team
4.8	**African-American**		
4.66	**Latino/a**		
4.4	**Caucasian**		
4	**Asian/Pacific Islander**		
4	**Other**		

Lab Exercise

- You may use any SPSS dataset
  - For this class, my 424 class, or any other
- Create one cross-tabulation
  - Any two (nominal and/or ordinal) variables
  - Use the handout @ Canvas
- Print it & provide the 3 key interpretations
  - Compare across
  - Main diagonal
  - Interpret chi²'s p

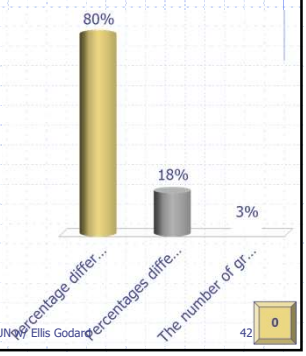
SOC497 @ CSUN w/ Ellis Godard 41

Q5. Chi-square indirectly measures…

A. Percentage differences across each row

B. Percentages differences up and down each column

C. The number of grains of sand in Long Beach, times the length in inches from this class to the Pub



SOC497 @ CSUN w/ Ellis Godard 42