Chi-Square Tests

(Source: Levine et al)

Contingency Tables

Contingency Tables

- Useful in situations comparing multiple population proportions
- Used to classify sample observations according to two or more characteristics
- Also called a cross-classification table.

Contingency Table Example

Left-Handed vs. Gender

Dominant Hand: Left vs. Right

Gender: Male vs. Female

- 2 categories for each variable, so this is called a 2 x 2 table
- Suppose we examine a sample of 300 children

Contingency Table Example

(continued)

Sample results organized in a contingency table:

sample size = n = 300:

120 Females, 12 were left handed

180 Males, 24 were left handed

Gender	Hand Preference		
	Left	Right	
Female	12	108	120
Male	24	156	180
	36	264	300

χ^2 Test for the Difference Between Two Proportions

H_0: $p_1 = p_2$ (Proportion of females who are left handed is equal to the proportion of males who are left handed)

H_1: $p_1 \neq p_2$ (The two proportions are not the same – hand preference is not independent of gender)

- If H_0 is true, then the proportion of left-handed females should be the same as the proportion of left-handed males
- The two proportions above should be the same as the proportion of left-handed people overall

The Chi-Square Test Statistic

The Chi-square test statistic is:

$$\boxed{\chi^2_{CAL} = \sum_{all\ \text{cells}} \frac{(f_o - f_e)^2}{f_e}}$$

- where:
  f_o = observed frequency in a particular cell
  f_e = expected frequency in a particular cell if H_0 is true

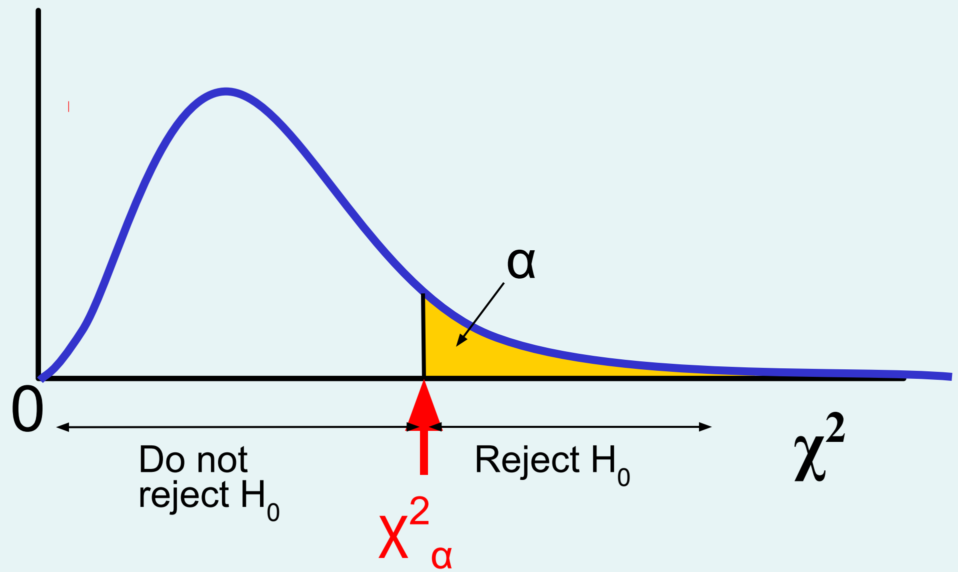
χ^2_{CAL} for the 2 x 2 case has 1 degree of freedom

(Assumed: each cell in the contingency table has expected frequency of at least 5)

Decision Rule

The χ^2_{CAL} test statistic approximately follows a chi-squared distribution with one degree of freedom

Decision Rule:
If $\chi^2_{CAL} > \chi^2_{\alpha}$, reject H_0, otherwise, do not reject H_0



Computing the Average Proportion

The average proportion is:

$$\bar{p} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{X}{n}$$

120 Females, 12 were left handed

180 Males, 24 were left handed

Here:

$$\bar{p} = \frac{12 + 24}{120 + 180} = \frac{36}{300} = 0.12$$

i.e., based on all 300 children the proportion of left handers is 0.12, that is, 12%

Finding Expected Frequencies

- To obtain the expected frequency for left handed females, multiply the average proportion left handed ($\bar{p}$) by the total number of females
- To obtain the expected frequency for left handed males, multiply the average proportion left handed ($\bar{p}$) by the total number of males

---

If the two proportions are equal, then

P(Left Handed | Female) = P(Left Handed | Male) = .12

i.e., we would expect (.12)(120) = 14.4 females to be left handed
(.12)(180) = 21.6 males to be left handed

Observed vs. Expected Frequencies

Gender	Hand Preference		
	Left	Right	
Female	Observed = 12 Expected = 14.4	Observed = 108 Expected = 105.6	120
Male	Observed = 24 Expected = 21.6	Observed = 156 Expected = 158.4	180
	36	264	300

The Chi-Square Test Statistic

Gender	Hand Preference: Left	Hand Preference: Right	
Female	Observed = 12 Expected = 14.4	Observed = 108 Expected = 105.6	120
Male	Observed = 24 Expected = 21.6	Observed = 156 Expected = 158.4	180
	36	264	300

The test statistic is:

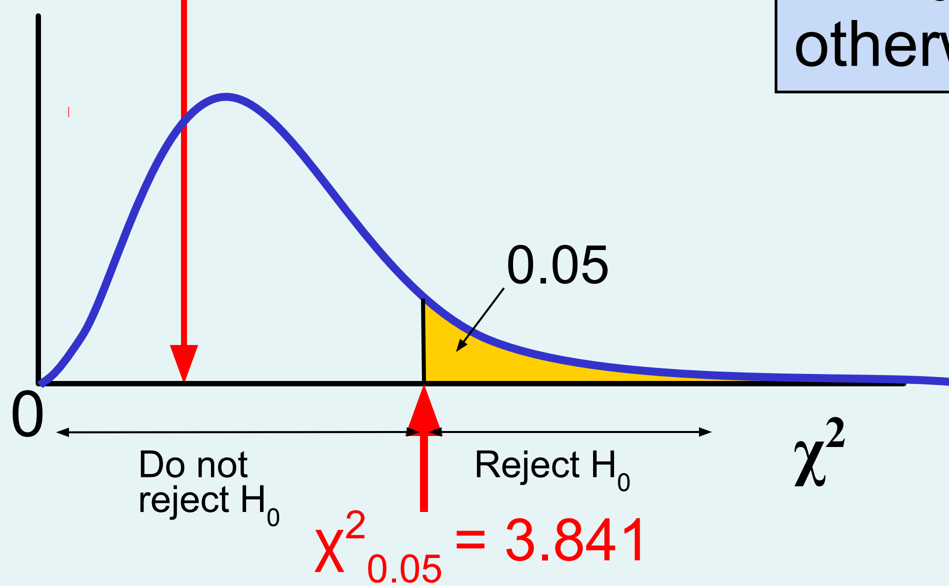
$$\chi^2_{CAL} = \sum_{\text{all cells}} \frac{(f_o - f_e)^2}{f_e}$$

$$= \frac{(12-14.4)^2}{14.4} + \frac{(108-105.6)^2}{105.6} + \frac{(24-21.6)^2}{21.6} + \frac{(156-158.4)^2}{158.4} = 0.7576$$

Decision Rule

The test statistic is $\chi = 0.7576$; $\chi^2_{0.05}$ with 1 d.f. $= 3.841$

Decision Rule:
If χ^2_{CAL} > 3.841, reject H_0,
otherwise, do not reject H_0



Here,
$\chi^2_{CAL} = 0.7576 < \chi^2_{0.05} = 3.841$,
so we do not reject H_0 and conclude that there is not sufficient evidence that the two proportions are different at $\alpha = 0.05$

χ^2 Test of Independence

- Similar to the χ^2 test for equality of more than two proportions, but extends the concept to contingency tables with r rows and c columns

H_0: The two categorical variables are independent
(i.e., there is no relationship between them)

H_1: The two categorical variables are dependent
(i.e., there is a relationship between them)

χ^2 Test of Independence

(continued)

The Chi-square test statistic is:

$$\chi^2_{CAL} = \sum_{all\ \text{cells}} \frac{(f_o - f_e)^2}{f_e}$$

- where:

  f_o = observed frequency in a particular cell of the r x c table
  f_e = expected frequency in a particular cell if H_0 is true

χ^2_{CAL} for the r x c case has (r - 1)(c - 1) degrees of freedom

(Assumed: each cell in the contingency table has expected frequency of at least 1)

Expected Cell Frequencies

- Expected cell frequencies:

$$f_e = \frac{\text{row total} \times \text{column total}}{n}$$

Where:

row total = sum of all frequencies in the row

column total = sum of all frequencies in the column

n = overall sample size

Decision Rule

- The decision rule is

If $\chi^2_{CAL} > \chi^2_{\alpha}$, reject H_0,
otherwise, do not reject H_0

Where χ^2_{α} is from the chi-squared distribution with (r – 1)(c – 1) degrees of freedom

Example

DCOVA

- The meal plan selected by 200 students is shown below:

Class Standing	Number of meals per week			Total
	20/week	10/week	none	
Fresh.	24	32	14	70
Soph.	22	26	12	60
Junior	10	14	6	30
Senior	14	16	10	40
Total	70	88	42	200

Example

DCOVA

(continued)

- The hypothesis to be tested is:

H_0: Meal plan and class standing are independent
(i.e., there is no relationship between them)

H_1: Meal plan and class standing are dependent
(i.e., there is a relationship between them)

Example: Expected Cell Frequencies

(continued)

DCOVA

Observed:

Class Standing	Number of meals per week			Total
	20/wk	10/wk	none	
Fresh.	24	32	14	70
Soph.	22	26	12	60
Junior	10	14	6	30
Senior	14	16	10	40
Total	70	88	42	200

Expected cell frequencies if H_0 is true:

Class Standing	Number of meals per week			Total
	20/wk	10/wk	none	
Fresh.	24.5	30.8	14.7	70
Soph.	21.0	26.4	12.6	60
Junior	10.5	13.2	6.3	30
Senior	14.0	17.6	8.4	40
Total	70	88	42	200

Example for one cell:

$$f_e = \frac{\text{row total} \times \text{column total}}{n}$$

$$= \frac{30 \times 70}{200} = 10.5$$

Example: The Test Statistic

(continued)

DCOVA

- The test statistic value is:

$$\chi^2_{CAL} = \sum_{\mathit{all}\ \text{cells}} \frac{(f_o - f_e)^2}{f_e}$$

$$= \frac{(24-24.5)^2}{24.5} + \frac{(32-30.8)^2}{30.8} + \cdots + \frac{(10-8.4)^2}{8.4} = 0.709$$

$\chi^2_{0.05}$ = 12.592 from the chi-squared distribution with (4 – 1)(3 – 1) = 6 degrees of freedom

Example: Decision and Interpretation

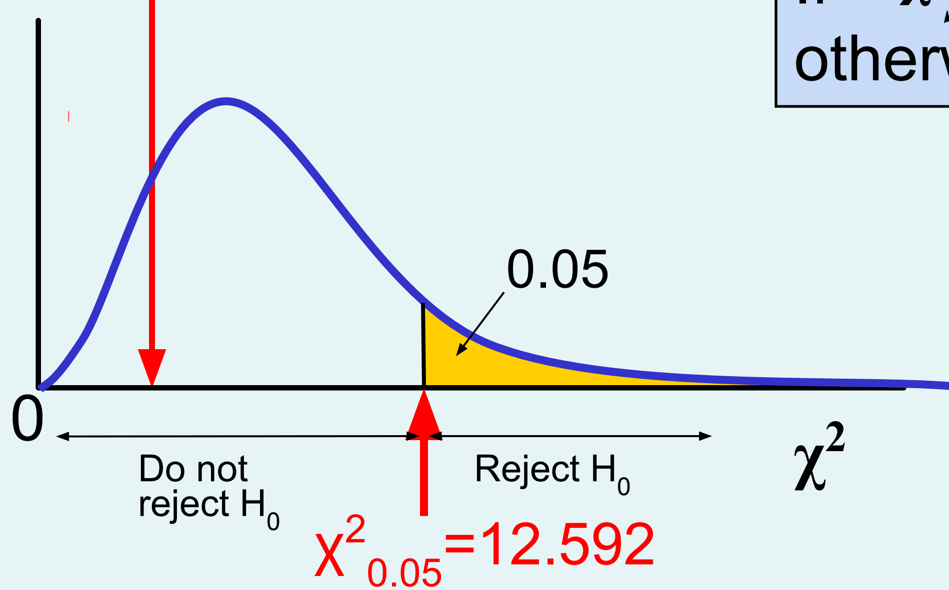
DCOVA

(continued)

The test statistic is $\chi^2_{CAL} = \boxed{0.709}$; $\chi^2_{0.05}$ with 6 d.f. $= 12.592$

Decision Rule:

If $\chi^2_{STAT} > 12.592$, reject H_0, otherwise, do not reject H_0



Here,
$\chi^2_{CAL} = 0.709 < \chi^2_{0.05} = 12.592$,
so do not reject H_0

Conclusion: there is not sufficient evidence that meal plan and class standing are related at $\alpha = 0.05$