Section 15.1: Compare Two Groups by Ranking

15.1 Tanning experiment

a)

Treatment	Ranks					
Lotion	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
Studio	(3,4)	(2,4)	(2,3)	(1,4)	(1,3)	(1,2)

b)

Lotion mean rank	1.5	2.0	2.5	2.5	3.0	3.5
Studio mean rank	3.5	3.0	2.5	2.5	2.0	1.5
Difference of mean ranks	–2.0	–1.0	0.0	0.0	1.0	2.0

c)

Difference between Mean Ranks	Probability
–2.0	1/6
–1.0	1/6
0.0	2/6
1.0	1/6
2.0	1/6

15.2 Test for tanning experiment

a) The P-value is 1/6 = 0.167; if the treatments had identical effects, the probability would be 0.167 of getting a sample like we observed, or even more extreme, in this direction. It is plausible that the null hypothesis is correct, and that the studio does not lead to better results than the lotion.

b) The P-value is 2/6 = 0.33; if the treatments had identical effects, the probability would be 0.33 of getting a sample like we observed, or even more extreme, in either direction. It is plausible that the null hypothesis is correct and that the treatments do not lead to different results.

c) It is a waste of time to conduct this experiment if we plan to use a 0.05 significance level because the smallest possible P-value is 0.17.

15.3 Comparing clinical therapies

a) and b) together

Treatment	Ranks						
Therapy 1	(1,2,3)	(1,2,4)	(1,2,5)	(1,2,6)	(1,3,4)	(1,3,5)	(1,3,6)
Therapy 2	(4,5,6)	(3,5,6)	(3,4,6)	(3,4,5)	(2,5,6)	(2,4,6)	(2,4,5)
Therapy 1 mean rank	2.0	2.33	2.67	3.0	2.67	3.0	3.33
Therapy 2 mean rank	5.0	4.67	4.33	4.0	4.33	4.0	3.67
Difference of mean ranks	–3.0	–2.33	–1.67	–1.0	–1.67	–1.0	–0.33

Treatment	Ranks						
Therapy 1	(1,4,5)	(1,4,6)	(1,5,6)	(2,3,4)	(2,3,5)	(2,3,6)	(2,4,5)
Therapy 2	(2,3,6)	(2,3,5)	(2,3,4)	(1,5,6)	(1,4,6)	(1,4,5)	(1,3,6)
Therapy 1 mean rank	3.33	3.67	4.0	3.0	3.33	3.67	3.67
Therapy 2 mean rank	3.67	3.33	3.0	4.0	3.67	3.33	3.33
Difference of mean ranks	–0.33	0.33	1.0	–1.0	–0.33	0.33	0.33

15.3 (continued)

Treatment	Ranks					
Therapy 1	(2,4,6)	(2,5,6)	(3,4,5)	(3,4,6)	(3,5,6)	(4,5,6)
Therapy 2	(1,3,5)	(1,3,4)	(1,2,6)	(1,2,5)	(1,2,4)	(1,2,3)
Therapy 1 mean rank	4.0	4.33	4.0	4.33	4.67	5.0
Therapy 2 mean rank	3.00	2.67	3.0	2.67	2.33	2.0
Difference of mean ranks	1.0	1.67	1.0	1.67	2.33	3.0

c)

Difference Between Mean Ranks	Probability
–3.00	1/20
–2.33	1/20
–1.67	2/20
–1.00	3/20
–0.33	3/20
0.33	3/20
1.00	3/20
1.67	2/20
2.33	1/20
3.00	1/20

d) The P-value is 4/20 = 0.20 (2/20 for each tail); if the treatments had identical effects, the probability would be 0.10 of getting a sample like we observed, or even more extreme, in either direction. It is plausible that the null hypothesis is correct and that the treatments do not lead to different results.

15.4 Body mass reduction and smoking

a) H_0: Body mass indices are same for individuals who smoke and who do not; H_a: Body mass indices are higher for individuals who do not smoke and who do.

b) Ranks of smokers: 1, 3, 6, 2, 5, 10, 4; Ranks of non-smokers: 9, 12, 13, 7, 14, 11, 8; Sum of ranks of smokers: 31; Sum of ranks of non-smokers: 74; Mean rank for smokers: 4.43; Mean rank for non-smokers: 10.57

c) If the body mass indices were the same for individuals who smoke as for individuals who do not, the probability would be 0.002 of getting a sample like we observed, or even more extreme, in this direction. We have strong evidence that individuals who smoke have lower body mass indices than individuals who do not.

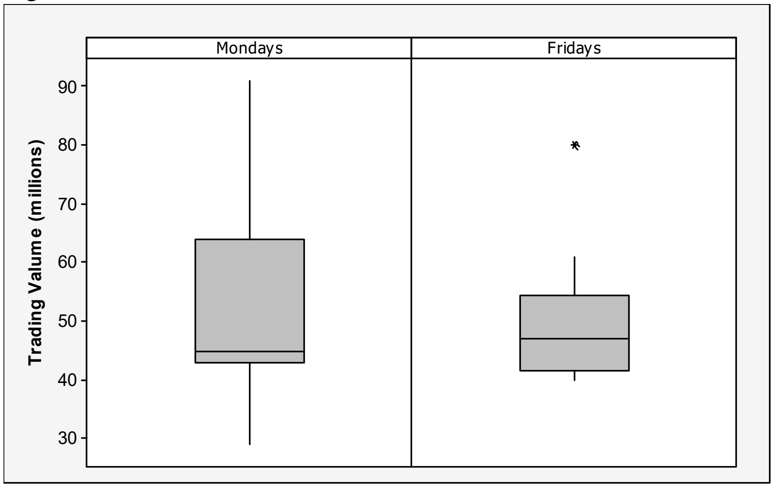
15.5 Estimating smoking effect

a) The point estimate of –2.00 is an estimate of the difference between the population median body mass index for individuals who smoke and individuals who do not.

b) The confidence interval of (–3.3, –0.7) estimates that the population median body mass index for individuals who smoke is between 3.3 and 0.7 pounds below the population median body mass index for individuals who do not smoke. Because all of the values in the interval are below 0, this interval supports the hypothesis that the median body mass index for individuals who smoke is lower than individuals who do not.

15.6 Trading volumes

a) The box plots suggest that the Monday trading volumes are heavily skewed to the right and are more variable than the Friday trading volumes. The Friday trading volumes are slightly right skewed with one extreme large outlier.



b) H_0: Identical population distributions for trading volumes of General Electric shares on Mondays and Fridays; H_a: Different expected values for the sample mean ranks. From technology: $W = 134.5$ and the P-value is 0.902.

c) (–11.0, 13.0); This interval estimates that the population median trading volume for Mondays is between 11 million below and 13 million above the population median trading volume on Fridays. Because 0 is included in the interval, it is plausible that the median trading volume is the same for Mondays and Fridays.

d) The assumption is that there are independent random samples from two groups. The confidence interval requires an extra assumption: that the population distributions for the two groups have the same shape.

15.7 Teenage anorexia

a) The estimated difference between the population median weight change for the cognitive behavioral treatment group and the population median weight change for the control group is 3.05.

b) The confidence interval of (–0.6, 8.1) estimates that the population median weight change for the cognitive-behavioral group is between 0.6 below and 8.1 above the population median weight change for the treatment group. Because 0 falls in the confidence interval, it is plausible that there is no difference between the population medians for the two groups.

c) The P-value is 0.11 for testing against the alternative hypothesis of different expected mean ranks. If the null hypothesis were true, the probability would be 0.11 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that the population distributions are identical.

Section 15.2: Nonparametric Methods for Several Groups and for Matched Pairs

15.8 How long do you tolerate being put on hold?

a) H_0: Identical population distributions for the three groups; H_a: Population distributions not all identical

b) $H = 7.38$; its approximate sampling distribution is the chi-squared distribution with $df = g - 1 = 3 - 1 = 2$.

c) The P-value is 0.025, if the null hypothesis were true, the probability would be 0.025 of getting a test statistic at least as extreme as the value observed. We have strong evidence that the population distributions are not all identical.

d) To find out which pairs of groups significantly differ, we could follow up the Kruskal-Wallis test with a Wilcoxon test to compare each pair of groups. Or, we could find a confidence interval for the difference between the population medians for each pair.

15.9 What's the best way to learn French?

a) Group 1 ranks: 2, 5, 6
Group 2 ranks: 1, 3.5
Group 3 ranks: 3.5, 7, 8
The mean rank for Group 1 = (2 + 5 + 6)/3 = 4.33

b) The P-value is 0.209, if the null hypothesis were true, the probability would be 0.21 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that the population median quiz score is the same for each group.

15.10 Tea versus Coffee

a) $se = \sqrt{(0.50)(0.50)/n} = \sqrt{(0.50)(0.50)/54} = 0.068;\ \ z = (\hat{p} - 0.50)/se = (0.556 - 0.50)/0.068 = 0.82$

b) The P-value is 0.41, if the null hypothesis was true, the probability would be 0.41 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that $p = 0.50$.

15.11 Cell phones and reaction times

a) The observations are dependent samples. All students receive both treatments.

b) The sample proportion is 26/32 = 0.8125.

c) $se = \sqrt{(0.50)(0.50)/n} = \sqrt{(0.50)(0.50)/32} = 0.088;\ \ z = (\hat{p} - 0.50)/se = (0.8125 - 0.50)/0.088 = 3.55;$ The P-value is 0.0002. If the null hypothesis were true, the probability would be 0.0002 of getting a test statistic at least as extreme as the value observed. We have strong evidence that the population proportion if drivers who have a faster reaction time when not using a cell phone is greater than 0.50.

d) The parametric method would be the matched-pairs *t*-test. The sign test uses merely the information about *which* response is higher and *how many*, not the quantitative information about *how much* higher. This is a disadvantage compared to the matched-pairs *t* test which analyzes the mean of the differences between the two responses.

15.12 Sign test for GRE scores

$P(2) = \frac{3!}{2!(3-2)!}(0.50)^2(0.50)^1 = 0.375;$ The more extreme result that all three people score higher on the writing portion is P(3) = $(0.50)^3$ = 0.125. The P-value is the right-tail probability of the observed result and the more extreme one, that is, 0.375 + 0.125 = 0.50. In summary, the evidence is not strong that the population median change in score is positive (but we can't get a small P-value with such a small *n* for this test).

15.13 Does exercise help blood pressure?

H_0: $p = 0.50$; H_a: $p > 0.50$; all three subjects show a decrease. P(3) = $(0.50)^3$ = 0.125; if the null hypothesis were true, the probability would be 0.125 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that walking does not lower blood pressure (but we can't get a small P-value with such a small *n* for this test).

15.14 More on blood pressure

a) H_0: Population median of difference scores is 0; H_a: Population median of difference scores is > 0.

b)

	Sample								Rank of
Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Absolute Value**
1	20	20	–20	20	–20	20	–20	–20	2
2	25	25	25	–25	25	–25	–25	25	3
3	15	–15	15	15	–15	–15	15	–15	1
	Sum of ranks for positive differences								
	6	5	4	3	3	2	1	0	

15.14 (continued)

c) The P-value is 1/8 = 0.125; if the null hypothesis were true, the probability would be 0.125 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that walking does not lower blood pressure (but we can't get a small P-value with such a small *n* for this test).

15.15 More on cell phones

a) H_0: Population median of difference scores is 0; H_a: Population median of difference scores is > 0.

b)

	Sample								Rank of
Subject	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Rank of Absolute Value**
1	32	–32	32	32	32	–32	–32	–32	1
2	67	67	–67	67	67	–67	67	67	2
3	75	75	75	–75	–75	75	–75	75	3
4	150	150	150	150	–150	150	150	–150	4
Sum of ranks for positive differences									
	10	9	8	7	6	7	6	5	

Subject	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**Rank of Absolute Value**
1	32	32	32	–32	–32	–32	32	–32	1
2	–67	–67	67	–67	–67	67	–67	–67	2
3	–75	75	–75	–75	75	–75	–75	–75	3
4	150	–150	–150	150	–150	–150	–150	–150	4
Sum of ranks for positive differences									
	5	4	3	4	3	2	1	0	

c) The P-value is 1/16 = 0.06; if the null hypothesis were true, the probability would be 0.06 of getting a test statistic at least as extreme as the value observed. There is some, but not strong, evidence that cell phones tend to impair reaction times.

15.16 Use all data on cell phones

a) H_0: Population Median of Difference Scores = 0.
H_a: Population Median of Difference Scores not = 0.

b) MINITAB found the reported value by determining the sum of ranks for the positive differences.

c) The two-sided P-value is 0.000. If the null hypothesis were true, the probability would be close to 0 of getting a test statistic at least as extreme as the value observed. We have very strong evidence that the population median of the differences is not 0.

d) The estimated median is 47.25. This estimates that the population median difference between the reaction times of those not using cell phones and those using cell phones was 47.25.

Chapter Problems: Practicing the Basics

15.17 Smartphone sales

a) Sales with discount: ranks are 4, 6, 5; mean is 5. Sales without discount: ranks are 1, 2, 3; mean is 2.

b)

Treatment	Ranks
Sales with discount	(1, 2, 3) (1,2,4) (1,2,5) (1,2,6) (1,3,4) (1,3,5) (1,3,6) (1,4,5) (1,4,6) (1,5,6)
Sales without discount	(4,5,6) (3,5,6) (3,4,6) (3,4,5) (2,5,6) (2,4,6) (2,4,5) (2,3,6) (2,3,5) (2,3,4)

Treatment	Ranks
Sales with discount	(2,3,4) (2,3,5) (2,3,6) (2,4,5) (2,4,6) (2,5,6) (3,4,5) (3,4,6) (3,5,6) (4,5,6)
Sales without discount	(1,5,6) (1,4,6) (1,4,5) (1,3,6) (1,3,5) (1,3,4) (1,2,6) (1,2,5) (1,2,4) (1,2,3)

c) There are only two ways in which the ranks are as extreme as in this sample: sales with discount with 1, 2, 3 and sales without discount with 4, 5, 6, or sales with discount with 4, 5, 6 and sales with discount 1, 2, 3.

d) The P-value is 0.10 because out of 20 possibilities, only two are this extreme. 2 / 20 = 0.10.

15.18 Comparing smartphone sales

a) The results would not change. This illustrates that the analysis does not take the magnitude of the sample scores into account and is not affected by outliers.

b) Kruskal–Wallis test

15.19 Telephone holding times

a)

Group	Ranks	Mean Rank
Muzak	1, 2, 4, 5, 3	3.0
Classical	9, 8, 7, 10, 6	8.0

b) There are only two cases this extreme, that in which Muzak has ranks 1–5 as it does here, and that in which Muzak has ranks 6–10. Thus, the P-value is the probability that one of these two cases would occur out of the 252 possible allocations of rankings. If the treatments had identical effects, the probability would be 0.008 of getting a sample like we observed or even more extreme. This is below a typical significance level such as 0.05; therefore, we can reject the null hypothesis.

15.20 Treating alcoholics

a) The two-sample *t* test assumes that the population distribution is normal. The test is robust with respect to this assumption with a two-sided test, but these researchers planned to conduct a one-sided test.

b)

Group	Ranks	Mean Rank
Control	13, 23, 18, 12, 22, 19, 16, 21, 5, 15, 11, 20	16.250
Treated	9, 2, 4, 7, 17, 10, 3, 6, 14, 8, 1	7.364

c) The P-value of 0.001 tells us that if the treatments had identical effects, the probability would be 0.001 of getting a sample like we observed or even more extreme. We have strong evidence that the treatment with social skills training reduced drinking.

d) The confidence interval is (186.0, 713.0). We infer that the population median alcohol consumption for the control group is between 186 and 713 centiliters more than for the treated group.

15.21 Comparing tans

a) Kruskal-Wallis test

b) There are several possible examples, but all would have one group with ranks 1–3, one with ranks 4–6 and one with ranks 7–9.

15.22 Comparing therapies for anorexia

a) H_0: Identical population distributions for the three anorexia treatment groups;
H_a: Population distributions not all identical.

b) Test statistic: 9.07; chi-squared distribution with $df = g - 1 = 3 - 1 = 2$
From Minitab:

```
Kruskal-Wallis Test on weight change
treatment       N   Median  Ave Rank      Z
cogchange      29   1.4000      37.0   0.15
controlchange  26  -0.3500      28.4  -2.46
famchange      17   9.0000      48.1   2.61
Overall        72               36.5
H = 9.07  DF = 2  P = 0.011
```

c) The P-value is 0.011; if the null hypothesis were true, the probability would be 0.011 of getting a test statistic at least as extreme as the value observed. We have strong evidence that the population distributions for the three treatments for anorexia are not all identical.

15.23 Internet versus cell phones

a) (i) H_0: Population proportion $p = 0.50$ who use a cell phone more than the Internet; H_a: $p \neq 0.50$.

(ii) $se = \sqrt{(0.50)(0.50)/n} = \sqrt{(0.50)(0.50)/39} = 0.080$; $z = (\hat{p} - 0.50)/se = (0.897 - 0.50)/0.080 = 4.96$

(iii) The P-value is 0.000; if the null hypothesis were true, the probability would be near 0 of getting a test statistic at least as extreme as the value observed. We have extremely strong evidence that a majority of countries have more cell phone use than Internet use.

b) This would not be relevant if the data file were comprised only of countries of interest to us. We would know the population parameters so inference would not be relevant.

15.24 Browsing the Internet

a) We would use the Kruskal-Wallis test because there are three political affiliations.
H_0: Identical population distributions for the 3 groups; H_a: Population distributions not all identical.

b) $H = 4.55$
From Minitab:

```
Kruskal-Wallis Test on BrowseInternet
PoliticalAff   N  Median  Ave Rank      Z
1              8   30.00      31.7   0.30
2             36   30.00      26.5  -1.96
3             15   60.00      37.5   1.96
Overall       59              30.0
H = 4.43  DF = 2  P = 0.109
H = 4.55  DF = 2  P = 0.103  (adjusted for ties)
```

c) The P-value is 0.10, if the null hypothesis were true, the probability would be 0.10 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that time spent browsing the Internet is independent of political affiliation.

15.25 GPAs

a) We could use the sign-test for matched pairs or the Wilcoxon signed-ranks test.

b) One reason for using a nonparametric method is if we suspected that the population distributions were not normal, for example, possibly highly skewed, because we have a one-sided alternative, and parametric methods are not then robust.

15.25 (continued)

c) From Minitab:

```
Wilcoxon Signed Rank Test: CGPA-HSGPA
Test of median = 0.000000 versus median < 0.000000
                N for   Wilcoxon          Estimated
             N   Test  Statistic      P      Median
CGPA-HSGPA  59     55      268.5  0.000     -0.1850
```

If the null hypothesis were true, the probability would be near 0 of getting a test statistic at least as extreme as the value observed. We have very strong evidence that population median high school GPA is higher than population median college GPA.

15.26 Sign test about the GRE workshop

a) H_0: Population proportion $p = 0.50$ who score better on the GRE; H_a: $p > 0.50$

$P(2) = \frac{3!}{2!(3-2)!}(0.50)^2(0.50)^1 = 0.375$; The more extreme result that all three people would score higher has probability $P(3) = (0.50)^3 = 0.125$. The P-value is the right-tail probability of the observed result and the more extreme one, that is, 0.375 + 0.125 = 0.50. If the null hypothesis were true, the probability would be 0.50 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that the GRE score difference is not positive.

b) The results are identical. Outliers do not have an effect on this nonparametric statistical method.

15.27 Wilcoxon signed-rank test about the GRE workshop

a) H_0: Population median of difference scores is 0.; H_a: Population median of difference scores is > 0.

Possible Samples with Absolute Difference Values of Sample

Subject	1	2	3	4	5	6	7	8	Rank of Absolute Value
1	5.5	5.5	–5.5	5.5	–5.5	5.5	–5.5	–5.5	3
2	–0.5	–0.5	–0.5	0.5	–0.5	0.5	0.5	0.5	1
3	1.5	–1.5	1.5	1.5	–1.5	–1.5	1.5	–1.5	2
Sum of ranks for positive differences									
	5	3	2	6	0	4	3	1	

The rank sum is 5 one-eighth of the time, and is more extreme (i.e., 6) one-eighth of the time. Thus, the P-value is 2/8 = 0.25. If the null hypothesis were true, the probability would be 0.25 of getting a test statistic at least as extreme as the value observed. It is plausible that the null hypothesis is correct and that the population median of difference scores is not positive.

b) The P-value is smaller than in Example 8. Outliers do not have an effect on this nonparametric statistical method.

Chapter Problems: Concepts and Investigations

15.28 Student survey

The one-page reports will be different for each student, but should include the following findings from technology:

From Minitab:

```
Mann-Whitney Test and CI: Newspaper_F, Newspaper_M
              N  Median
Newspaper_F  31   3.000
Newspaper_M  29   3.000
Point estimate for ETA1-ETA2 is -0.000
95.1 Percent CI for ETA1-ETA2 is (-2.000,1.000)
W = 892.5
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at 0.4374
The test is significant at 0.4323 (adjusted for ties)
```

15.29 Why nonparametrics?

There are many possible situations. One example is a situation in which the population distribution is likely to be highly skewed and the researcher wants to use a one-sided test.

15.30 Why matched pairs?

With a crossover design, we get a score for each treatment for every subject, whereas with an independent samples design, we must assign subjects to one treatment only.

15.31 Complete the analogy

Kruskal-Wallis

15.32 Complete the analogy

Sign test for matched pairs or the Wilcoxon signed-ranks test

15.33 True or false

False

15.34 Multiple choice

The best answer is (c).

♦♦15.35 Mann-Whitney statistic

a) The proportions are calculated by pairing up the subjects in every possible way, and then counting the number of pairs for which the tanning studio gave a better tan. For the first set of ranks in the chart (lotion: 1,2,3 and studio: 4,5), the possible pairs are as follows with lotion first: (1,4) (1,5) (2,4) (2,5) (3,4) (3,5). In none of these pairs did the studio have the higher rank; therefore, the proportion is 0/6.

b)

Proportion	Probability
0/6	1/10
1/6	1/10
2/6	2/10
3/6	2/10
4/6	2/10
5/6	1/10
6/6	1/10

c) The P-value is 2/10. The probability of an observed sample proportion of 5/6 or more extreme (i.e., 6/6) is 2/10 = 0.20.

♦♦15.36 Rank-based correlation

a) The Spearman rank correlation is not affected by an outlier. The largest score in a data set of 30 observations would receive a rank of 30 whether it was bigger than the second largest score by 1 or by 100.

b) The null hypothesis would include the value 0.

♦♦15.37 Nonparametric regression

The nonparametric estimate of the slope is not strongly affected by a regression outlier because we are taking the median of all slopes. The median is not susceptible to outliers. The ordinary slope, on the other hand, takes the magnitude of all observations into account.