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. Effect size can be determined by calculating a ***rank-biserial correlation coefficient*** (r_{rb}) and interpreting it similarly to Pearson *r*. 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 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). Like Mann-Whitney *U*, you can use r_{rb} for measuring the size of the effect. 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 considered 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 ***Kruskal-Wallis test*** and measure the effect size using ε^2. Rank the scores for the whole data set and then separately sum the ranks for each group. Use these summed ranks in the Kruskal-Wallis formula. Compare that observed *H* value with the critical value from the χ^2 distribution (Table E). If the Kruskal-Wallis test leads to rejection of the null, then evaluate the significance of each paired difference using Dunn's test and critical values from Table K. The parametric counterpart to the *Kruskal-Wallis test with Dunn's test follow-up* is one-way ANOVA, followed by Tukey *HSD* tests.

To find the *direction and 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 size of the relationship can be evaluated, like Pearson *r*, by examining the size of the r_s. 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 Kruskal-Wallis test?

a. three independent samples
b. class standings at three different high schools
c. three repeated measures samples
d. none of the above are correct, Kruskal-Wallis test can be used in all of the analyses above

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 discrimination in salaries at Old State U. based on being a parent, a statistician began with 15 professors who identified themselves as parents. On the basis of degree, discipline, and years of experience, each was matched with a professor who did not identify as a parent. Because salaries at Old State were severely skewed, 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 of the team for 10 schools in the conference was determined. Those average heights for the 10 teams were ranked from tallest to shortest. Conference ranks were then identified.
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. Because the reading scores in the three groups did not have equal variances, they converted the reading comprehension scores to ranks before conducting their data analysis.

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 different 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 significantly 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 (mathematics and statistics, electrical engineering, and accounting) and asked each of them to report their starting salary. She then assigned each salary a rank, with the highest salary being ranked 1. She found a test statistic of 5.82. The table that follows summarize her hypothetical data (loosely based on averages from Wells Fargo; retrieved from https://collegesteps.wf.com/your-major-your-earning-potential/ on December 7, 2023). Identify which test statistic she calculated. Set alpha at .05 and answer the following question: Should this student expect to earn a significantly different starting salary based on her choice of major?

	Electrical Engineering	Mathematics and Statistics	Accounting
Sum of ranks:	26	35	59

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 2023 report (retrieved from https://www.usnews.com/news/best-countries/rankings on December 7, 2023). 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	3
2. Canada	2
3. Sweden	1
4. Australia	5
5. United States	10
6. Japan	9
7. Germany	6
8. New Zealand	7
9. United Kingdom	8
10. Netherlands	4

a. Which nonparametric test is appropriate for these data?
b. Write a null and alternative hypothesis.
c. Identify the appropriate critical value for this test.
d. Calculate the appropriate test statistic.
e. Make a decision about the null hypothesis.
f. Write an interpretation about the relationship between quality of life and overall country ranking.

2. This problem uses the same 2023 "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	3
2. Canada	5
3. Sweden	6
4. Australia	2
5. United States	8
6. Japan	7
7. Germany	10
8. New Zealand	1
9. United Kingdom	9
10. Netherlands	4

a. Which nonparametric test is appropriate for these data?
b. Write a null and alternative hypothesis.
c. Identify the appropriate critical value for this test.
d. Calculate the appropriate test statistic.
e. Make a decision about the null hypothesis.
f. Write an interpretation about the relationship between adventure ranking and overall country ranking.

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 (to increase or decrease helpfulness) 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 researcher has found that willingness to help is generally a positively skewed variable so decides to do a nonparametric test. The hypothetical data that follow will allow you to draw a conclusion about mental health literacy interventions that mirrors Lo et al. (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. Write a null and alternative hypothesis.
c. What is the test value for these data?
d. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
e. Calculate an effect size index.
f. 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.

Time for 200-Meter Dash	
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. Write a null and alternative hypothesis.
c. Assigning ranks with the fastest time as rank #1, what is the test value for these data?
d. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
e. Calculate an effect size index.
f. Write a conclusion about the effect of off-season 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 consumer advocate decided to use a nonparametric test to analyze his data. Find the consumer advocate's data below.

Cleanliness Ratings			
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

a. Which nonparametric test is appropriate for these data?
b. Write a null and alternative hypothesis.
c. Assign ranks with the cleanest score as rank #24, what is the test value for these data?
d. Set alpha at .05. Do you retain or reject the null hypothesis?
e. Calculate any necessary follow-up tests.
f. Calculate an effect size index.
g. Write a conclusion about the relative cleanliness of the four store chains.

6. Students in a 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 IQ scores: one with glasses and one without glasses.

Model	Estimated IQ	
	Without Glasses	With Glasses
1	121	127
2	97	104
3	114	110
4	131	139
5	118	120
6	91	86
7	127	140
8	114	125

a. Which nonparametric test is appropriate for these data?
b. Write a null and alternative hypothesis.
c. What is the test value for these data?
d. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
e. Calculate an effect size index.
f. Write a conclusion about the effect of wearing glasses on IQ judgements.
g. Make a statement about power for this analysis.

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. Fifteen students were divided into three 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 researcher used a nonparametric test because test scores tend to be negatively skewed. Their test scores follow:

% of Study Time Spent in Self-Recitation		
0%	40%	80%
85	89	98
75	84	97
72	87	99
65	92	91
67	93	90

a. Which nonparametric test is appropriate for these data?
b. Write a null and alternative hypothesis.
c. Assign ranks with the best test score a rank of 15, what is the test value for these data?
d. Set alpha at .05. Do you retain or reject the null hypothesis?
e. Calculate any necessary follow-up tests.
f. Calculate an effect size index.
g. Write a conclusion about the effectiveness of self-recitation.

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	1	Japan	37
Sweden	4	Singapore	7
Germany	22	China	61
United Kingdom	45	South Korea	74
Netherlands	14	Qatar	71
Norway	5	Thailand	9
France	44	India	36
Denmark	6	Saudi Arabia	84
Finland	3	Turkey	56
Italy	47	Israel	83

a. Which nonparametric test is appropriate for these data?
b. Write a null and alternative hypothesis.
c. Assign ranks with the best (lowest) “Open for Business” scores as 1, what is the test value for these data?
d. Set alpha at .05 (two-tailed). Do you retain or reject the null hypothesis?
e. Calculate an effect size index.
f. Write a conclusion about the difference in openness to business between European and Asian countries.

ANSWERS

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 Kruskal-Wallis test can only be used when all of the samples are independent.
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. Kruskal-Wallis test followed by Dunn's test (if necessary)

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 different cardiovascular fitness than control participants ($U = 20$, $p < .05$). To determine which group had better cardiovascular fitness you would need to examine the medians or means of the ranked data.

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 change IQ scores, $T = 2$, $p < .05$. To determine the direction of this change, the means or medians of IQ scores for before and after the 3rd grade should be examined.
   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 Kruskal-Wallis test, because the researcher is comparing more than two independent groups on a ranked variable. The appropriate critical value from Table E is 5.991 ($df = 3 – 1 = 2$). Although Electrical engineering majors had the highest overall salaries, followed by mathematics and statistics, then accounting, these differences in salary were not statistically significant. Because this *H* test was not significant, she should not conduct Dunn's test here. Thus, this student should not expect to earn a significantly different starting salary based on her choice of these three majors.

Problems

1. a. Spearman r_s

b. H_0: There is no relationship between quality of life and overall country ranking; $\rho_s = 0$
H_1: There is a relationship between quality of life and overall country ranking; $\rho_s \neq 0$

c. $r_{s.05}$ (10 pairs) = .648

d. $r_s = .50$

e. retain

f. Spearman's correlation coefficient was conducted to examine the relationship between quality of life and overall country ranking. A nonparametric test was selected because the data were ranked. Although the relationship was large and 25% of the variance in quality of life ranking was associated with overall country rating, quality of life ranking was not significantly related to overall country ranking for the top 10 countries according to U. S. News and World Report, r_s (10 pairs) = .50, $p > .05$.

Explanation:

Country Ranking	Quality of Life Ranking	D	D^2
1. Switzerland	3	-2	4
2. Canada	2	0	0
3. Sweden	1	2	4
4. Australia	5	-1	1
5. United States	10	-5	25
6. Japan	9	-3	9
7. Germany	6	1	1
8. New Zealand	7	1	1
9. United Kingdom	8	1	1
10. Netherlands	4	6	36
			$\Sigma D^2 = 82$

Remember for Spearman r_s keep differences that are 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 0, r_s would be equal to 1.

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6(82)}{10(99)} = 1 - \frac{492}{990} = 1 - .497 = .50$$

Be very careful working this formula: A common mistake is to forget the last step of the math (here, that is 1 - .497).

$r_{s.05} = .648$

Because the test value, .50, 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 while by the metrics you've learned in this course is big, it is based on a small sample size.

2. a. Spearman r_s

b. H_0: There is no relationship between quality of life and overall country ranking; $\rho_s = 0$
H_1: There is a relationship between quality of life and overall country ranking; $\rho_s \neq 0$

c. $r_{s.05}$ = (10 pairs) = .648

d. r_s = .26

e. retain

f. Spearman's correlation coefficient was conducted to examine the relationship between adventure ranking and overall country ranking. A nonparametric test was selected because the data were ranked. The relationship was medium and only 4.4% of the variance in adventure ranking was associated with overall country ranking, adventure ranking was not significantly related to overall country ranking for the top 10 countries according to U. S. News and World Report, r_s (10 pairs) = .21, p > .05.

Explanation:

Country Ranking	Adventure Ranking	D	D^2
1. Switzerland	3	-2	4
2. Canada	5	-3	9
3. Sweden	6	-3	9
4. Australia	2	2	4
5. United States	8	-3	9
6. Japan	7	-1	1
7. Germany	10	-3	9
8. New Zealand	1	7	49
9. United Kingdom	9	0	0
10. Netherlands	4	6	36
			$\Sigma D^2 = 130$

Remember for Spearman r_s keep differences that are 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 0, r_s would be equal to 1.

$$r_s = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6(130)}{10(99)} = 1 - \frac{780}{990} = 1 - .788 = .21$$

Be very careful working this formula: A common mistake is to forget the last step of the math (here, that is 1 - .788).

$r_{s.05}$ = .648

Because the test value, .21, 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.

3. a. Wilcoxon signed-rank *T* test

b. H_0: $\Sigma R_{pos} = \Sigma R_{neg}$, H_1: $\Sigma R_{pos} \neq \Sigma R_{neg}$

c. $T = 1$

d. reject

e. $r_{rb} = .928$

f. A Wilcoxon signed-rank *T* test was used to investigate if a mental health literacy intervention changed willingness to help. A nonparametric test was chosen because helpfulness is often a positively skewed variable. Willingness to help scores were significantly higher in the mental health literacy intervention group than the control group, $T = 1$, $p < .05$. Six of the eight students increased their willingness to help; the average increase in helping was 3.71. This difference was large ($r_{rb} = .928$); with 86.1% of the variance in willingness to help associated with the intervention.

Explanation:

Pair	Willingness to Help Scores: Control Group	Willingness to Help Scores: Intervention Group	*D*	Rank	Signed Rank
1	21	23	-2	2	-2
2	12	18	-6	6	-6
3	17	22	-5	5	-5
4	23	23	0	eliminated	eliminated
5	17	16	1	1	1
6	21	24	-3	3	-3
7	19	27	-8	7	-7
8	14	18	-4	4	-4
					$\Sigma R_{pos} = 1$
					$\Sigma R_{neg} = -27$
					$T = 1$
		Check: 27 + 1 = 28; and ((7)(8))/2 = 28			$N = 7$

$T_{.05} = 2$

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.

$$r_{rb} = \left(\frac{\Sigma R_{largest}}{\Sigma R_{total}}\right) - \left(\frac{\Sigma R_{smallest}}{\Sigma R_{total}}\right) = \left(\frac{27}{28}\right) - \left(\frac{1}{27}\right)$$

$= 0.964 - 0.036 = .928$

$.928^2 = .861 = 86.1\%$

4. a. Mann-Whitney *U* test

b. H_0: The distribution of ranks for bicycle training is NOT different from traditional training.
H_1: The distribution of ranks for bicycle training is different from traditional training.

c. $U = 20.5$

d. reject

e. $r_{rb} = .59$

d. A Mann-Whitney *U* test was used to investigate if off-season bicycle training or traditional training is better. A nonparametric test was chosen because running times on a track team are often positively skewed. Low value ranks meant faster running time. Those who participated in the bicycle training ran a 200-meter dash significantly faster (mean rank = 7.55) than those who participated in traditional training (mean rank = 13.45), $U = 20.5$, $p < .05$. This difference was large ($r_{rb} = .590$); with 34.81% of the variance in running time associated with the type of training.

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
	$\Sigma R_{\text{Bike}} = 75.5$		$\Sigma R_{\text{Trad}} = 134.5$

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

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

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. Because the fastest runner was assigned a rank of 1, we can conclude that the bicycle training group (mean rank = 7.55) was faster than the traditional training group (mean rank = 13.45).

$$r_{rb} = 1 - \frac{2(20.5)}{(10)(10)} = 1 - \frac{41}{100} = 1 - .410 = .590$$

5. a. Kruskal-Wallis Test

b. H_0: $\Sigma R_{\text{ChainA}} = \Sigma R_{\text{ChainB}} = \Sigma R_{\text{ChainC}} = \Sigma R_{\text{ChainD}}$

H_1: $\Sigma R_{\text{ChainA}} \neq \Sigma R_{\text{ChainB}} \neq \Sigma R_{\text{ChainC}} \neq \Sigma R_{\text{ChainD}}$ or any inequality

c. $H = 12.78$

d. reject

e. Dunn's test is necessary here.

	Chain A = 11.833	Chain B = 4.5	Chain C = 15.333
Chain B = 4.5	1.79		
Chain C = 15.333	-0.86	-2.65	
Chain D = 18.333	1.59	-3.38*	-0.73

f. $\varepsilon^2 = 0.55$

g. A Kruskal-Wallis test was used to analyze the relative cleanliness of four different supermarket chains. A nonparametric test was selected because the scores on this cleanliness scale tend to be skewed. There was a large ($\varepsilon^2 = 0.55$) significant difference in cleanliness among the chains, $H(3) = 12.78$, $p < .05$. Dunn's test showed that Chain D ($\bar{X} = 42.83$) was rated as significantly cleaner than Chain B ($\bar{X} = 31.5$), Dunn's test = -3.38, $p < .05$. There were no significant differences between any of the other chains ($p > .05$).

Explanation:

Chain Cleanliness							
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	34	8
46	20	31	5	41	15	49	23
38	12	33	7	42	16	40	14
44	18	28	3	47	21	48	22
37	11	32	6	36	10	45	19
	$\Sigma R_A = 71$		$\Sigma R_B = 27$		$\Sigma R_C = 92$		$\Sigma R_D = 110$

Check: 71 + 27 + 92 + 110 = 300; $\frac{N(N+1)}{2} = \frac{24(25)}{2} = 300$

$$H = \frac{12}{24(24+1)}\left(\sum\left[\frac{71^2}{6} + \frac{27^2}{6} + \frac{92^2}{6} + \frac{110^2}{6}\right]\right) - 3(24+1)$$

$$= \frac{12}{600}(\Sigma[840.167 + 121.5 + 1410.667 + 2016.667]) - 75$$

$$= .02(4389.001) - 75 = 87.780 - 75 = 12.78$$

$df = K - 1 = 4 - 1 = 3$; critical value from the χ^2 table = 7.815

$$\varepsilon^2 = 12.78\left(\frac{24+1}{24^2-1}\right) = 12.78\left(\frac{25}{575}\right) = .55 \quad large$$

$$Dunn's = \frac{|\bar{R}_1 - \bar{R}_2|}{\sqrt{\frac{24(24+1)}{12}\left(\frac{1}{6}+\frac{1}{6}\right)}} = \frac{|\bar{R}_1 - \bar{R}_2|}{4.087}$$

	Chain A = 11.833	Chain B = 4.5	Chain C = 15.333
Chain B = 4.5	(11.833 – 4.5)/4.087 1.79		
Chain C = 15.333	(11.833 – 15.333)/4.087 -0.86	(4.5 – 15.333)/4.087 -2.65	
Chain D = 18.333	(11.833 – 18.333)/4.087 -1.59	(4.5 – 18.333)/4.087 -3.38*	(15.333 – 18.333)/4.087 -0.73

Dunn's$_{.05}$ (6, 23) = 2.875

6. a. Wilcoxon signed-rank *T* test
   b. H_0: $\Sigma R_{\text{pos}} = \Sigma R_{\text{neg}}$, H_1: $\Sigma R_{\text{pos}} \neq \Sigma R_{\text{neg}}$
   c. $T = 5$
   d. retain
   e. $r_{\text{rb}} = .722$
   f. A Wilcoxon signed-rank *T* test was used to investigate if wearing glasses changed IQ estimates of models. A nonparametric test was chosen because IQ estimates when the models were wearing glasses had much more variability than when they did not. Six of the eight students gave higher scores to the model wearing glasses, and this difference was large ($r_{\text{rb}} = .722$); with 52.1% of the variance in IQ associated with wearing glasses or not. However, this difference in IQ estimates was not significant, $T = 5$, $p > .05$.
   g. We should repeat the experiment with a larger *N* to see if more power leads to different outcomes.

Explanation:

Pair	IQ Estimates: Without Glasses	IQ Estimates: With Glasses	*D*	Rank	Signed Rank
1	121	127	-6	4	-4
2	97	104	-7	5	-5
3	114	110	4	2	2
4	131	139	-8	6	-6
5	118	120	-2	1	-1
6	91	86	5	3	3
7	127	140	-13	8	-8
8	114	125	-11	7	-7
					$\Sigma R_{pos} = 5$
					$\Sigma R_{\text{neg}} = -31$
					$T = 5$
		Check: 5 + 31 = 36; and ((8)(9))/2 = 36			$N = 8$

$T_{.05} = 3$

Because the test value, 5, is greater than the critical value, 3 (Table J), we retain the null hypothesis and conclude that estimates of IQ are not significantly different for models wearing or not wearing glasses.

$$r_{rb} = \left(\frac{\Sigma R_{largest}}{\Sigma R_{total}}\right) - \left(\frac{\Sigma R_{smallest}}{\Sigma R_{total}}\right) = \left(\frac{31}{36}\right) - \left(\frac{5}{36}\right)$$

= 0.861 – 0.139 = .722

$.722^2 = .521 = 52.1\%$

7. a. Kruskal-Wallis Test
   b. H_0: $\Sigma R_{0\%} = \Sigma R_{40\%} = \Sigma R_{80\%}$
      H_1: $\Sigma R_{0\%} \neq \Sigma R_{40\%} \neq \Sigma R_{80\%}$ or any inequality
   c. $H = 10.26$
   d. reject
   e. Dunn's test is necessary here.

	0% = 3.2	**40% = 8.6**
40% = 8.6	**1.908**	
80% = 12.2	**3.180***	**1.272**

   f. $\varepsilon^2 = 0.55$
   g. A Kruskal-Wallis test was used to analyze the relative effectiveness of spending increasing amounts of time in self-recitation while studying. A nonparametric test was selected because test scores tend to be negatively skewed. There was a large ($\varepsilon^2 = 0.73$) significant difference in test scores among the different amounts of time in self-recitation, $H(2) = 10.26$, $p < .05$. Dunn's test showed that 80% of study time spent in self-recitation ($\bar{X} = 95$) led to significantly better test scores than not using self-recitation at all ($\bar{X} = 72.8$), Dunn's test = -3.18, $p < .05$. There were no significant differences between any of the other groups ($p > .05$).

Explanation:

0%	Rank	40%	Rank	80%	Rank
85	6	89	8	98	14
75	4	84	5	97	13
72	3	87	7	99	15
65	1	92	11	91	10
67	2	93	12	90	9
	$\Sigma R_{0\%} = 16$		$\Sigma R_{40\%} = 43$		$\Sigma R_{80\%} = 61$

Check: 16 + 43 + 61 = 120; $\frac{N(N+1)}{2} = \frac{15(16)}{2} = 120$

$$H = \frac{12}{15(15+1)}\left(\sum\left[\frac{16^2}{5} + \frac{43^2}{5} + \frac{61^2}{5}\right]\right) - 3(15+1)$$

$$= \frac{12}{240}(\Sigma[51.2 + 369.8 + 744.2]) - 48$$

$$= .05(1165.2) - 48 = 58.26 - 48 = 10.26$$

$df = K - 1 = 3 - 1 = 2$; critical value from the χ^2 table = 5.991

$$\varepsilon^2 = 10.26\left(\frac{15+1}{15^2 - 1}\right) = 10.26\left(\frac{16}{224}\right) = .73 \quad large$$

$$Dunn's = \frac{|\bar{R}_1 - \bar{R}_2|}{\sqrt{\frac{15(15+1)}{12}\left(\frac{1}{5} + \frac{1}{5}\right)}} = \frac{|\bar{R}_1 - \bar{R}_2|}{2.83}$$

	0% = 3.2	40% = 8.6
40% = 8.6	(8.6 – 3.2)/2.83 1.908	
80% = 12.2	(12.2 – 3.2)/2.83 3.180*	(12.2 – 8.6)/2.83 1.272

Dunn's$_{.05}$ (3, 14) = 2.694

8. a. Mann-Whitney *U* test
   b. H_0: The distribution of ranks for European and Asian countries is NOT different.
      H_1: The distribution of ranks for European and Asian countries is different.
   c. $U = 16$
   d. reject
   e. $r_{rb} = .68$
   f. A Mann-Whitney *U* test was used to investigate if European or Asian countries were more “open for business” according to U.S. News and World Report. A nonparametric test was chosen because the open for business scores were based on ranks. Low value ranks meant more friendly to business. The top-rated European countries have significantly more favorable business conditions than the top-rated Asian countries, $U = 16$, $p < .05$. This difference was large ($r_{rb} = .680$); with 46.2% of the variance in open for business scores associated with geographic location.

Explanation:

“Open for Business” Scores					
European Countries		**Rank**	**Asian Countries**		**Rank**
Switzerland	1	1	Japan	37	11
Sweden	4	3	Singapore	7	6
Germany	22	9	China	61	16
United Kingdom	45	13	South Korea	74	18
Netherlands	14	8	Qatar	71	17
Norway	5	4	Thailand	9	7
France	44	12	India	36	10
Denmark	6	5	Saudi Arabia	84	20
Finland	3	2	Turkey	56	15
Italy	47	14	Israel	83	19
		$\Sigma R_{EU} = 71$			$\Sigma R_{AS} = 139$

$$U = (N_1)(N_2) + \frac{N_2(N_2+1)}{2} - \sum R_1$$
$$= (10)(10) + \frac{10(11)}{2} - 71 = 100 + 55 - 71 = 84$$

$$U = (N_1)(N_2) + \frac{N_2(N_2+1)}{2} - \sum R_2$$
$$= (10)(10) + \frac{10(11)}{2} - 139 = 100 + 55 - 139 = 16$$

Check: 84 + 16 = 100; (10)(10) = 100

$U_{.05} = 23$

Because the test value, 16 (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, 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. Because the most open for business country was assigned a rank of 1, we can conclude that the European countries (mean rank = 7.1) were more open than the Asian countries (mean rank = 13.9).

$$r_{rb} = 1 - \frac{2(16)}{(10)(10)} = 1 - \frac{32}{100} = 1 - .320 = .680$$

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