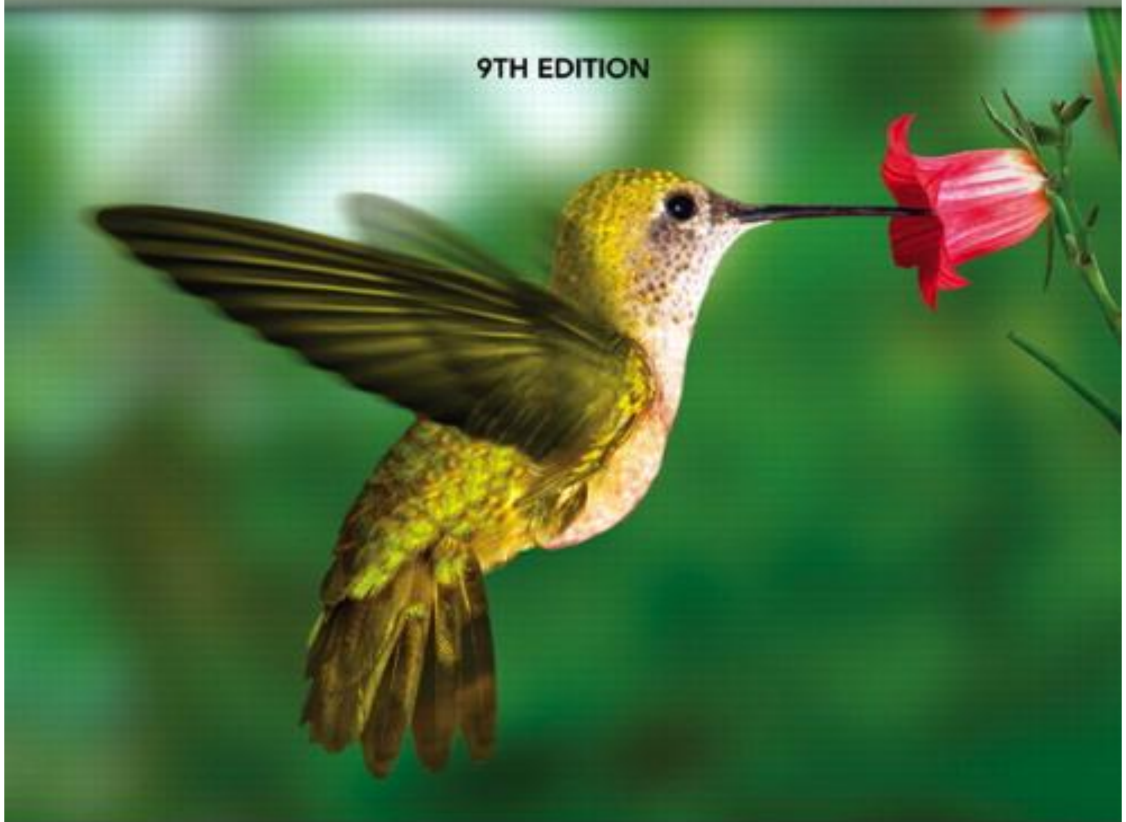
Introductory STATISTICS

9TH EDITION



Neil WEISS

Chapter 16

Analysis of Variance (ANOVA)

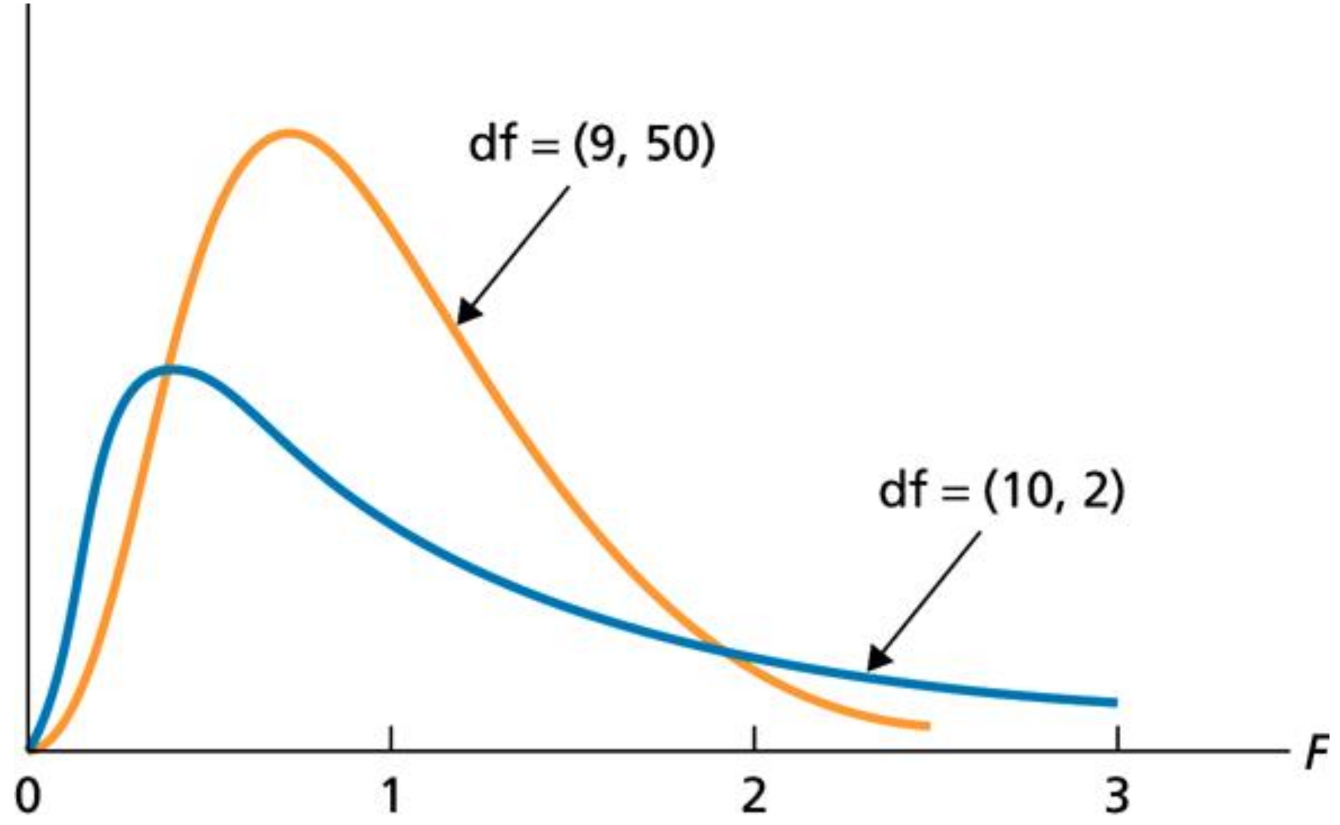
PEARSON

Section 16.1

The *F*-Distribution

Figure 16.1

Two different *F*-curves



Section 16.2

One-Way ANOVA: The Logic

Table 16.1 & Figure 16.3

Sample data from Populations 1 and 2

Sample from Population 1	21	37	11	20	8	23
Sample from Population 2	24	31	29	40	9	17

Dotplots for sample data in Table 16.1

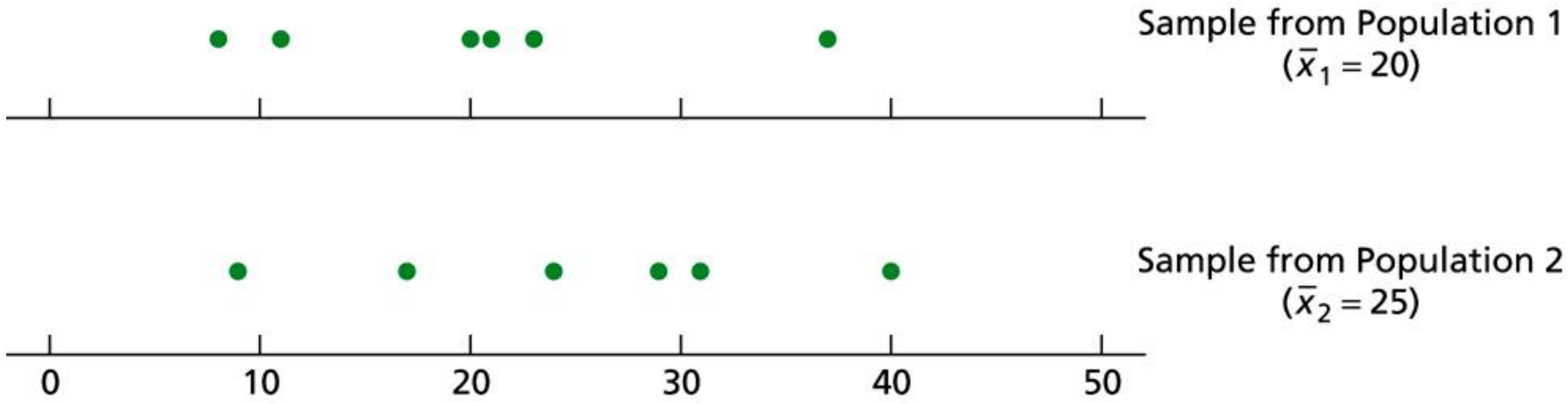


Table 16.2 & Figure 16.4

Sample data from Populations 1 and 2

Sample from Population 1	21	21	20	18	20	20
Sample from Population 2	25	28	25	24	24	24

Dotplots for sample data in Table 16.2

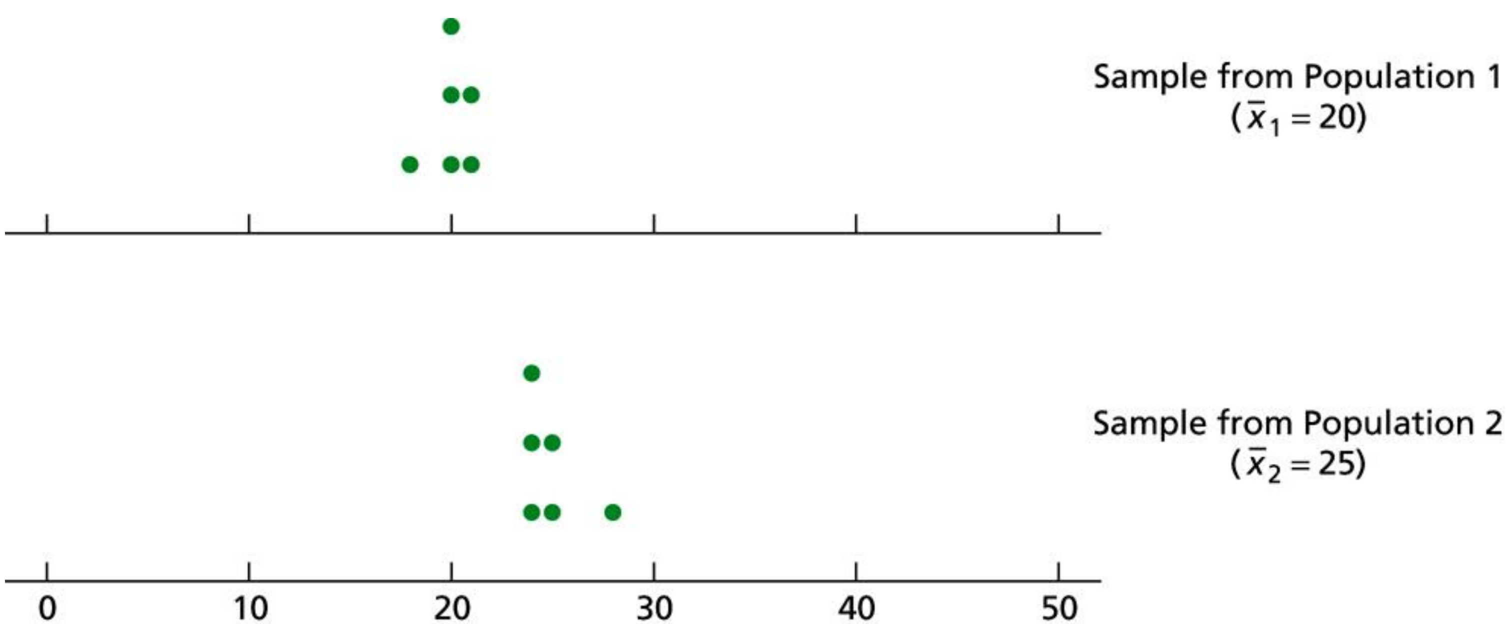
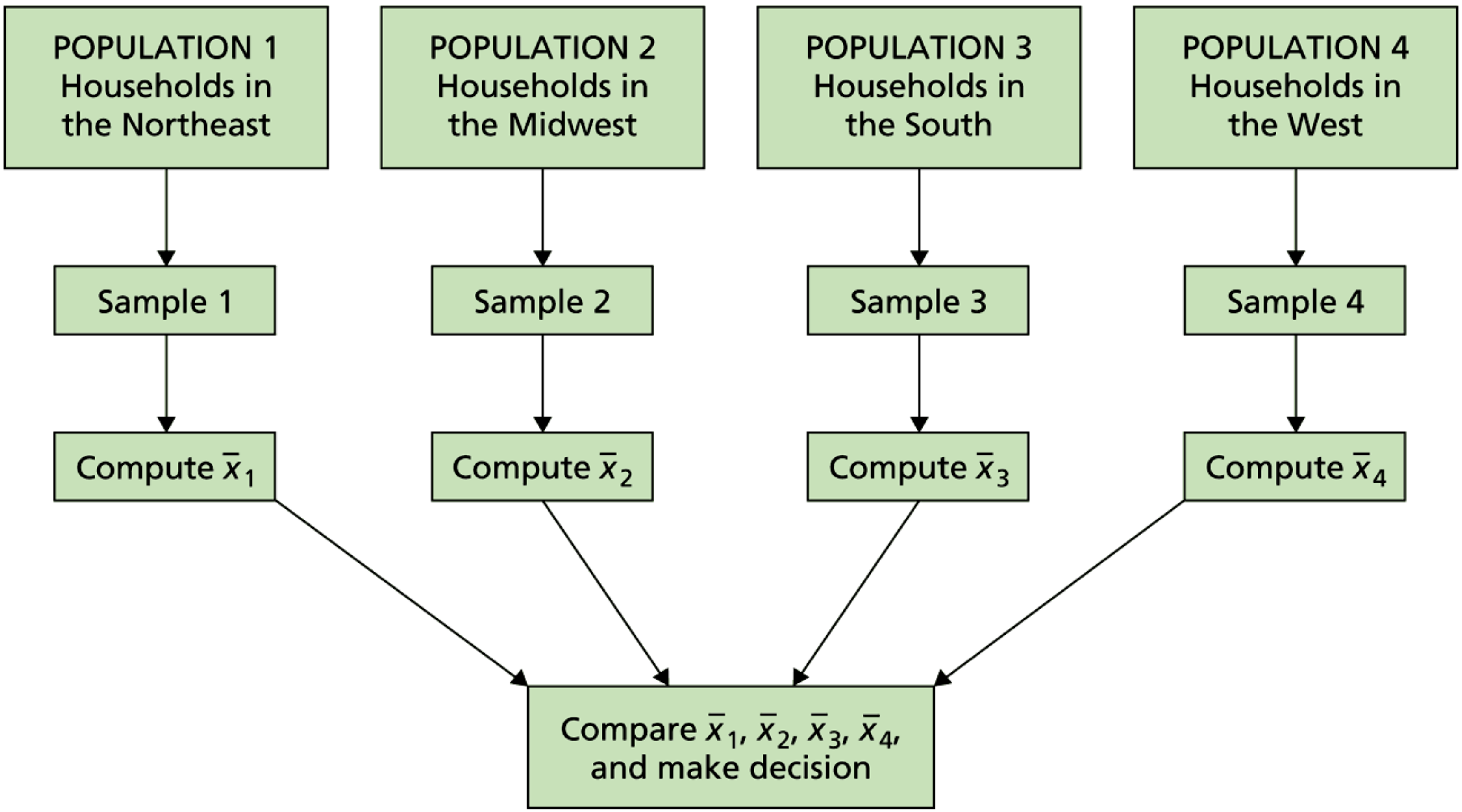


Figure 16.5

Process for comparing four population means



Section 16.3

One-Way ANOVA: The Procedure

Table 16.4

ANOVA table format for a one-way analysis of variance

Source	**df**	***SS***	***MS* = *SS*/df**	***F*-statistic**
Treatment	$k-1$	$SSTR$	$MSTR = \frac{SSTR}{k-1}$	$F = \frac{MSTR}{MSE}$
Error	$n-k$	SSE	$MSE = \frac{SSE}{n-k}$	
Total	$n-1$	SST		

Procedure 16.1

One-Way ANOVA Test

Purpose To perform a hypothesis test to compare k population means, $\mu_1, \mu_2, \ldots, \mu_k$

Assumptions

1. Simple random samples
2. Independent samples
3. Normal populations
4. Equal population standard deviations

Step 1 **The null and alternative hypotheses are, respectively,**

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

H_a: **Not all the means are equal.**

Step 2 **Decide on the significance level, α.**

Step 3 **Compute the value of the test statistic**

$$F = \frac{MSTR}{MSE}$$

and denote that value F_0. To do so, construct a one-way ANOVA table:

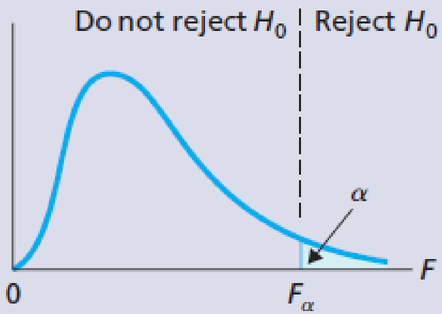
Source	df	SS	$MS = SS/\text{df}$	F-statistic
Treatment	$k - 1$	$SSTR$	$MSTR = \frac{SSTR}{k-1}$	$F = \frac{MSTR}{MSE}$
Error	$n - k$	SSE	$MSE = \frac{SSE}{n-k}$	
Total	$n - 1$	SST		

Procedure 16.1 (cont.)

CRITICAL-VALUE APPROACH	OR	P-VALUE APPROACH

CRITICAL-VALUE APPROACH

Step 4 The critical value is F_α with df $= (k-1, n-k)$. Use Table VIII to find the critical value.

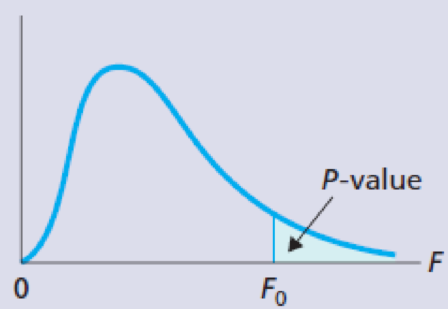


Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

OR

P-VALUE APPROACH

Step 4 The F-statistic has df $= (k-1, n-k)$. Use Table VIII to estimate the P-value or obtain it exactly by using technology.



Step 5 If $P \le \alpha$, reject H_0; otherwise, do not reject H_0.

Step 6 Interpret the results of the hypothesis test.

Section 16.4

Multiple Comparisons

Procedure 16.2

Tukey Multiple-Comparison Method

Purpose To determine the relationship among k population means $\mu_1, \mu_2, \ldots, \mu_k$

Assumptions

1. Simple random samples
2. Independent samples
3. Normal populations
4. Equal population standard deviations

Step 1 **Decide on the family confidence level, $1 - \alpha$.**

Step 2 **Find q_α for the q-curve with parameters $\kappa = k$ and $\nu = n - k$, where n is the total number of observations.**

Step 3 **Obtain the endpoints of the confidence interval for $\mu_i - \mu_j$:**

$$(\bar{x}_i - \bar{x}_j) \pm \frac{q_\alpha}{\sqrt{2}} \cdot s\sqrt{(1/n_i) + (1/n_j)},$$

where $s = \sqrt{MSE}$. Do so for all possible pairs of means with $i < j$.

Step 4 **Declare two population means different if the confidence interval for their difference does not contain 0; otherwise, do not declare the two population means different.**

Step 5 **Summarize the results in Step 4 by ranking the sample means from smallest to largest and by connecting with lines those whose population means were not declared different.**

Step 6 **Interpret the results of the multiple comparison.**

Section 16.5

The Kruskal-Wallis Test

Table 16.10

Number of miles driven (1000s) last year for independent samples of cars, buses, and trucks

Cars	Buses	Trucks
19.9	1.8	24.6
15.3	7.2	37.0
2.2	7.2	21.2
6.8	6.5	23.6
34.2	13.3	23.0
8.3	25.4	15.3
12.0		57.1
7.0		14.5
9.5		26.0
1.1		

Table 16.11

Results of ranking the combined data from Table 16.10

Cars	Rank	Buses	Rank	Trucks	Rank	
19.9	16	1.8	2	24.6	20	
15.3	14.5	7.2	7.5	37.0	24	
2.2	3	7.2	7.5	21.2	17	
6.8	5	6.5	4	23.6	19	
34.2	23	13.3	12	23.0	18	
8.3	9	25.4	21	15.3	14.5	
12.0	11			57.1	25	
7.0	6			14.5	13	
9.5	10			26.0	22	
1.1	1					
	9.850		9.000		19.167	← *Mean ranks*

Procedure 16.3

Kruskal–Wallis Test

Purpose To perform a hypothesis test to compare k population means, $\mu_1, \mu_2, \ldots, \mu_k$

Assumptions

1. Simple random samples
2. Independent samples
3. Same-shape populations
4. All sample sizes are 5 or greater

Step 1 The null and alternative hypotheses are, respectively,

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$
$$H_a: \text{Not all the means are equal.}$$

Step 2 Decide on the significance level, α.

Step 3 Compute the value of the test statistic

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

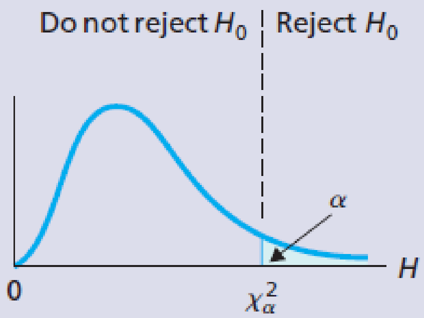
and denote that value H_*. Here, n is the total number of observations and $R_1, R_2, \ldots, R_k$ denote the sums of the ranks for the sample data from Populations 1, 2, ..., k, respectively. To obtain H, first construct a work table to rank the data from all the samples combined.

Procedure 16.3 (cont.)

CRITICAL-VALUE APPROACH	OR	P-VALUE APPROACH

CRITICAL-VALUE APPROACH

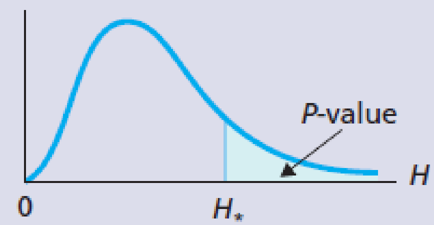
Step 4 The critical value is χ^2_α with df $= k - 1$. Use Table VII to find the critical value.



Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

P-VALUE APPROACH

Step 4 The H-statistic has df $= k - 1$. Use Table VII to estimate the P-value or obtain it exactly by using technology.



Step 5 If $P \le \alpha$, reject H_0; otherwise, do not reject H_0.

Step 6 Interpret the results of the hypothesis test.