Chi square test of independence

- *Eyeball* differences between percentages: large enough to be "important"
- Better: Are they *statistically significant*?
- *Statistical significance*: are observed differences significantly different from zero that they could not occur by chance?

Chi square test of independence

- Statistical significance often used as measure of substantive importance
- But, statistical significance and importance are theoretically distinct
- Some differences can be *statistically significant* but *not very important substantively*

Chi square test of independence

- Use chi square test of independence for cross classification tables
- Chi square test: are two variables *statistically associated* ?
- Are differences in sample likely to persist in the population?

Association

- *Association*: two variables covary, either positively or negatively
- Test for *significant* association: compare observed frequencies to a model that assumes *statistical independence*

Statistical independence

Belief in life after death	Religious affiliation (hypothetical)			
	Protestant	Catholic	Jewish	Other
Yes	120	90	30	60
No	80	60	20	40
	200	150	50	100

Statistical association (statistical dependence)

Percentage believing in life after death by religious affiliation				
Belief in life after death	Religious affiliation			
	Protestant	Catholic	Jewish	Other
Yes	75.0	86.7	10.0	15.0
N	(200)	(150)	(50)	(100)

Chi square test of independence

- Do differences by religious affiliation reflect true differences in the population?
- Are observed differences large enough that we're sure they're not due purely to chance?
- Weird sample?

Chi square test of independence

- Chi square tests for independence between two nominal (or ordinal) variables
- $H_{o:}$ statistical independence (no differences across religious affiliation)
- H_a : statistical dependence (association between religious affiliation and attitudes toward life after death)

Chi square test of independence

- Chi square: comparison between frequencies observed in cells and the numbers you would expect if variables were statistically independent

$$f_0 = observed\ frequencies$$

$$fe = \exp ected\ frequencies$$

Chi square test of independence

$$f_e = \frac{r\ x\ c}{n}$$

where, r = row total
c = column total
n = # of cases

Calculating chi square: observed frequencies

Belief in life after death	Religious affiliation				
	Prot.	Cath.	Jewish	Other	Row totals
Yes	150	130	5	15	300
No	50	20	45	85	200
Column totals	200	150	50	100	500

Calculating expected frequencies: for Catholics who say yes

$$f_e = \frac{rxc}{n} = \frac{(300)(150)}{500}$$

$$f_e = 90$$

Calculating chi square: expected frequencies

Belief in life after death	Religious affiliation				
	Prot.	Cath.	Jewish	Other	Row totals
Yes	150 (120)	130 (90)	5 (etc.)	15 (etc.)	300
No	50 (etc.)	20 (etc.)	45 (etc.)	85 (40)	200
Column totals	200	150	50	100	500

Chi square

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$= (150\text{-}120)^2/120 + (50\text{-}80)^2/80 + \ldots (85\text{-}40)^2/40$

$$\chi^2 = 199.65$$

Chi square: evaluation

- If H_0 of no association is *true*, then f_o and f_e will be close and the chi square value small
- If H_0 of no association is *false*, f_o and f_e should be relatively farther apart, and hence the chi square value larger
- Chi square value = 0 when $f_o = f_e$

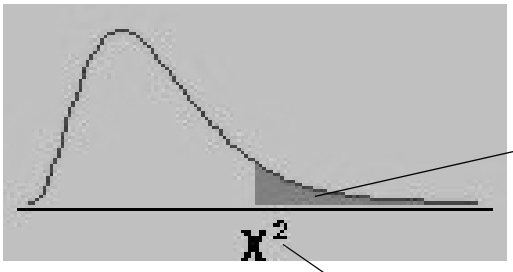
Chi square evaluation made easy

- How big is $\chi^2 = 199.65$?
- Evaluate relative to degrees of freedom for the table (a measure of the number of rows and columns)
- Also sensitive to sample size (the larger the N the greater the statistic)

How big is $\chi^2 = 199.65$?

- Decide on how confident you want to be that the null hypothesis (H_0) is false
- Typically, either 95% or 99% confident
- If 95%, then possibility of error is .05
- If 99%, then possibility of error is .01
- Let's assume .05 level of error

Chi square distribution: how big is $\chi^2 = 199.65$?



Note: χ^2 never negative

Chi square distribution: how big is $\chi^2 = 199.65$?

- GSS provides p values for you
- Want p <= .05 (indicates significance)
- p = probability, when H_0 is true, of getting a value at least as large as the observed χ^2

Chi square test of independence

Percentage believing in life after death by religious affiliation				
Belief in life after death	Religious affiliation			
	Protestant	Catholic	Jewish	Other
Yes	75.0	86.7	10.0	15.0
	(200)	(150)	(50)	(100)
	$\chi^2 : p <= .0001$			

Interpretation

Substantively:

Religious affiliation is *significantly* associated with belief in life after death