SUMMARY

The statistical tests discussed prior to this chapter (such as the z-test and the t-test) are based on certain assumptions. The major assumptions are these:

1. The scores constituting the raw data, or the x's, are *independent* of one another.
2. The x's are drawn from a *normal distribution.*
3. The scores from the different conditions are drawn from populations having *equal variances.*

Violation of the independence assumption has a rather serious effect, but most experiments can be arranged so that this assumption is not violated. Violations of the other two assumptions are less serious. But if one of these assumptions has been strongly violated, it is proper to analyze data using a nonparametric test. Nonparametric tests do not make any assumptions about the parameters of the populations from which samples were drawn. Most parametric tests have nonparametric tests corresponding to them.

DIGRESSION

DIGRESSION 17-1

When the n's Are Large

When there are more than five observations in each of the various samples, the statistic H is distributed as a $\chi^2(J - 1)$, where J, as usual, is the number of groups. This makes things easy, for instead of using Table E-7 in Appendix E, we can determine whether an observed H is significant by reference to Table E-5. If the observed value of H is equal to or larger than the value of χ^2 given in Table E-5, H_0 may be rejected. So if we had 10 observations in each of our samples instead of only 4, we would have required an H[crit $\chi^2(2)$] of 5.99 or greater in order to reject the null hypothesis at the 0.05 level of significance.

PROBLEMS

1. A developmental psychologist is interested in whether kindergarten boys and girls are different from each other in terms of how happy they are. A happiness test is given to 5 boys and 5 girls, and the 10 children are ranked from first to tenth in terms of their total happiness score. Are boys significantly happier than girls?

Ranking of happiness	1	2	3	4	5	6	7	8	9	10
Sex of child	G	G	B	G	G	B	B	G	B	B

2. An experimenter is testing the hypothesis that a person's expectations can influence the outcome of an experiment. Nine rats, all from the same strain, are given to psychology students. Some students are told that these rats were bred to be bright but others are told that the rats were bred to be dull. Each student takes a single rat and tests its ability to learn a maze, carefully recording the number of errors made. Do they provide support for the hypothesis that relative to a student who expects the rat to be dull, a student who expects his rat to be bright finds the rat making fewer errors?

Errors Made by Rats That Are Expected to Be	
Bright	Dull
6 8 5 12	10 2 14 9 4

3. Two researchers are interested in the question of whether psychologists or political scientists tend to have larger book collections. They collect some data from 9 political scientists and 12 psychologists; they visit the offices of these individuals and simply count the number of books on the shelves. The observations are given below. Determine whether there is a significant difference between the size of collection using a nonparametric test.

Political Scientists	Psychologists
87	131
72	94
65	77
54	88
67	116
76	90
73	87
82	76
104	95
	164
	127
	77

4. Two political researchers are interested in whether the Watergate political scandal had any effect on the voting practices of Democrats versus Republicans. They hypothesize that Republicans would be so discouraged that they would be less likely to vote in local and national elections. The voting behavior of three Democrats and four Republicans is examined. The number of times that each of these people voted in the previous three years is given below. Test the hypothesis that Republicans vote less often than Democrats.

Democrats	Republicans
4	2
3	1
1	0
	0

5. A medical scientist, Dr. Greenwald, investigated the effect of a new drug on patients with insomnia. She took three groups of patients and administered drug A to one group, drug B to a second group, and no drug to the third group. After a week the patients reported the average number of hours they had been able to sleep. A total of 15 patients participated in the experiment, 5 per group. When the data were in, Dr. Greenwald ranked them from best to worst (1–15) in terms of ability to sleep.

   Use a nonparametric procedure to analyze her data:

Drug A	Drug B	No Drug
3	1	2
6	5	4
11	9	7
12	10	8
15	14	13

6. Dr. Smith is interested in whether people can more accurately detect the pitch of a tone when it is presented to the right or left ear rather than to the top of the head. Six individuals hear tones presented to the right ear, six hear tones presented to the left ear, and six hear tones presented over the top of the heads. Does nonparametric statistical analysis support the contention that the placement of the tone makes a difference? Analyze the following data, which are the number of errors that each subject made.

Right Ear	Left Ear	Top of Head
0	1	3
0	2	5
0	3	2
1	2	1
0	2	4
0	1	5

7. Many Americans are currently seeking to improve their health by jogging. Dr. Jones wishes to study the effects of jogging on well-being. He devises a test designed to measure subjective well-being, a test in which people can receive a score between 0 and 40. He finds 36 individuals to participate in his study. One-third are instructed to jog 2 miles per day, one-third 4 miles per day, and one-third are not given a treatment. At the end of the week his well-being test is administered, and the following scores are obtained.

Use a nonparametric procedure to determine whether the groups are different from each other.

Control	2 Miles per Day	4 Miles per Day
6	31	13
11	7	32
12	9	31
20	11	30
24	16	28
21	19	29
18	17	25
15	11	26
14	22	26
10	23	27
8	27	26
14	26	19

8. Many statisticians suggest that when there are ties in the rankings, the value of H should be corrected. The formula to be used for correcting is

$$C = 1 - \frac{\sum_{i}^{T} (t_i^3 - t_i)}{N^3 - N}$$

where T is the number of sets of tied observations and t_i is the number tied in any set i. In the previous problem there were four sets of two tied observations, one set of three ties, and one set of four tied observations. Thus,

$$C = 1 - \left(\frac{4[(2)^3 - 2] + [(3)^3 - 3]}{(36)^3 - 36} + [(4)^3 - 4] \right) = 0.997$$

To correct the value of H, we divide it by C:

$$H^1 = \frac{H}{C} = \frac{13.81}{0.997} = 13.84$$

Now suppose that the first subject score of the 4-mile-per-day condition had received a 7 instead of a 13. What would this do to the value of H? Correct this value for ties.

9. Ten Triumphs and nine Pontiacs were driven from Seattle to San Francisco. The total number of hours it took each of the cars to travel the distance is given here:

Triumphs	16.5	16.6	16.7	16.1	17.7	17.4	16.2	16.4	17.1	16.9
Pontiacs	16.3	17.9	17.8	16.8	17.0	17.2	17.3	17.5	17.6	

a. Suppose you had no preconceived notions about whether drivers of Triumphs drive faster than drivers of Pontiacs. Based on a nonparametric

test of these data, would you conclude that one type of person was a faster driver?

b. Assume that the n's in this case were large enough so that the normal approximation could be used. Would your conclusion be any different?

c. Compare the above results with those based on a t-test.

10. A student taking a statistics course was required to think up a question to be answered, collect some data, and then analyze the data as part of a course requirement. The question in which she was most interested was whether her favorite Rolling Stones album was cheaper if bought in a record store in California, Oregon, or Washington. She wrote letters to a number of stores in the three states asking what they charged for the album. Responses came from 18 stores in California, 17 in Oregon, and 18 in Washington. The prices each store charged for the album are given here:

Cost for Rolling Stones Album

California		Oregon		Washington	
3.92	4.08	4.70	1.56	2.43	2.52
3.45	3.67	4.15	4.29	2.98	3.10
2.00	2.94	4.55	1.74	3.04	3.53
2.28	5.90	3.31	2.17	4.94	3.06
3.50	2.18	2.13	1.97	3.15	2.57
4.25	5.39	4.68	4.69	2.46	3.48
2.38	2.74	2.68	2.87	3.34	5.94
3.02	3.49	2.36	3.17	2.39	2.51
3.26	2.70	3.93		2.27	1.61