Chapter 9

NONPARAMETRIC METHODS

REVIEW OF KEY CONCEPTS

SECTION 9.1 Types of Data

1. **Cardinal data** Data on a scale where addition is meaningful (e.g., change in 3 inches for height). There are two types of cardinal data:
   a. **Ratio-scale data** Cardinal data on a scale where ratios between values are meaningful (e.g., serum-cholesterol levels).
   b. **Interval-scale data** Cardinal data where the zero point is arbitrary. For such data, ratios are not meaningful (e.g., Julian dates; we can calculate the number of days between two dates, but we can't say that one date is twice as large as another date).
2. **Ordinal data** Categorical data that can be ordered, but the increment between specific values is arbitrary (e.g., degree of redness of an eye as determined on a photograph in comparison with reference photographs).
3. **Nominal data** Categorical data where the categories are not ordered (e.g., ethnic group).

In this chapter, we are concerned with nonparametric methods (methods that do not assume a specific underlying distribution for our data) and will apply these methods to either cardinal or ordinal data.

SECTION 9.2 The Sign Test

9.2.1 Large-Sample Test

Suppose we measure the change in periodontal status among 28 adults with periodontal disease after a dental-education program. The distribution of change scores is as follows:

	Category	Frequency
+3	greatest improvement	4
+2	moderate improvement	5
+1	slight improvement	6
0	no change	5
–1	slight decline	4
–2	moderate decline	2
–3	greatest decline	2

If we eliminate the "no change" people, then 15 people improved and 8 declined. Suppose we wish to determine whether more people improved than declined with the program. We wish to test the hypothesis

$$H_0: p = \frac{1}{2} \quad \text{vs.} \quad H_1: p \neq \frac{1}{2}$$

where p = probability of improvement given that a person has changed. We use the test statistic

$$z = \frac{\left|C - \frac{n}{2}\right| - \frac{1}{2}}{\sqrt{\frac{n}{4}}} \sim N(0,\ 1) \text{ under } H_0$$

The p-value $= 2 \times [1 - \Phi(z)]$. This is a special case of the one-sample binomial test, where we test the hypothesis H_0: $p = p_0$ versus H_1: $p \neq p_0$ and $p_0 = .5$. The criterion for using this test is that n = number of people who have changed is ≥ 20.

In this example, $C = 15$ and the test statistic is

$$z = \frac{\left|15 - \frac{23}{2}\right| - \frac{1}{2}}{\sqrt{23/4}} = \frac{3}{2.398} = 1.251 \sim N(0,\ 1)$$
$$p = 2 \times [1 - \Phi(1.251)] = .21$$

Thus, there is no significant change in periodontal status after the program.

9.2.2 Small-Sample Test

If $n < 20$, then the following small-sample test must be used. If x = number of people who improve among n people who have changed, then

$$p\text{-value} = 2 \times \sum_{k=0}^{x} \binom{n}{k} \left(\frac{1}{2}\right)^n \qquad \text{if } x < \frac{n}{2}$$
$$= 2 \times \left[\sum_{k=x}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^n\right] \qquad \text{if } x > \frac{n}{2}$$
$$= 1 \qquad \text{if } x = \frac{n}{2}$$

SECTION 9.3 The Wilcoxon Signed-Rank Test

Suppose we want to take account of the magnitude of change as well as its direction. In particular, the positive-change scores seem greater than the negative-change scores.

Procedure

Order the change scores by absolute value and rank the absolute values:

Absolute value of change score	Positive	Negative	Total	Rank	Average rank
1	6	4	10	1–10	5.5
2	5	2	7	11–17	14.0
3	4	2	6	18–23	20.5

Assign the average rank to each observation in a group (which equals the average of the highest and lowest ranks within a group). For example, in the group with absolute value of change score = 1, the lowest rank = 1 and the highest rank = 10; the average rank $= (1+10)/2 = 5.5$ is assigned to each subject in the group. Compute the rank sum in the lst group $\equiv R_1$.

$$R_1 = 6 \times 5.5 + 5 \times 14.0 + 4 \times 20.5 = 33 + 70 + 82 = 185$$

Compare the observed rank sum for people with positive scores with the expected rank sum under the null hypothesis that the rank sum is the same for people with positive and negative scores. If most of the changes are in the positive direction, particularly for change scores with large absolute value, then the rank sum for people with positive scores will be large; if most of the changes are in the negative direction, particularly for change scores with large absolute value, then it will be small; if the changes are equally dispersed between positive and negative change scores, particularly for change scores with large absolute value, then the rank sum will be close to $(1 + 2 + \cdots + n)/2 = n(n+1)/4$. Under H_0,

$$E(R_1) = \frac{n(n+1)}{4}$$

$$Var(R_1) = \frac{n(n+1)(2n+1)}{24} - \sum_{i=1}^{g}\left(\frac{t_i^3 - t_i}{48}\right)$$

where $t_i =$ number of observations in the ith tied group. We use the test statistic

$$z = \frac{|R_1 - E(R_1)| - .5}{sd(R_1)} \sim N(0,\ 1) \text{ under } H_0$$

The p-value $= 2 \times \Phi(z)$ if $z < 0$ or $= 2 \times [1 - \Phi(z)]$ if $z \geq 0$. We will only use this test if n = number of nonzero change scores is ≥ 16. In this case,

$$n = 23$$

$$E(R_1) = \frac{23(24)}{4} = 138$$

$$\text{Var}(R_1) = \frac{23(24)(47)}{24} - \left[\frac{(10^3 - 10) + (7^3 - 7) + (6^3 - 6)}{48}\right]$$

$$= 1081 - \frac{990 + 336 + 210}{48} = 1081 - \frac{1536}{48} = 1081 - 32 = 1049$$

$$z = \frac{|185 - 138| - \frac{1}{2}}{\sqrt{1049}} = \frac{46.5}{32.4}$$

$$= 1.436 \sim N(0,\ 1) \text{ under } H_0,\ \text{two - tailed } p = .15$$

Thus, periodontal status has not significantly changed.

Notice that we obtained a slightly lower p-value using the signed-rank test, where both the magnitude and the direction of the difference scores are taken into account than with the sign test, where only the direction, but not the magnitude of the difference scores are taken into account. The signed-rank test is more appropriate in this case.

If the number of nonzero change scores is ≤ 15, then we refer to Table 11 (Appendix, text). Upper and lower two-tailed critical values are provided for $\alpha = .10$, .05, .02, and .01, for each $n \leq 15$. Results are significant at level α if $R_1 \geq$ upper critical value or $R_1 \leq$ lower critical value. For example, if we have 10 untied pairs and $R_1 = 49$, then since the upper critical value at $\alpha = .05$ is 47 and the upper critical value at $\alpha = .02$ is 50, it follows that $.02 \leq p < .05$.

SECTION 9.4 The Wilcoxon Rank-Sum Test

Since plasma aldosterone has been related to blood-pressure (BP) level in adults, and black adults generally have higher mean BP levels than white adults, it is of interest to compare the aldosterone distributions of black and white children.

In Figure 5.23 (in Chapter 5, text), we present the distribution of plasma-aldosterone concentrations for 53 white and 46 black children. The distributions are very skewed, and we are reluctant to use a *t* test to compare the distributions. Instead, we use the Wilcoxon rank-sum test. For simplicity, we will group the distributions in 200 pmol/liter groups as follows:

Aldosterone group[a]	White	Black	Total	Rank range	Average rank
0–199	12	28	40	1–40	20.5
200–399	17	10	27	41–67	54.0
400–599	15	5	20	68–87	77.5
600+	9	3	12	88–99	93.5
Total	53	46	99		

[a] pmol/liter.

We compute the rank range and average rank in each group over the combined sample (white + black children). We then compute the rank sum in the 1st sample (white children). If the rank sum is large, then white children have higher levels than black children; if it is small, then white children have lower levels. We have

$$R_1 = 12(20.5) + \cdots + 9(93.5) = 3168$$

Under the null hypothesis that the average aldosterone rank is the same in each group,

$$E(R_1) = n_1\left(\frac{1+n_1+n_2}{2}\right)$$

$$Var(R_1) = \frac{n_1 n_2}{12}\left[n_1+n_2+1-\frac{\sum(t_i^3-t_i)}{(n_1+n_2)(n_1+n_2-1)}\right]$$

where $t_i =$ number of tied observations in the ith tied group. We have the test statistic

$$T = \frac{|R_1 - E(R_1)| - .5}{sd(R_1)} \sim N(0,\ 1) \text{ under } H_0$$

The p-value $= 2 \times \Phi(T)$ if $T < 0$, or $= 2 \times [1 - \Phi(T)]$ if $T \geq 0$. This test should only be used if n_1 and n_2 are each ≥ 10.

In this case,

$$E(R_1) = 53\left(\frac{1+53+46}{2}\right) = 2650$$

$$Var(R_1) = \frac{53(46)}{12}\left[100 - \frac{\left(40^3-40\right)+\cdots+\left(12^3-12\right)}{99(98)}\right] = 203.167\left(100-\frac{93,312}{9702}\right) = 203.167(90.382) = 18,362.65$$

$$sd(R_1) = \sqrt{18,362.65} = 135.51$$

$$T = \frac{|3,168-2,650|-.5}{135.51} = \frac{517.5}{135.51} = 3.82 \sim N(0,\ 1) \text{ under } H_0$$

The p-value $= 2 \times [1 - \Phi(3.82)] < .001$. Thus, since $R_1 > E(R_1)$, white children have significantly higher aldosterone levels than black children.

If either n_1 or n_2 are < 10, then we refer to Table 12 (Appendix, text). For each combination of n_1 = minimum of the two sample sizes = 4, …, 9 and n_2 = maximum of the two sample sizes = 4, …, 50, two-tailed critical values are provided for $\alpha = .10$, .05, .02, and .01. The results are significant at level α if $R_1 \leq$ lower critical value or $R_1 \geq$ upper critical value. Suppose $n_1 = 5$, $n_2 = 10$ and $R_1 = 54$. The critical values for $\alpha = .05$ are 23 and 57. Since $R_1 > 23$ and $R_1 < 57$, it follows that $p > .05$ and there is no significant difference between the two groups.

PROBLEMS

Ophthalmology
Suppose an ophthalmologist reviews fundus photographs of 30 patients with macular degeneration both before and 3 months after receiving a laser treatment. To assess the efficacy of treatment, each patient is rated as improved, remained the same, or declined.

9.1 If 20 patients improved, 7 declined, and 3 remained the same, then assess whether or not patients undergoing this treatment are showing significant change from baseline to 3 months afterward. Report a p-value.

Suppose that the patients are divided into two groups according to initial visual acuity (VA). Of the 14 patients with VA 20-40 or better, 8 improved, 5 declined, and 1 stayed the same. Of the 16 patients with VA worse than 20–40, 12 improved, 2 declined, and 2 stayed the same,

9.2 Assess the results in the subgroup of patients with VA of 20–40 or better.

9.3 Assess the results in the subgroup of patients with VA worse than 20–40.

Diabetes
An experiment was conducted to study responses to different methods of taking insulin in patients with type I diabetes. The percentages of glycosolated hemoglobin initially and 3 months after taking insulin by nasal spray are given in Table 9.1 [1].

Table 9.1 Percentages of glycosolated hemoglobin (HgbA1c) before and 3 months after taking insulin by nasal spray

Patient number	Before	3 months after
1	11.0	10.2
2	7.7	7.9
3	5.9	6.5
4	9.5	10.4
5	8.7	8.8
6	8.6	9.0
7	11.0	9.5
8	6.9	7.6

Source: Reprinted with permission of the *New England Journal of Medicine*, **312**(17), 1078–1084, 1985.

9.4 Perform a t test to compare the percentages of glycosolated hemoglobin before and 3 months after treatment.

9.5 Suppose normality is not assumed. Perform a nonparametric test corresponding to the t test in Problem 9.4.

9.6 Compare your results in Problems 9.4 and 9.5.

Pathology
Refer to Table 6.1 (Chapter 6, Study Guide).

9.7 Suppose normality is not assumed. What nonparametric test can be used to compare total heart weight of males with left-heart disease with that of normal males?

9.8 Implement the test in Problem 9.7 and report a p-value.

Cardiology
Propranolol is a standard drug given to ease the pain of patients with episodes of unstable angina. A new drug for the treatment of this disease is tested on 30 pairs of patients who are matched on a one-to-one basis according to age, sex, and clinical condition and are assessed as to the severity of their pain. Suppose that in 15 pairs of patients, the patient with the new drug has less pain; in 10 pairs of patients, the patient with propranolol has less pain; and in 5 pairs of patients, the pain is about the same with the two drugs.

9.9 What is the appropriate test to use here?

9.10 Perform the test in Problem 9.9 and report a p-value.

Ophthalmology
Table 8.15 (in Chapter 8, text) presents data giving the median gray levels in the lens of the human eye for 6 cataractous and 6 normal people.

9.11 What nonparametric test could be used to compare the median gray levels of normal and cataractous eyes?

9.12 Carry out the test in Problem 9.11 and report a p-value.

Psychiatry
Suppose we are conducting a study of the effectiveness of lithium therapy for manic-depressive patients. The study is carried out at two different centers, and we want to determine if the patient populations are comparable at baseline. A self-rating questionnaire about their general psychological well-being is administered to the prospective patients at the two different centers. The outcome measure on the questionnaire is a four-category scale: (1) = feel good; (2) = usually feel good, once in a while feel nervous; (3) = feel nervous half the time; (4) = usually feel nervous. Suppose the data at the two different centers are as follows:

Center 1	3, 4, 1, 1, 3, 2, 3, 4, 4, 3, 2, 4, 4, 4
Center 2	1, 2, 1, 3, 2, 4, 1, 2, 1, 3, 1, 2, 2, 2, 1, 3

9.13 What type of data does this type of scale represent?

9.14 Why might a parametric test not be useful with this type of data?

9.15 Assess if there is any significant difference in the responses of the two patient populations using a nonparametric test.

Psychiatry
Much attention has been given in recent years to the role of transcendental meditation in improving health, particularly in lowering blood pressure. One hypothesis that emerges from this work is that transcendental meditation might also be useful in treating psychiatric patients with symptoms of anxiety. Suppose that a protocol of meditational therapy is administered once a day to 20 patients with anxiety. The patients are given a psychiatric exam at baseline and at a follow-up exam 2 months later. The degree of improvement is rated on a 10-point scale, with 1 indicating the most improvement and 10 the least improvement. Similarly, 26 comparably affected patients with anxiety are given standard psychotherapy and are asked to come back 2 months later for a follow-up exam. The results are given in Table 9.2.

Table 9.2 Degree of improvement in patients with anxiety who are treated with transcendental meditation or psychotherapy

Meditation		Psychotherapy	
d^a	f^b	d^a	f^b
1	3	1	0
2	4	2	2
3	7	3	5
4	3	4	3
5	2	5	8
6	1	6	4
7	0	7	2
8	0	8	1
9	0	9	1
10	0	10	0
	20		26

[a] d = degree of improvement.
[b] f = frequency.

9.16 Why might a parametric test not be useful here?

9.17 What nonparametric test should be used to analyze these data?

9.18 Compare the degree of improvement between the two groups using the test in Problem 9.17.

SOLUTIONS

9.1 Use the sign test. If the people who have remained the same are ignored, then 27 people have either improved or declined. Thus, the normal-theory version of the test can be used. We have that C = number of patients improved = 20, $n = 27$. The exact p-value is given by

$$p = 2\times\left[1-\Phi\left(\frac{C-\frac{n}{2}-0.5}{\sqrt{\frac{n}{4}}}\right)\right] = 2\times\left[1-\Phi\left(\frac{20-\frac{27}{2}-0.5}{\sqrt{\frac{27}{4}}}\right)\right]$$
$$= 2\times\left[1-\Phi\left(\frac{6.5-0.5}{2.598}\right)\right] = 2\times\left[1-\Phi\left(\frac{6.0}{2.598}\right)\right]$$
$$= 2\times\left[1-\Phi(2.309)\right] = 2\times(1-.9895) = .021$$

9.2 In the subgroup of patients with better visual acuity, 13 patients have either improved or declined. Thus, the exact binomial test must be used. We have $C = 8$, $n = 13$. Refer to the exact binomial tables (Table 1, Appendix, text) under $n = 13$, $p = .5$ to compute

$$p = 2\times\sum_{k=8}^{13}\binom{13}{k}\left(\frac{1}{2}\right)^{13}$$
$$= 2\times(.1571+.0873+.0349+.0095+.0016+.0001)$$
$$= 2\times.2905 = .581$$

Thus, there is no significant change in the fundus photographs for the subgroup of patients with VA 20–40 or better.

9.3 Fourteen patients have either improved or declined. Thus, the exact binomial test must again be used. We have $C = 12$, $n = 14$. Refer to the exact binomial tables under $n = 14$, $p = .5$ to compute

$$p = 2\times\sum_{k=12}^{14}\binom{14}{k}\left(\frac{1}{2}\right)^{14}$$
$$= 2\times(.0056+.0009+.0001)$$
$$= 2\times.0066 = .013$$

Thus, the fundus photographs of persons with VA of worse than 20–40 at baseline have significantly improved.

9.4 We use a paired t test, where d_i = (HgbA1c before – HgbA1c 3 months after) for the i^{th} subject and $\bar{d} = -0.075$, $s_d = 0.821$, $n = 8$. We have the test statistic

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{-0.075}{\frac{0.821}{\sqrt{8}}} = \frac{-0.075}{0.290} = -0.26 \sim t_7 \text{ under } H_0$$

Since $t_{7,\,.975} = 2.365 > |t|$, we accept H_0 at the 5% level.

9.5 Since each patient is being used as his or her own control, we use the Wilcoxon signed-rank test. First, rank the difference scores by absolute value as follows:

$\lvert d_i \rvert$	Neg d_i	f_i	Pos d_i	f_i	Number of patients with absolute value	Range of ranks	Average rank
1.5	–1.5	0	1.5	1	1	8	8.0
0.9	–0.9	1	0.9	0	1	7	7.0
0.8	–0.8	0	0.8	1	1	6	6.0
0.7	–0.7	1	0.7	0	1	5	5.0
0.6	–0.6	1	0.6	0	1	4	4.0
0.4	–0.4	1	0.4	0	1	3	3.0
0.2	–0.2	1	0.2	0	1	2	2.0
0.1	–0.1	1	0.1	0	1	1	1.0
		6		2			

We then compute the rank sum of the positive differences $= R_1 = 6 + 8 = 14$. Since there are fewer than 16 nonzero differences, we refer to the small sample table for the signed-rank test (Table 11, Appendix, text). We note that for n = 8, $\alpha = .10$, the two-tailed upper and lower critical values are 31 and 5, respectively. Since 5 < 14 < 31, it follows that $p > .10$ and there is no significant difference in the % of glycosolated hemoglobin before and 3 months after taking insulin by nasal spray.

9.6 The same conclusions were reached in Problems 9.4 and 9.5.

9.7 Since the groups are not matched, we can use the Wilcoxon rank-sum test.

9.8 We combine the data from the two groups and assign ranks based on the combined sample.

Value	Left heart disease males	Normal males	Combined sample	Range of ranks	Average rank
245	0	1	1	1	1.0
270	0	1	1	2	2.0
285	1	0	1	3	3.0
290	0	1	1	4	4.0
300	0	2	2	5-6	5.5
310	1	1	2	7-8	7.5
325	1	0	1	9	9.0
340	0	1	1	10	10.0
350	0	1	1	11	11.0
360	0	1	1	12	12.0
375	1	0	1	13	13.0
405	0	1	1	14	14.0
425	1	0	1	15	15.0
450	2	0	2	16-17	16.5
460	1	0	1	18	18.0
495	1	0	1	19	19.0
615	1	0	1	20	20.0
760	1	0	1	21	21.0

We compute the rank sum in the left heart disease group. This is given by

$$R_1 = 3.0 + 7.5 + \cdots + 21.0 = 158.5.$$

Since both n_1 and n_2 are ≥ 10, we can use the normal approximation method. We have

$$T = \frac{\left|158.5 - \frac{11(22)}{2}\right| - .5}{\sqrt{\frac{11(10)}{12}\left[22 - \frac{2(2^2-1)+2(2^2-1)+2(2^2-1)}{21(20)}\right]}}$$

$$= \frac{|158.5 - 121| - .5}{\sqrt{9.167(22 - \frac{18}{420})}} = \frac{37.0}{\sqrt{201.274}} = \frac{37.0}{14.187}$$

$$= 2.608 \sim N(0,\ 1) \text{ under } H_0$$

Thus, the p-value

$$2 \times \left[1 - \Phi\left(2.608\right)\right] = 2 \times \left(1 - .9954\right) = .009.$$

Therefore, the mean total heart weight for left heart disease males is significantly higher than for normal males. This is the same conclusion we reached in Problem 8.7 using the independent-samples t test on these data.

9.9 The appropriate test to use here is the sign test because we only know which treated patient in a pair has less severe pain, but not how much less severe it is.

9.10 We test the hypothesis H_0: $p = 1/2$ versus H_1: $p \neq 1/2$, where p = probability that the patient in a matched pair taking the new drug has less severe pain than the patient taking propranolol. Since the number of nonzero differences $= 25 \geq 20$, we can use the normal-theory method given in Equation 9.1 and Equation 9.2 (text, Chapter 9). We have the test statistic

$$z = \frac{\left|C - \frac{n}{2}\right| - .5}{\sqrt{\frac{n}{4}}} = \frac{\left(\left|15 - \frac{25}{2}\right| - .5\right)}{\sqrt{\frac{25}{4}}} = \frac{2.0}{2.50}$$
$$= 0.80 \sim N(0,\ 1) \text{ under } H_0$$

The two-sided p-value is given by

$$p = 2 \times [1 - \Phi(0.80)] = 2 \times (1 - .7881) = .424,$$

which is not statistically significant. Thus, there is no significant difference between the two treatments.

9.11 The Wilcoxon rank-sum test.

9.12 We combine the two samples and rank-order the observations in the combined sample. We have

Value	Cataractous people	Normal people	Total	Rank
106	1	0	1	1.0
136	1	0	1	2.0
140	1	0	1	3.0
145	0	1	1	4.0
149	1	0	1	5.0
158	0	1	1	6.0
161	1	0	1	7.0
167	0	1	1	8.0
171	1	0	1	9.0
177	0	1	1	10.0
182	0	1	1	11.0
185	0	1	1	12.0
	6	6	12	

The rank sum in the cataractous group is given by $R_1 = 1.0 + 2.0 + \cdots + 9.0 = 27.0$. We use the exact test due to the small sample sizes. We refer to Table 12 (Appendix, text) under $n_1 = 6$, $n_2 = 6$, $\alpha = .05$ and note that $T_l = 26$, $T_r = 52$. Since 26 < 27 < 52, it follows that $p \geq .05$ and there is no significant difference between the two groups.

9.13 This type of scale is an example of ordinal data because we can only say that patients with a rating of 1 feel better than those with a rating of 2, but not how much better they feel.

9.14 A parametric test such as the t test would not be useful here because we would have to assume specific cardinal values for the four ratings that would then be added and multiplied.

9.15 We have the following grouped data:

	Score				
Center	1	2	3	4	Total
1	2	2	4	6	14
2	6	6	3	1	16

We will use the Wilcoxon rank-sum test. First, we pool the data from both samples together and assign a rank to each individual based on his or her rank in the combined sample. We have

Score	Frequency	Rank range	Average rank
1	8	1-8	4.5
2	8	9-16	12.5
3	7	17-23	20.0
4	7	24-30	27.0

We compute the rank sum for center 1 as follows:

$$R_1 = 2 \times 4.5 + 2 \times 12.5 + 4 \times 20.0 + 6 \times 27.0 = 276.$$

We use the normal approximation method. We have

$$E(R_1) = \frac{14(31)}{2} = 217$$
$$Var(R_1) = \frac{14(16)}{12} \times \left[31 - \frac{8(63) + 8(63) + 7(48) + 7(48)}{30(29)}\right]$$
$$= 18.667(29.069) = 542.621$$
$$sd(R_1) = 23.294$$

Thus,

$$T = \frac{|R_1 - E(R_1)| - .5}{sd(R_1)} = \frac{|276 - 217| - .5}{23.294} = \frac{58.5}{23.294} = 2.511 \sim N(0,\ 1)$$

Thus, $p = 2 \times [1 - \Phi(2.511)] = .012$. Therefore, there is a significant difference between the patient populations at the two centers, with the patients at center 1 having significantly more anxiety at baseline.

9.16 A parametric test would not be useful here because this is an ordinal scale, with 1 indicating more improvement than 2 but not 1 unit more, so that addition and subtraction are not meaningful with this scale.

9.17 The Wilcoxon rank-sum test should be used to analyze these data because we are comparing two independent groups.

9.18 We will perform the Wilcoxon rank-sum test based on these data. We compute ranks as follows:

Value	Medita-tion	Psycho-therapy	Total	Rank range	Average rank
1	3	0	3	1–3	2.0
2	4	2	6	4–9	6.5
3	7	5	12	10–21	15.5
4	3	3	6	22–27	24.5
5	2	8	10	28–37	32.5
6	1	4	5	38–42	40.0
7	0	2	2	43–44	43.5
8	0	1	1	45	45.0
9	0	1	1	46	46.0
	20	26	46		

We compute the rank sum in the meditation group:

$$R_1 = 3 \times 2 + 4 \times 6.5 + 7 \times 15.5 + 3 \times 24.5 + 2 \times 32.5 + 1 \times 40 = 319.$$

The corresponding expected value and variance of the rank sum are given as follows:

$$E(R_1) = \frac{20(47)}{2} = 470$$

$$Var(R_1) = \frac{20(26)}{12} \times \left[47 - \frac{3(3^2 - 1) + \cdots + 2(2^2 - 1)}{46(45)} \right]$$

$$= \frac{520}{12} \times \left(47 - \frac{3{,}276}{2{,}070} \right) = 1968.087$$

$$sd(R_1) = 44.36$$

The test statistic is obtained as follows:

$$T = \frac{|319 - 470| - .5}{44.36} = \frac{150.5}{44.36} = 3.392 \sim N(0,\ 1) \text{ under } H_0$$

The p-value is given by

$$2 \times [1 - \Phi(3.392)] = 2 \times (1 - .9997) < .001.$$

This result is very highly significant and indicates that there is a significant difference in the degree of improvement between the two treatment groups, with the meditation group showing more improvement.

REFERENCE

[1] Salzman, R., Manson, J. E., Griffing, G. T., Kimmerle, R., Ruderman, N., McCall, A., Stoltz, E. I., Mullin, C., Small, D., Armstrong, J., & Melby, J. C. (1985). Intranasal aerosolized insulin: Mixed-meal studies and long term use in type I diabetes. *New England Journal of Medicine*, **312**(17), 1078–1084.