CHAPTER 15

More Nonparametric Tests

Summary

The chapter reviews four different *nonparametric tests*, also known as ***distribution-free tests***. Unlike *parametric tests*, ***nonparametric tests*** do not require assumptions about the parameters of sampled populations. Specifically, *nonparametric tests* do not assume that populations are normally distributed or that the variances are equal. Thus, if quantitative data do not meet the assumptions of normality and homogeneity of variance that parametric tests require, reducing the data to ranks and using a nonparametric test is a good option. Nonparametric tests are also different from parametric tests because the null hypothesis is that the population *distributions* are identical rather than the population *means*. For example, the null hypothesis for the chi square test—which is the nonparametric test you learned in the last chapter—involves *frequency counts*. Nonparametric tests are similar to parametric tests in they both use the same hypothesis-testing logic (NHST). The four nonparametric tests in this chapter use data that are ranks or reduced to ranks. You should be sure you understand how to assign ranks to scores for each of these tests, paying particular attention about what to do with identical scores.

To test the hypothesis that two population distributions are identical when data are from a two-group, *independent-samples design*, use the ***Mann-Whitney U test****.* Rank the scores for the entire data set, then separately sum the ranks for each group. Using these sums in a formula, calculate two ***U values****.* Use the smaller *U.* For designs with both *N*s less than 20, *U* is evaluated for significance with Table H. For larger samples, convert the smaller *U* to a *z* and then a *z test* (Table C) based on the smaller *U* is used. The parametric counterpart to the Mann-Whitney U test is the *independent-samples t test.*

To test the hypothesis that two population distributions are identical when data are from a two-group, *paired-samples design*, use the ***Wilcoxon signed-rank T test****.* The difference between pairs of scores is found. The *absolute value of the differences* are then ranked, with the smallest difference ranked 1 (unless the there are differences of 0). Each rank receives the algebraic sign of its difference. Sum the positive ranks, then sum the negative ranks. Select the smaller of the absolute values of the two sums to serve as the ***T value*** that is evaluated for significance. If the number of pairs is less than 50, *T* is evaluated with Table J. For larger samples, convert the *T* to a *z* and then conduct a *z* test (Table C). The parametric counterpart to the *Wilcoxon signed-rank T test* is the *paired-samples t test.*

Be aware: Contrary to other statistical tests you have studied, calculated values for both *U* and *T* must be *less* than tabled critical value to be significant.

To test the hypothesis that population distributions are identical when the data are from an independent-samples design with *more than two levels* of the independent variable, use the ***Wilcoxon-Wilcox multiple comparisons test****.* Rank the scores for the whole data set and then separately sum the ranks for each group. Find the difference for each pair of groups. Each difference is evaluated for significance with a table of critical differences for the Wilcoxon-Wilcox multiple comparisons test, Table K. The parametric counterpart to the *Wilcoxon-Wilcox multiple comparisons test* is one-way ANOVA, followed by Tukey HSD tests.

To find the *degree of relationship* between two variables that consist of ranks, use the ***Spearman*** r_s. Rank the scores within one group and then rank the scores within the other group. Find the difference between the ranks for each pair. Square the differences, sum them, and use this sum of the squared differences in the formula for r_s. To test the hypothesis that two populations have a correlation coefficient of .00, calculate r_s and evaluate its significance using Table L (for $N \leq 16$) or Table A (for $N > 16$). The parametric counterpart to the *Spearman* r_s is the Pearson product-moment correlation coefficient (*r*).

Multiple-Choice Questions

1. The nonparametric tests in the text are based on sampling distributions of ________________.

a. means
b. mean differences
c. ranks
d. variances

2. The nonparametric test that corresponds in design to the independent-samples *t* test is the ________________.

a. Mann-Whitney *U* test
b. Wilcoxon signed-rank *T* test
c. Wilcoxon-Wilcox multiple-comparisons test
d. r_s

3. Suppose you find that three people tied for the top score (1) in a Mann-Whitney *U* test. The correct procedure is to ________________.

a. assign a rank of 1 to all three
b. assign a rank of 2 to all three
c. assign a rank of 3 to all three
d. randomly determine which scores get ranks 1, 2, and 3

4. Nonparametric tests are used rather than a *t* test or an ANOVA when ________________.

a. the researcher does not know the specific value of the population parameters
b. the data are in the form of ranks
c. the dependent variable is normally distributed with equal variance in the sampled populations
d. the researcher wants to avoid using NHST

5. The null hypothesis for testing the significance of r_s is that the population correlation coefficient is ________________.

a. .00
b. 1.00
c. -1.00
d. none of the above

6. When a sample size is large, the *U* value from a Mann-Whitney *U* test is evaluated using the ________________.

a. *t* distribution
b. normal distribution
c. *F* distribution
d. χ^2 distribution

7. Which of the following statistical tests is most similar to the Wilcoxon signed-rank *T* test?

a. paired-samples *t* test
b. one-way repeated-measures ANOVA
c. χ^2 test
d. Pearson product-moment correlation

8. To test for a significant difference between paired samples, use a ________________.

a. Mann-Whitney *U* test
b. Wilcoxon signed-rank *T* test
c. Wilcoxon-Wilcox multiple-comparisons test
d. r_s

9. Which of the following ranked data could *not* be analyzed with a Wilcoxon-Wilcox multiple-comparisons test?

a. three independent samples
b. class standing at three different high schools
c. $N_1 = 10$, $N_2 = 20$, $N_3 = 30$, for three independent samples
d. four independent samples

10. What is the rank of 4 in the following distribution? 1, 2, 2, 3, 3, 4, 4, 5, 5.

a. 4
b. 7
c. 6.5
d. 6

11. In which test(s) does it matter which value is given a rank of 1?

a. Mann-Whitney *U* test
b. Wilcoxon signed-rank *T* test
c. both a. and b.
d. neither a. nor b.

12. The Wilcoxon signed-rank *T* test and Spearman r_s differ in terms of ________________.

a. what scores are being ranked
b. the null hypotheses they test
c. whether or not the data are ranked
d. both a. and b.

13. Suppose you found, for 26 people in your dorm, a Spearman r_s of .38 between the number of breakfasts eaten during the term and grade point average. You may conclude for a two-tailed test that there is ________________.

a. no significant relationship
b. a significant relationship at the .05 level
c. a significant relationship at the .01 level
d. a significant relationship at the .001 level

14. If you have one difference score of 0 it should be kept in the analysis for ________________.

a. Wilcoxon signed-rank *T* test
b. Spearman r_s
c. both a. and b.
d. neither a. nor b.

15. One common use of statistics is to rank a set of cities so the Number 1 city is the "best" and the last, highest ranked, city is the "worst." Suppose you thought that large cities would have higher ranks and that small cities would generally have lower ranks. Which statistic should you use to test your idea?

a. Mann-Whitney *U* test
b. Wilcoxon signed-rank *T*
c. Wilcoxon-Wilcox multiple comparisons test
d. Spearman r_s

16. The ________________ the value of *U* and the ________________ the value of *T*, the more likely you are to reject the null hypothesis.

a. larger, larger
b. larger, smaller
c. smaller, smaller
d. smaller, larger

Short Answer and Interpretation

1. Which nonparametric test should be used to analyze data from the following studies?

a. Countries are ranked on per-capita health care costs. Suppose you had the per-capita health care costs for 15 countries as well as the life expectancy ranks for the same countries. What test would determine whether there was a relationship between health care costs and life expectancy?
b. To determine if there was sex discrimination in salaries at Old State U., a statistician began with 15 female professors. On the basis of degree, discipline, and years of experience, each was matched with a male professor. Salary differences were calculated and ranked for each pair.
c. To produce high frustration, some participants were forced to wait 15 minutes for a late participant before beginning an experiment. Low-frustration participants started on time. During the experiment, aggression was measured and ranked as the dependent variable.
d. To find out if there is a relationship between a basketball team's height and its conference rank, the average height and rank were determined for 10 schools in the same conference.
e. In a before-and-after study, participants rated their views on immigration policies. Between conditions, they read the personal story of a single immigrant. Rating differences were calculated and ranked for each pair.
f. Julia and Davis were interested in the effects of music on reading comprehension in 5th graders. They played instrumental classical music, instrumental jazz music, or no music to three different groups of 5th graders while they were reading. Later the students took a test over the material they read.

2. A certified personal trainer designed a cardiovascular fitness program that gradually increased duration and intensity of exercise over a 5-month period. She recruited 20 individuals with low levels of daily physical activity and randomly assigned half of them to participate in the training and half of them to continue daily life as usual. At the end of the 5-month training, she asked all participants to complete one lap around a track (0.25 miles) and ranked participants by completion time as one measure of cardiovascular fitness. Her analyses revealed a test value of 20. Identify which test statistic she calculated. Set alpha at .05 (two-tailed) and answer the following question: Did participants who completed the program have better cardiovascular fitness than control participants?

3. Jae was interested in the relationship between size and status within rats. He ranked nine rats by their weight and their standing in the dominance hierarchy. Rankings closer to 1 represent heavier rats and more dominant rats, respectively. His analyses revealed a test value of .53. Identify which test statistic he calculated. Set alpha at .05 (two-tailed) and answer the following question: Is there a significant relationship between size and dominance in rats?

4. Do IQ scores change after an extra year of education? Assume that a researcher administered an IQ test to eight children in third grade at both the beginning and end of the school year. After finding each child's IQ difference between tests, she ranked the differences and found a test statistic of 2. Identify which test statistic she calculated. Set alpha at .05 (two-tailed) and answer the following question: Do IQ scores change (and, if so, how) after an additional year of education? The correct answer will mirror the conclusion drawn by Ritchie and Tucker-Drob (2018).

5. A first-year undergraduate student wanted to know if starting job salaries differed based on college major. She sampled five recent graduates from each of the three majors she was most interested in (computer science, engineering, and business) and asked each of them to report their starting salary. She then assigned each salary a rank, with the highest salary being ranked 1. The tables that follow summarize her hypothetical data (loosely based on averages from Time Magazine; retrieved from https://tinyurl.com/kg53ddp on July 20, 2018). Identify which test statistic she calculated. Set alpha at .05 (two-tailed) and answer the following question: Should this student expect to earn a significantly different starting salary based on her choice of major?

	Engineering	Computer Science	Business
Sum of ranks:	26	35	59

Differences in ranks for the three majors follow.

		Major	
		Engineering	Computer Science
Major	Computer Science	9	
	Business	33	24

Problems

1. In recent years, U. S. News and World Report has released results from their research on the "Best Countries." This report ranks countries on a wide-variety of variables (e.g., quality of life, entrepreneurship, power) and also provides an overall ranking. The table that follows presents the 10 most highly-rated countries from the 2018 report (retrieved from https://tinyurl.com/yav95zbv on July 19, 2018). Within the top 10, each country was then ranked by its quality of life rating here. Quality of life ratings were based on factors such as job market quality, safety, and quality of education and health care. The researcher is interested in whether there is a relationship between a country's overall ranking and their ranking on quality of life.

Country Ranking	Quality of Life Ranking
1. Switzerland	4
2. Canada	1
3. Germany	6
4. United Kingdom	7
5. Japan	8
6. Sweden	2
7. Australia	3
8. United States	10
9. France	9
10. Netherlands	5

a. Which nonparametric test is appropriate for these data?
b. What is the test value for these data?
c. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
d. Write a conclusion about the relationship between quality of life rank and overall rank.

2. This problem uses the same 2018 "Best Countries" dataset from U. S. News and World Report. Within the top 10, each country was ranked by its adventure rating here. Adventure ratings were based on the extent to which a country has a pleasant climate and can be described as fun and "sexy." The researcher is interested in whether there is a relationship between a country's overall ranking and their adventure ranking.

Country Ranking	Adventure Ranking
1. Switzerland	4
2. Canada	5
3. Germany	10
4. United Kingdom	9
5. Japan	8
6. Sweden	6
7. Australia	1
8. United States	7
9. France	2
10. Netherlands	3

a. Which nonparametric test is appropriate for these data?
b. What is the test value for these data?
c. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
d. Write a conclusion about the relationship between adventure rank and overall rank. Include a sentence that compares the relationship between adventure rank and overall rank to the relationship between quality of life rank and overall rank (from Problem 1).

3. A university administrator designed a mental health literacy intervention as a way to increase students' willingness to help people diagnosed with mental illness. To assess the effectiveness of the program, the administrator formed two matched groups of students based on year in school and background in psychology. Only one group completed the program, but both groups completed a willingness to help measure. Higher scores indicate a greater willingness to help people diagnosed with mental illness. The hypothetical data that follow will allow you to draw a conclusion about mental health literacy interventions that mirrors Lo, Gupta, and Keating (2017).

	Willingness to Help Scores	
Pair	Control Group	Intervention Group
1	21	23
2	12	18
3	17	22
4	23	23
5	17	16
6	21	24
7	19	27
8	14	18

a. Which nonparametric test is appropriate for these data?
b. What is the test value for these data?
c. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
d. Write a conclusion about the success or failure of the program.

4. During the off season, sprinters on a track team were randomly assigned to a bicycle training group or a traditional training group. After the off season, all sprinters ran a 200 meter dash; times are shown below.

Bicycle Training	Traditional Training
21.1	21.1
21.5	21.6
21.3	20.6
20.8	22.4
21.2	21.9
20.7	20.5
20.3	21.7
21.4	22.6
20.9	21.8
20.4	22.1

a. Which nonparametric test is appropriate for these data?
b. What is the test value for these data?
c. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
d. Write a conclusion about the effect of training type on 200-meter dash performance.

5. A consumer advocate compared the cleanliness of four chains of supermarkets by devising a 50-point rating scale and inspecting six stores of each chain. The higher the score, the cleaner the store. Past research indicated that the population scores on this cleanliness scale were severely skewed, therefore the researcher decided to use a nonparametric test to analyze his data. Set alpha at .05, analyze the data with the appropriate test, and write a conclusion about the relative cleanliness of these four chains.

Chain A	Chain B	Chain C	Chain D
35	29	39	50
26	27	43	33
46	31	41	49
38	34	42	32
44	28	47	48
37	40	36	45

6. Students in general psychology course observed pictures of 20 people they did not know and estimated their IQs. Only eight of those pictures are of interest in this experiment and all eight of those individuals were wearing glasses; the other 12 pictures were mixed in to prevent the psychology students from realizing the question being addressed. One week later, the same students observed pictures of the same 20 individuals and, again, judged their IQs. This time, the original eight people of interest were not wearing glasses. The IQ estimates for the psychology students were averaged so that each person of interest received two scores: one with glasses and one without glasses.

Without Glasses	With Glasses
121	127
97	104
114	110
131	139
118	120
91	86
127	140
114	125

a. Which nonparametric test is appropriate for this experiment?
b. What is the test value for these data?
c. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
d. Write a conclusion about the effect of wearing glasses on ratings of IQ.

7. Students face the practical question of how to use study time efficiently. Gates conducted an early study on this topic (1917), which was used as a model for the hypothetical data that follow. Twenty students were divided into five groups. The students all studied an article on dinosaurs, but each group spent a different percentage of the time in “self-recitation” (looking away from the article and mentally reciting what had been read). Afterward, each student took a 100-point test on the material in the article. The scores were ranked, with the best score ranked as 1. The sums of the ranks follow. Set alpha at .05, analyze the data with the appropriate test, and write a conclusion that the data support.

	Percentage of Study Time Spent in Self-Recitation				
	0%	20%	40%	60%	80%
Sum of ranks:	73	54	44.5	27.5	11

8. Returning to the U. S. News and World Report “Best Countries” rankings, imagine that two young entrepreneurs were deciding whether to base their company in Europe or Asia. Using the “Best Countries” report, the entrepreneurs selected the 10 highest-ranked countries from each continent and then gathered the “open for business” score for each (which was based on factors such as having a favorable tax environment and transparent government practices). Their data follow. Because these data were originally based on ranks, they assumed the scores would not meet the assumptions of a parametric test. Use these “open for business” scores to determine which continent is likely to be a better place to start their new company.

“Open for Business” Scores			
European Countries		Asian Countries	
Switzerland	2	Japan	26
Germany	19	Singapore	13
United Kingdom	20	China	33
Sweden	5	South Korea	25
France	27	United Arab Emirates	67
Netherlands	9	India	29
Denmark	4	Russia	80
Norway	8	Thailand	15
Finland	6	Israel	64
Italy	45	Malaysia	12

a. Which nonparametric test is appropriate for these data?
b. What is the test value for these data?
c. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
d. Write a conclusion about the difference in openness to business between European and Asian Countries.

Multiple-Choice Questions

1. c
2. a
3. b

   Explanation: For the Mann-Whitney *U* test, the Wilcox signed-rank *T* test, and the Wilcoxon-Wilcox multiple comparisons test, tied scores all get assigned the same rank, which is the mean of the ranks that would have been assigned, had there been no ties. Here, (1 + 2 + 3) / 3 = 2. Thus, all three of the top scores would be assigned the rank of 2. A point that sometimes confuses students is what to do next. The fourth score in the list (assuming it was not also tied with another score) would be assigned a 4 *not 3*. Thus, when you continue after a set of tied scores, assign the rank of the next scores as if the preceding scores were not tied.
4. b

   Explanation: If the assumptions of normality and equal variance are met, parametric tests (e.g., *t* test, ANOVA) should be used over nonparametric tests. Ranks are not normally distributed because there is just one score for every value.
5. a
6. b
7. a
8. b
9. c

   Explanation: The Wilcoxon-Wilcox multiple comparisons test can only be used when sample size is equal across groups.
10. c

    Explanation: Tied scores all get assigned the same rank, which is the mean of the ranks that would have been assigned, had there been no ties. Here, the 4s fall in places 6 and 7, respectively. Thus, both scores would be assigned the rank of 6.5.
11. b

    Explanation: In a Wilcoxon signed-rank *T* test, a rank of 1 is always assigned to the smallest difference, unless there is a difference of 0. In the Mann-Whitney *U* test, it does not matter whether 1 is assigned to the lowest or highest score.
12. d

    Explanation: All nonparametric tests reviewed in this chapter are for ranked data. In the Wilcoxon signed-rank *T* test, difference scores are calculated, then ranked; in Spearman r_s, scores are ranked, then difference scores are calculated. Furthermore, the Wilcoxon signed-rank *T* test tests the null hypothesis that two paired populations have identical distributions, whereas the Spearman r_s tests the null hypotheses that the correlation between two ranked populations is .00.
13. a

    Explanation: Because $N > 16$, the r_s value should be compared to two-tailed critical values located in Table A. For these data, $df = N – 2 = 26 – 2 = 24$. For $\alpha = .05$, the r_s value would need to be greater than .3883 to be statistically significant, which it is not. The critical values get larger as alpha gets smaller, so $r_s = .38$ is not statistically significant at any of the levels provided.
14. b

    Explanation: Difference scores are always kept in analyses for Spearman r_s, because zero means there is a perfect correspondence between the two ranks. In a Wilcoxon signed-rank *T* test, one difference score of 0 gets dropped from the analysis and *N* gets reduced by 1.
15. a

    Explanation: This research question involves the comparison of rankings between two independent groups: large cities and small cities. Thus, a Mann-Whitney *U* test is most appropriate.
16. c

    Explanation: Unlike most tests reviewed in the textbook, test values for *U* and *T* must be *smaller* than critical values to reject the null hypothesis.

Short Answer and Interpretation

1. a. Spearman r_s
   b. Wilcoxon signed-rank *T* test
   c. Mann-Whitney *U* test
   d. Spearman r_s
   e. Wilcoxon signed-rank *T* test
   f. Wilcoxon-Wilcox multiple-comparisons test

2. The appropriate test statistic for this design is the Mann-Whitney *U* test, because two independent groups are being compared on a ranked variable. The appropriate critical value from Table H is 23 ($N_1 = 10$, $N_2 = 10$). Because 20 < 23, we reject the null hypothesis and conclude that participants who completed the program had significantly better cardiovascular fitness than control participants ($p < .05$).

3. The appropriate test statistic for this design is the Spearman r_s test, because the researcher is interested in the relationship between two ranked variables. The appropriate critical value from Table L is .700. Because .53 < .700, we should retain the null hypothesis and conclude that there is no significant relationship between size and dominance in rats, $r_s = .53$, $p > .05$.

4. The appropriate test statistic for this design is a Wilcoxon signed-rank *T* test, because the researcher is interested in comparing two paired (repeated measures) samples. The appropriate critical value from Table J is 3. Because 2 < 3, we should reject the null hypothesis and conclude that one year of schooling does significantly increase IQ scores, $T = 2$, $p < .05$.
Explanation: Though it can be difficult to remember, Wilcoxon signed-rank *T* test values that are less than the critical value lead to rejection of the null.

5. The appropriate test statistic for this design is a Wilcoxon-Wilcox multiple comparisons test, because the researcher is comparing more than two independent groups on a ranked variable. The appropriate critical value from Table K is 33.1. Although Engineering majors had the highest overall ranks, followed by computer science, then business, none of the rank differences were statistically significant. Thus, this student should not expect to earn a significantly different starting salary based on her choice of major.

Problems

1. a. Spearman r_s
   b. r_s = .41
   c. retain
   d. Quality of life rank is not significantly related to overall rank for the top 10 countries according to U. S. News and World Report, $r_s = .41$, $p > .05$.

Explanation:

Country Ranking	Quality of Life Ranking	*D*	D^2
1. Switzerland	4	-3	9
2. Canada	1	1	1
3. Germany	6	-3	9
4. United Kingdom	7	-3	9
5. Japan	8	-3	9
6. Sweden	2	4	16
7. Australia	3	4	16
8. United States	10	-2	4
9. France	9	0	0
10. Netherlands	5	5	25
			$\Sigma D^2 = 98$

Remember for Spearman r_s keep differences equal to zero because they indicate that there is a perfect correlation. If you examine the formula for r_s, you will see if all the differences were perfect, r_s would be equal to 1.

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6(98)}{10(99)} = 1 - \frac{588}{990} = 1 - .594 = .41$$

Be very careful working this formula: A common mistake is to forget the last step of the math (here, that is 1 - .594).

$r_{s.05} = .648$

Because the test value, .41, is less than the critical value, .648 (Table L), we retain the null hypothesis and conclude that the correlation between quality of life rank and overall rank is not significantly different from .00. Note, also, that interpretation of the meaning and magnitude of a correlation is context dependent and that this correlation is based on a small sample size.

2. a. Spearman r_s
   b. r_s = -.43
   c. retain
   d. Adventure rank is not significantly related to overall rank for the top 10 countries according to U. S. News and World Report, $r_s = -.43$, $p > .05$. The magnitude of the relationship between quality of life rank and overall rank $r_s = .41$ is smaller than the magnitude of relationship between adventure rank and overall rank $r_s = -.43$. Although, quality of life rank is positively related to overall rank, whereas adventure rank is negatively related to overall rank.

Explanation:

Country Ranking	Adventure Ranking	*D*	D^2
1. Switzerland	4	-3	9
2. Canada	5	-3	9
3. Germany	10	-7	49
4. United Kingdom	9	-5	25
5. Japan	8	-3	9
6. Sweden	6	0	0
7. Australia	1	6	36
8. United States	7	1	1
9. France	2	7	49
10. Netherlands	3	7	49
			$\Sigma D^2 = 236$

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6(236)}{10(99)} = 1 - \frac{1416}{990} = 1 - 1.430 = -.43$$

$r_{s.05} = .648$

Because the test value, -.43, is less than the critical value, .648 (Table L), we retain the null hypothesis and conclude that the correlation between adventure rank and overall rank is not significantly different from .00. Note, also, that interpretation of the meaning and magnitude of a correlation is context dependent and that this correlation is based on a small sample size. However, given that these correlations are drawn from similar contexts, a direct comparison of the magnitude of -.43 and .59 is acceptable.

3. a. Wilcoxon signed-rank *T* test
b. $T = 1$
c. reject
d. According to the data, willingness to help scores were significantly higher in the mental health literacy intervention group than the control group, $T = 1$, $p < .05$

Explanation:

	Willingness to Help Scores			
Pair	Control Group	Intervention Group	*D*	Signed rank
1	21	23	-2	-2
2	12	18	-6	-6
3	17	22	-5	-5
4	23	23	0	Eliminated
5	17	16	1	1
6	21	24	-3	-3
7	19	27	-8	-7
8	14	18	-4	-4

$\Sigma_{positive} = 1$
$\Sigma_{negative} = -27$
$T = 1$
$N = 7$ (Remember the pair with the difference score of 0 was eliminated)
Check: 1 + 27 = 28; $\frac{N(N+1)}{2} = \frac{7(8)}{2} = 28$
$T_{.05} = 2$

Because the test value, 1, is less than the critical value, 2 (Table J), we reject the null hypothesis and conclude that willingness to help scores were higher in the intervention group than the control group.

4. a. Mann-Whitney *U* test
   b. $U = 20.5$
   c. reject
   d. According to the data, those who participated in the bicycle training ran a 200-meter dash significantly faster than those who participated in traditional training, $U = 20.5$, $p < .05$

   Explanation:

Bicycle Training	Rank	Traditional Training	Rank
21.1	8.5	21.1	8.5
21.5	13	21.6	14
21.3	11	20.6	4
20.8	6	22.4	19
21.2	10	21.9	17
20.7	5	20.5	3
20.3	1	21.7	15
21.4	12	22.6	20
20.9	7	21.8	16
20.4	2	22.1	18
	$\Sigma R_1 = 75.5$		$\Sigma R_2 = 134.5$

$$U = (N_1)(N_2) + \frac{N_2(N_2+1)}{2} - \sum R_1$$
$$= (10)(10) + \frac{10(10+1)}{2} - 75.5 = 100 + 55 - 75.5 = 79.5$$

$$U = (N_1)(N_2) + \frac{N_2(N_2+1)}{2} - \sum R_2$$
$$= (10)(10) + \frac{10(10+1)}{2} - 134.5 = 100 + 55 - 134.5 = 20.5$$

Check: 79.5 + 20.5 = 100; (10)(10) = 100

$U_{.05} = 23$

Because the test value, 20.5 (the smaller of the two *U*s), is less than the critical value, 23 (Table H), we reject the null hypothesis and conclude that, on average, those who participated in bicycle training ran significantly faster in a 200-meter dash than those who participated in traditional training.

5. A Wilcoxon-Wilcox multiple-comparisons test was used to analyze the relative cleanliness of four different supermarket chains. Chain B was found to be significantly less clean than Chain D, $p < .05$. No other chains significantly differed from one another on cleanliness. Differences in rank sums for each comparison are provided in the table that follows.

		Chain		
		A	B	C
Chain	B	35		
	C	21	56	
	D	30	65*	9

*$p < .05$; for $\alpha = .05$, the critical value is 62.9 ($K = 4$; $N = 6$).

Explanation:

Chain A	Rank	Chain B	Rank	Chain C	Rank	Chain D	Rank
35	9	29	4	39	13	50	24
26	1	27	2	43	17	33	7
46	20	31	5	41	15	49	23
38	12	34	8	42	16	32	6
44	18	28	3	47	21	48	22
37	11	40	14	36	10	45	19
	$\Sigma R_A = 71$		$\Sigma R_B = 36$		$\Sigma R_C = 92$		$\Sigma R_D = 101$

Check: 71 + 36 + 92 + 101 = 300; $\frac{N(N+1)}{2} = \frac{24(25)}{2} = 300$

6. a. Wilcoxon signed-rank *T* test
   b. $T = 5$
   c. retain
   d. According to the data, IQ ratings when wearing glasses were not significantly different from IQ ratings when not wearing glasses, $T = 5$, $p > .05$

Explanation:

Without Glasses	With Glasses	*D*	Signed rank
121	127	-6	-4
97	104	-7	-5
114	110	4	2
131	139	-8	-6
118	120	-2	-1
91	86	5	3
127	140	-13	-8
114	125	-11	-7

$\Sigma_{positive} = 5$
$\Sigma_{negative} = -31$
$T = 5$
$N = 8$
Check: 5 + 31 = 36; $\frac{N(N+1)}{2} = \frac{8(9)}{2} = 36$
$T_{.05} = 3$
Because the test value, 5, is not less than the critical value, 3 (Table J), we retain the null hypothesis and conclude that wearing glasses has no effect on IQ ratings.

7. A Wilcoxon-Wilcox multiple-comparisons test was used to analyze the relative effectiveness of spending increasing amounts of time in self-recitation while studying. Although spending more time in self-recitation resulted in higher test scores, only spending 80 percent of study time in self-recitation resulted in significantly higher scores than spending no time in self-recitation, $p < .05$. No other amounts of self-recitation study time significantly differed from one another on test scores. Differences in rank sums for each comparison are provided in the table that follows.

		Percentage of Time			
		0	20	40	60
Percentage of Time	20	19			
	40	28.5	9.5		
	60	45.5	26.5	17	
	80	62*	43	33.5	16.5

*$p < .05$; for $\alpha = .05$, the critical value is 45.6 ($K = 5$; $N = 4$).

8. a. Mann-Whitney *U* test
   b. $U = 18$
   c. reject
   d. The top-rated European countries have significantly more favorable business conditions than the top-rated Asian countries, according to data from U. S. News and World Report, $U = 18$, $p < .05$.

Explanation:

European Countries	Score	Rank	Asian Countries	Score	Rank
Switzerland	2	1	Japan	26	13
Germany	19	10	Singapore	13	8
United Kingdom	20	11	China	33	16
Sweden	5	3	South Korea	25	12
France	27	14	United Arab Emirates	67	19
Netherlands	9	6	India	29	15
Denmark	4	2	Russia	80	20
Norway	8	5	Thailand	15	9
Finland	6	4	Israel	64	18
Italy	45	17	Malaysia	12	7
		$\Sigma R_1 = 73$			$\Sigma R_2 = 137$

$$U = (N_1)(N_2) + \frac{N_2(N_2 + 1)}{2} - \sum R_1$$
$$= (10)(10) + \frac{10(11)}{2} - 73 = 100 + 55 - 73 = 82$$

$$U = (N_1)(N_2) + \frac{N_2(N_2 + 1)}{2} - \sum R_2$$
$$= (10)(10) + \frac{10(11)}{2} - 137 = 100 + 55 - 137 = 18$$

Check: 82 + 18 = 100; (10)(10) = 100

$U_{.05} = 23$

Because the test value, 18, is less than the critical value, 23 (Table H), we reject the null hypothesis and conclude that the top-rated European countries have more favorable business conditions than the top-rated Asian countries, according to data from U. S. News and World Report.

References

Lo, K., Gupta, T., & Keating, J. L. (2017). Interventions to promote mental health literacy in university students and their clinical educators. A systematic review of randomized control trials. *Health Professions Education.* Advance online publication. doi: 10.1016/j.hpe.2017.08.001

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*. Advance online publication. doi: 10.1177/0956797618774253