# *Statistics for Managers Using Microsoft Excel*

7th Edition

## Chapter 12

Chi-Square Tests and Nonparametric Tests

# Learning Objectives

**In this chapter, you learn:**

- How and when to use the chi-square test for contingency tables
- How to use the Marascuilo procedure for determining pairwise differences when evaluating more than two proportions
- How and when to use nonparametric tests

# Contingency Tables

DCOVA

## 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

DCOVA

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)*

DCOVA

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 |

# $\chi^2$ Test for the Difference Between Two Proportions

DCOVA

$H_0$: $\pi_1 = \pi_2$ (Proportion of females who are left handed is equal to the proportion of males who are left handed)

$H_1$: $\pi_1 \neq \pi_2$ (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

DCOVA

The Chi-square test statistic is:

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

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

$\chi^2_{STAT}$ **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)

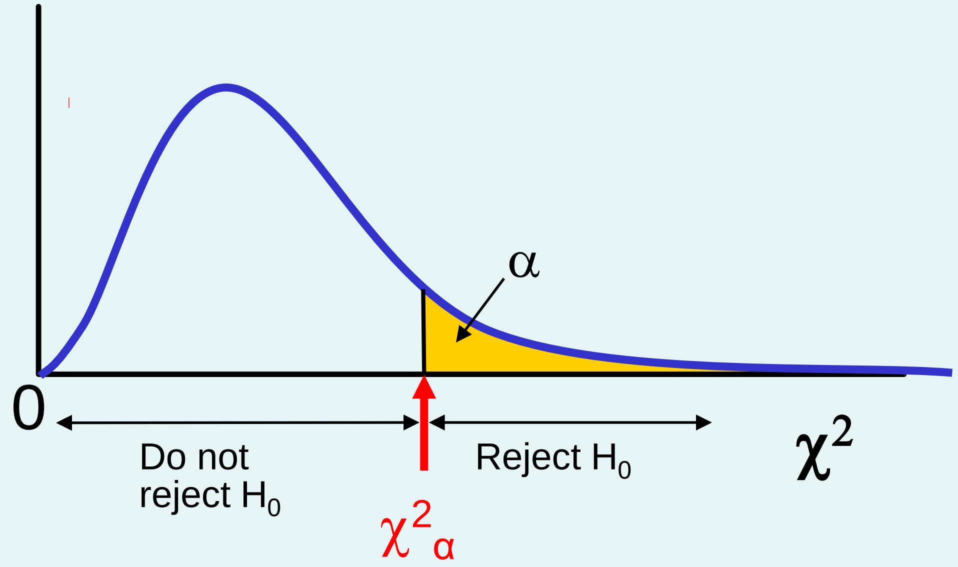
# Decision Rule

DCOVA

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

**Decision Rule:**

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



# Computing the Average Proportion

DCOVA

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 180 children the proportion of left handers is 0.12, that is, 12%**

# Finding Expected Frequencies

DCOVA

- 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

DCOVA

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

# The Chi-Square Test Statistic

DCOVA

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

The test statistic is:

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

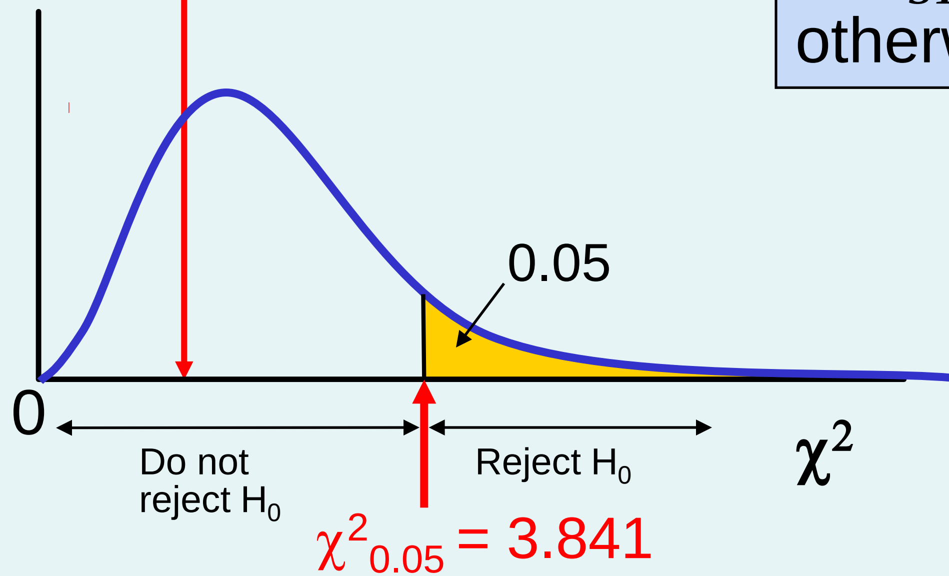
$$= \frac{(12-14.4)^2}{14.4} + \frac{(108-105.6)^2}{105.6} + \frac{(24-21.6)^2}{21.6} + \frac{(156-158.4)^2}{158.4} = 0.7576$$

# Decision Rule

DCOVA

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

**Decision Rule:**
If $\chi^2_{STAT} > 3.841$, reject $H_0$, otherwise, do not reject $H_0$




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

# $\chi^2$ Test for Differences Among More Than Two Proportions

DCOVA

- Extend the $\chi^2$ test to the case with more than two independent populations:

$H_0$: $\pi_1 = \pi_2 = \cdots = \pi_c$

$H_1$: Not all of the $\pi_j$ are equal ($j = 1, 2, \cdots, c$)

# The Chi-Square Test Statistic

DCOVA

The Chi-square test statistic is:

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

- Where:

  $f_o$ = observed frequency in a particular cell of the 2 x c table

  $f_e$ = expected frequency in a particular cell if $H_0$ is true

$\chi^2_{STAT}$ **for the 2 x c case has (2 - 1)(c - 1) = c - 1 degrees of freedom**

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

# Computing the Overall Proportion

DCOVA

The overall proportion is:

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

- Expected cell frequencies for the c categories are calculated as in the 2 x 2 case, and the decision rule is the same:

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

Where $\chi^2_{\alpha}$ is from the chi-squared distribution with c – 1 degrees of freedom

# The Marascuilo Procedure

DCOVA

- Used when the null hypothesis of equal proportions is rejected
- Enables you to make comparisons between all pairs
- Start with the observed differences, $p_j - p_{j'}$, for all pairs (for $j \neq j'$) then compare the absolute difference to a calculated critical range

# The Marascuilo Procedure

*(continued)*

DCOVA

- Critical Range for the Marascuilo Procedure:

$$\text{Critical range} = \sqrt{\chi^2_\alpha}\sqrt{\frac{p_j(1-p_j)}{n_j} + \frac{p_{j'}(1-p_{j'})}{n_{j'}}}$$

  - (Note: the critical range is different for each pairwise comparison)

- A particular pair of proportions is significantly different if

$|p_j - p_{j'}|$ > critical range for j and j'

# Marascuilo Procedure Example

DCOVA

A University is thinking of switching to a trimester academic calendar. A random sample of 100 administrators, 50 students, and 50 faculty members were surveyed

| Opinion | Administrators | Students | Faculty |
|---|---|---|---|
| Favor | 63 | 20 | 37 |
| Oppose | 37 | 30 | 13 |
| Totals | 100 | 50 | 50 |

Using a 1% level of significance, which groups have a different attitude?

# Chi-Square Test Results

DCOVA

$H_0$: $\pi_1 = \pi_2 = \pi_3$

$H_1$: Not all of the $\pi_j$ are equal (j = 1, 2, 3)

**Chi-Square Test: Administrators, Students, Faculty**

| | Admin | Students | Faculty | Total |
|---|---|---|---|---|
| **Favor** | 63 | 20 | 37 | 120 |
| | 60 | 30 | 30 | |
| **Oppose** | 37 | 30 | 13 | 80 |
| | 40 | 20 | 20 | |
| **Total** | 100 | 50 | 50 | 200 |

Expected

Observed

$$\chi^2_{STAT} = 12.792 > \chi^2_{0.01} = 9.2103 \text{ so reject } H_0$$

# Marascuilo Procedure: Solution

DCOVA

Excel Output:

compare

| Marascuilo Procedure | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| | Sample | Sample | | Absolute | Std. Error | Critical | |
| Group | Proportion | Size | Comparison | Difference | of Difference | Range | Results |
| 1 | 0.63 | 100 | 1 to 2 | 0.23 | 0.084445249 | 0.2563 | Means are not different |
| 2 | 0.4 | 50 | 1 to 3 | 0.11 | 0.078606615 | 0.2386 | Means are not different |
| 3 | 0.74 | 50 | 2 to 3 | 0.34 | 0.092994624 | 0.2822 | Means are different |

At 1% level of significance, there is evidence of a difference in attitude between students and faculty

# $\chi^2$ Test of Independence

DCOVA

- Similar to the $\chi^2$ 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)

# $\chi^2$ Test of Independence

*(continued)*

The Chi-square test statistic is:

DCOVA

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

- 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

$\chi^2_{STAT}$ 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

DCOVA

- 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

DCOVA

- The decision rule is

If $\chi^2_{STAT} > \chi^2_{\alpha}$, reject $H_0$,

otherwise, do not reject $H_0$

Where $\chi^2_{\alpha}$ 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:

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

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

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

# Example: Decision and Interpretation

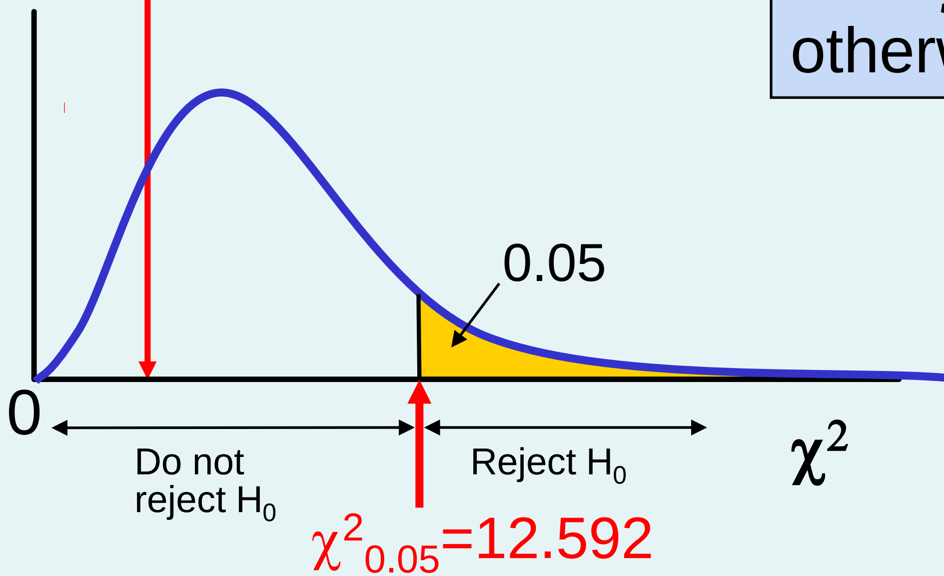
DCOVA

*(continued)*

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

**Decision Rule:**

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




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

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