(11-3) χ^2 TESTS FOR INDEPENDENCE

(always right-tailed)

Ex 700 random likely U.S. voters are polled.

Contingency Table or Two-Way Frequency Table

It doesn't matter which var. is the row var. and which is the col. var., as we'll see in our formulas. Just be consistent!

		Gender of Voter	
		Female	Male
Candidate	Smith	250	200
	Jones	150	100

4 cells

2 "by" 2 table
\# rows, \# columns

Note 60% of the females in the poll support Smith, while about 67% of the males do. Could the actual %s be the same?

Test the claim that candidate preference is independent of voter gender at the .05 significance level. ☆

Sol'n

H_0: ☆ (Claim)

H_0 always assumes indep.

H_A: Candidate preference depends on voter gender.

Note: Statistical dependence does not imply causality; see Ch. 10

$\alpha = 0.05$

Compute row, column totals

"O" Table
Observed freq.

	Female	Male	
Smith	250	200	450
Jones	150	100	250
	400	300	700 = n

Expected frequencies under H_0

For each cell, $E = \dfrac{(\text{its row total})(\text{its column total})}{n}$

n (grand total)

Why? Ex

	(F) Female
(S) Smith	

$$E = \left(\begin{matrix}\#\text{voters}\\ \text{in poll}\end{matrix}\right) \cdot P(S \text{ and } F)$$

$$= n \cdot P(S) \cdot P(F) \quad \text{by indep. under } H_0$$

$$= \cancel{700} \cdot \left(\frac{450}{\cancel{700}}\right) \cdot \left(\frac{400}{700}\right) \quad \text{(based on table)}$$

$$= \frac{(450)(400)}{700}$$

row col. n

$$\approx 257.1429$$

or 3 dec. places, like in χ^2 table. I used 4 because we're adding

"E" Table

	Female	Male
Smith	$\frac{(450)(400)}{700} \approx 257.1429$	$\frac{(450)(300)}{700} \approx 192.8571$
Jones	$\frac{(250)(400)}{700} \approx 142.8571$	$\frac{(250)(300)}{700} \approx 107.1429$

Make sure each ≥ 5. ✓

Test statistic

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\begin{aligned} \# df &= (\# \text{rows} - 1)(\# \text{columns} - 1) \\ &= (2-1)\quad(2-1) \quad \text{here} \\ &= 1\ df \end{aligned}$$

Sorry; at least you remember it's E down here →

$$\chi^2 \approx \frac{(250-257.1429)^2}{257.1429} \quad S,F$$

$$+ \frac{(200-192.8571)^2}{192.8571} \quad S,M$$

$$+ \frac{(150-142.8571)^2}{142.8571} \quad J,F$$

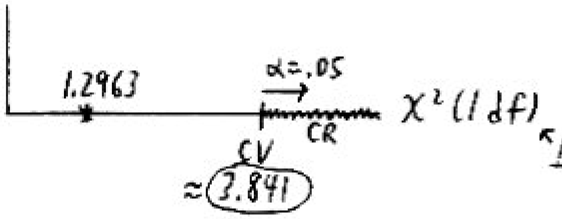
$$+ \frac{(100-107.1429)^2}{107.1429} \quad J,M$$

$$\begin{aligned} &\approx 0.1984 \\ &\quad +0.2646 \\ &\quad +0.3571 \\ &\quad +0.4762 \end{aligned}$$

$$= 1.2963$$

Find CV, CR

always right-tailed



Note looks like

Decision

We do not reject H_0.

Final Conclusion

There is <u>insufficient</u> evidence <u>against</u> the claim that ☆.

(Think: H_0 is found "not guilty.")

A <u>test of homogeneity</u> is similar, except that either the row or column totals are predetermined.