12/4/18
Wave more to

Lecture #19: ANOVA

- Narrow intervals increase chance of making a mistake → must make them wider
- L-271,272 (see table)

$$\hat{SE}(\bar{y}_i - \bar{y}_j) = \hat{\sigma}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

RMSE = $\sqrt{MSW}$

$$\hat{SE}(\bar{y}_1 - \bar{y}_2) = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \boxed{0.27}$$

$\bar{y}_1$: no treatment; $\bar{y}_2$: fertilizer; $\hat{\sigma}$: 0.42619; n_1, n_2: 5

$-0.022 \pm (3.008)(0.27) = \boxed{(-0.83, +0.79)}$

- Is 0 in this interval? Yes, not stat sig.

Categorical Data Analysis L-290

L-295 → case study by CDC

1 = NS
0 = S

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

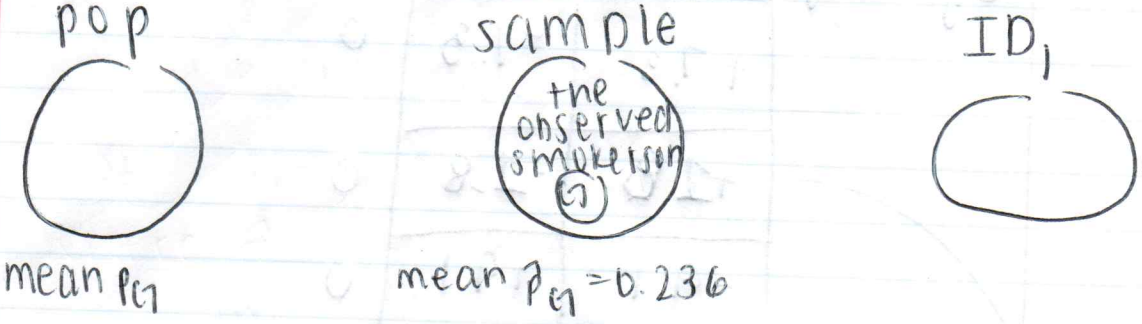
$\hat{p}_I = \frac{27}{122} = 22.1\% = \hat{p}(NS|I)$

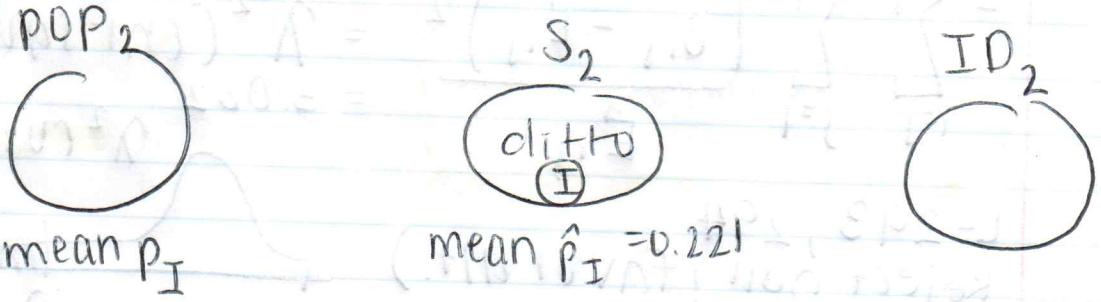
$\hat{p}_P = \frac{57}{320} = 0.178 = 17.8\% = \hat{p}(NS|P)$

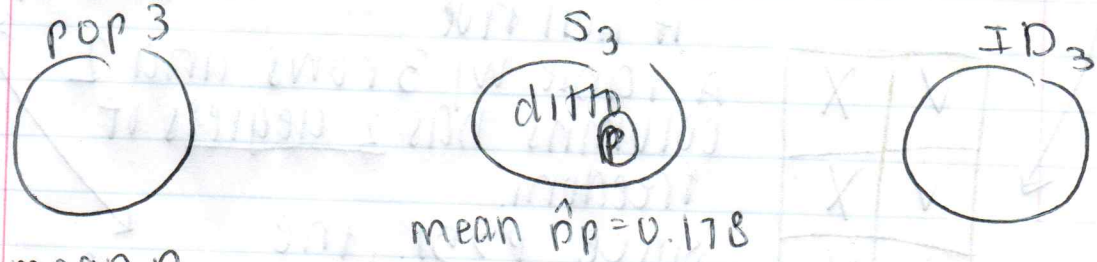
This diff is large in pract terms. Stat sig?

$\frac{0.236 - 0.178}{0.178} \times 100\% = 33\%$ increase in chance of not-smoking when switch from patch to gum.

Karl Pearson (~1900)







null: $p_G = p_I = p_P$ ← if null true, (A) method and (B) smoking are independent

$P(A \text{ and } B) \overset{\text{independent}}{=} P(A) \cdot P(B)$

L-292

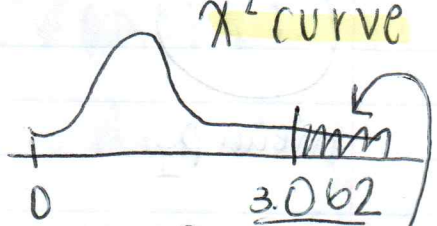
$\hat{O}_{ij} - \hat{E}_{ij}$ = residuals

+7.3	-7.3	0
+1.8	-1.8	0
-9.1	+9.1	0
0	0	0

$$\frac{(\hat{O}_{11} - \hat{E}_{11})^2}{\hat{E}_{11}} + \frac{(\hat{O}_{12} - \hat{E}_{12})^2}{\hat{E}_{12}} + \ldots + \frac{(\hat{O}_{32} - \hat{E}_{32})^2}{\hat{E}_{32}}$$

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

L-293, 294
Reject null (favor alt.) if χ^2 is big



long run hist of χ^2 if null true

use JMP → P = 0.2163 ≈ 22%

✓	X
✓	X
X	X

A table w/ 3 rows and 2 columns has 2 degrees of freedom.

since P > 5%, the differences are not stat sig.
↳ not enough data

L-304 - what n should have been