

Inferential Statistics and Probability  
a Holistic Approach

---

Chapter 11  
Chi-square Tests for  
Categorical Data

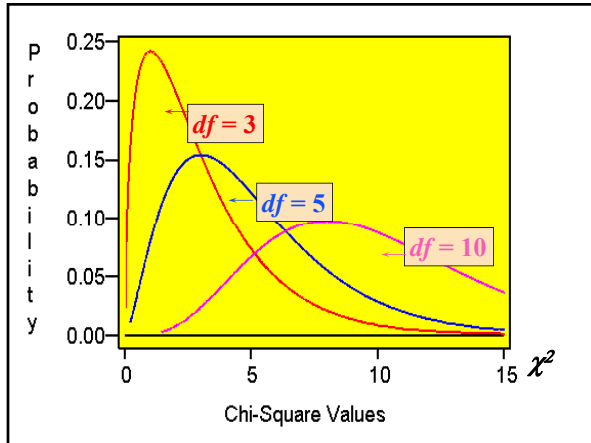
This Course Material by Maurice Geraghty is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. Conditions for use are shown here: <https://creativecommons.org/licenses/by-sa/4.0/>

1

### Characteristics of the Chi-Square Distribution

- The major characteristics of the chi-square distribution are:
  - It is positively skewed
  - It is non-negative
  - It is based on degrees of freedom
  - When the degrees of freedom change a new distribution is created

2



### Goodness-of-Fit Test: Equal Expected Frequencies

- Let  $O_i$  and  $E_i$  be the observed and expected frequencies respectively for each category.
- $H_0$ : there is no difference between Observed and Expected Frequencies
- $H_a$ : there is a difference between Observed and Expected Frequencies
- The test statistic is:  $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$
- The critical value is a chi-square value with  $(k-1)$  degrees of freedom, where  $k$  is the number of categories

4

### EXAMPLE 1

The following data on absenteeism was collected from a manufacturing plant. At the .01 level of significance, test to determine whether there is a difference in the absence rate by day of the week.

Day	Frequency
Monday	95
Tuesday	65
Wednesday	60
Thursday	80
Friday	100

5

### EXAMPLE 1 *continued*

- Assume equal expected frequency:  $(95+65+60+80+100)/5=80$

Day	O	E	$(O-E)^2/E$
Mon	95	80	2.8125
Tues	65	80	2.8125
Wed	60	80	5.0000
Thur	80	80	0.0000
Fri	100	80	5.0000
<b>Total</b>	<b>400</b>	<b>400</b>	<b>15.625</b>

6

### EXAMPLE 1 *continued*

$H_0$ : there is no difference between the observed and the expected frequencies of absences.

- $H_a$ : there is a difference between the observed and the expected frequencies of absences.
- Test statistic:  $\chi^2 = \sum(O-E)^2/E = 15.625$
- Decision Rule: reject  $H_0$  if test statistic is greater than the critical value of 13.277. (4 df,  $\alpha = .01$ )
- Conclusion: reject  $H_0$  and conclude that there is a difference between the observed and expected frequencies of absences.

7

### Goodness-of-Fit Test: Unequal Expected Frequencies

#### EXAMPLE 2

The U.S. Bureau of the Census (2000) indicated that 54.4% of the population is married, 6.6% widowed, 9.7% divorced (and not re-married), 2.2% separated, and 27.1% single (never been married).

A sample of 500 adults from the San Jose area showed that 270 were married, 22 widowed, 42 divorced, 10 separated, and 156 single.

At the .05 significance level can we conclude that the San Jose area is different from the U.S. as a whole?

8

### EXAMPLE 2 *continued*

Status	O	E	$\sum \frac{(O-E)^2}{E}$
Married	270	272	0.015
Widowed	22	33	3.667
Divorced	42	48.5	0.871
Separated	10	11	0.091
Single	156	135.5	3.101
Total	500	500	7.745

9

### EXAMPLE 2 *continued*

- **Design:**  $H_0: p_1 = .544, p_2 = .066, p_3 = .097, p_4 = .022, p_5 = .271$   
 $H_a$ : at least one  $p_i$  is different
- $\alpha = .05$
- Model: Chi-Square Goodness of Fit,  $df = 4$
- $H_0$  is rejected if  $\chi^2 > 9.488$
- **Data:**  $\chi^2 = 7.745$ , Fail to Reject  $H_0$
- **Conclusion:** Insufficient evidence to conclude San Jose is different than the US Average

10

### Contingency Table Analysis

- **Contingency table** analysis is used to test whether two traits or variables are related.
- Each observation is classified according to two variables.
- The usual hypothesis testing procedure is used.
- The *degrees of freedom* is equal to: (number of rows-1)(number of columns-1).
- The expected frequency is computed as: Expected Frequency = (row total)(column total)/grand total

11

### EXAMPLE 3

- In May 2014, Colorado became the first state to legalize the recreational use of marijuana.
- A poll of 1000 adults were classified by gender and their opinion about legalizing marijuana
- At the .05 level of significance, can we conclude that gender and the opinion about legalizing marijuana for recreational use are dependent events?

Marijuana should be	Men	Women	Total
Legal	270	230	500
Not Legal	205	245	450
Unsure	25	25	50
<b>Total</b>	<b>500</b>	<b>500</b>	<b>1000</b>

12

