CHAPTER 15

ALTERNATIVES TO *t* AND *F*

OBJECTIVES

After completing this chapter, you should

- be able to compute and use nonparametric alternatives to parametric tests, such as the *t* test for independent samples (nonparametric alternative—Mann-Whitney *U* test), the *t* test for dependent samples (nonparametric alternative—Wilcoxon test) and the one-way ANOVA (nonparametric alternative—Kruskal-Wallis test).

CHAPTER REVIEW

Nonparametric alternatives to *t* and *F* are useful when the data are ordinal scale, ease of computation is desired, or the samples are small and have unequal *N*s. Small samples often have a large amount of variability, which makes resulting *t* ratios and *F* ratios relatively small. By avoiding a measure of variability, a nonparametric test may be more powerful than the corresponding parametric test with small sample sizes and outliers (extreme scores).

The *Mann–Whitney U test* is a useful alternative to the two-sample *t* test for independent samples when the measurement level is ordinal and the normality assumption cannot be made. The M–W assumes that the samples are independent, that there is an underlying continuous measurement scale, and that the measurement scale actually used is at least ordinal. The tested hypothesis is that the populations from which the samples are drawn are identical in shape. If the populations are identical, then if we rank the combined observations from two samples, the scores from the two samples should be evenly mixed in the combined ranking. If the scores are not evenly mixed, then they probably come from different populations.

The computed U is a measure of how evenly mixed the scores from the combined samples are, with a low value indicating a lack of mixing. Thus, unlike the previous significance tests, the null hypothesis is rejected if U (or U') is equal to or *less* than values from Table H (see Appendix 2). With samples containing more than 20 observations, U is converted to a z score. The null hypothesis is rejected if the computed score is equal to or *larger* than 1.96 (two-tailed test at the 5% level) or 1.64 (one-tailed test at the 5% level).

The *Wilcoxon matched-pairs signed-ranks test* is a nonparametric alternative to the t test for dependent samples. The test assumes that the subjects are randomly and independently selected, that there is at least ordinal scale measurement, and that it is possible to rank-order the difference scores (difference between a pair of scores). The tested hypothesis is that the distributions of the populations under each condition are identical.

First, the difference between each pair of scores is found, with 0 differences discarded. Next, the absolute values of the difference scores are rank-ordered, but the signs are retained. The ranks with the less frequently occurring sign are summed, and the resulting T is compared with table values in a significance test. If the population distributions are really identical, there will usually be about the same number of positive and negative differences, and the sums of the ranks for the positive and negative differences will not be very different. However, if the distributions are dissimilar, there will be many more differences of one sign than of the other. Thus, the smaller the sum of the less frequently occurring ranks, the more likely that the population distributions are different. If the computed T is equal to or *smaller* than the critical value in Table I (see Appendix 2), the null hypothesis is rejected. For samples larger than 25, the distribution of T is approximately normal, and T is converted to a z score. H_0 is rejected at the 5% level if z is 1.96 or larger in a two-tailed test (1.64 in a one-tailed test).

The *Kruskal–Wallis one-way analysis of variance by ranks* is an extension of the Mann–Whitney test for comparing more than two independent groups. A statistic, H, is computed, and if there are three groups with at least five subjects per group, H is distributed approximately as χ^2. After a significant H, further testing with the M–W test will reveal the between-group differences.

SYMBOLS

Symbol	Stands For
U or M–W U or U'	statistic computed for the Mann–Whitney (M–W) test of significance
N_1, N_2	number of subjects in the first and second groups, respectively
R_1, R_2	sum of the ranks of the scores in the first and second groups, respectively
d	differences between pairs of scores in the Wilcoxon test
T	sum of the ranks of the scores with the less frequent sign (Wilcoxon test)
H or K–W H	statistic computed for the Kruskal–Wallis (K–W) test
N_i	number of observations in a particular sample
R_i	sum of the ranks for a particular sample
K	number of samples

FORMULAS

Formula 15-1. *Computational formula for the Mann–Whitney U test*

$$U = N_1 N_2 + \frac{N_1(N_1 + 1)}{2} - R_1$$

N_1 is the number of observations in the first sample, N_2 is the number of observations in the second sample, and R_1 is the sum of the ranks of the scores in the first sample.

Formula 15-2. Equation for U′

$$U' = N_1 N_2 - U$$

The smaller of U and U' is used in the test of significance.

Formula 15-4. Equation for converting large-sample U to a z score

$$z = \frac{U - \dfrac{N_1 N_2}{2}}{\sqrt{\dfrac{(N_1)(N_2)(N_1 + N_2 + 1)}{12}}}$$

U or U' is converted to a z score when sample sizes are larger than $N = 20$.

Formula 15-5. Equation for converting a large-sample T to a z score

$$z = \frac{T - \dfrac{N(N+1)}{4}}{\sqrt{\dfrac{N(N+1)(2N+1)}{24}}}$$

T is the sum of the ranks with the less frequently occurring sign. With samples of 25 or more, T is converted to a z score.

Formula 15-6. Computational formula for the Kruskal–Wallis test

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{N_i} - 3(N+1)$$

N_i is the number of observations for a particular sample, N is the total number of observations, and R_i is the sum of the ranks for a particular sample. With sample sizes of at least 5 and at least three samples, H is distributed approximately as χ^2 with $df = K - 1$, where K is the number of samples.

TERMS TO DEFINE AND/OR IDENTIFY

Mann-Whitney test

Wilcoxon test

Kruskal-Wallis test

FILL-IN-THE-BLANK ITEMS

Introduction

A (1) _____ test is one in which population parameters such as μ and σ are not involved; they are also called (2) _____ _____ because no particular distribution is assumed. The tests discussed in this chapter are useful when (3) _____ of the parametric tests are likely to be violated and when the level of measurement is less than (4) _____ scale.

The Mann–Whitney *U* Test

The M–W test is a useful alternative to the (5) _____ _____

_____ _____ samples. The assumptions for the M–W are that the samples are (6) _____, that there is an underlying continuous scale of measurement, and that the measurement scale used is at least (7) _____ scale. The hypothesis tested is that the populations contributing to the samples are (8) _____ in shape.

 In the M–W test, the scores from the two samples are (9) _____, and a statistic, the smaller of U or (10) _____, is computed. If the ranks are not evenly mixed, the samples probably come from different (11) _____. For samples with 20 or fewer subjects, the computed U (or U') is compared with critical values in Table (12) _____. If the computed statistic is (13) _____ than the critical value, H_0 is rejected. For samples larger than 20, U (or U') is converted to a (14) _____, and H_0 is rejected if z is (15) _____ or larger (two-tailed test at 5% level).

The Wilcoxon Matched-Pairs Signed-Ranks Test

The Wilcoxon test is a nonparametric alternative to the *t* test for (16) _____ samples. The assumptions are that the subjects must be (17) _____ and independently selected, that the

measurement scale must be at least (18) _____, and that we must be able to rank-order the difference scores. The null hypothesis is that the population distributions are (19) _____.

Computation begins with finding the (20) _____ between each pair of scores, discarding all (21) _____ differences. The difference scores are (22) _____ ordered on the basis of absolute magnitude, and the (23) _____ of the differences is retained. The sum of the differences with the (24) _____ frequently occurring sign is found and called *T*. *T* is compared with critical values in Table I, and if *T* is equal to or (25) _____ than the critical value, H_0 is rejected. With large samples, $N = 25$ or greater, the distribution of *T* is approximately (26) _____, and the *T* score is converted to a (27) _____.

The Kruskal–Wallis One-Way ANOVA

The K–W test is an extension of the (28) _____ test and is used for comparing more than two groups when the assumptions underlying the (29) _____ cannot be met. At least (30) _____ scale measurement is required.

To perform the test, the combined groups are (31) _____, and the sum of the (32) _____ for each group is found. For three or more samples with at least five subjects each, the computed *H* is distributed approximately as (33) _____ _____, with $df = $ (34) _____, where *K* is the number of groups.

Further Testing After a Significant *H*

After a significant result is found with the K–W test, the (35) _____ test can be used to make further group comparisons.

Troubleshooting Your Computations

Both the M–W and the K–W tests require (36) _____ the combined scores from lowest to highest. The rank of the highest score should be (37) _____ unless the top scores are tied.

The value obtained for either U or H should be a (38) _____ number. In using the M–W test,

the smaller of U and (39) _____ is used in the significance test.

With the Wilcoxon test, be sure to discard all (40) _____ differences. The difference

scores are ranked in terms of (41) _____ value. Both the computed T and the computed U

must be equal to or (42) _____ than the critical table values for H_0 to be rejected.

PROBLEMS

1. For each of the following, tell which significance test would be most appropriate.

    a. two small, independent samples; ordinal scale data

    b. two dependent samples; interval scale data

    c. three or more independent samples; ordinal scale data

    d. two dependent samples; ordinal scale data

    e. two independent samples; frequency data

2. Seven children from families in which there is only one child and six children with at least one sibling are rated for willingness to share toys with another child. Each child is given a rating from 0 (no sharing) to 10 (virtually complete sharing) during a 20-minute observation period. Use the appropriate test to compare the groups, and tell what your decision means in the context of the problem.

Only Child	Child With Sibling(s)
5	10
3	9
2	9
2	7
1	4
0	2
0	

3. Thirty-one randomly selected rats are assigned to one of three different experimental diets. After 30 days on the diets, each animal is given a test of irritability to handling. In the test, the behavior is rated from 0 to 15 with a higher score reflecting greater irritability. The scores are shown here. Do an overall significance test. If a significant result is obtained, do all pairwise comparisons. Tell what each decision means in the context of the problem.

Diet A	Diet B	Diet C
6	14	4
5	12	4
5	12	3
4	11	2
3	9	2
2	7	1
1	7	1
0	5	0
0	4	0
0	1	0
0		

4. A trained speech analyst has received brief taped excerpts of the speech of 18 parents. Ten of the parents have schizophrenic children, and the remaining 8 have nonschizophrenic children. Without knowing whether the parent has a schizophrenic child, the analyst has rated the excerpts from 0 to 20 for defectiveness of speech. The groups do not differ on variables such as IQ, age, education, or social class. Is there evidence for a difference in the speech patterns of the parents of the schizophrenic children?

Parent of Schizophrenic	Parent of Nonschizophrenic
16	12
15	11
13	10
12	9
9	9
7	5
5	4
3	2
3	
2	

5. A self-rating scale was used to measure attitudes toward risk taking before and after alcohol consumption for 12 persons. A high score indicates a positive attitude toward risk taking; a low score indicates greater concern. Compare the before and after ratings.

Person	Rating Before	Rating After
A	14	17
B	14	19
C	13	14
D	12	9
E	11	12
F	9	9
G	9	15
H	8	7
I	5	9
J	4	8
K	2	1
L	2	5

6. Twenty-four students are selected randomly from a large introductory psychology class and assigned randomly to one of two treatment groups. Half are given an alcohol-flavored drink, and the other half receive a drink containing an ounce of alcohol. Ten minutes later, each student fills out a self-rating scale measuring attitudes toward risk taking. Assume the data are ordinal scale at best. The results are shown here. Compare the two groups. As before, a high score indicates a positive attitude toward risk taking.

Alcohol Group	No-Alcohol Group
14	19
14	17
13	15
12	14
11	12
9	9
9	9
8	9
5	8
4	7
2	5
2	1

7. Matched pairs of parents have written letters to a child in a state mental institution. One member of each pair has a schizophrenic child, and the other member has a nonschizophrenic child in the hospital. One letter from each parent has been rated for double-bind statements (incompatible ideas and feelings) on a scale from 1 to 7, with 7 reflecting a high incidence of double-bind statements. Use the appropriate test to compare the groups.

Pair	Parent of Schizophrenic	Parent of Nonschizophrenic
A	7	3
B	5	5
C	3	6
D	2	6
E	6	5
F	1	3
G	2	1
H	4	6
I	3	1
J	7	1
K	3	7
L	1	5

8. An investigator wants to see whether creativity (divergent thinking) can be taught. In one class, the teacher specifically rewards divergent responses during a 1-hour daily art period. In a second class, a 1-hour art period is held, but no effort is made to reward divergent responses. In a third class, a 1-hour study hall is given while the other classes have the art period. At the end of the year, 10 students are randomly selected from each class and given a standard test of creativity on which they receive a score from 1 to 50. A higher score indicates greater creativity. Assume the data are ordinal scale at best. Compare the classes with an overall test. If the result is significant, do all pairwise comparisons, and tell what your conclusions mean in the context of the problem.

Class 1	Class 2	Class 3
48	41	42
46	40	25
43	27	24
40	25	22
39	13	18
38	11	17
35	10	9
28	10	8
27	9	8
15	7	5

USING SPSS—EXAMPLES AND EXERCISES

SPSS provides procedures for computing the Mann–Whitney test, the Wilcoxon test, and the Kruskal–Wallis test, as well as other types of "nonparametric" techniques. Because the M–W, Wilcoxon, and K–W tests involve the median as the preferred measure of central tendency and the ranking of data, SPSS Boxplots provides a useful way to display the results.

Example—Mann–Whitney Test: We will use SPSS to work Problem 6, which is a M–W problem. The steps are as follows:

1. Start SPSS, name variables **group** and **risk,** and enter the data.
2. Select *Analyze>Nonparametric Tests>2-Independent-Samples*.
3. Move **risk** into the Test Variable List box, and move **group** into the Grouping Variable box.
4. Define the groups: 1 for Alcohol and 2 for No-Alcohol>*Continue*.
5. In the Test Type box, be sure the *Mann–Whitney U* is checked—it should be by default. Then click *OK* and the output should appear in the Viewer window.
6. For boxplots of the groups, select *Graphs>Boxplot>Simple>Summaries of groups of cases>Define*. Move **risk** into the Variable box and move **group** into the Category Axis box; then click *OK*.

Notes on Reading the Output

1. You should find the output easy to read and understand. The *U* value and *p* value [Asymp. Sig. (2-tailed)] are given in the Test Statistics box.
2. The results indicate that there is no significant difference between the two alcohol treatment groups in self-rated risk-taking attitude: $U = 58.5$, $p = .433$.
3. Examination of the boxplots shows that the groups are similar. The heavy line within the box indicates the position of the median, which is virtually identical for the two groups.

```
NPAR TESTS
  /M-W= risk   BY group(1 2)
  /MISSING ANALYSIS.
```

NPar Tests

Mann-Whitney Test

Ranks

	GROUP	N	Mean Rank	Sum of Ranks
RISK	1.00	12	11.38	136.50
	2.00	12	13.63	163.50
	Total	24		

Test Statistics[b]

	RISK
Mann-Whitney U	58.500
Wilcoxon W	136.500
Z	-.784
Asymp. Sig. (2-tailed)	.433
Exact Sig. [2*(1-tailed Sig.)]	.443[a]

a. Not corrected for ties.

b. Grouping Variable: GROUP

```
EXAMINE
  VARIABLES=risk BY group /PLOT=BOXPLOT/STATISTICS=NONE/NOTOTAL
  /MISSING=REPORT.
```
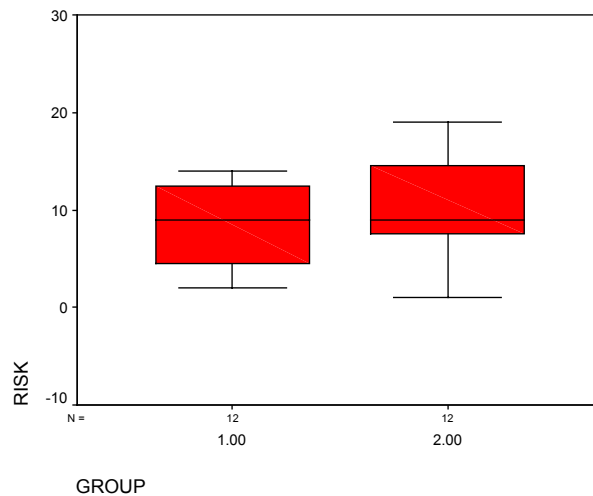
Explore

GROUP

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	GROUP	N	Percent	N	Percent	N	Percent
RISK	1.00	12	100.0%	0	.0%	12	100.0%
	2.00	12	100.0%	0	.0%	12	100.0%

RISK

Example—Wilcoxon Test: We will use SPSS to work Problem 5. The steps are as follows:
1. Start SPSS and name variables **before** and **after.** Enter the data.
2. Select *Analyze>Nonparametric Tests>2 Related Samples.*
3. In the dialog box, move the two variables into the Test Pairs List box and check *Wilcoxon* in the Test Type box, then click *OK*. The results should appear in the output Viewer window.
4. You can obtain the boxplots in the same manner as for the M–W except click Summaries of separate variables and move both variables into the Boxes Represents box.

Notes on Reading the Output

1. The sum of the ranks with the less frequent sign is 11 ($T = 11$), which is the smaller of the two values in the Ranks box under Sum of Ranks. The Test Statistics box gives T converted to a z score and the p value for z. The results indicate that the null hypothesis of identical population distributions is rejected, $T = 11, p = .049$.
2. To determine the direction of the group differences, examine the N and Mean Rank of the AFTER-BEFORE differences given in the Ranks box. There are more positive than negative ranks, which indicates that the After group is greater than the Before group in risk.
3. In this particular example, the boxplot does not clearly show the After group to be greater in risk.

```
NPAR TEST
  /WILCOXON=before  WITH after (PAIRED)
  /MISSING ANALYSIS.
```

NPar Tests

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
AFTER - BEFORE	Negative Ranks	3[a]	3.67	11.00
	Positive Ranks	8[b]	6.88	55.00
	Ties	1[c]		
	Total	12		

a. AFTER < BEFORE

b. AFTER > BEFORE

c. BEFORE = AFTER

Test Statistics[b]

	AFTER - BEFORE
Z	-1.971[a]
Asymp. Sig. (2-tailed)	.049

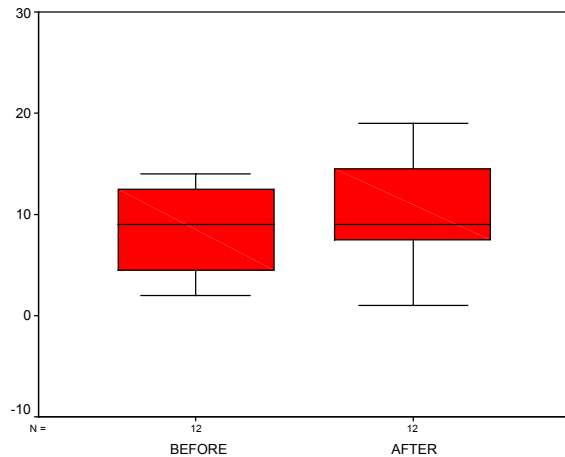a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

```
EXAMINE
  VARIABLES=before after /COMPARE
VARIABLE/PLOT=BOXPLOT/STATISTICS=NONE/NOTOTAL
  /MISSING=LISTWISE .
```

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
BEFORE	12	100.0%	0	.0%	12	100.0%
AFTER	12	100.0%	0	.0%	12	100.0%

Example—Kruskal–Wallis Test: We will use SPSS to work Problem 3, a K–W example. The steps are as follows:
1. Start SPSS and name variables **group** and **irritate.** Enter the data, coding the groups 1, 2, and 3 for the diet groups A, B, and C, respectively.
2. Select *Analyze>Nonparametric Tests>K Independent Samples.*
3. In the dialog box, move **irritate** into the Test Variable List box and move **group** into the Grouping Variable box. Click *Define Range*, indicate the minimum (1) and maximum (3) range for the grouping variable, then click *Continue*.
4. Be sure *Kruskal–Wallis H* is checked in the Test Type box (it should be by default), then *OK*, and the results should appear in the output Viewer window.
5. Obtain the boxplot of groups by adapting the instructions given in the M–W example to this problem.
6. Looking ahead, we see that the test is significant, so we will need to perform pairwise M–W tests to determine the group differences. Follow the steps in the M–W example and compare groups 1 and 2, 1 and 3, and 2 and 3.

Notes on Reading the Output

1. The output provides the mean ranks for each group in the Ranks box. In the Test Statistics box, SPSS does not provide the exact *H* value; instead, a chi-square equivalent is given with a *p* value.
2. The results indicate that the null hypothesis is rejected, and the samples come from different populations. We conclude that the diet groups are significantly different in terms of irritability by the K–W test, χ^2 (2, $N = 31$) = 13.657, $p = .001$. The mean ranks in ascending order are as follows: Diet C = 10.95, Diet A = 12.77, and Diet B = 24.6.
3. The boxplot indicates that Group 2 (Diet B) is higher than either of the other two groups, which do not appear to differ. Further testing is needed to confirm this observation.
4. Pairwise group comparisons: Comparison 1 versus 2 (Diet A vs. Diet B) indicates that Diet B animals have higher irritability than Diet A animals, $U = 12$, $p = .002$. Comparison 1 versus 3 (Diet A vs. Diet C) showed no significant difference in irritability, $U = 47.5$, $p = .59$. Comparison 2 versus 3 (Diet B vs. Diet C) indicated that Diet B rats had higher irritability than Diet C rats, $U = 7.0$, $p = .001$. These results can be summarized in terms of irritability as follows: Diet C = Diet A < Diet B.

```
NPAR TESTS
  /K-W=irritate   BY group(1 3)
  /MISSING ANALYSIS.
```

NPar Tests

Kruskal-Wallis Test

Ranks

	GROUP	N	Mean Rank
IRRITATE	1.00	11	12.77
	2.00	10	24.60
	3.00	10	10.95
	Total	31	

Test Statistics[a,b]

	IRRITATE
Chi-Square	13.657
df	2
Asymp. Sig.	.001

a. Kruskal Wallis Test

b. Grouping Variable: GROUP

```
EXAMINE
  VARIABLES=irritate BY group /PLOT=BOXPLOT/STATISTICS=NONE/NOTOTAL
  /MISSING=REPORT.
```

Explore

GROUP

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	GROUP	N	Percent	N	Percent	N	Percent
IRRITATE	1.00	11	100.0%	0	.0%	11	100.0%
	2.00	10	100.0%	0	.0%	10	100.0%
	3.00	10	100.0%	0	.0%	10	100.0%

IRRITATE
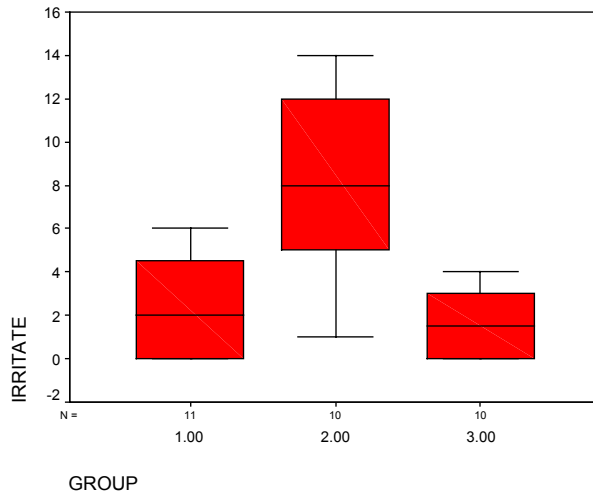

```
NPAR TESTS
  /M-W= irritate   BY group(1 2)
  /MISSING ANALYSIS.
```

NPar Tests

Mann-Whitney Test

Ranks

	GROUP	N	Mean Rank	Sum of Ranks
IRRITATE	1.00	11	7.09	78.00
	2.00	10	15.30	153.00
	Total	21		

Test Statistics[b]

	IRRITATE
Mann-Whitney U	12.000
Wilcoxon W	78.000
Z	-3.046
Asymp. Sig. (2-tailed)	.002
Exact Sig. [2*(1-tailed Sig.)]	.002[a]

a. Not corrected for ties.

b. Grouping Variable: GROUP

```
NPAR TESTS
  /M-W= irritate   BY group(1 3)
  /MISSING ANALYSIS.
```

NPar Tests

Mann-Whitney Test

Ranks

	GROUP	N	Mean Rank	Sum of Ranks
IRRITATE	1.00	11	11.68	128.50
	3.00	10	10.25	102.50
	Total	21		

Test Statistics[b]

	IRRITATE
Mann-Whitney U	47.500
Wilcoxon W	102.500
Z	-.541
Asymp. Sig. (2-tailed)	.589
Exact Sig. [2*(1-tailed Sig.)]	.605[a]

a. Not corrected for ties.

b. Grouping Variable: GROUP

```
NPAR TESTS
  /M-W= irritate   BY group(2 3)
  /MISSING ANALYSIS.
```

NPar Tests

Mann-Whitney Test

Ranks

	GROUP	N	Mean Rank	Sum of Ranks
IRRITATE	2.00	10	14.80	148.00
	3.00	10	6.20	62.00
	Total	20		

Test Statistics[b]

	IRRITATE
Mann-Whitney U	7.000
Wilcoxon W	62.000
Z	-3.269
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.000[a]

a. Not corrected for ties.

b. Grouping Variable: GROUP

Exercises Using SPSS

1. Work Problem 2, using SPSS to perform the Mann–Whitney test. Also provide boxplots of the data and write a conclusion.
2. Use SPSS to perform the Wilcoxon test on the data in Problem 7. Also provide boxplots of the data and write a conclusion.
3. Use SPSS to perform the K–W test on the data in Problem 8. Also provide boxplots of the data. If overall differences are obtained, do all pairwise comparisons using the M–W test; interpret the results in the context of the problem.

CHECKING YOUR PROGRESS: A SELF-TEST

1. Match the nonparametric test with its parametric alternative.

    _____Mann–Whitney   **a.** one-way between-subjects ANOVA

    _____Kruskal–Wallis   **b.** *t* test for dependent samples

    _____Wilcoxon   **c.** *t* test for independent samples

2. Three groups of rats are fed different diets for 4 weeks. Each animal's latency (rounded to the nearest whole second) to leave a lighted platform is recorded, and the results are shown here. Because of the unequal and small sample sizes and the large amount of within-group variability, the one-way ANOVA may not be appropriate. Use the nonparametric alternative to perform an overall test. If the overall test is significant, do all pairwise comparisons, and tell what your conclusions mean in the context of the problem.

Group 1	Group 2	Group 3
30	43	13
25	33	12
23	30	10
16	28	8
12	25	7
10	15	6
8	14	5
5	10	

3. Fifteen students, rated as extreme introverts on the Myers–Briggs Type Indicator, are given 5 hours of assertiveness training. The MBTI is again administered. Did assertiveness training affect the introversion score on the MBTI?

Student	Score Before Training	Score After Training
1	27	27
2	33	27
3	35	27
4	45	35
5	47	43
6	31	35
7	35	37
8	31	25
9	29	15
10	27	11
11	35	31
12	43	45
13	39	27
14	35	37
15	31	25