read: L-312 → L-320
today: L-290 →

this time: categorical data analysis

next time: decision; wrap-up

take home final handed out on Thu 7 Jun, due by 11:59 pm Fri 15 Jun 18

HW 4 due by 11:59 pm on Fri 8 Jun 18 on canvas

Ashley Tai
AMS 7
5 Jun 18

CDC study (2004)

NS: nonsmoker
S: smoker
G: gum
I: inhaler
P: nicotine patch

Categorical Data Analysis

(outcome) smoking	(predictor) method	
N	G	↑
:	:	59
N	G	↓
Y	G	↑
:	:	191
Y	G	↓
N	P	57
Y	P	263
N	I	27
Y	I	95

contingency table

$\hat{O}_{ij}$	NS	S	
G	59	191	250
I	27	95	122
P	57	263	320
	143	549	692

$$\hat{p}_G = \frac{59}{250} = 0.236 = 23.6\% = \hat{P}(NS \mid G)$$

$$\hat{p}_I = \frac{27}{122} = 0.221 = 22.1\% = \hat{P}(NS \mid I)$$

$$\hat{p}_P = \frac{57}{320} = 0.178 = 17.8\% = \hat{P}(NS \mid P)$$

* Practsig:

$$\frac{23.6\% - 17.8\%}{17.8\%} = 33\%$$

Karl Pearson (1800s)

NULL: $p_G = p_I = p_P$

ALT: not so

(observed frequencies) vs. (expected frequencies if null true)

* If null true, method & success are INDEPENDENT

* multiply row & column proportions

	NS	S	
G	0.075	0.287	$\frac{250}{692} = 0.361$
I	0.036	0.140	$\frac{122}{692} =$
P	0.096	0.367	

$$\frac{(\hat{O}_{11} - \hat{E}_{11})}{\hat{E}_{11}} + \frac{(\hat{O}_{12} - \hat{E}_{12})}{\hat{E}_{12}} + \cdots \frac{(\hat{O}_{32} - \hat{E}_{32})}{\hat{E}_{32}}$$

$$\boxed{\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(\hat{O}_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}} = 3.062$$

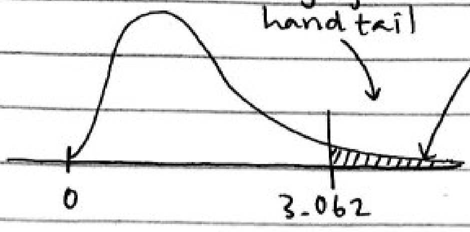
Chi-square

J

I		
	+7.3 ✓	−7.3 ×
	+1.8 ✓	−1.8 ×
	−9.1 ×	+9.1 ×

right # df $= (I-1)(J-1)$
$= (3-1)(2-1) = \boxed{2}$

long run hist. of χ^2 if null true

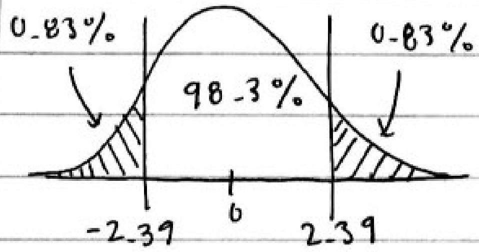


* since $p \doteq 22\% > 5\%$, not statsig
* practsig, but not statsig → not enough data

Bonferroni Method

unadjusted → $\hat{SE}(\hat{p}_G - \hat{p}_I) = \sqrt{\frac{(0.236)(1-0.236)}{250} + \frac{(0.221)(1-0.221)}{122}} = 2.39$

* make 3 statements with overall 95% CI by making each at $(1 - \frac{0.05}{3})\% = 98.3\%$



$(\hat{p}_G - \hat{p}_I) \pm 2.39\ \hat{SE}(\hat{p}_G - \hat{p}_I)$

$(\hat{p}_G - \hat{p}_P) \pm 2.39\ \hat{SE}(\hat{p}_G - \hat{p}_P)$

$(\hat{p}_I - \hat{p}_P) \pm 2.39\ \hat{SE}(\hat{p}_I - \hat{p}_P)$

Decision Theory

- how well can humans make decisions in situations where they do not have enough information?