Statistics 2
Chapter 9

Nonparametric Methods: Chi-Square Applications

Chapter 9

Nonparametric Methods: Chi-Square Applications

GOALS

When you have completed this chapter, you will be able to:

ONE

List the characteristics of the Chi-Square Distribution.

TWO

Conduct a test of hypothesis comparing an observed set of frequencies to an expected set of frequencies.

THREE

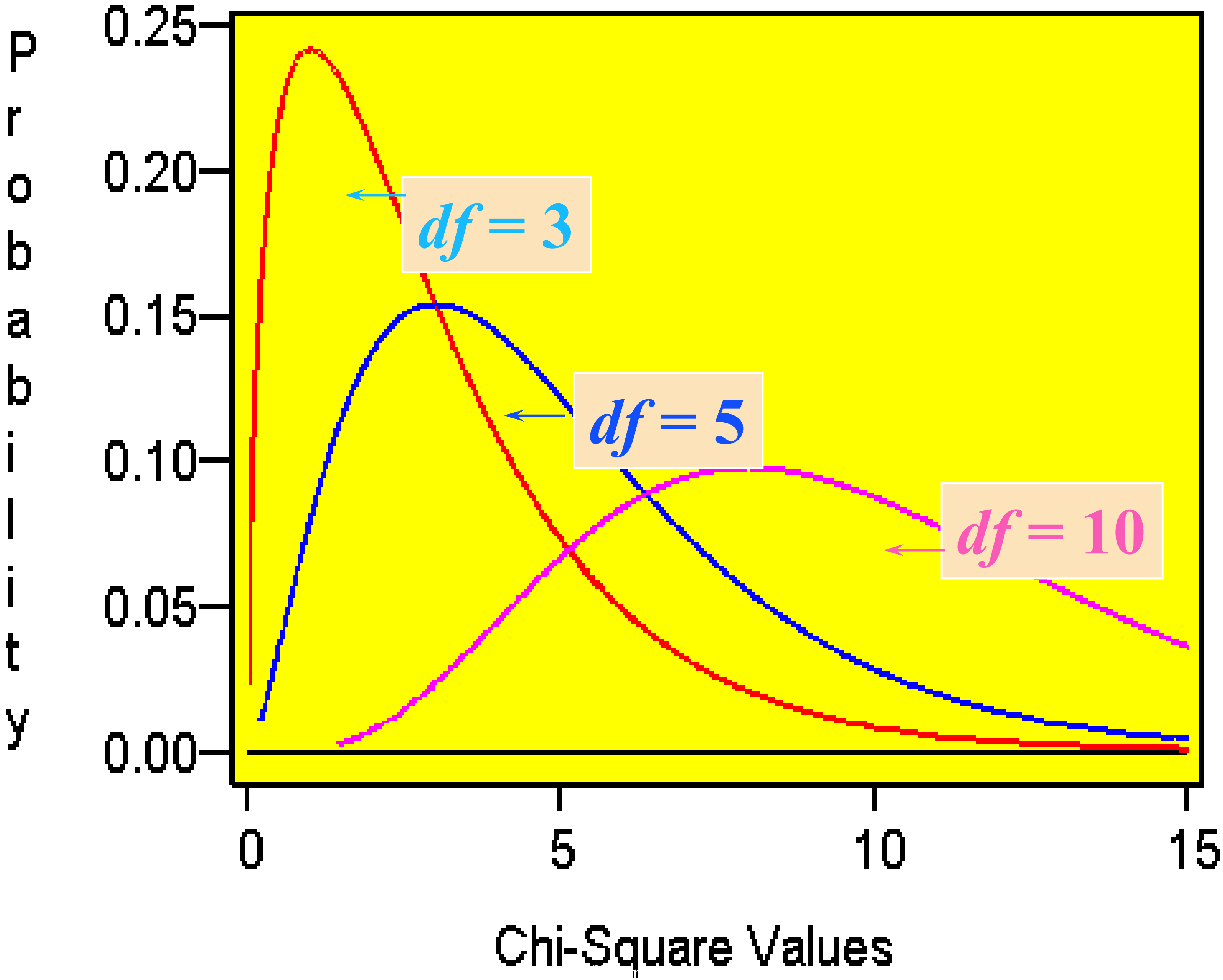
Conduct a test of hypothesis for normality using the Chi-square distribution.

FOUR

Conduct a hypothesis test to determine whether two classification criteria are related.

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


Probability
0.25
0.20
0.15
0.10
0.05
0.00
df = 3
df = 5
df = 10
0
5
10
15
Chi-Square Values

Goodness-of-Fit Test: Equal Expected Frequencies

- Let f_0 and f_e be the observed and expected frequencies respectively.
- H_0 : no difference between f_0 and f_e
- H_1 : there is a difference between f_0 and f_e
- The test statistic is:

$$x^2 = \Sigma\left[\frac{\left(f_0 - f_e\right)^2}{f_e}\right]$$

- The critical value is a chi-square value with (k-1) degrees of freedom, where k is the number of categories

EXAMPLE 1

- The following data on absenteeism was collected from a manufacturing plant. At the .05 level of significance, test to determine whether there is a difference in the absence rate by day of the week

Day	Frequency
Monday	120
Tuesday	45
Wednesday	60
Thursday	90
Friday	130

EXAMPLE 1 *continued*

- Assume equal expected frequency: (120+45+60+90+130)/5=89.
- Using these numbers, the computed test statistic is 42.4719.
- The degrees of freedom is (5-1)=4.
- Therefore, the critical value is 9.488

EXAMPLE 1 *continued*

- H_0 : there is no difference between the observed and the expected frequencies of absences.
- H_1 : there is a difference between the observed and the expected frequencies of absences.
- Test statistic: chi-square=60.8
- Decision Rule: reject H_0 if test statistic is greater than the critical value.
- Conclusion: reject H_0 and conclude that there is a difference between the observed and expected frequencies of absences.

14-8

Goodness-of-Fit Test: Unequal Expected Frequencies EXAMPLE 2

- The U.S. Bureau of the Census indicated that 63.9% of the population is married, 7.7% widowed, 6.9% divorced (and not re-married), and 21.5% single (never been married). A sample of 500 adults from the Philadelphia area showed that 310 were married, 40 widowed, 30 divorced, and 120 single. At the .05 significance level can we conclude that the Philadelphia area is different from the U.S. as a whole?

EXAMPLE 2 *continued*

Status	f_0	f_e	$(f_0-f_e)^2/f_e$
Married	310	319.5	.2825
Widowed	40	38.5	.0584
Divorced	30	34.5	.5870
Single	120	107.5	1.4535
Total	500		2.3814

EXAMPLE 2 *continued*

- Step 1: H_0 : The distribution has not changed.
- H_1 : The distribution has changed.
- Step 2: H_0 is rejected if $x^2 > 7.815, df = 3, \alpha = .05$
- Step 3: $x^2 = 2.3824$
- Step 4: H_0 is rejected. The distribution has changed.

Goodness-of-Fit Test for Normality

- *Purpose:* To test whether the observed frequencies in a frequency distribution match the theoretical normal distribution.
- *Procedure:* Determine the mean and standard deviation of the frequency distribution.
  - Compute the z-value for the lower class limit and the f_e upper class limit for each class.
  - Determine f_0 for each category f_e
  - Use the chi-square goodness-of-fit test to determine if coincides with .

EXAMPLE 3

- A sample of 500 donations to the Arthritis Foundation is reported in the following frequency distribution. Is it reasonable to conclude that the distribution is normally distributed with a mean of $10 and a standard deviation of $2? Use the .05 significance level.
- *Note*: To compute f_e for the first class, first compute the probability for this class. P(X<6)=P[Z<(6-10)/2]=.0228. Thus f_e is (.0228)(500)=11.4

EXAMPLE 3 *continued*

amount spent	f_0	area	f_e	$(f_0-f_e)^2/f_e$
<$6	20	.02	11.40	6.49
$6-8	60	.14	67.95	.93
$8-10	140	.34	170.65	5.50
$10-12	120	.34	170.65	15.03
$12-14	90	.14	67.95	7.16
>$14	70	.02	11.40	301.22
Total	500		500	336.33

EXAMPLE 3 *continued*

- Step 1: H_0 : The distribution is normal.
- H_1 : The distribution is not normal.
- Step 2: H_0 is rejected if $x^2 > 11.07, df = 5, \alpha = .05$
- Step 3: $x^2 = 336.33$
- Step 4: H_0 is rejected . The distribution is not normal

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

EXAMPLE 4

- Is there a relationship between the location of an accident and the sex of the person involved in the accident? A sample of 150 accidents reported to the police were classified by type and gender. At the .05 level of significance, can we conclude that gender and the location of the accident are related?

EXAMPLE 4 *continued*

Sex	Work	Home	Other	Total
Male	60	20	10	90
Female	20	30	10	60
Total	80	50	20	150

Note: The expected frequency for the work-male intersection is computed as (90)(80)/150=48. Similarly, you can compute the expected frequencies for the other cells.

EXAMPLE 4 *continued*

- Step 1: H_0 : Gender and location are not related. H_1 : Gender and location are related.
- Step 2: H_0 is rejected if $x^2 > 5.991, df = 2, \alpha = .05$
- Step 3: $x^2 = 16.667$
- Step 4: H_0 is rejected. Gender and location are related.