Chapter 16
Nonparametrics

16.1 Introduction

Most of the statistical procedures we've introduced in the previous 15 chapters require an assumption about the form of a probability distribution, often the Normal distribution. When such assumptions are unjustified, the consequences of the procedure are dubious at best. In situations where distributional assumptions cannot be justified, even after a well-chosen data transformation, the analyst should consider another approach.

Nonparametric statistical procedures are ones that do not require the assumption of a specific probability distribution. (In contrast, procedures that do make distributional assumptions are referred to as parametric procedures.) In exchange for not requiring a detailed distributional assumption, nonparametric testing procedures have less power, and nonparametric confidence intervals are wider than their parametric analogues in the same problem situation using the same error control. In addition, the parameter(s) of interest in a nonparametric procedure may not be identical to those of corresponding parametric procedures. For example, we may use a nonparametric procedure for comparing two population medians as a substitute for a parametric procedure for comparing two means. For these reasons, parametric procedures are preferred if their assumptions are reasonably well met.

In the previous chapters we have discussed a number of procedures for inferring about one or more population means. Those procedures required assumptions that underlying populations are normally distributed, at least approximately, and that when inferring about the means of two or more populations, these populations have a common variance. Often transformations such as those discussed in Chapter 4 can be used to make such assumptions tenable once the transformations have been applied. Sometimes this is not possible, for example, when the data contain outliers whose elimination cannot be justified. In such instances, a nonparametric approach

R.M. Heiberger, B. Holland, *Statistical Analysis and Data Display*,
Springer Texts in Statistics, DOI 10.1007/978-1-4939-2122-5_16

may be considered. Our discussion here is limited to hypothesis tests. Analogous confidence interval estimates cannot be described succinctly and are discussed in Lehmann (1998) and Desu and Raghavarao (2003).

Many nonparametric procedures use statistics based on *ranks* of the data rather than the data themselves. Such procedures require only that the data be on at least an ordered scale. In contrast, parametric procedures generally carry the more stringent requirement that the data be measured on an interval or ratio scale. The various scale types are defined in Section 2.1.

We've actually encountered some nonparametric procedures in earlier chapters. For example, the two-sample Kolmogorov–Smirnov goodness-of-fit test discussed in Section 5.9 does not require that we specify the natures of the two populations being sampled. In this chapter we introduce some additional commonly used nonparametric procedures. Our examples include checks of the assumptions of competing parametric procedures and comparisons of the nonparametric and parametric results.

16.2 Sign Test for the Location of a Single Population

The example in Section 5.2.1 discusses a parametric approach to a problem involving `data(vocab)`, concerning whether $\mu = 10$ is consistent with a random sample of 54 test scores. Since Figure 5.3 shows that the sample had one high outlier, there was at least some doubt that the assumptions underlying the parametric analysis were correct.

The nonparametric approach here is the *sign test* to assess the hypothesis that the population median equals 10. If the population is not symmetric so that its mean and median are not identical, then the analogy between the nonparametric and parametric inferences is imperfect. We look again at the data in Table 16.1.

The logic of this nonparametric test stems from the insight that if the population median η is indeed 10, we would expect half of the sample values not exactly equal to 10 to fall on either side of 10. If appreciably more than half exceed 10, this suggests that the median exceeds 10. Let n be the number of sample items different from 10 and m be the number of these exceeding 10. The formal test of

$$H_0\colon \eta = 10$$

vs

$$H_1\colon \eta > 10$$

is based on the distribution of a binomial random variable X with n trials and success probability .5. The test has p-value $= P(X \geq m)$. The two-sided test, with the same null hypothesis but having alternative hypothesis

Table 16.1 $p = P(X \geq 49) = P(X \geq 48.5) = 1 - \mathcal{F}_{\text{Bi}}(48.5 \mid 50, .5) \approx 4_{10}^{-14}$

```
> data(vocab)

> table(vocab)
vocab
 9 10 11 12 13 14 15 16 17 19
 1  4 13  7  9  9  4  5  1  1

> table(vocab$score - 10)

-1  0  1  2  3  4  5  6  7  9
 1  4 13  7  9  9  4  5  1  1

> table( sign(vocab$score-10) )

-1  0  1
 1  4 49

> pbinom(48.5, 50, .5, lower=FALSE)
[1] 4.53e-14

> 1 - pbinom(48.5, 50, .5)
[1] 4.53e-14
```

$$H_1 : \eta \neq 10$$

has p-value $2P\big(X \geq \max(n - m, m)\big)$.

This test is called the sign test because it is based on the arithmetic signs of the differences between the data and the null hypothesized median. Some authors refer to it as the binomial test.

In `data(vocab)`, 4 of the 54 scores equalled the null value η=10, and n=50 scores were not exactly equal to the null value of η=10. We observed m=49 scores that exceeded 10 and $n-m$=1 score less than 10. If the null were true, then X comes from the distribution Bin(n = 50, p = .5) with discrete density shown in Figure 16.1. The one-sided p-value is the probability of observing 49 or more (that is, 49 or 50) larger scores from this distribution. This probability is the sum of the probabilities for the bars at 49 and 50. Therefore, the one-sided p-value is

$$P(X \geq 49) = 1 - \mathcal{F}_{\text{Bi}}(48 \mid 50, .5)$$

using the binomial distribution with n = 50 and p = .5. This number $p \approx 4_{10}^{-14}$ is calculated in R by

```
pbinom(48.5, 50, .5, lower.tail=FALSE)
```

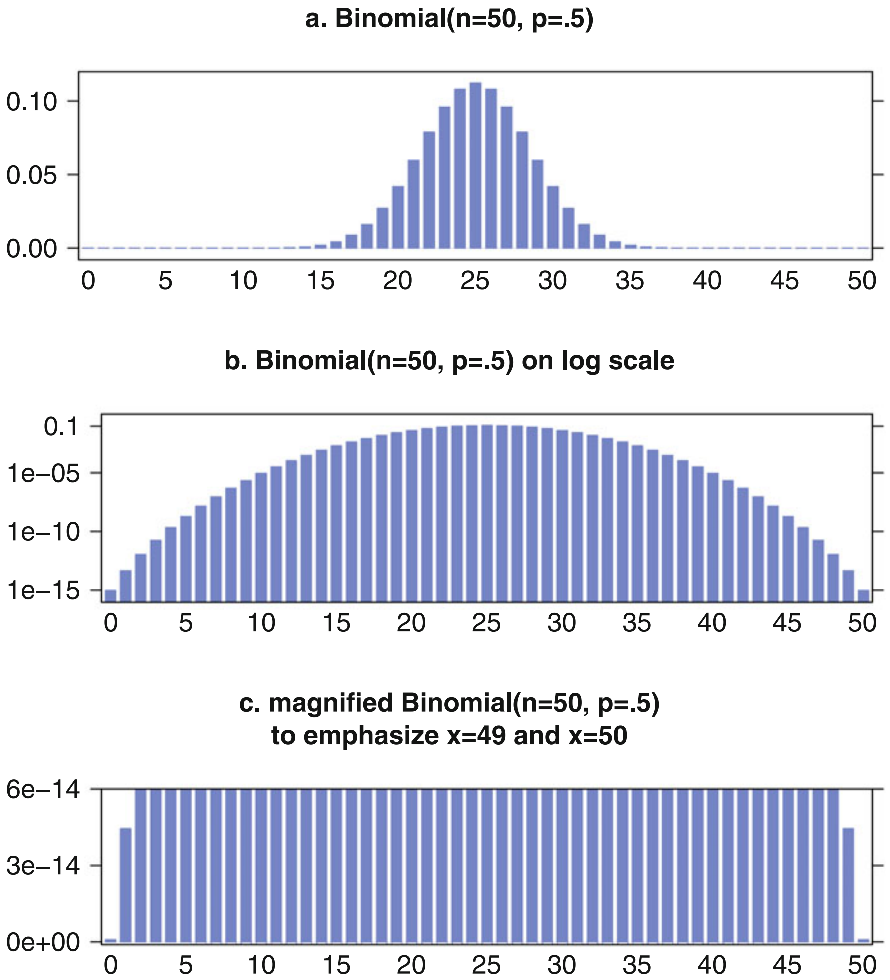

Fig. 16.1 Panel a shows the Bin($n = 50, p = .5$) discrete density. Panel b shows the same discrete density on a log scale. Panel c magnifies (and truncates) the probability scale to emphasize that there are nonzero values at $x = 49$ and $x = 50$. The p-value for the sign test is the probability that a random selection from this distribution would yield the observed $x = 49$ or the larger $x = 50$. From the graph we see that this is the very small number $p \approx 4_{10}{}^{-14}$.

$p \approx 4_{10}{}^{-14}$ is overwhelming evidence that the population median η exceeds 10, consistent with the strong parametrically based conclusion in Section 5.2.1 that the population mean μ exceeds 10.

The procedure we just developed is an example of an *exact test* or of a *randomization test*. In this form of testing procedure, a discrete model for the distribution of the observed statistic under the null hypothesis is postulated. Then the probability of obtaining the observed value (or larger) from that model is calculated and used as the p-value of the test.

16.3 Comparing the Locations of Paired Populations

Just as the parametric paired t-test is equivalent to a one-sample t-test on the pair differences, the locations of paired populations can be compared by inferring about the median of the population consisting of the differences (in a consistent direction) between the two individual populations.

16.3.1 Sign Test

Exercise 5.13 requests a comparison of the population means of pre- and post-treatment measurements based on a random sample of 48 patients. However, the density and histogram plots of the post-treatment measurements in Figure 16.2 show a number of both low and high outliers, suggesting that the post population may not be close to normally distributed, and calling into question the validity of an analysis based on a two-sample t-test. Therefore, a nonparametric approach should be considered.

We show the sign test for paired observations in Table 16.2. Analogous to Section 16.2, let n be the number of pairs of observations, excluding tied pairs. For testing against the one-sided alternative hypothesis that the post-treatment median is less than the pretreatment median, let m be the number of sample pairs where the post-treatment measurement is less than the pretreatment measurement. Then for binomially distributed X with n trials and success probability 0.5, the p-value is $P(X \leq m)$. For these data, $n = 46, m = 15$ and we use R to calculate p =`pbinom(15,46,.5)` $\approx$ .0129. This p-value indicates moderate evidence that in the population of patients from which this sample was selected, the median post-treatment angle is less than the median pretreatment angle. For the analogous parametric paired t-test requested in Exercise 5.13, the p-value is considerably less than 0.01. Because we question the validity of the t-test, we believe the p-value from the t-test is way too small.

Table 16.2 Sign test applied to relative rotation angle data. The conclusion of the test is that the median post-treatment angle is less than the median pretreatment angle.

```
> table(sign(har1$Post - har1$Pre))

-1  0  1
31  2 15

> pbinom(15, 46, .5, lower=TRUE)
[1] 0.01295
```

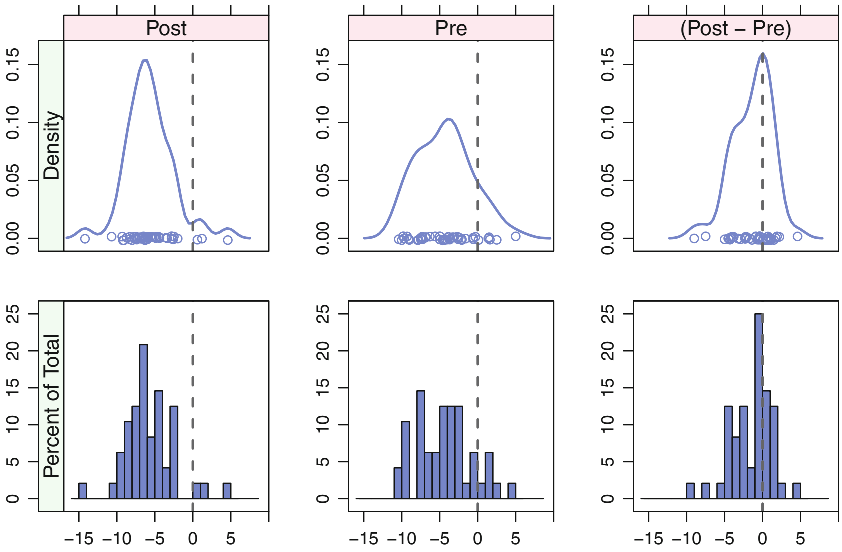

Fig. 16.2 We show the density plot of the `Post` and `Pre` scores, and their difference `(Post-Pre)`, in the top set of panels. We show the histogram of the same variables in the bottom panels. The individual jittered points in the density plots may be thought of as a vertical view of the histogram. Both graphical displays show that the data does not appear to be normally distributed. The `Post` scores have outliers are well outside the vicinity of the bulk of the data. The `Pre` scores have a relatively flat distribution. `Post` scores are generally smaller than `Pre` scores. The `(Post-Pre)` differences are predominantly negative.

16.3.2 Wilcoxon Signed-Ranks Test

The sign test in Section 16.3.1 uses the arithmetic sign of differences between members of each pair but ignores the magnitude of these differences. Such information can be crucial, particularly if the samples contain outlying values. The Wilcoxon signed-ranks test uses information on the magnitude of the differences and therefore is more sensitive than the sign test. The Wilcoxon test assumes an interval or ratio measurement scale involving continuous variables. If the data are ordinal but not interval, the Wilcoxon test cannot be used.

The construction of the signed-rank test for `data(har1)` is depicted in Table 16.3. For testing hypotheses involving the difference between the medians of two populations, say the X = `har1$Pre` population and Y = `har1$Post` population, let $D_i = X_i - Y_i$ = `har$diff` be the observed difference for the i^{th} pair. Let $n = 46$ be the number of such differences that are not zero; hereafter ignore any zero differences. Rank the n nonzero absolute differences $(|D_i|, i = 1, \ldots, n)$ = `har$abs` in ascending order from 1 to n and store the ranks in `har$rank`. (If there are ties, assign the average rank. For example, if the first and second largest absolute differences are both equal to .3, assign rank 1.5 to both absolute differences.) Copy the ranks for the positive differences into `har$prnk`. Then the test statistic is the sum of the ranks in `har$prnk`. The idea here is that if the population medians are equal, we expect half of the differences to be positive and so the test statistic should be approximately one half the sum of the ranks of the nonzero differences, or $n(n + 1)/4$, where n is now the number of nonzero differences. A test statistic much different from this value suggests that the difference between the population medians is nonzero. The signed rank distribution is discussed in Section J.3.7. We usually show the normal approximation. The variance is related to the variance of the discrete uniform distribution (see Appendix J.3.1). We outline the calculation of the variance in Exercise 16.6.

Figure 16.3 shows the differences and the ranks of the absolute values of these differences used in a Wilcoxon signed-ranks test for the data in `data(har1)`. The positive differences (between the pre- and post-treatment angles) in this example generally have larger absolute values than the negative differences. The magnitudes of these differences is relevant to the test result—a large positive difference is greater evidence that pre exceeds post than a small positive difference. These magnitudes are ignored by the sign test but accounted for by the Wilcoxon signed-ranks test.

Assuming one-sided alternative hypothesis as in Section 16.2, this test is run in R with

```
wilcox.test(X, Y, alternative="greater",
            paired=TRUE, exact=TRUE)
```

If restrictive conditions are met, R calculates the p-value based on the exact distribution of the test statistic. These conditions are $n < 50$ and no tied ranks. Otherwise, as in the analysis of the `data(har1)` dataset, the p-value is based on a normal approximation. The results of the `wilcox.test` command applied to `data(har1)` are shown in Table 16.5.

The fact that the p-value for the Wilcoxon test is much less than that of the sign test for the same data demonstrates that the signed-ranks test is more powerful than the sign test.

Table 16.3 Detail for construction of the Wilcoxon signed-ranks test applied to relative rotation angle data. The `diff` column contains the differences `Pre - Post`. The `abs` column contains the absolute value of the differences. `NA` values were assigned to the 0-valued differences. This has the effect of placing them last in the sort order. Only nonzero differences are used in the test. In this example with 48 observations, there are $n = 46$ nonzero differences. The `rank` column contains the ranks of the absolute differences. The `prnk` column contains the same ranks, but only for rows that have positive differences. The test statistic in Tables 16.5 and 16.4 is the sum of the `prnk` column.

```
> har <- data.frame(diff=har1$Pre - har1$Post)

> har$abs <- abs(har$diff)

> har$abs[har$abs==0] <- NA

> har$rank <- rank(har$abs)

> har$rank[har$diff == 0] <- 0

> har$prnk <- har$rank    ## rank for positive differences

> har$prnk[har$diff < 0] <- 0

> har[order(har$abs),]            ## manually edit into columns
   diff abs rank prnk           diff abs rank prnk
4   0.3 0.3  1.5  1.5        11 -1.9 1.9 25.0  0.0
22  0.3 0.3  1.5  1.5        33  2.1 2.1 26.0 26.0
13 -0.4 0.4  3.5  0.0        18  2.2 2.2 27.0 27.0
35  0.4 0.4  3.5  3.5        36  2.2 2.2 28.5 28.5
1   0.4 0.4  5.0  5.0        44 -2.2 2.2 28.5  0.0
17 -0.4 0.4  7.0  0.0        46  2.3 2.3 30.0 30.0
21 -0.4 0.4  7.0  0.0        25  2.3 2.3 31.0 31.0
39  0.4 0.4  7.0  7.0        5   2.9 2.9 32.0 32.0
19  0.5 0.5  9.5  9.5        43  3.2 3.2 33.0 33.0
45  0.5 0.5  9.5  9.5        6   3.2 3.2 34.0 34.0
16  0.7 0.7 11.5 11.5        23  3.8 3.8 35.0 35.0
26 -0.7 0.7 11.5  0.0        31  3.9 3.9 36.0 36.0
12 -0.7 0.7 13.0  0.0        47  4.1 4.1 37.0 37.0
15 -0.8 0.8 14.0  0.0        40  4.2 4.2 38.0 38.0
42  0.9 0.9 15.0 15.0        38  4.3 4.3 39.0 39.0
37  0.9 0.9 16.0 16.0        14  4.5 4.5 41.0 41.0
20 -1.0 1.0 17.0  0.0        24  4.5 4.5 41.0 41.0
28 -1.1 1.1 18.0  0.0        30  4.5 4.5 41.0 41.0
27 -1.1 1.1 19.0  0.0        34 -4.6 4.6 43.0  0.0
48 -1.1 1.1 20.0  0.0        29  5.0 5.0 44.0 44.0
10  1.4 1.4 21.0 21.0        8   7.5 7.5 45.0 45.0
32 -1.5 1.5 22.0  0.0        41  9.0 9.0 46.0 46.0
2   1.7 1.7 23.0 23.0        3   0.0  NA  0.0  0.0
7  -1.7 1.7 24.0  0.0        9   0.0  NA  0.0  0.0
```

Table 16.4 The sum of the positive ranks from Table 16.3 is first calculated. Then we construct the normal approximation. These values match the values in Table 16.5.

```
> ## calculate the statistic
> sum.prnk <- sum(har$prnk)

> n <- sum(har$diff != 0)

> sum.prnk
[1] 808.5

> n
[1] 46

> ## normal approximation
> mean.prnk <- n*(n+1)/4

> numerator <- sum.prnk - mean.prnk

> var.prnk1 <- n * (n + 1) * (2 * n + 1)/24

> NTIES <- table(har$abs[1:46]) ## non-zero differences

> TIEadjustment <- sum(NTIES^3 - NTIES)/48

> var.afterTIES <- var.prnk1 - TIEadjustment

> z <- (numerator - .5) / sqrt(var.afterTIES)

> z
[1] 2.923

> p.val <- pnorm(z, lower.tail = FALSE)

> p.val
[1] 0.001733
```

Table 16.5 Wilcoxon signed-ranks test applied to relative rotation angle data. The conclusion of the test is that the median post-treatment angle is less than the median pretreatment angle.

```
> wilcox.test(har1$Pre, har1$Post, alternative="greater",
+             paired=TRUE, exact=FALSE)

Wilcoxon signed rank test with continuity correction

data:  har1$Pre and har1$Post
V = 808.5, p-value = 0.001734
alternative hypothesis: true location shift is greater than 0
```

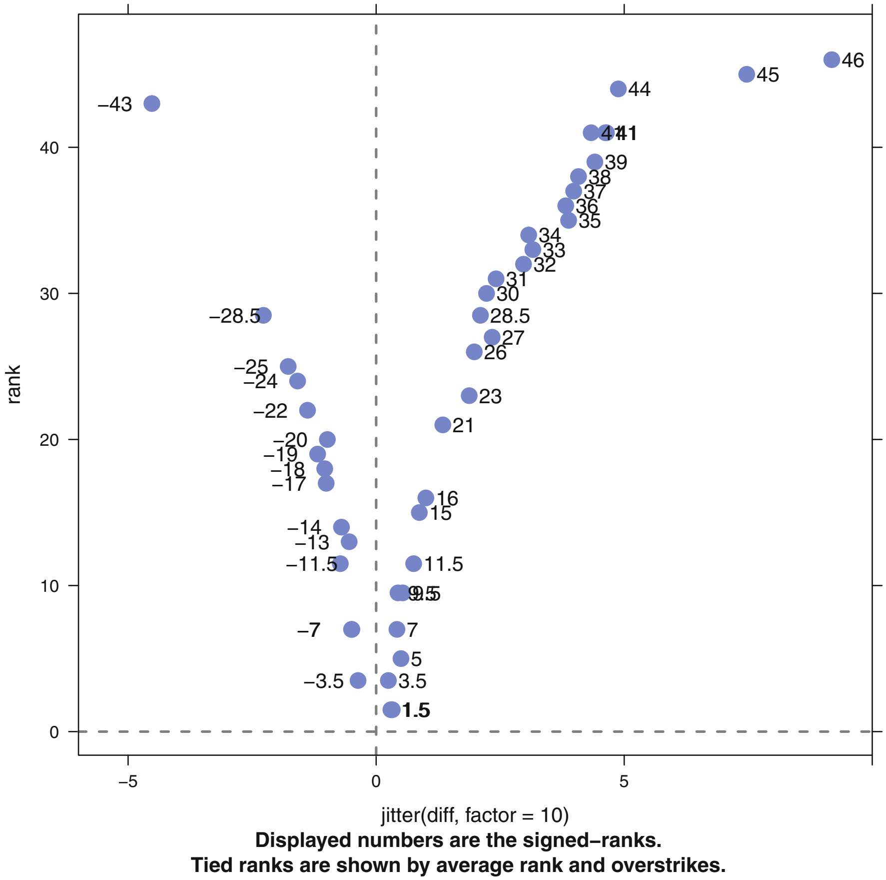

Fig. 16.3 The differences and ranks from Table 16.3 used in the Wilcoxon test of `data(har1)`. Observe that the positive differences between pre- and post-treatment angles tend to have larger magnitudes than the negative differences. The Wilcoxon test, which takes account of these magnitudes, has more power than the sign test, which ignores the magnitudes. The test statistic is the sum of the ranks for positive differences (the sum of the displayed numbers in the upper-right quadrant).

16.4 Mann–Whitney Test for Two Independent Samples

Some authors refer to this test as the Wilcoxon–Mann–Whitney test because Frank Wilcoxon initiated its development. It is analogous to the parametric two-sample t-test but compares medians rather than means. We wish to compare the medians of two populations based on independent random samples from each. It is assumed that the measurement scale is at least ordinal. Combine the two samples and then rank the resulting observations in ascending order. As in Section 16.3.2, if there are tied observations, each should be assigned the average rank. The test statistic

Fig. 16.4 Balance data. The extreme outlier in the `age=="old"` group makes the two-sample t-test invalid. See Table 16.6.

T is the sum of the ranks of the smaller sample. Assume the smaller sample is assigned index 1 and the larger sample index 2. Denoting the observed value of T as t_{calc}, the p-value depends on the form of the alternative hypothesis. If the alternative hypothesis is $\eta_1 > \eta_2$, then p-value $= P(T > t_{\text{calc}})$. For the alternative $\eta_1 < \eta_2$, the p-value is $= P(T < t_{\text{calc}})$. For the two-sided alternative, p-value $= 2\min\big(P(T > t_{\text{calc}}),\ P(T < t_{\text{calc}})\big)$.

The data in `data(balance)`, taken from Teasdale et al. (1993), illustrate this test. These authors sought to compare the forward/backward sway in mm of 9 elderly and 8 young subjects who took part in an investigation of their reaction times. We see in Figure 16.4 that one of the `elderly` measurements is an extreme outlier, and if it is retained in the analysis the two-sample t-test is invalid. The alternative hypothesis is that the median sway of elderly subjects exceeds the median sway of young subjects. The R analysis uses `wilcox.test` as in Section 16.3.2.

In Table 16.6 we list the two samples and their ranks after they are combined. The analysis with the R function `wilcox.test` is in Table 16.7 and "by hand" with R in Table 16.8. The Wilcoxon distribution is described in Section J.3.8. We conclude that there is moderate evidence that, on average, elderly subjects have greater sway than young subjects.

Table 16.6 Mann–Whitney test applied to balance data. See Figure 16.4.

old		young	
sway	ranks	sway	ranks
19	6.5	25	13.5
30	16.0	21	9.5
20	8.0	17	4.5
19	6.5	15	3.0
29	15.0	14	1.5
25	13.5	14	1.5
21	9.5	22	11.0
24	12.0	17	4.5
50	17.0		

49.0 = rank sum, smaller sample

Table 16.7 Mann–Whitney test applied to balance data by R. R does not compute the exact test when some of the ranks are tied. The ranks are plotted in Figure 16.5.

```
> wilcox.test(balance$sway[balance$age=="young"],
+             balance$sway[balance$age=="old"],
+             alternative="less", exact=FALSE)

Wilcoxon rank sum test with continuity correction

data:  balance$sway[balance$age == "young"] and
       balance$sway[balance$age == "old"]
W = 13, p-value = 0.01494
alternative hypothesis: true location shift is less than 0
```

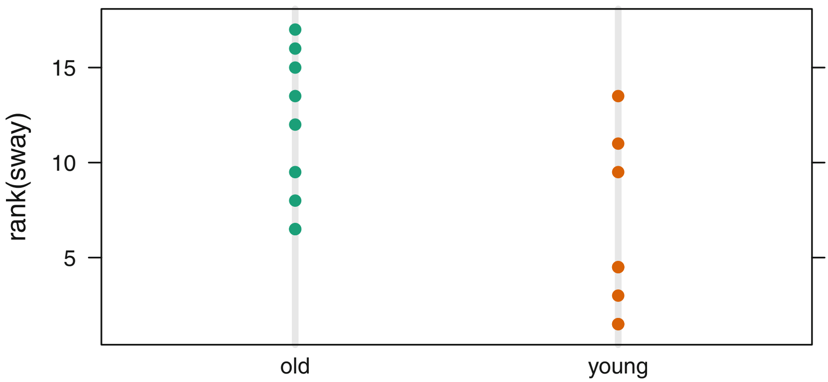

Fig. 16.5 Balance data. The plot of the ranks shows a distinct difference in the medians ranks of the two age groups.

Table 16.8 Mann–Whitney test applied "by hand" to balance data.

```
> data(balance)

> unname(n.old <- table(balance$age)["old"])
[1] 9

> unname(n.young <-  table(balance$age)["young"])
[1] 8

> (r <- rank(balance$sway))
 [1]  6.5 16.0  8.0  6.5 15.0 13.5  9.5 12.0 17.0 13.5  9.5
[12]  4.5  3.0  1.5  1.5 11.0  4.5

> (NTIES <- table(r))
r
 1.5    3  4.5  6.5    8  9.5   11   12 13.5   15   16   17
   2    1    2    2    1    2    1    1    2    1    1    1

> unname(tie.adjustment <-
+          sum(NTIES^3 - NTIES) /
+          ((n.old + n.young) * (n.old + n.young - 1)))
[1] 0.1103

> unname(SIGMA <-
+          sqrt((n.old * n.young/12) *
+               ((n.old + n.young + 1) - tie.adjustment)))
[1] 10.36

> unname(STATISTIC.old <-
+          c(W = sum(r[1:9]) - n.old * (n.old + 1)/2))
[1] 59

> unname(z.old <- STATISTIC.old - n.old * n.young/2)
[1] 23

> unname(z.old <- (z.old - .5)/SIGMA)
[1] 2.172

> unname(pnorm(z.old, lower.tail = FALSE))
[1] 0.01494

> unname(STATISTIC.young <-
+          c(W = sum(r[10:17]) - n.young * (n.young + 1)/2))
[1] 13

> unname(z.young <- STATISTIC.young - n.old * n.young/2)
[1] -23

> unname(z.young <- (z.young + .5)/SIGMA)
[1] -2.172

> unname(pnorm(z.young))
[1] 0.01494
```

16.5 Kruskal–Wallis Test for Comparing the Locations of at Least Three Populations

This test is the nonparametric analogue of the F-test for equality of the means of several populations. A natural generalization of the Mann–Whitney test, it tests the null hypothesis that all k populations have a common median. It is assumed that the measurement scale is at least ordered and that the k random samples are mutually independent.

Suppose n_i is the size of the sample from population i and $n. = \sum_i n_i$. Rank all $n.$ observations from 1 to $n.$ and let R_i be the sum of the ranks from sample i. As usual, in case of ties, assign average ranks to the tied values. The test statistic is

$$T = \frac{12}{n.(n. + 1)} \sum_{i=1}^{k} \frac{[R_i - n_i(n. + 1)/2]^2}{n_i} \tag{16.1}$$

The idea behind this formula is that if the null hypothesis is exactly true, the expected sum of the ranks of sample i is $E(R_i) = n_i(n. + 1)/2$.

This test is conducted in R with the command

```
kruskal.test(y ~ g, dataframe)
```

where `y` is the numeric vector of sample observations and `g` is the same sized factor indicating the population number from which the observation came. As usual `dataframe` is the `data.frame` containing the observations.

Exercise 6.2 refers to an experiment comparing the pulse rates of workers while performing six different tasks `data(pulse)`. From Figure 16.6 we see that the normality assumption required for a standard one-way analysis of variance is somewhat questionable as most tasks seem to show a uniform distribution of pulses. We therefore investigate the data analysis via the Kruskal–Wallis test.

Table 16.9 Kruskal–Wallis rank sum test applied to pulse rate data by R.

```
> kruskal.test(pulse ~ task, data=pulse)

Kruskal-Wallis rank sum test

data:  pulse by task
Kruskal-Wallis chi-squared = 16, df = 5, p-value = 0.007
```

Table 16.9 shows the chi-square approximation to the distribution of the Kruskal–Wallis statistic yields a p-value of .007. This is not too dissimilar from the one-way

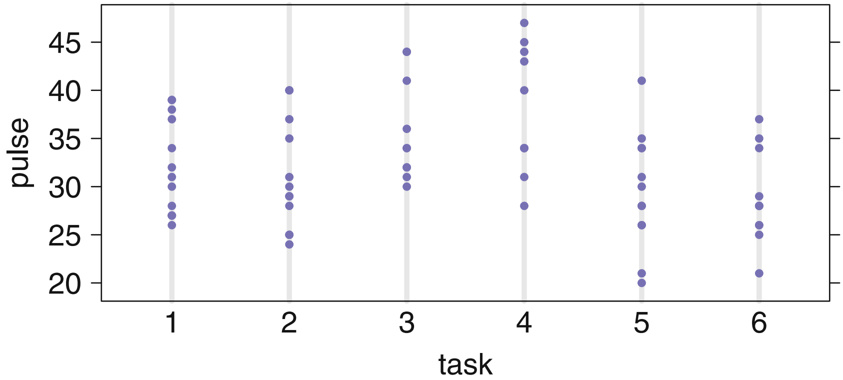

Fig. 16.6 Pulse data. The uniform distributions within the groups make the one-way ANOVA inappropriate. See Table 16.9.

ANOVA p-value of .0015. We conclude, unsurprisingly, that the pulse rate of workers differs according to the task performed immediately prior to the pulse reading.

16.6 Exercises

16.1. A study of Darwin (1876), discussed in Hand et al. (1994) with the data in `data(darwin)`, compared the ultimate heights of plants grown from otherwise comparable seedlings that were either cross-fertilized or self-fertilized. Compare the heights using the Wilcoxon signed-ranks procedure. Would a paired t-test have been appropriate?

16.2. High levels of carbon monoxide transfer are a risk factor for contracting pneumonia. Ellis et al. (1987), also in Hand et al. (1994), studied the levels of carbon monoxide transfer in 7 chicken pox patients who were smokers. They were measured upon hospital admission and one week later. The data are in the file `data(pox)`.

a. Verify the inappropriateness of a paired t-test for these data. Discuss your reasoning.
b. Analyze using the Wilcoxon signed-ranks procedure.

16.3. Simpson et al. (1975), also in Chambers et al. (1983), studied the amount of rainfall (measured in acre-feet) following cloudseeding. They seeded 26 clouds with silver nitrate and 26 clouds with a placebo seeding (that is, the airplane went up but didn't release the silver nitrate). The data are contained in the file `data(seeding)`. Assume these data represent independent random samples.

a. Use the Mann–Whitney test to assess whether cloudseeding impacted rainfall.

b. Construct the ladder of powers graph (see Figure 4.19) to find a power transformation that makes the histograms of post-transformed samples symmetric and bell-shaped. Redo part a using the two-sample t-test instead of the Mann–Whitney test.

c. Discuss which of these procedures is preferred.

16.4. VanVliet and Gupta (1973), later cited by Hand et al. (1994), compared the birthweights in kg of 50 infants having severe idiopathic respiratory distress syndrome. Twenty-seven (27) of these infants subsequently died and the remainder lived. The data appear in the file `data(distress)`. Perform a nonparametric test to address whether deaths from this cause are associated with low birthweight.

16.5. Exercise 6.8 requested a comparison of the mean disintegration times of four types of pharmaceutical tablets. The data are in file `data(tablet1)`. Compare the results of a Kruskal–Wallis test with the conclusion of the F-test for that exercise.

16.6. Calculate the variance of the null distribution for the Wilcoxon signed-ranks statistic, assuming no ties. The set of all signed ranks, not just the positive values, in the null distribution consists of the integers $\{-n$ to $-1, 1$ to $n\}$. This has mean 0 and variance proportional to the $\sum i^2$. The variance of just the positive values is 1/4 times the variance of all of them. The mean of just the positive signed ranks is proportional to $\sum i$.