Chapter 20 Nonparametric Hypothesis Testing of Nominal Data

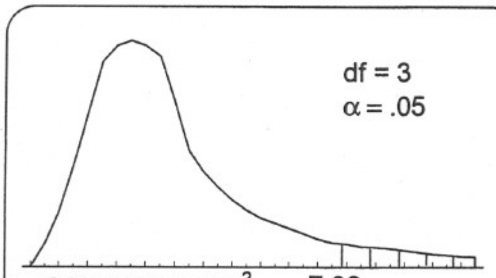
I. **Introduction**

A. **Parametric statistics** is the name given to much of the material covered through chapter 19.
  1. Parametric tests involve a population parameter for which the test statistic has a known distribution (shape).
  2. Measurement (data) sophistication is of an interval or ratio level. (see page 2)

B. **Nonparametric statistics** are used when the requirements of parametric statistics are not fulfilled.
  1. Data is considered **distribution-free** because the distribution of the sample statistic may be unknown.
  2. Nominal and ordinal data can be tested.

C. **Count data (categorical data)**
  1. In this chapter, sample observations (counts) are grouped into categories and compared to some expected count (frequency). A small difference between the actual and expected frequencies indicates a match.
  2. Applications
     a. Determining brand preference by age, gender, etc.
     b. Measuring the success of an advertising campaign or training program.

D. **The chi-square distribution** (pronounced "kigh" square)
  1. The chi-square distribution is like the t distribution because there is a family of curves, one for each degree of freedom.
  2. The distribution becomes more normal as the degrees of freedom increase. Chi-square is the ratio of $(n-1)s^2$ to σ^2.



II. **Goodness of fit tests for a one categorical variable**

A. Linda is interested in determining if consumers at her four stores are giving equal acceptance to the low sales price of a new hit music video.

Music Video Sales	Store A	Store B	Store C	Store D	Totals
Sample sales, f_0	8	22	19	11	60
Expected sales, f_e	15	15	15	15	60

B. The 5-step approach to hypothesis testing
  1. H_0 : sales are equally distributed

     H_1 : sales are not equally distributed
  2. The significance level is .05.
  3. Chi-square is the test statistic.

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

f_0 is an observed frequency of a category.
f_e is an expected frequency of a category. It should be ≥ 5 when using the continuous chi-square distribution for a discrete problem.
Equal acceptance means f_e= 60/4 = 15.
k is the number of categories.
There are k - 1 degrees of freedom for a goodness of fit problem.

  4. The decision rule:
     If χ^2 from the test statistic is beyond the critical value, the difference is high and the null hypothesis is rejected.
  5. Apply the decision rule for this one-tail test.

Store	f_0	f_e	f_0-f_e	$(f_0-f_e)^2$	$\frac{(f_0-f_e)^2}{f_e}$
A	8	15	-7	49	49/15 = 3.27
B	22	15	7	49	49/15 = 3.27
C	19	15	4	16	16/15 = 1.07
D	11	15	-4	16	16/15 = 1.07
			0		χ^2 = 8.68

df = k - 1 = 4 - 1 = 3 → χ = 7.82
Reject H_0 because 8.68 > 7.82.
Sales are not equally distributed.

Chi-Square					
Degrees	Right-tail area				
of freedom	.10	.05	.025	.01	.005
1	2.71	3.84	5.02	6.64	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.82	9.35	11.35	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75

Page ST 6 has a more complete chi-square table.

Note: This procedure can be used to test unequal expected frequencies. Suppose Store A usually has 40% of company sales and the 3 other stores each have 20%. Store A would be expected to have 24 sales (.40 x 60) and the other stores would be expected to have 12 sales (.20 x 60).
Note: Opinions vary on the exact lower limit for f_e.