Chapter 3

The One-Sample Location Problem

INTRODUCTION

The procedures of this chapter are designed for statistical analyses in which primary interest is centered on the location (median) of a population. We encounter two types of data for which such analyses are important. The first of these, referred to as *paired replicates data*, represents pairs of "pretreatment" and "posttreatment" observations; here, we are concerned with a shift in location due to the application of the "treatment." The second type of data, referred to as *one-sample data*, consists of observations from a single population about whose location we wish to make inferences.

In Sections 3.1–3.3, procedures are considered for analyzing paired replicates data using signed ranks. In particular, Section 3.1 presents a distribution-free signed rank test; Section 3.2, a point estimator associated with the signed rank statistic; and Section 3.3, a related distribution-free confidence interval. In Section 3.7, these procedures are applied to some one-sample data. An asymptotically distribution-free test for symmetry of the underlying population (one of the assumptions in Sections 3.1–3.3 and 3.7) is considered in Section 3.9. A distribution-free test for exchangeability of the paired replicates data is discussed in Section 3.10.

Procedures for analyzing paired replicates data using signs are discussed in Sections 3.4–3.6. A distribution-free sign test is considered in Section 3.4, a point estimator associated with the sign statistic in Section 3.5, and a related distribution-free confidence interval in Section 3.6. These sign procedures are applied to some one-sample data in Section 3.8.

The asymptotic relative efficiencies for translation alternatives of the procedures based on the signed rank statistic and those based on the sign statistic with respect to their normal theory counterparts based on the sample mean are discussed in Section 3.11.

PAIRED REPLICATES ANALYSES BY WAY OF SIGNED RANKS

Data. We obtain $2n$ observations, two observations on each of n subjects (blocks, patients, etc.).

Nonparametric Statistical Methods, Third Edition. Myles Hollander, Douglas A. Wolfe, Eric Chicken.

Subject i	X_i	Y_i
1	X_1	Y_1
2	X_2	Y_2
.	.	.
.	.	.
.	.	.
n	X_n	Y_n

Assumptions

A1. We let $Z_i = Y_i - X_i$, for $i = 1, \ldots, n$. The differences $Z_1, \ldots, Z_n$ are mutually independent.

A2. Each $Z, i = 1, \ldots, n$, comes from a continuous population (not necessarily the same one) that is symmetric about a common median θ. If F_i represents the distribution function for $Z_i, i = 1, \ldots n$, this assumption requires that

$$F_i(\theta + t) + F_i(\theta - t) = 1, \text{ for every } t \text{ and } i = 1, \ldots, n.$$

The parameter θ is referred to as the *treatment effect.*

3.1 A DISTRIBUTION-FREE SIGNED RANK TEST (WILCOXON)

Hypothesis

The null hypothesis of interest here is that of zero shift in location due to the treatment, namely,

$$H_0 : \theta = 0. \tag{3.1}$$

This null hypothesis asserts that each of the distributions (not necessarily the same) for the differences (posttreatment minus pretreatment observations) is symmetrically distributed about 0, corresponding to no shift in location due to the treatment.

Procedure

To compute the Wilcoxon signed rank statistic T^+, form the absolute values $|Z_1|, \ldots, |Z_n|$ of the differences and order them from least to greatest. Let R_i denote the rank of $|Z_i|, i = 1, \ldots, n$, in this ordering. Define indicator variables $\psi_i, i = 1, \ldots, n$, where

$$\psi_i = \begin{cases} 1, & \text{if } Z_i > 0, \\ 0, & \text{if } Z_i < 0, \end{cases} \tag{3.2}$$

and obtain the n products $R_1\psi_1, \ldots, R_n\psi_n$. The product $R_i\psi_i$ is known as the *positive signed rank of* Z_i. It takes on the value zero if Z_i is negative and is equal to the rank of $|Z_i|$ when Z_i is positive. The Wilcoxon signed rank statistic T^+ is then the sum of the positive signed ranks, namely,

$$T^+ = \sum_{i=1}^{n} R_i\psi_i. \tag{3.3}$$

a. *One-Sided Upper-Tail Test.* To test

$$H_0 : \theta = 0$$

versus

$$H_1 : \theta > 0$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } T^+ \geq t_\alpha; \text{ otherwise do not reject,} \tag{3.4}$$

where the constant t_α is chosen to make the type I error probability equal to α.

b. *One-Sided Lower-Tail Test.* To test

$$H_0 : \theta = 0$$

versus

$$H_2 : \theta < 0$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } T^+ \leq \frac{n(n+1)}{2} - t_\alpha; \text{ otherwise do not reject.} \tag{3.5}$$

c. *Two-Sided Test.* To test

$$H_0 : \theta = 0$$

versus

$$H_3 : \theta \neq 0$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } T^+ \geq t_{\alpha/2} \text{ or } T^+ \leq \frac{n(n+1)}{2} - t_{\alpha/2}; \text{ otherwise do not reject.} \tag{3.6}$$

This two-sided procedure is the two-sided symmetric test with $\alpha/2$ probability in each tail of the null distribution of T^+.

The tests can be performed using the `R` command `wilcox.test` (see Example 3.1). The t_α critical values can be obtained from the `R` command `psignrank` (see Comment 5).

Large-Sample Approximation

The large-sample approximation is based on the asymptotic normality of T^+, suitably standardized. We first need to know the expected value and variance of T^+ when the null hypothesis is true. When H_0 is true, the expected value and variance of T^+ are

$$E_0(T^+) = \frac{n(n+1)}{4} \tag{3.7}$$

and

$$\text{var}_0(T^+) = \frac{n(n+1)(2n+1)}{24}, \tag{3.8}$$

respectively. These expressions for $E_0(T^+)$ and $\text{var}_0(T^+)$ are verified by direct calculations in Comment 6 for the special case of $n = 3$. General derivations of both expressions are presented in Comment 7.

The standardized version of T^+ is

$$T^* = \frac{T^+ - E_0(T^+)}{\{\text{var}_0(T^+)\}^{1/2}} = \frac{T^+ - \left\{\dfrac{n(n+1)}{4}\right\}}{\{n(n+1)(2n+1)/24\}^{1/2}}. \tag{3.9}$$

When H_0 is true, T^* has, as n tends to infinity, an asymptotic $N(0,1)$ distribution (see Comment 7 for indications of the proof). The normal theory approximation for procedure (3.4) is

$$\text{Reject } H_0 \text{ if } T^* \geq z_\alpha; \text{ otherwise do not reject;} \tag{3.10}$$

the normal theory approximation for procedure (3.5) is

$$\text{Reject } H_0 \text{ if } T^* \leq -z_\alpha; \text{ otherwise do not reject;} \tag{3.11}$$

and the normal theory approximation for procedure (3.6) is

$$\text{Reject } H_0 \text{ if } |T^*| \geq z_{\alpha/2}; \text{ otherwise do not reject.} \tag{3.12}$$

Ties

If there are zero values among the Z's, discard the zero values and redefine n to be the number of nonzero Z's. If there are ties among the (nonzero) $|Z|$'s, assign each of the observations in a tied group the average of the integer ranks that are associated with the tied group. After computing T^+ with these average ranks for nonzero Z's, use procedure (3.4), (3.5), or (3.6). Note, however, that this test associated with tied $|Z|$'s is only approximately, and not exactly, of significance level α. (To get an exact level α test even in this tied setting, see Comment 11.)

When applying the large-sample approximation, an additional factor must be taken into account. Although ties in the nonzero $|Z|$'s do not affect the null expected value of T^+, its null variance is reduced to

$$\text{var}_0(T^+) = (24)^{-1}\left[n(n+1)(2n+1) - \frac{1}{2}\sum_{j=1}^{g} t_j(t_j-1)(t_j+1)\right], \tag{3.13}$$

where g denotes the number of tied groups of nonzero $|Z|$'s and t_j is the size of the tied group j. We note that an untied observation is considered to be a tied "group" of size 1. In particular, if there are no ties among the $|Z|$'s, then $g = n$ and $t_j = 1$ for $j = 1, \ldots, n$. In this case, each term in (3.13) of the form $t_j(t_j - 1)(t_j + 1)$ reduces to zero, and the variance expression in (3.13) reduces to the usual null variance of T^+ when there are no ties, as given in (3.8). Note that the term $(48)^{-1}\sum_{j=1}^{g} t_j(t_j-1)(t_j+1)$ represents the reduction in the null variance of T^+ due to the presence of tied nonzero Z's.

As a consequence of the effect that ties have on the null variance of T^+, the following modification is needed to apply the large-sample approximation when there are tied nonzero Z's. Compute T^+ using average ranks and set

$$T^* = \frac{T^+ - \left\{\dfrac{n(n+1)}{4}\right\}}{\{\text{var}_0(T^+)\}^{1/2}}, \tag{3.14}$$

where $\text{var}_0(T^+)$ is now given by display (3.13). With this modified value of T^*, approximations (3.10), (3.11), or (3.12) can be applied.

EXAMPLE 3.1 ***Hamilton Depression Scale Factor IV.***

The data in Table 3.1 are a portion of the data obtained by Salsburg (1970). These data, based on nine patients who received tranquilizer T, were taken from a double-blind clinical trial involving two tranquilizers. The measure used was the Hamilton (1960) depression scale factor IV (the "suicidal" factor). The X (pre) value was obtained at the first patient visit after initiation of therapy, whereas the Y (post) value was obtained at the second visit after initiation of therapy. The patients had been diagnosed as having mixed anxiety and depression.

In this example, an improvement due to tranquilizer T corresponds to a reduction in factor IV values. Hence, we apply test (3.5), which is designed to detect the alternative $\theta < 0$. One obtains the value of T^+ by first calculating the nine $Z_i = Y_i - X_i$ differences, then ranking from least to greatest the nine absolute values $|Z_1|, \ldots, |Z_9|$, and finally adding the ranks of the $|Z|$'s that emanated from positive Z difference.

To perform the test using R, set

```
pre<-c(1.83, .50, 1.62, 2.48, 1.68, 1.88, 1.55, 3.06, 1.30),

post<-c(.878, .647, .598, 2.05, 1.06, 1.29, 1.06, 3.14, 1.29).
```

Then apply `wilcox.test(pre,post, paired=TRUE, alterative = "less")` to obtain $T^+ = 5$ with a P-value of .02.

Table 3.1 The Hamilton Depression Scale Factor IV Values

Patient i	X_i	Y_i
1	1.83	0.878
2	0.50	0.647
3	1.62	0.598
4	2.48	2.05
5	1.68	1.06
6	1.88	1.29
7	1.55	1.06
8	3.06	3.14
9	1.30	1.29

Source: D. S. Salsburg (1970).

i	Z_i	$\lvert Z_i \rvert$	R_i	ψ_i	$R_i \psi_i$
1	−0.952	0.952	8	0	0
2	0.147	0.147	3	1	3
3	−1.022	1.022	9	0	0
4	−0.430	0.430	4	0	0
5	−0.620	0.620	7	0	0
6	−0.590	0.590	6	0	0
7	−0.490	0.490	5	0	0
8	0.080	0.080	2	1	2
9	−0.010	0.010	1	0	0
					$T^+ = 5$

For the large-sample approximation, we find (since there are no ties) from (3.9) that

$$T^* = \frac{5 - (9(10)/4)}{\{9(10)(19)/24\}^{1/2}} = -2.07.$$

From `pnorm(-2.07)=.0192`, the smallest significance level at which we can reject H_0 in favor of $\theta < 0$ using the normal approximation is .0192. Both the exact test and the large-sample approximation indicate that there is strong evidence that tranquilizer T does lead to patient improvement, as measured by a reduction in the Hamilton scale factor IV values.

EXAMPLE 3.2 ***Government versus Private Sector Salaries.***

In an annual survey to determine whether federal pay scales were commensurate with private sector salaries, government and private workers were matched as closely as possible (with respect to type of job, educational background, years experience, etc.) and the salaries of the matched pairs were obtained. The data in Table 3.2 are the annual salaries (in dollars) for 12 such matched pairs, as reported by McClave and Benson (1978).

Letting X correspond to the government worker's salary and Y to the matched private sector salary, the tabular presentation of the associated positive signed ranks (using average ranks to break ties) is as follows:

Table 3.2 Annual Salaries

Pair i	Private	Government
1	12,500	11,750
2	22,300	20,900
3	14,500	14,800
4	32,300	29,900
5	20,800	21,500
6	19,200	18,400
7	15,800	14,500
8	17,500	17,900
9	23,300	21,400
10	42,100	43,200
11	16,800	15,200
12	14,500	14,200

Source: J. T. McClave and G. Benson (1978).

i	z_i	$\lvert Z_i \rvert$	R_i	ψ_i	$R_i\psi_i$
1	750	750	5	1	5
2	1400	1400	9	1	9
3	−300	300	1.5	0	0
4	2400	2400	12	1	12
5	−700	700	4	0	0
6	800	800	6	1	6
7	1300	1300	8	1	8
8	−400	400	3	0	0
9	1900	1900	11	1	11
10	−1100	1100	7	0	0
11	1600	1600	10	1	10
12	300	300	1.5	1	1.5

To test H_0 versus the alternative that government workers are generally paid less than their counterparts in the private sector, we use the signed rank test of $H_0 : \theta = 0$ versus $H_0 : \theta > 0$. From the signed rank computational array, we see that

$$T^+ = 5 + 9 + 12 + 6 + 8 + 11 + 10 + 1.5 = 62.5.$$

Using the R command `psignrank(62,12,lower.tail=F)`, we find that the smallest significance level at which these data lead to rejection of $H_0 : \theta = 0$ in favor of $H_1 : \theta > 0$ (i.e., the one-sided P-value) is $\alpha = .0320$. Hence, there is moderate evidence to indicate that federal government workers (at least in the type of jobs considered in this survey) are, indeed, paid less than their private sector counterparts. (We point out that the P-value for these data is only approximate, due to the tied \$300 absolute differences. For a discussion of how to obtain the exact conditional P-value in this case, see Comment 11.)

For the normal approximation with the data in Table 3.2, we need to use the ties-corrected version of T^* given in (3.14). For the salary data, we have $g = 11$ and (arbitrarily labeling the tied groups in the order of increasing ranks) $t_1 = 2, t_2 = t_3 = \ldots = t_{10} = t_{11} = 1$. Using the ties-corrected formula (3.13) for $\text{var}_0(T^+)$, we obtain

$$T^* = \frac{62.5 - \dfrac{12(13)}{4}}{\left\{\dfrac{12(12+1)(2(12)+1) - \frac{1}{2}(2)(1)(3)}{24}\right\}^{1/2}} = \frac{62.5 - 39}{\left\{\dfrac{3897}{24}\right\}^{1/2}} = 1.84.$$

To find the P-value associated with this normal approximation, we obtain `1-pnorm(1.84)=.0329`, which is in good agreement with the value of .0320 obtained without using the normal approximation.

Comments

1. *Motivation for the Test.* When θ is greater than 0, there will tend to be a large proportion of positive Z differences and they will tend to have the larger absolute values. Hence, when θ is greater than 0, we would expect a higher proportion of positive signed ranks with relatively large sizes, leading to a big value of T^+. This suggests rejecting H_0 in favor of $\theta > 0$ for large values of T^+ and motivates procedures (3.4) and (3.10). Similar rationales lead to procedures (3.5), (3.6), (3.11), and (3.12).

2. *Assumptions.* There is no requirement that the individual X_i and Y_i be independent, only that the pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, and therefore the resulting differences $Z_1, \ldots, Z_n$, be mutually independent. Indeed, in most applications, the individual X_i and Y_i are dependent. For paired replicates data, the symmetry part of Assumption A2 is often inherently satisfied. In particular, if each X_i and $Y_i, i = 1, \ldots, n$, arise from populations differing only in location (i.e., the only treatment "effect" is a change in location), then the $(Z_i - \theta)$'s come from populations that are symmetric about zero. (This is, in fact, true under more general conditions.)
3. *Testing θ Equal to Some Specified Nonzero Value.* Procedures (3.4), (3.5), and (3.6) and the corresponding normal approximations (3.10), (3.11), and (3.12) are for testing θ equal to zero. To test $\theta = \theta_0$, where θ_0 is some specified nonzero number, subtract θ_0 from each of the differences $Z_1, \ldots, Z_n$ to form a modified sample $Z_1' = Z_1 - \theta_0, \ldots, Z_n' = Z_n - \theta_0$. Then compute T^+ as the sum of the positive signed ranks for these Z_i's. Procedures (3.4), (3.5), and (3.6) and their corresponding large-sample approximations (3.10), (3.11), and (3.12) are then applied as previously described.
4. *Equivalent Form.* It may appear that some of the information in the ranking of the sample Z-differences is being lost by using only the positive signed ranks to compute T^+. Such is not the case. If we define T^- to be the sum of ranks (of the absolute values) corresponding to the negative Z observations, then $T^- = \sum_{i=1}^{n}(1 - \psi_i)R_i$. It follows that $T^+ + T^- = \sum_{i=1}^{n} R_i = n(n+1)/2$. Thus, the test procedures defined in procedures (3.4), (3.5), and (3.6) and the corresponding approximations (3.10), (3.11), and (3.12) could equivalently be based on $T^- = [n(n+1)/2] - T^+$.
5. *Derivation of the Distribution of T^+ under H_0 (No Ties Case).* Let B be the number of positive Z's and let $r_1 < \cdots < r_B$ denote the ordered ranks of the absolute values of these positive Z's. Then the null (H_0) distribution can be obtained directly from the representation $T^+ = \sum_{i=1}^{B} r_i$. Under the assumption that the underlying Z_i distributions are all continuous, the probabilities are zero that there are ties among the absolute values of the Z's or that any of the Z's are exactly zero. In addition, under H_0, these underlying Z_i distributions are all symmetric about $\theta = 0$. It follows that under H_0, each of the 2^n possible outcomes for the ordered configuration $(r_1, \ldots, r_B)$ occurs with equal probability $\left(\frac{1}{2}\right)^n$. For example, in the case of $n = 3$, the $2^3 = 8$ possible outcomes for $(r_1, \ldots, r_B)$ and associated values of T^+ are given in the following table.

B	$(r_1, r_2, \ldots, r_B)$	Probability under H_0	$T^+ = \sum_{i=1}^{B} r_i$
0		$\frac{1}{8}$	0
1	$r_1 = 1$	$\frac{1}{8}$	1
1	$r_1 = 2$	$\frac{1}{8}$	2
1	$r_1 = 3$	$\frac{1}{8}$	3
2	$r_1 = 1, r_2 = 2$	$\frac{1}{8}$	3
2	$r_1 = 1, r_2 = 3$	$\frac{1}{8}$	4
2	$r_1 = 2, r_2 = 3$	$\frac{1}{8}$	5
3	$r_1 = 1, r_2 = 2, r_3 = 3$	$\frac{1}{8}$	6

Thus, for example, the probability is $\frac{2}{8}$ under H_0 that T^+ is equal to 3, since $T^+ = 3$ when either of the exclusive outcomes $B = 1,\ r_1 = 3$ or $B = 2, (r_1 = 1, r_2 = 2)$ occurs and each of these outcomes has null probability $\frac{1}{8}$. Simplifying, we obtain the null distribution.

Possible value of T^+	Probability under H_0
0	$\frac{1}{8}$
1	$\frac{1}{8}$
2	$\frac{1}{8}$
3	$\frac{2}{8}$
4	$\frac{1}{8}$
5	$\frac{1}{8}$
6	$\frac{1}{8}$

The probability, under H_0, that T^+ is greater than or equal to 5, for example, is therefore

$$P_0(T^+ \geq 5) = P_0(T^+ = 5) + P_0(T^+ = 6)$$
$$= .125 + .125 = .25.$$

This agrees with what is obtained from `psignrank(4,3,lower.tail=F)` which gives, for $n = 3$, the probability under the null hypothesis that $T^+ > 4$.

Note that we have derived the null distribution of T^+ without specifying the forms of the underlying Z populations under H_0 beyond the point of requiring that they be continuous and symmetric about zero. This is why the test procedures based on T^+ are called *distribution-free procedures*. From the null distribution of T^+ we can determine the critical value t_α and control the probability α of falsely rejecting H_0 when H_0 is true, and this error probability does not depend on the specific forms of the underlying continuous and symmetric (about 0) Z distributions.

6. *Calculation of the Mean and Variance of T^+ under the Null Hypothesis H_0.* In displays (3.7) and (3.8) we presented formulas for the mean and variance of T^+ when the null hypothesis is true. In this comment, we illustrate a direct calculation of $E_0(T^+)$ and $\text{var}_0(T^+)$ in the particular case of $n = 3$, using the null distribution of T^+ obtained in Comment 5. (Later, in Comment 7, we present general derivations of $E_0(T^+)$ and $\text{var}_0\ (T^+)$.) The null mean, $E_0(T^+)$, is obtained by multiplying each possible value of T^+ with its probability under H_0. Thus,

$$E_0(T^+) = 0(.125) + 1(.125) + 2(.125) + 3(.25) + 4(.125) + 5(.125)$$
$$+ 6(.125) = 3.$$

This is in agreement with what we obtain using (3.7), namely,

$$E_0(T^+) = \frac{n(n+1)}{4} = \frac{3(3+1)}{4} = 3.$$

A check on the expression for $\text{var}_0(T^+)$ is also easily performed, using the well-known fact that

$$\text{var}_0(T^+) = E_0[(T^+)^2] - \{E_0(T^+)\}^2.$$

The value of $E_0[(T^+)^2]$, the second moment of the null distribution of T^+, is again obtained by multiplying possible values (in this case, values of $(T^+)^2$) by the corresponding probabilities under H_0. We find

$$E_0[(T^+)^2] = [(0+1+4)(.125) + 9(.25) + (16+25+36)(.125)] = 12.5.$$

Thus,

$$\text{var}_0(T^+) = 12.5 - (3)^2 = 3.5,$$

which agrees with what we obtain using (3.8) directly, namely,

$$\text{var}_0(T^+) = \frac{3(3+1)(2(3)+1)}{24} = 3.5.$$

7. *Large-Sample Approximation.* In view of the representation $T^+ = \sum_{i=1}^{B} r_i$, it follows from the discussion in Comment 5 that $T^+ \underline{\underline{d}} \sum_{i=1}^{n} V_i$, where the symbol $\underline{\underline{d}}$ means "has the same distribution as" and $V_1, \ldots, V_n$ are mutually independent dichotomous random variables with probability distributions

$$P(V_i = i) = P(V_i = 0) = \frac{1}{2},$$

for $i = 1, \ldots, n$. From this distributionally equivalent form, we can immediately use well-known expressions for the mean and variance of a sum of mutually independent random variables to obtain

$$E_0(T^+) = E\left[\sum_{i=1}^{n} V_i\right] = \sum_{i=1}^{n} E[V_i] \tag{3.15}$$

and

$$\text{var}_0(T^+) = \text{var}\left(\sum_{i=1}^{n} V_i\right) = \sum_{i=1}^{n} \text{var}(V_i). \tag{3.16}$$

Since V_i is a dichotomous variable, we have, for $i = 1, \ldots, n$, that

$$E_0(V_i) = i\left(\frac{1}{2}\right) + 0\left(\frac{1}{2}\right) = \frac{i}{2}$$

and

$$\text{var}_0(V_i) = E_0(V_i^2) - [E_0(V_i)]^2 = \left[i^2\left(\frac{1}{2}\right) + 0^2\left(\frac{1}{2}\right)\right] - \left[\frac{i}{2}\right]^2$$
$$= \frac{i^2}{2} - \frac{i^2}{4} = \frac{i^2}{4}.$$

Using these results, along with the closed-form expressions for the sum of the first n positive integers and the sum of the squares of the first n positive integers, in (3.15) and (3.16), we obtain

$$E_0(T^+) = \frac{1}{2}\sum_{i=1}^{n} i = \frac{1}{2}\left[\frac{n(n+1)}{2}\right] = \frac{n(n+1)}{4}$$

and

$$\text{var}_0(T^+) = \frac{1}{4}\sum_{i=1}^{n} i^2 = \frac{1}{4}\left[\frac{n(n+1)(2n+1)}{6}\right] = \frac{n(n+1)(2n+1)}{24},$$

which agree with the general expressions stated in (3.7) and (3.8), respectively.

Also using the distributional equality between T^+ and $\sum_{i=1}^{n} V_i$, the asymptotic normality of the standardized form

$$T^* = \frac{T^+ - E_0(T^+)}{\{\text{var}_0(T^+)\}^{1/2}} = \frac{T^+ - \dfrac{n(n+1)}{4}}{\left\{\dfrac{n(n+1)(2n+1)}{24}\right\}^{1/2}}$$

follows from standard theory for sums of mutually independent, but not identically distributed, random variables, such as the Liapounov central limit theorem (cf. Randles and Wolfe (1979, p. 423)). Asymptotic normality results are also obtainable under general alternatives to H_0. See, for example, the Hoeffding (1948a) U-statistic theorem as stated and applied to the Wilcoxon signed rank statistic on pages 82–85 of Randles and Wolfe (1979).

8. *Symmetry of the Distribution of T^+ under the Null Hypothesis.* When H_0 is true, the distribution of T^+ is symmetric about its mean $n(n+1)/4$. (See Comment 5 for verification of this when $n = 3$.) This implies that

$$P_0(T^+ \le x) = P_0\left(T^+ \ge \frac{n(n+1)}{2} - x\right), \tag{3.17}$$

for $x = 0, 1, \ldots, n(n+1)/2$.

9. *Zero Z Values.* We have recommended dealing with zero values among the Z's by discarding them and redefining n to be the number of nonzero Z's. This approach is satisfactory as long as the zero values are a very small percentage of the Z differences. If, however, there is a relatively large number of zero Z's, it would be advisable to consider an appropriate statistical procedure designed

for analyzing such discrete data. See, for example, Chapter 10 or a book on categorical data analysis, such as Agresti (2013).

We should also point out that there are methods other than elimination that have been proposed for dealing with zero Z values. One could use individual randomization (e.g., flipping a fair coin) to decide whether each of the zero Z values is to be counted as positive or negative in the construction of T^+. (Although this approach maintains many of the nice properties of T^+ that hold when there are no zeros, it introduces extraneous randomness that could quite easily have a direct effect on the outcome of any subsequent inferences based on such a modified T^+.) A second alternative approach in the case of the one-sided test procedures (3.4), (3.5), (3.10), and (3.11) is to be conservative about rejecting the null hypothesis H_0; that is, we could count all the zero Z values as if they were in favor of not rejecting H_0. Thus, for example, in applying either procedure (3.4) or (3.10) to test H_0 against the alternative $\theta > 0$, we would treat all of the zero Z's as if they were negative (in favor of not rejecting H_0) in the calculation of T^+. (In the case of procedures (3.5) and (3.11), zero Z's would be considered positive in the calculation of T^+.) Any rejection of H_0 with this conservative approach to deal with zero Z values could then be viewed as providing strong evidence in favor of the appropriate alternative. For a more detailed discussion of methods for handling zero observations, see Pratt (1959).

10. *Tied Nonzero Absolute Z Values.* Methods for dealing with tied nonzero absolute Z values other than using average ranks have been discussed in the literature. These include analogs to the randomization and conservative approaches mentioned in Comment 9 with regard to zero Z values. For further discussion of these alternative methods for dealing with tied nonzero absolute Z's, see Pratt (1959).

11. *Exact Conditional Distribution of T^+ with Ties Among the Nonzero Absolute Z Values.* To have a test with exact significance level even in the presence of tied absolute Z's (assuming there are no zero Z values or they have been discarded and n reduced accordingly), one considers all 2^n possible outcomes for the ordered configuration $(r_1, \ldots, r_B)$, where B represents the number of positive Z's as in Comment 5 but where $r_1 < \cdots < r_B$ now denote the ordered ranks of the absolute values of the positive Z's using average ranks to break the ties. As in Comment 5, it still follows that under H_0, each of the 2^n possible outcomes for the ordered configurations $(r_1, \ldots, r_B)$ based on using average ranks to break ties occurs with the probability $(\frac{1}{2})^n$. For each such configuration, the value of T^+ is computed and the results are tabulated. We illustrate this construction for $n = 4$ and the data $Z_1 = -12$, $Z_2 = -10$, $Z_3 = 10$, $Z_4 = 12$. Using average ranks to break ties, the associated absolute value ranks are $R_1 = 3.5, R_2 = 1.5, R_3 = 1.5$, and $R_4 = 3.5$. Thus, $B = 2$ and the ordered ties-broken-ranks for the positive Z's are $r_1 = 1.5$ and $r_2 = 3.5$, leading to an attained value of $T^+ = 5$. To assess the significance of T^+, we obtain its conditional distribution by considering the $2^4 = 16$ equally likely (under H_0) possible values of $(r_1, \ldots, r_B)$ for the given tied rank vector (1.5,1.5,3.5,3.5). These 16 values of $(r_1, \ldots, r_B)$ and associated values of T^+ are shown in the following table.

B	$(r_1, r_2, \ldots, r_B)$	Probability under H_0	Value of T^+
0		$\frac{1}{16}$	0
1	$r_1 = 1.5$	$\frac{1}{16}$	1.5
1	$r_1 = 1.5$	$\frac{1}{16}$	1.5
1	$r_1 = 3.5$	$\frac{1}{16}$	3.5
1	$r_1 = 3.5$	$\frac{1}{16}$	3.5
2	$r_1 = 1.5, r_2 = 1.5$	$\frac{1}{16}$	3
2	$r_1 = 1.5, r_2 = 3.5$	$\frac{1}{16}$	5
2	$r_1 = 1.5, r_2 = 3.5$	$\frac{1}{16}$	5
2	$r_1 = 1.5, r_2 = 3.5$	$\frac{1}{16}$	5
2	$r_1 = 1.5, r_2 = 3.5$	$\frac{1}{16}$	5
2	$r_1 = 3.5, r_2 = 3.5$	$\frac{1}{16}$	7
3	$r_1 = 1.5, r_2 = 1.5, r_3 = 3.5$	$\frac{1}{16}$	6.5
3	$r_1 = 1.5, r_2 = 1.5, r_3 = 3.5$	$\frac{1}{16}$	6.5
3	$r_1 = 1.5, r_2 = 3.5, r_3 = 3.5$	$\frac{1}{16}$	8.5
3	$r_1 = 1.5, r_2 = 3.5, r_3 = 3.5$	$\frac{1}{16}$	8.5
4	$r_1 = 1.5, r_2 = 1.5, r_3 = 3.5, r_4 = 3.5$	$\frac{1}{16}$	10

This yields the null tail probabilities

$$
\begin{aligned}
P_0(T^+ \geq 10) &= \frac{1}{16}, \\
P_0(T^+ \geq 8.5) &= \frac{3}{16}, \\
P_0(T^+ \geq 7) &= \frac{4}{16}, \\
P_0(T^+ \geq 6.5) &= \frac{6}{16}, \\
P_0(T^+ \geq 5) &= \frac{10}{16}, \\
P_0(T^+ \geq 3.5) &= \frac{12}{16}, \\
P_0(T^+ \geq 3) &= \frac{13}{16}, \\
P_0(T^+ \geq 1.5) &= \frac{15}{16}, \\
P_0(T^+ \geq 0) &= 1.
\end{aligned}
$$

This distribution is called the *conditional distribution* or the *permutation distribution of* T^+, given the set of tied ranks {1.5,1.5,3.5,3.5}. For the particular observed value $T^+ = 5$, we have $P_0(T^+ \geq 5) = \frac{10}{16}$, so that such a value does not indicate a deviation from H_0 in the direction of $\theta > 0$.

12. *Some Power Results for the Wilcoxon Signed Rank Test.* We consider the upper-tail α-level test of $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ given by procedure (3.4). Under the additive shift model (see Assumption A2) and common underlying distribution $F_1 \equiv F_2 \equiv \cdots \equiv F_n \equiv F$ for the Z differences, the power, or probability of correctly rejecting H_0, for median θ_0 values "near" the null hypothesis value of 0 can be approximated by

$$\text{Power} \doteq \Phi(A_F), \tag{3.18}$$

where $\Phi(A_F)$ is the area under a standard normal density to the left of the point

$$A_F = \left\{ \frac{n(n-1)f^*(0) + nf(0)}{[n(n+1)(2n+1)/24]^{1/2}} \right\} \theta - z_\alpha, \tag{3.19}$$

where $f(0)$ is the common density function, evaluated at 0, for the Z differences and $f^*(0)$ is the density function, also evaluated at 0, of the sum of two independent random variables drawn from the Z population having distribution F (cf. Lehmann (1975, 167 and 403)).

When F is normal with standard deviation σ, we have $f(0) = (\sigma\sqrt{2\pi})^{-1}$ and $f^*(0) = (2\sigma\sqrt{\pi})^{-1}$. Under this setting, A_F in (3.19) reduces to

$$A_{\text{normal}} = \left\{ \frac{(n(n-1)/2) + n/\sqrt{2}}{[n(n+1)(2n+1)/24]^{1/2}} \right\} \frac{\theta}{\sigma\sqrt{\pi}} - z_\alpha. \tag{3.20}$$

Thus, when F is normal, the approximate power for the additive shift model depends on θ and σ only through their ratio θ/σ. This implies, for example, that the approximate power for the pair $(\theta = .5, \sigma = 4)$ is the same as the approximate power for the pair $(\theta = 1, \sigma = 8)$.

For the purpose of illustration, suppose that the additive shift model holds, with the common underlying population F taken to be normal with variance $\sigma^2 = 4$ and treatment effect $\theta = 1.5$. For the case where $n = 10$ and $\alpha = .053$, the test rejects H_0 if and only if $T^+ \geq 44$. Substituting the appropriate values in (3.20), we obtain

$$A_{\text{normal}} = \left\{ \frac{10(9)/2 + 10/\sqrt{2}}{[10(11)(21)/24]^{1/2}} \right\} \frac{1.5}{2\sqrt{\pi}} - 1.62 = .61$$

Thus, the approximate power of this test at $\theta = 1.5$ (and $\sigma^2 = 4$) is

$$\text{Power} \doteq \texttt{pnorm}(.61) = .73.$$

This compares with the exact power of .70 as given in Table 1 of Klotz (1963). Additional exact power values for the one-sided Wilcoxon signed rank test and sample sizes 5(1)10 can be found in Klotz (1963) for normal shift alternatives and in Arnold (1965) for shifted t-distributions with $\nu = \frac{1}{2}$, 1, 2, and 4 degrees of freedom.

13. *Sample Size Determination.* The Wilcoxon signed rank test detects a more general class of alternatives than the location-shift alternatives associated with model Assumption A2. When $Z_1, \ldots, Z_n$ are a random sample from a single continuous, symmetric population F, the one-sided upper-tail test defined by procedure (3.4) is consistent (i.e., has power tending to 1 as n tends to infinity) against those F populations for which $\eta > \frac{1}{2}$, with

$$\eta = P(Z_1 + Z_2 > 0), \tag{3.21}$$

where Z_1 and Z_2 are independent and identically distributed as F. The parameter η is the probability that a Z_1 randomly selected from the continuous and symmetric F will be greater than the negative of a second independent Z_2 also randomly selected from the same distribution F.

Noether (1987) shows how to determine an approximate sample size n so that the α-level one-sided test given by procedure (3.4) will have approximate power $1 - \beta$ against an alternative value of η greater than $\frac{1}{2}$. This approximate value of n is

$$n \doteq \frac{(z_\alpha + z_\beta)^2}{3\left(\eta - \frac{1}{2}\right)^2}. \tag{3.22}$$

As an illustration of the use of (3.22), suppose we are testing H_0 and we desire to have an upper-tail level $\alpha = .025$ test with power $1 - \beta$ of at least .95 against an alternative for which $\eta = P(Z_1 + Z_2 > 0) = .8$ (recall that under $H_0, \eta = .5$). Since $z_\alpha = z_{.025} = \texttt{qnorm}(.975) = 1.96$ and $Z_\beta = Z_{.05} = \texttt{qnorm}(.95) = 1.65$, we find that the approximate required sample size for the alternative $\eta = .8$ is

$$n \doteq \frac{(1.96 + 1.65)^2}{3(.8 - .5)^2} = 48.3.$$

To be conservative, we would take $n = 49$.

14. *Consistency of the T^+ Test.* Under the assumption that $Z_1, \ldots, Z_n$ is a random sample from a single continuous population F, the consistency of the tests based on T^+ depends on the parameter

$$\eta^* = P(Z_1 + Z_2 > 0) - \frac{1}{2},$$

where Z_1 and Z_2 are independent and identically distributed as F. The test procedures defined by (3.4), (3.5), and (3.6) are consistent against the classes of alternatives corresponding to $\eta^\star >, <$, and $\neq 0$, respectively.

Properties

1. *Consistency.* For our consistency statement we strengthen Assumption A2 to require that each Z has the same continuous population that is symmetric about θ. Then the tests defined by (3.4), (3.5), and (3.6) are consistent against the alternatives $\theta >, <$, and $\neq 0$, respectively. (See also Comment 14.)
2. *Asymptotic Normality.* See Randles and Wolfe (1979, pp. 83–85).
3. *Efficiency.* See Section 3.11.

Problems

1. The data in Table 3.3 are a subset of the data obtained by Kaneto, Kosaka, and Nakao (1967). The experiment investigated the effect of vagal nerve stimulation on insulin secretion. The subjects were mongrel dogs with varying body weights. Table 3.3 gives the amount of immunoreactive insulin in pancreatic venous plasma just before stimulation of the left vagus nerve (X) and the amount measured 5 min after stimulation (Y) for seven dogs. Test the hypothesis of no effect against the alternative that stimulation of the vagus nerve increases the blood level of immunoreactive insulin.
2. Change the value of X_3, in Table 3.1, from 1.62 to 16.2. What effect does this outlying observation have on the calculations performed in Example 3.1? What does this suggest about the relative insensitivity of the signed rank tests to outliers? Construct an example in which changing one observation has a marked effect on the final decision regarding rejection or acceptance of H_0.
3. Let $T^- = \sum_{i=1}^{n} R_i(1 - \psi_i)$, where $\psi_i = 1$ if $Z_i > 0$, and 0 otherwise. Verify directly, or illustrate using the data of Table 3.1, the equation $T^+ + T^- = n(n+1)/2$.
4. August, Hung, and Houck (1974) studied collagen metabolism in children deficient in growth hormone before and after growth hormone therapy. The data in Table 3.4 are the values of heat-insoluble hydroxyproline in the skin of children before and 3 months after growth hormone

Table 3.3 Blood Levels of Immunoreactive Insulin (μU/ml)

Dog i	X_i	Y_i
1	350	480
2	200	130
3	240	250
4	290	310
5	90	280
6	370	1450
7	240	280

Source: A. Kaneto, K. Kosaka, and K. Nakao (1967).

Table 3.4 Heat-Insoluble Hydroxyproline Micromoles per Gram of Dry Weight

Child i	Before	After
1	349	425
2	400	533
3	520	362
4	490	628
5	574	463
6	427	427
7	435	449

Source: G. P. August, W. Hung, and J. C. Houck (1974).

therapy. Can we conclude on the basis of these data that growth hormone therapy increases heat-insoluble hydroxyproline in the skin?

5. Assume that the additive shift model (see Assumption A2) holds with the common underlying distribution $F_1 \equiv F_2 \equiv \cdots \equiv F_n \equiv F$. If we have 15 observations and F is normal with variance 16, what is the approximate power of the level $\alpha = .076$ test of $H_0 : \theta = 0$ versus the alternative $\theta > 0$ when the treatment effect is $\theta = 1.25$?

6. For arbitrary number of observations n, what are the smallest and largest possible values for T^+? Justify your answers.

7. Consider the case $n = 8$ and use the `R` command `psignrank(0:18,8,lower.tail=T)` to produce the lower-tail probabilities of the null distribution of T^+. What are the possible α values between .05 and .10? Compare the $\alpha = .055$ test of $H_0 : \theta = 0$ versus $H_2 : \theta < 0$ with the corresponding $\alpha = .055$ test based on the large-sample approximation.

8. Consider a level $\alpha = .05$ test of $H_0 : \theta = 0$ versus the alternative $\theta > 0$ based on T^+ and let η be as given in (3.21). If our data $Z_1, \ldots, Z_n$ are a random sample from a single continuous, symmetric distribution $F(\cdot)$, how many observations n will we need to collect in order to have an approximate power of at least .84 against an alternative for which $\eta = .7$?

9. Suppose $n = 5$ and we observe the data $Z_1 = -1.3, Z_2 = 2.4, Z_3 = 1.3, Z_4 = 1.3$, and $Z_5 = 2.4$. What is the conditional probability distribution of T^+ under $H_0 : \theta = 0$ when average ranks are used to break ties among the absolute values of the Z's? How extreme is the observed value of T^+ in this conditional null distribution?

10. Apply the large-sample approximation test of $H_0 : \theta = 5$ versus $H_1 : \theta > 5$ based on T^+ to the beak-clapping data in Table 3.5. What is the P-value?

11. Consider procedure (3.6) with n observations for testing H_0: $\theta = 0$ versus $H_1 : \theta \neq 0$. If your critical region consists of the four values $T^+ = 0, 1, [n(n+1)/2] - 1, n(n+1)/2$, what is the significance level for your test?

12. Apply the one-sided upper-tail test based on T^+ to the data on Stanford Profile Scales of hypnotic susceptibility in Table 3.6. What is the P-value obtained?

13. For the case $n = 5$ untied Z observations, use the representation for T^+ discussed in Comment 5 to obtain the form of the exact null (H_0) distribution of T^+.

14. Let Z_1 and Z_2 be independent, identically distributed continuous random variables with a common probability distribution that is symmetric about 0. What is the value of η^* in Comment 14 for this setting?

15. Consider the test of H_0: $\theta = 0$ versus $H_1 : \theta > 0$ based on T^+ for the following $n = 10$ Z observations: $Z_1 = 2.5, Z_2 = 3.7, Z_3 = 0, Z_4 = -0.6, Z_5 = 4.7, Z_6 = 0, Z_7 = 1.4, Z_8 = 0, Z_9 = 1.9$, $Z_{10} = 5.2$. Compute the P -values for the competing T^+ procedures based on either (i) discarding the zero Z values and reducing n accordingly, as recommended in the Ties portion of this section, or (ii) treating the zero Z values in a conservative manner, as presented in Comment 9. Discuss the results.

16. Suppose you desire an upper-tail test of $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ based on T^+ and you want the test to have $\alpha = .05$ and a power of at least .90 when the distribution of $Z = Y - X$ is N(.5,1). Find the approximate required sample size.

17. What are the possible values of T^+ when $n = 8$? Suppose you are testing $H_0 : \theta = 0$ versus $H_2 : \theta < 0$ and you want your α level to be between .05 and .10. What are the tests available?

3.2 AN ESTIMATOR ASSOCIATED WITH WILCOXON'S SIGNED RANK STATISTIC (HODGES–LEHMANN)

Procedure

To estimate the treatment effect θ, form the $M = n(n+1)/2$ averages $(Z_i + Z_j)/2$, for $i \le j = 1, \ldots, n$. The estimator of θ associated with the Wilcoxon signed rank statistic T^+ is (see Comment 15)

$$\hat{\theta} = \text{median} \left\{ \frac{Z_i + Z_j}{2}, i \le j = 1, \ldots, n \right\}. \tag{3.23}$$

Let $W^{(1)} \le \cdots \le W^{(M)}$ denote the ordered values of $(Z_i + Z_j)/2$. Then if M is odd, say $M = 2k+1$, we have $k = (M-1)/2$ and

$$\hat{\theta} = W^{(k+1)}, \tag{3.24}$$

the value that occupies position $k+1$ in the list of the ordered $(Z_i + Z_j)/2$ averages. If M is even, say $M = 2k$, then $k = M/2$ and

$$\hat{\theta} = \frac{W^{(k)} + W^{(k+1)}}{2}. \tag{3.25}$$

That is, when M is even, $\hat{\theta}$ is the average of the two $(Z_i + Z_j)/2$ values that occupy positions k and $k+1$ in the ordered list of the M $(Z_i + Z_j)/2$ averages. The $(Z_i + Z_j)/2$ averages are known as the *Walsh averages* (see Walsh (1949)).

EXAMPLE 3.3 ***Continuation of Example 3.1.***

To estimate θ for the Hamilton depression scale factor IV data in Table 3.1, we first calculate the $M = 9(10)/2 = 45$ $(Z_i + Z_j)/2$ averages. To facilitate this computation and eventual ordering of these averages, we first obtain the individual ordered Z values, denoted by $Z^{(1)} \le \cdots \le Z^{(9)}$, and place them as the headings for a 9×9 upper triangular array. The (i, j)th entry in this array then corresponds to the value of the sum $(Z^{(i)} + Z^{j})$, for $i \le j = 1, 2, \ldots, 9$. This representation for the Hamilton depression scale factor IV data is as follows:

	−1.022	−0.952	−0.620	−0.590	−0.490	−0.430	−0.010	0.080	0.147
−1.022	−2.044	−1.974	−1.642	−1.612	−1.512	−1.452	−1.032	−0.942	−0.875
−0.952		−1.904	−1.572	−1.542	−1.442	−1.382	−0.962	−0.872	−0.805
−0.620			−1.240	−1.210	−1.110	−1.050	−0.630	−0.540	−0.473
−0.590				−1.180	−1.080	−1.020	−0.600	−0.510	−0.443
−0.490					−0.980	−0.920	−0.500	−0.410	−0.343
−0.430						−0.860	−0.440	−0.350	−0.283
−0.010							−0.020	0.070	0.137
0.080								0.160	0.227
0.147									0.294

Thus, for example, the entry in the fourth row and sixth column of the array $(i = 4, j = 6)$ is $Z^{(4)} + Z^{(6)} = -.590 - .430 = -1.020$. The remaining 44 entries are calculated similarly. The ordered $Z^{(i)} + Z^{(j)}$ sums are then obtained by observation in this array, moving carefully from the upper left (the (1, 1) entry) across and down the array to the lower right (the (9,9) entry). The ordered $Z^{(i)} + Z^{(j)}$ sums for these data are as follows: $-2.044, -1.974, -1.904, -1.642, -1.612, -1.572, -1.542, -1.512, -1.452, -1.442, -1.382, -1.240, -1.210, -1.180, -1.110, -1.080, -1.050, -1.032, -1.020, -.980, -.962, -.942, -.920, -.875, -.872, -.860, -.805, -.630, -.600, -.540, -.510, -.500, -.473, -.443, -.440, -.410, -.350, -.343, -.283, -.020, .070, .137, .160, .227, .294$. The ordered values of the $(Z_i + Z_j)/2$ averages, namely, $W^{(1)} \leq \cdots \leq W^{(45)}$, then correspond to these ordered $Z^{(i)} + Z^{(j)}$ sums divided by 2. As $M = 45$ is odd, we use (3.24) with $k = (45 - 1)/2 = 22$ to obtain the estimate $\hat{\theta} = W^{(23)} = -.920/2 = -.460$ for the treatment effect θ. Thus, we estimate that a typical patient of the type included in this study will have a drop in the Hamilton depression scale factor IV value of roughly .460 due to treatment with tranquilizer T.

We can use the `R` command `owa` to compute the ordered Walsh averages and the Hodges–Lehmann estimator. Use `owa(pre,post)`.

[1]	−1.0220	−0.9870	−0.9520	−0.8210	−0.8060	−0.7860	−0.7710	−0.7560	−0.7260
[10]	−0.7210	−0.6910	−0.6200	−0.6050	−0.5900	−0.5550	−0.5400	−0.5250	−0.5160
[19]	−0.5100	−0.4900	−0.4810	−0.4710	−0.4600	−0.4375	−0.4360	−0.4300	−0.4025
[28]	−0.3150	−0.3000	−0.2700	−0.2550	−0.2500	−0.2365	−0.2215	−0.2200	−0.2050
[37]	−0.1750	−0.1715	−0.1415	−0.0100	0.0350	0.0685	0.0800	0.1135	0.1470

Comments

15. *Motivation for the Hodges–Lehmann Estimator.* The Hodges–Lehmann estimator $\hat{\theta}$, defined by (3.23), is associated with the Wilcoxon signed rank test. When $\theta = 0$, the distribution of the statistic T^+ is symmetric about its mean, $n(n+1)/4$ (see Comment 8). A natural estimator of θ is the amount $\hat{\theta}$ (say) that should be subtracted from each Z_i so that the value of T^+, when applied to the shifted sample $Z_1 - \hat{\theta}, \ldots, Z_n - \hat{\theta}$, is as close to $n(n+1)/4$ as possible. Roughly speaking, we estimate θ by the amount $(\hat{\theta})$ that the Z sample should be shifted in order that $Z_1 - \hat{\theta}, \ldots, Z_n - \hat{\theta}$ appears (when "viewed" by the signed rank statistic T^+) as a sample from a population with median 0. (Under Assumptions A1 and A2, each of the $Z_1 - \theta, \ldots, Z_n - \theta$ variables is from a population with median 0.)

    The Hodges–Lehmann method can be applied to a large class of statistics containing T^+. However, the forms of the resulting estimators for other members of this class are not always as convenient for calculation as is $\hat{\theta}$. See Hodges and Lehmann (1983) for an expository article on their method.

16. *Sensitivity to Gross Errors.* The estimator $\hat{\theta}$ is relatively insensitive to outliers. This is not the case with the classical estimator $\bar{Z} = \sum_{i=1}^{n} Z_i/n$. Thus the use of $\hat{\theta}$ provides protection against gross errors.

17. *The Walsh Averages.* Each of the $n(n+1)/2$ averages $(Z_i + Z_j)/2, i \leq j = 1, \ldots, n$, is called a *Walsh average* (see Walsh (1949)). If we define W^+ to be the number of positive Walsh averages, then (when there are no ties among the

$|Z|$'s and none of the Z's is zero) the statistic W^+ is identical to T^+ (3.3). (See Problem 22.) This result is due to Tukey (1949).

18. *Zero and Tied Absolute Z's.* Note that in calculating the estimator $\hat{\theta}$ we use *all* of the Z differences in computing the $(Z_i + Z_j)/2$ averages. Although we recommend (see Ties in Section 3.1) discarding the zero Z values (and reducing n accordingly) prior to applying the signed rank test to the data, it is not necessary to do so when calculating $\hat{\theta}$. In fact, the zero Z values contain important information about the magnitude of the treatment effect. This is also the case when we consider (Section 3.3) confidence intervals and bounds for θ.
19. *Pseudomedian.* A pseudomedian (cf. Høyland (1965)) of a distribution F is defined to be a median of the distribution of $(Z_1 + Z_2)/2$, where Z_1 and Z_2 are independent, each with the same distribution F. We assume here that our F is such that both the median and the pseudomedian of F are unique. The estimator $\hat{\theta}$ (3.23) is a consistent estimator of the pseudomedian, which in general may differ from the median θ. However, when F is symmetric as assumed in this section, the median and the pseudomedian coincide.

Properties

1. *Standard Deviation of $\hat{\theta}$.* For the asymptotic standard deviation of $\hat{\theta}$ (3.23), see Hodges and Lehmann (1963), Lehmann (1963c), and Comment 24.
2. *Asymptotic Normality.* See Hodges and Lehmann (1963) and Ramachandramurty (1966a).
3. *Efficiency.* See Hodges and Lehmann (1963), Bickel (1965), Høyland (1968), Gastwirth and Rubin (1969), and Section 3.11.

Problems

18. Consider the data of Table 3.2. Using the X and Y associations from Example 3.2, estimate θ for the salary data of that example.

19. Estimate θ for the blood-level data of Table 3.3.

20. Change the value of X_3, as given in Table 3.1, from 1.62 to 16.2. How does this affect the value of $\bar{Z} = \sum_{i=1}^{9} Z_i/9$? How does it affect the estimate of θ given by $\hat{\theta}$? Interpret these calculations in light of Comment 16.

21. Estimate θ for the heat-insoluble hydroxyproline data of Table 3.4.

22. Verify directly, or illustrate using the data of Table 3.1, that (when there are no ties among the absolute values of the Z's and none of the Z's is zero) T^+ is equal to the number of positive Walsh averages W^+. (See Comment 17.)

23. (a) What happens to $\hat{\theta}$ when we add a number b to each of the sample values $Z_1, \ldots, Z_n$?
(b) What happens to $\hat{\theta}$ when we multiply each sample value $Z_i, \ldots, Z_n$ by a number d?
(c) Let k be a positive integer such that $n > 2k$. What happens to $\hat{\theta}$ when we discard the k largest and the k smallest Z values from the sample?

24. (a) Do we need to calculate all of the $n(n+1)/2$ Walsh averages in order to compute the value of $\hat{\theta}$? Explain.

25. Explain why the Hodges–Lehmann estimator is less influenced by outlying observations than is the sample mean of the Z's.

26. Use R to obtain the Hodges–Lehmann estimator for the salary data of Table 3.2.

3.3 A DISTRIBUTION-FREE CONFIDENCE INTERVAL BASED ON WILCOXON'S SIGNED RANK TEST (TUKEY)

Procedure

For a symmetric two-sided confidence interval for θ, with confidence coefficient $1-\alpha$, set

$$C_\alpha = \frac{n(n+1)}{2} + 1 - t_{\alpha/2}, \tag{3.26}$$

where $t_{\alpha/2}$ is the upper $(\alpha/2)$th percentile point of the null distribution of T^+. The percentile points can be found using the R function `psignrank`.

The $100(1-\alpha)\%$ confidence interval (θ_L, θ_U) for θ that is associated with the two-sided Wilcoxon signed rank test (see Comment 20) of $H_0 : \theta = 0$ is then given by

$$\theta_L = W^{(C_\alpha)}, \theta_U = W^{(M+1-C_\alpha)} = W^{(t_{\alpha/2})}, \tag{3.27}$$

where $M = n(n+1)/2$ and $W^{(1)} \le \cdots \le W^{(M)}$ are the ordered values of the $(Z_i + Z_j)/2$ averages, $1 \le i \le j \le n$, used in computing the point estimator $\hat{\theta}$ (3.23); that is, θ_L is the $(Z_i + Z_j)/2$ average (i.e., the Walsh average; see Comment 17) that occupies position C_α in the list of M ordered $(Z_i + Z_j)/2$ averages. The upper end point θ_U is the $(Z_i + Z_j)/2$ average that occupies the position $M + 1 - C_\alpha = t_{\alpha/2}$ in this ordered list. With θ_L and θ_U given by display (3.27), we have

$$P_\theta(\theta_L < \theta < \theta_U) = 1 - \alpha \text{ for all } \theta. \tag{3.28}$$

(For upper or lower confidence bounds for θ associated with the appropriate one-sided Wilcoxon signed rank tests of H_0: $\theta = 0$, see Comment 21.)

Large-Sample Approximation

For large n, the integer C_α may be approximated by

$$C_\alpha \approx \frac{n(n+1)}{4} - z_{\alpha/2}\left\{\frac{n(n+1)(2n+1)}{24}\right\}^{1/2}. \tag{3.29}$$

In general, the value of the right-hand side of (3.29) is not an integer. To be conservative, take C_α to be the largest integer that is less than or equal to the right-hand side of (3.29).

EXAMPLE 3.4 ***Continuation of Examples 3.1 and 3.3.***

Consider the Hamilton depression scale factor IV data of Table 3.1. We illustrate how to obtain the 96% confidence interval for θ. With $1-\alpha = .96, \alpha/2 = .02$. From `psignrank(0:22,9,lower.tail=T)`, we find $P_0(T^+ \le 6) = P_0(T^+ \ge 40) = .02$. Thus, $t_{.02} = 40$. From (3.26), it follows that

$$C_{.04} = \left[\frac{9(9+1)}{2} + 1 - 40\right] = 6.$$

Using these values of $C_{.04} = 6$ and $t_{.02} = 40$ in display (3.27), we see that

$$\theta_{\mathrm{L}} = W^{(6)} = -.786 \text{ and } \theta_{\mathrm{U}} = W^{(40)} = -.010.$$

The value $\theta_{\mathrm{L}} = -.786$ is the sixth smallest Walsh average and can be found from the list of ordered Walsh averages at the end of Example 3.3. Similarly, $\theta_{\mathrm{U}} = -.010$ is the sixth largest Walsh average (or 40^{th} ordered).

If we choose to apply the large-sample approximation, we find from approximation (3.29) that

$$C_{.04} \approx \left[\frac{9(9+1)}{4}\right] - 2.05\left\{\frac{9(9+1)(2(9)+1)}{24}\right\}^{1/2} = 5.2.$$

Thus, with a conservative approach and the large-sample approximation, we set $C_{.04} = 5$ and find that

$$(\theta_{\mathrm{L}}, \theta_{\mathrm{U}}) = (W^{(5)}, W^{(41)}) = (-.806, .035)$$

is the approximate 96% confidence interval for θ.

The exact 96% confidence internal can be found from the `R` command `wilcox.test(post-pre,conf.int=T, conf.level=.96)` yielding $(-.786, -.010)$.

Comments

20. *Relationship of Confidence Interval to Two-Sided Test.* The 100 $(1-\alpha)$% confidence interval for θ given by display (3.27) can be obtained from the two-sided signed rank test as follows. The confidence interval $(\theta_{\mathrm{L}}, \theta_{\mathrm{U}})$ consists of those θ_0 values for which the two-sided α-level test of $\theta = \theta_0$ (see Comment 3) does not reject the hypothesis $\theta = \theta_0$. The confidence interval given by display (3.27) was defined by way of a graphical procedure by Lincoln Moses (who attributed it to John Tukey) in Chapter 18 of Walker and Lev (1953). See Lehmann (1986, p. 90) for a general result relating confidence intervals and acceptance regions of tests, and see Lehmann (1963c) for the specific result involving the signed rank test.

21. *Confidence Bounds.* In many settings, we are interested only in making one-sided confidence statements about the parameter θ; that is, we wish to assert with specified confidence that θ is no larger (or, in other settings, no smaller) than some upper (lower) confidence bound based on the sample data. To obtain such one-sided confidence bounds for θ, we proceed as follows. For the specified confidence coefficient $1-\alpha$, set

$$C_\alpha^* = \frac{n(n+1)}{2} + 1 - t_\alpha, \tag{3.30}$$

where t_α is the upper αth percentile point of the null distribution of T^+.
The $100(1-\alpha)$% lower confidence bound θ_{L}^* for θ that is associated with the one-sided Wilcoxon signed rank test of $H_0 : \theta = 0$ against the alternative $H_1 : \theta > 0$ is then given by

$$(\theta_{\mathrm{L}}^*, \infty) = (W^{(C_\alpha^*)}, \infty), \tag{3.31}$$

where, as before, $M = n(n+1)/2$ and $W^{(1)} \le \cdots \le W^{(M)}$ are the ordered values of the $(Z_i + Z_j)/2$ averages, $1 \le i \le j \le n$. With θ_{L}^* given by display (3.31), we have

$$P_\theta(\theta_{\mathrm{L}}^* < \theta < \infty) = 1 - \alpha \text{ for all } \theta. \tag{3.32}$$

The corresponding $100(1-\alpha)\%$ upper confidence bound θ_{U}^* for θ that is associated with the one-sided Wilcoxon signed rank test of $H_0 : \theta = 0$ against the alternative $H_1 : \theta < 0$ is given by

$$(-\infty, \theta_{\mathrm{U}}^*) = (-\infty, W^{(M+1-C_\alpha^*)}) = (-\infty, W^{(t_a)}), \tag{3.33}$$

where C_α^* is given in (3.30). It follows that

$$P_\theta(-\infty < \theta < \theta_{\mathrm{U}}^*) = 1 - \alpha \text{ for all } \theta. \tag{3.34}$$

For large n, the integer C_α^* may be approximated by

$$C_\alpha^* \approx \frac{n(n+1)}{4} - z_\alpha \left\{ \frac{n(n+1)(2n+1)}{24} \right\}^{1/2}. \tag{3.35}$$

As with C_α (3.29) and the confidence interval for θ, the value of the right-hand side of (3.35) is not an integer. To be conservative, take C_α^* to be the largest integer that is less than or equal to the right-hand side of (3.35).

The $100(1-\alpha)\%$ lower and upper confidence bounds θ_{L}^* (3.31) and θ_{U}^* (3.33) are related to the acceptance regions of the one-sided Wilcoxon signed rank tests of $H_0 : \theta = \theta_0$ against the alternatives $\theta > \theta_0$ and $\theta < \theta_0$, respectively, in the same way that the confidence interval $(\theta_{\mathrm{L}}, \theta_{\mathrm{U}})$ is related to the acceptance region of the two-sided Wilcoxon signed rank test of $H_0 : \theta = \theta_0$. (See Comment 20.)

22. *Zero and Tied Absolute Z's.* Note that in calculating the confidence interval $(\theta_{\mathrm{L}}, \theta_{\mathrm{u}})$ from display (3.27) or the confidence bounds θ_{L}^* (3.31) or θ_{U}^* (3.33) for θ, we use *all* the Z differences in computing the $(Z_i + Z_j)/2$ averages. This is in common with our recommendation (see Comment 18) for computing the point estimator $\hat{\theta}$ (3.23), but different from the recommended policy (see Ties in Section 3.1) of discarding the zero Z values (and reducing n accordingly) prior to applying the signed rank test to the data. However, if there are zero Z's in the data, the equivalence (discussed in Comments 20 and 21) between the acceptance regions of the one-sided and two-sided signed rank tests and the appropriate confidence bound and confidence interval, respectively, are no longer valid. In addition, in cases with tied absolute Z's, the nominal confidence coefficient $1-\alpha$ used in displays (3.27), (3.31), and (3.33) is no longer exact. (See Comment 11.)

23. *Midpoint of Confidence Interval as an Estimator.* The midpoint of the interval (3.27), namely, $[W^{(C_\alpha)} + W^{(M+1-C_\alpha)}]/2$, suggests itself as a reasonable estimator of θ. (Note that this actually yields a class of estimators depending on the value of α.) In general, this midpoint is not the same as $\hat{\theta}$ (3.23). Lehmann (1963c) has also derived an asymptotically distribution-free confidence interval centered at $\hat{\theta}$. This asymptotically distribution-free confidence interval is based on the assumption that each of the n Z_i's comes from the *same* continuous population that is symmetric about θ. This assumption is more restrictive than Assumption A2.

24. *Estimating the Asymptotic Standard Deviation of* $\hat{\theta}$. Replace Assumption A2 by the stronger Assumption $A2'$: each Z comes from the *same* continuous population that is symmetric about θ. Then, it follows from Lehmann (1963c) that the statistic $(\theta_U - \theta_L)/2z_{\alpha/2}$, where (θ_L, θ_U) is the $100(1-\alpha)\%$ confidence interval for θ defined by display (3.27), is a consistent estimator for the asymptotic standard deviation of the point estimator $\hat{\theta}$ (3.23).

Properties

1. *Distribution-Freeness.* For populations satisfying Assumptions Al and A2, (3.28) holds. Hence, we can control the coverage probability to be $1-\alpha$ without having more specific knowledge about the forms of the underlying Z distributions. Thus, (θ_L, θ_U) is a distribution-free confidence interval for θ over a very large class of populations.
2. *Efficiency.* See Lehmann (1963c) and Section 3.11.

Problems

27. For the blood-level data of Table 3.3, obtain a confidence interval for θ with the exact confidence coefficient .954.

28. For the heat-insoluble hydroxyproline data of Table 3.4, obtain a confidence interval for θ with the exact confidence coefficient .922.

29. For the blood-level data of Table 3.3 and $\alpha = .078$, calculate the point estimator of θ defined in Comment 23. Compare with the value of $\hat{\theta}$ obtained in Problem 19.

30. Use the results of Example 3.4 to obtain an estimate of the asymptotic standard deviation of $\hat{\theta}$ for the Hamilton depression scale factor IV data of Table 3.1 (see Comment 24).

31. For the Hamilton depression scale factor IV data of Table 3.1, find an upper confidence bound for θ with the exact confidence coefficient .973 (see Comment 21).

32. For the salary data of Table 3.2, use (3.31) in Comment 21 and find a lower confidence bound for θ with approximate confidence coefficient .936. Why is the confidence coefficient only approximate and not exact?

33. Consider the $1-\alpha$ confidence interval for θ defined by display (3.27). Let $Z_{(1)} \le \cdots \le Z_{(n)}$ be the ordered Z's. Show that when $\alpha = 2/2^n$,

$$\theta_L = Z_{(1)} \text{ and } \theta_U = Z_{(n)}.$$

34. Consider the $1-\alpha$ upper confidence bound for θ given in (3.33) in Comment 21. If $Z_{(1)} \le \cdots \le Z_{(n)}$ denote the ordered Z's and $\alpha = 2/2^n$, show that

$$\theta_U^* = \frac{Z_{(n-1)} + Z_{(n)}}{2}.$$

35. Consider the $1-\alpha$ confidence interval for θ defined by display (3.27). Let $Z_{(1)} \le \cdots \le Z_{(n)}$ be the ordered Z's. If $\alpha = 4/2^n$, express the length $(\theta_U - \theta_L)$ of the confidence interval in terms of $Z_{(1)}, \ldots, Z_{(n)}$.

36. How does varying α affect the length of the confidence interval defined by display (3.27)? How does it affect the point estimator defined in Comment 23?

37. Consider the blood-level data of Table 3.3. Obtain an approximate 95% confidence interval for θ using the large-sample approximation of this section. Compare this approximate confidence interval with the exact 95.4% confidence interval obtained in Problem 27.

38. Consider the salary data of Table 3.2. Use the large-sample approximation of this section to obtain an approximate 90% confidence interval for θ.

39. Consider the heat-insoluble hydroxyproline data of Table 3.4. Use the large-sample approximation to obtain an approximate 99% lower confidence bound for θ. (See Comment 21.)

40. Consider the case $n = 10$ and compare the length of the exact 95.2% confidence interval for θ given by display (3.27) with the length of the approximate 95.2% confidence interval for θ obtained using the large-sample approximation of this section.

41. Consider the case $n = 15$ and compare the exact 96.8% upper confidence bound for θ given by (3.33) with the approximate 96.8% upper confidence bound for θ obtained from the large-sample approximation in Comment 21.

42. Use (3.26) and (3.27) to show that, for a fixed value of n, as α decreases the width of the confidence interval increases. Explain this trade-off.

PAIRED REPLICATES ANALYSES BY WAY OF SIGNS

Data. We obtain $2n$ observations, two observations on each of the n subjects (blocks, patients, etc.)

Subject i	X_i	Y_i
1	X_1	Y_1
2	X_2	Y_2
$\vdots$	$\vdots$	$\vdots$
n	X_n	Y_n

Assumptions

B1. We let $Z_i = Y_i - X_i$, for $i = 1, \ldots, n$. The differences $Z_1, \ldots, Z_n$ are mutually independent.

B2. Each $Z_i, i = 1, \ldots, n$, comes from a continuous population (not necessarily the same) that has a common median θ. If F_i represents the distribution function for $Z_i, i = 1, \ldots, n$, this assumption requires that

$$F_i(\theta) = P(Z_i \leq \theta) = P(Z_i > \theta) = 1 - F_i(\theta), \text{ for } i = 1, \ldots, n. \tag{3.36}$$

The parameter θ is referred to as the *unknown treatment effect*.

3.4 A DISTRIBUTION-FREE SIGN TEST (FISHER)

Hypothesis

The null hypothesis of interest here is that of zero shift in location due to the treatment, namely,

$$H_0 : \theta = 0. \tag{3.37}$$

This null hypothesis asserts that each of the distributions (not necessarily the same) for the differences (posttreatment minus pretreatment observations) has median 0, corresponding to no shift in location due to the treatment.

Procedure

To compute the sign statistic B, define indicator variables $\psi_i, i = 1, \ldots, n$, where

$$\psi_i = \begin{cases} 1, & \text{if } Z_i > 0 \\ 0, & \text{if } Z_i < 0, \end{cases} \tag{3.38}$$

and set

$$B = \sum_{i=1}^{n} \psi_i. \tag{3.39}$$

The sign statistic B is the number of positive Z's.

a. *One-Sided Upper-Tail Test.* To test

$$H_0 : \theta = 0$$

versus

$$H_1 : \theta > 0,$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } B \geq b_{\alpha,1/2}; \text{ otherwise do not reject,} \tag{3.40}$$

where the constant $b_{\alpha,1/2}$ is chosen to make the type I error probability equal to α and is the upper αth percentile point for the binomial distribution with sample size n and $p = \frac{1}{2}$. Values of $b_{\alpha,1/2}$ are found with the `R` command `qbinom`.

b. *One-Sided Lower-Tail Test.* To test

$$H_0 : \theta = 0$$

versus

$$H_2 : \theta < 0,$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } B \leq n - b_{\alpha,1/2}; \text{ otherwise do not reject.} \tag{3.41}$$

c. *Two-Sided Test.* To test

$$H_0 : \theta = 0$$

versus

$$H_3 : \theta \neq 0,$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } B \geq b_{\alpha/2,1/2} \text{ or } B \leq n - b_{\alpha/2,1/2}; \text{ otherwise do not reject.} \tag{3.42}$$

This two-sided procedure is the two-sided symmetric test with $\alpha/2$ probability in each tail of the null distribution of B.

Large-Sample Approximation

The large-sample approximation is based on the asymptotic normality of B, suitably standardized. As the distribution of B under the null hypothesis H_0: $\theta = 0$ is binomial with parameters n and $p = \frac{1}{2}$, we know that

$$E_0(B) = \frac{n}{2} \tag{3.43}$$

and

$$\text{var}_0(B) = \frac{n}{4}. \tag{3.44}$$

The standardized version of B is then

$$B^* = \frac{B - E_0(B)}{\{\text{var}_0(B)\}^{1/2}} = \frac{B - (n/2)}{\{n/4\}^{1/2}}. \tag{3.45}$$

When H_0 is true, B^* has, as n tends to infinity, an asymptotic $N(0, 1)$ distribution. (See Comment 32 for indications of the proof.) The normal theory approximation for procedure (3.40) is

$$\text{Reject } H_0 \text{ if } B^* \geq z_\alpha; \text{ otherwise do not reject,} \tag{3.46}$$

the normal theory approximation for procedure (3.41) is

$$\text{Reject } H_0 \text{ if } B^* \leq -z_\alpha; \text{ otherwise do not reject,} \tag{3.47}$$

and the normal theory approximation for procedure (3.42) is

$$\text{Reject } H_0 \text{ if } |B^*| \geq z_{\alpha/2}; \text{ otherwise do not reject.} \tag{3.48}$$

Ties

If there are zero values among the Z's, discard the zero values and redefine n to be the number of nonzero Z's.

EXAMPLE 3.5 ***Beak-Clapping Counts.***

The data in Table 3.5 are a subset of the data obtained by Oppenheim (1968) in an experiment investigating light responsivity in chick embryos. The subjects were white leghorn chick embryos, and the behavioral response measured in the investigation was beak-clapping (i.e., the rapid opening and closing of the beak that occurs during the latter one-third of incubation in chick embryos). (Gottlieb (1965) had previously shown that changes in the rate of beak-clapping constituted a sensitive indicator of auditory responsiveness in chick embryos.) The embryos were placed in a dark chamber 30 min before the initiation of testing. Then ten 1-min readings were taken in the dark, and at the end of this 10-min period, a single reading was obtained for a 1-min period of illumination. Table 3.5 gives the average number of claps per minute during the dark period (X) and the corresponding rate during the period of illumination (Y) for 25 chick embryos.

Table 3.5 Beak-Clapping Counts per Minute

Embryo i	X_i (Dark period)	Y_i (Illumination)	$Z_i = Y_i - X_i$	ψ_i
1	5.8	5	−0.8	0
2	13.5	21	7.5	1
3	26.1	73	46.9	1
4	7.4	25	17.6	1
5	7.6	3	−4.6	0
6	23.0	77	54.0	1
7	10.7	59	48.3	1
8	9.1	13	3.9	1
9	19.3	36	16.7	1
10	26.3	46	19.7	1
11	17.5	9	−8.5	0
12	17.9	25	7.1	1
13	18.3	59	40.7	1
14	14.2	38	23.8	1
15	55.2	70	14.8	1
16	15.4	36	20.6	1
17	30.0	55	25.0	1
18	21.3	46	24.7	1
19	26.8	25	−1.8	0
20	8.1	30	21.9	1
21	24.3	29	4.7	1
22	21.3	46	24.7	1
23	18.2	71	52.8	1
24	22.5	31	8.5	1
25	31.1	33	1.9	1

Source: R. W. Oppenheim (1968).

As responsivity of a chick embryo to a light stimulus is expected to correspond to positive Z differences, we apply procedure (3.40), which is designed to detect the alternative $\theta > 0$. To implement the sign test, one may use `qbinom` directly in procedure (3.40). If setting $\alpha = .05$, the appropriate command is

```
qbinom(p=0.05, size=25, prob=1/2, lower.tail=F)
```

where the argument `p` is α and `size` and `prob` are n and p, respectively, in the binomial distribution. The resulting value is 17. Recall that the argument `lower.tail=F` provides probabilities that are strictly greater than a specified value. Therefore, to be consistent with (3.40), one must use the value $b_{\alpha,1/2} = 18$. Procedure (3.40) is then given by

$$\text{Reject } H_0 \text{ if } B \geq 18.$$

Note that the critical value 18 given by `R` results in a significance level of $\alpha = .022$, not .05. Now, the sample value of B can be obtained directly from the indicator variables $\psi_1, \ldots, \psi_{25}$ listed in Table 3.5.

We find that $B = \sum_{i=1}^{25} \psi_i$= (number of positive Z's) = 21. As this value of B is greater than the critical value 18, we reject H_0 in favor of $\theta > 0$ at the $\alpha = .05$ level.

(We note that the actual magnitudes of the Z differences are not needed to calculate B. We require only the information as to whether or not Y_i is larger than X_i, for $i = 1, \ldots, n$, and this information is contained entirely in the indicator variables $\psi_1, \ldots, \psi_n$. However, the actual magnitude of the Z_i's will be necessary in Sections 3.5 and 3.6

to obtain point and interval estimates, respectively, of θ associated with the sign test.) It is simpler to use the `R` command `SIGN.test` from package `BSDA` (Arnholt, 2012). Running

```
SIGN.test(y, x, alt=''greater''),
```

where `y` is the vector of illumination data and `x` is the vector of dark period data from Table 3.5, results in a test statistic of $B = 21$ (the output actually uses S as the name of the test statistic rather than B) and a P-value of .0005. The P-value may also be found using the `pbinom` command. Partial output from `SIGN.test` is shown below:

```
Dependent-samples Sign-Test

data: y and x
S = 21, p-value =.0004553
alternative hypothesis: true median difference is greater
   than 0
95 percent confidence interval:
7.4519 Inf
sample estimates:
median of x-y
17.6
```

For the large-sample approximation, we find from (3.45) that

$$B^* = \frac{21 - \left(\frac{25}{2}\right)}{\left(\frac{25}{4}\right)^{1/2}} = 3.40.$$

Thus, the smallest significance level at which we can reject H_0 in favor of $\theta > 0$ using the normal approximation (i.e., the approximate P-value) is .0003. Clearly, both the exact test and the large-sample approximation indicate that there is strong evidence that chick embryos are indeed responsive to a light stimulus, as measured by an increase in the frequency of beak-claps.

Comments

25. *Motivation for the Test.* When θ is greater than 0, there will tend to be a large number of positive Z differences, leading to a big value of B. This suggests rejecting H_0 in favor of $\theta > 0$ for large values of B and motivates the procedures (3.40) and (3.46). Similar rationales lead to procedures (3.41), (3.42), (3.47), and (3.48).
26. *Assumptions.* Assumption B2 is implied by Assumption A2, but the converse is not true. Thus, Assumption B2 is less stringent than Assumption A2—an advantage of the sign test over the signed rank test. We can, when testing $\theta = 0$, weaken Assumption B2 further to Assumption B2', namely, $P(Z_i < 0) = P(Z_i > 0) = \frac{1}{2}, i = 1, \ldots, n$, when θ is the hypothesized value 0. When testing $\theta = \theta_0$ (see Comment 28), for $\theta_0 \neq 0$, Assumption B2 can be replaced by the weaker Assumption B_2'', namely, $P(Z_i < \theta_0) = P(Z_i > \theta_0) = \frac{1}{2}, i = 1, \ldots, n$, when θ is the hypothesized value θ_0.

    We also note that there is no requirement that the individual X_i and Y_i be independent, only that the pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, and therefore the

resulting differences $Z_1, \ldots, Z_n$, be mutually independent. Indeed, in most applications, the individual X_i and Y_i are dependent.

27. *Binomial Test.* The test procedures based on the sign statistic B are actually special cases of the general binomial test procedures considered in Chapter 2. The sign test procedures are simply binomial procedures, with "success" corresponding to a positive Z difference, "failure" corresponding to a negative Z difference, and $p = P(\text{"success"}) = P(Z_i > 0)$ assuming the value $p_0 = \frac{1}{2}$ when the null hypothesis $H_0 : \theta = 0$ is true.

28. *Testing θ Equal to Some Specified Nonzero Value.* Procedures (3.40), (3.41), and (3.42) and the corresponding normal approximations in (3.46), (3.47), and (3.48) are for testing θ equal to zero. To test H_0: $\theta = \theta_0$, where θ_0 is some specified nonzero number, subtract θ_0 from each of the differences $Z_1, \ldots, Z_n$ to form a modified sample $Z_1' = Z_1 - \theta_0, \ldots, Z_n' = Z_n - \theta_0$. Then compute B as the number of these Z_i's that are positive. Procedures (3.40), (3.41), and (3.42) and their corresponding large-sample approximations in (3.46), (3.47), and (3.48) are then applied as previously described.

29. *Equivalent Form.* The statistic B (3.39) is the number of positive Z differences. If we define B^- to be the number of negative Z differences, then $B^- = \sum_{i=1}^{n}(1 - \psi_i) = n - \sum_{i=1}^{n} \psi_i = n - B$. Thus, the test procedures (3.40), (3.41), and (3.42) and (3.46), (3.47), and (3.48) could equivalently be based on $B^- = (n - B)$. (We point out that B^- also (as does B) has a binomial distribution with sample size n and $p = \frac{1}{2}$ when $H_0 : \theta = 0$ is true.)

30. *Derivation of the Distribution of B under H_0 (When There Are No Zero Z Values).* The null (H_0) distribution of B can be obtained directly from the representation $B = \sum_{i=1}^{n} \psi_i$. Under the assumption that the underlying Z_i distributions are all continuous, the probabilities are zero that any of the Z_i's are zero. Hence, under H_0, each of the 2^n possible outcomes for the configuration $(\psi_1, \ldots, \psi_n)$ occurs with the probability $\left(\frac{1}{2}\right)^n$. For example, in the case of $n = 3$, the $2^3 = 8$ possible outcomes for (ψ_1, ψ_2, ψ_3) and the associated values of B are given in the following table.

(ψ_1, ψ_2, ψ_3)	Probability under H_0	$B = \sum_{i=1}^{3} \psi_i$
(0,0,0)	$\frac{1}{8}$	0
(0,0,1)	$\frac{1}{8}$	1
(0,1,0)	$\frac{1}{8}$	1
(1,0,0)	$\frac{1}{8}$	1
(0,1,1)	$\frac{1}{8}$	2
(1,0,1)	$\frac{1}{8}$	2
(1,1,0)	$\frac{1}{8}$	2
(1,1,1)	$\frac{1}{8}$	3

Thus, for example, the probability is $\frac{3}{8}$ under H_0 that B is equal to 2, as $B = 2$ when any of the three exclusive outcomes $(\psi_1, \psi_2, \psi_3) = (0, 1, 1)$, $(1, 0, 1)$, or $(1, 0, 1)$ occurs, and each of these outcomes has null probability $\frac{1}{8}$. Simplifying, we obtain the null distribution.

Possible value of B	Probability under H_0
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

The probability, under H_0, that B is greater than or equal to 2, for example, is therefore

$$\begin{aligned} P_0(B \geq 2) &= P_0(B = 2) + P_0(B = 3) \\ &= .375 + .125 = .50. \end{aligned}$$

(We note that this null distribution of B could alternatively be obtained from the binomial probability distribution in (2.15) in Comment 2.7 by taking $P_0 = \frac{1}{2}$.)

Note that we have derived the null distribution of B without specifying the forms of the underlying Z populations under H_0 beyond the requirement that they be continuous and have common median 0. This is why the test procedures based on B are called *distribution-free procedures*. From the null distribution of B we can determine the critical value $b_{\alpha/2,1/2}$ and control the probability α of falsely rejecting H_0 when H_0 is true, and this error probability does not depend on the specific forms of the underlying continuous Z distributions with common median 0.

31. *Calculation of the Mean and Variance of B under the Null Hypothesis H_0.* In displays (3.43) and (3.44), we presented formulas for the mean and variance of B when the null hypothesis is true. In this comment, we illustrate a direct calculation of $E_0(B)$ and $\text{var}_0(B)$ in the particular case of $n = 3$, using the null distribution of B obtained in Comment 30. (Later, in Comment 32, we present the general derivations of $E_0(B)$ and $\text{var}_0(B)$.) The null mean, $E_0(B)$, is obtained by multiplying each possible value of B with its probability under H_0. Thus,

$$E_0(B) = 0(.125) + 1(.375) + 2(.375) + 3(.125) = 1.5.$$

This is in agreement with what we obtain using (3.43), namely,

$$E_0(B) = \frac{n}{2} = \frac{3}{2} = 1.5.$$

A check on the expression for $\text{var}_0(B)$ is also easily performed using the well-known fact that

$$\text{var}_0(B) = E_0(B^2) - \{E_0(B)\}^2.$$

The value of $E_0(B^2)$, the second moment of the null distribution of B, is again obtained by multiplying possible values (in this case, value of B^2) by the corresponding probabilities under H_0. We find

$$E_0(B^2) = 0(.125) + 1(.375) + 4(.375) + 9(.125) = 3.0.$$

Thus,

$$\text{var}_0(B) = 3.0 - (1.5)^2 = 0.75,$$

which agrees with what we obtain using (3.44) directly, namely,

$$\text{var}_0(B) = \frac{n}{4} = \frac{3}{4} = 0.75.$$

32. *Large-Sample Approximation.* Under Assumption B1, the variables $Z_1, \ldots, Z_n$ are mutually independent. The variable ψ_i is a function of Z_i only, for $i = 1, \ldots, n$, therefore $\psi_1, \ldots, \psi_n$ are also mutually independent variables. In view of the representation $B = \sum_{i=1}^{n} \psi_1$ in (3.39), we can immediately use well-known expressions for the mean and variance of a sum of mutually independent random variables to obtain

$$E_0(B) = E_0\left[\sum_{i=1}^{n} \psi_i\right] = \sum_{i=1}^{n} E_0(\psi_i) \tag{3.49}$$

and

$$\text{var}_0(B) = \text{var}_0\left(\sum_{i=1}^{n} \psi_i\right) = \sum_{i=1}^{n} \text{var}_0(\psi_i). \tag{3.50}$$

Now, under H_0, the ψ_i's are also identically distributed, each following the Bernoulli probability distribution with $p = \frac{1}{2}$. Thus, for $i = 1, \ldots, n$, we see that

$$E_0(\psi_i) = 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = \frac{1}{2}$$

and

$$\begin{aligned}\text{var}_0(\psi_i) &= E_0(\psi_i^2) - [E_0(\psi_i)]^2 \\ &= 0^2\left(\frac{1}{2}\right) + 1^2\left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.\end{aligned}$$

Using these results in (3.49) and (3.50), we obtain

$$E_0(B) = \sum_{i=1}^{n}\left(\frac{1}{2}\right) = \frac{n}{2}$$

and

$$\text{var}_0(B) = \sum_{i=1}^{n}\left(\frac{1}{4}\right) = \frac{n}{4},$$

which agree with the general expressions stated in (3.43) and (3.44).

The asymptotic normality of the standardized form

$$B^* = \frac{B - E_0(B)}{\{\text{var}_0(B)\}^{1/2}} = \frac{B - \frac{n}{2}}{\left\{\frac{n}{4}\right\}^{1/2}}$$

follows from standard central limit theory for sums of mutually independent, identically distributed random variables (cf. Randles and Wolfe (1979, p. 421)). Asymptotic normality results are also obtainable under general alternatives to H_0. (See Comment 35.)

33. *Symmetry of the Distribution of B under the Null Hypothesis.* When H_0 is true, the distribution of B is symmetric about its mean $n/2$. (See Comment 30 for verification of this when $n = 3$.) This implies that

$$P_0(B \leq x) = P_0(B \geq n - x), \tag{3.51}$$

for $x = 0, 1, \ldots, n$.

Equation (3.51) is used directly to convert upper-tail probabilities to lower-tail probabilities.

34. *Zero Z Values.* We have recommended discarding zero Z values and redefining n to be the number of nonzero Z's. This approach is satisfactory as long as the zero values do not represent a sizable percentage of the total number of Z differences. If, however, there is a relatively large number of zero Z's, it would be advisable to consider an appropriate statistical procedure designed specifically for analyzing such discrete data. See, for example, Chapter 10 or a book on categorical data analysis such as Agresti (2013).

We should also point out that there are methods other than elimination that have been proposed for dealing with zero Z values. One could use individual randomization (e.g., flipping a fair coin) to decide whether each of the zero Z values is to be counted as positive or negative in the computation of B. (Although this approach maintains many of the nice theoretical properties of B that hold when there are no zeros, it introduces extraneous randomness that could quite easily have a direct effect on the outcome of any subsequent inferences based on such a modified B.) A second alternative approach in the case of the one-sided test procedures in (3.40), (3.41), (3.46), and (3.47) is to be conservative about rejecting the null hypothesis H_0; that is, we could count all the zero Z values as if they were in favor of not rejecting H_0. Thus, for example, in applying either procedure (3.41) or (3.47) to test H_0 against the alternative $\theta < 0$, we would treat all the zero Z's as if they were "positive" (in favor of not rejecting H_0) in the calculation of B. (In the case of procedures (3.40) and (3.46), zero Z values would be considered "negative" in the calculation of B.) Any rejection of H_0 with this conservative approach to dealing with zero Z values could be viewed as providing strong evidence in favor of the appropriate alternative. For a more detailed discussion of methods for handling zero observations, see Pratt (1959).

35. *Some Power Results for the Sign Test.* We consider the upper-tail α-level test of H_0: $\theta = 0$ versus $H_1 : \theta > 0$ given by procedure (3.40). When we have a common underlying distribution with median θ and distribution function $F_1 \equiv F_2 \equiv \cdots \equiv F_n \equiv F$ for the Z differences, the sign statistic B has a binomial distribution with parameters n and $p_\theta = P_\theta(Z_1 > 0) = 1 - F(0)$. It follows (see (2.15) in Comment 2.7) that the exact power of the sign test procedure (3.40) against the alternative $\theta > 0$ is given by the expression

$$\text{Power}_\theta = \sum_{t=b_{\alpha,1/2}}^{n} \binom{n}{t} p_\theta^t (1 - p_\theta)^{n-t} = \sum_{t=b_{\alpha,1/2}}^{n} \binom{n}{t} [1 - F(0)]^t [F(0)]^{n-t}. \tag{3.52}$$

(As $p_\theta = 1 - F(0) > 1 - F(\theta) = 1 - \frac{1}{2} = \frac{1}{2}$ for all $\theta > 0$, it follows that $\text{Power}_\theta > \text{Power}_0 = \alpha$ for all alternatives $\theta > 0$.) Evaluation of Power_θ for a moderate sample size n and particular value of $\theta > 0$ (and associated $p_\theta = 1 - F(0) > \frac{1}{2}$) can thus be accomplished by direct computation.

For large sample sizes, we can make use of the standard central limit theorem for sums of mutually independent and identically distributed random variables to conclude that

$$\frac{B - np_\theta}{[np_\theta(1-p_\theta)]^{1/2}} = \frac{B - n(1-F(0))}{[n(1-F(0))(F(0))]^{1/2}} \tag{3.53}$$

has an asymptotic ($n \longrightarrow \infty$) standard normal distribution. Thus, for large n, we can approximate the exact power in (3.52) by

$$\begin{aligned}\text{Power}_\theta &\approx 1 - \Phi\left(\frac{b_{\alpha,1/2} - np_\theta}{[np_\theta(1-p_\theta)]^{1/2}}\right) \\ &= 1 - \Phi\left(\frac{b_{\alpha,1/2} - n(1-F(0))}{[n(1-F(0))(F(0))]^{1/2}}\right),\end{aligned} \tag{3.54}$$

where $\Phi(t)$ is the area under a standard normal density to the left of t.

We note that both the exact power (3.52) and the approximate power (3.54) against an alternative $\theta > 0$ depend on the common distribution only through the value of its distribution function $F(z)$ at $z = 0$. Thus, if two distributions have a common median $\theta > 0$ and distribution functions F_1 and F_2 such that $F_1(0) = F_2(0)$, then the exact power (3.52) of the sign test against the alternative $\theta > 0$ will be the same for both distributions F_1 and F_2. (The same is, of course, true for the approximate power in (3.54).)

For the purpose of illustration, consider the case where $n = 10$ and $\alpha = .05$. Then using `qbinom`, we see that $b_{.05,1/2} = 9$ and the test (3.40) rejects H_0 if and only if $B \geq 9$. If F is the distribution function for a probability distribution with median $\theta = 2$ (i.e., $F(2) = \frac{1}{2}$) and $F(0) = \frac{1}{4}$, then from (3.52), the exact power of the sign test in this setting is

$$\begin{aligned}\text{Power}_{\theta=2} &= \sum_{t=9}^{10} \binom{10}{t}\left(\frac{3}{4}\right)^t\left(\frac{1}{4}\right)^{10-t} \\ &= \texttt{pbinom(8,size = 10, prob = 3/4, lower.tail = F)} \\ &= .2440.\end{aligned}$$

The approximate power for the same setting is seen from (3.54) to be

$$\begin{aligned}\text{Power}_{\theta=2} &\approx 1 - \Phi\left(\frac{9 - 10\left(\frac{3}{4}\right)}{\left[10\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\right]^{1/2}}\right) \\ &= 1 - \Phi\left(\frac{1.5}{\left[\frac{30}{16}\right]^{1/2}}\right) = .1367.\end{aligned}$$

Thus, for n as small as 10 and $p_\theta = \frac{3}{4}$, the agreement between the exact power (3.52) and the approximate power (3.54) is not too good. (We note

that for an underlying normal distribution F with mean $\theta = 2$ and variance σ^2, the condition $F(0) = \frac{1}{4}$ corresponds to $\Phi((0-2)/\sigma) = \frac{1}{4}$, which in turn corresponds to $(-2/\sigma) = z_{.75} = -.675$, or $\sigma = (2/.675) = 2.96$. More generally, the test (3.40) for $n = 10$ and $\alpha = .05$ has an exact power of .2440 against any normal distribution with mean θ and variance σ^2 for which $-(\theta/\sigma) = z_{.75} = -6.75$, or $\theta = .675\sigma$. Although we specified $\alpha = .05$ in this example, the upper tail probability for $B \geq 9$ is actually .01.

36. *Sample Size Determination.* When $Z_1, \ldots, Z_n$ are a random sample from a single continuous F, the one-sided upper-tail test defined by procedure (3.40) is consistent (i.e., has power tending to 1 as n tends to infinity) against those F populations for which $p > \frac{1}{2}$, with

$$p = P(Z > 0), \tag{3.55}$$

where Z is distributed as F.

Noether (1987), among others, shows how to determine an approximate sample size n so that the α-level one-sided test given by procedure (3.40) will have approximate power $1 - \beta$ against an alternative value of p (3.55) greater than $\frac{1}{2}$. This approximate value of n is

$$n \doteq \frac{(z_\alpha + z_\beta)^2}{4\left(p - \frac{1}{2}\right)^2}. \tag{3.56}$$

As an illustration of the use of (3.56), suppose we are testing H_0 and we wish to have an upper-tail level $\alpha = .04$ test with power $1 - \beta$ of at least .975 against an alternative for which $p = P(Z > 0) = .7$ (recall that $p = .5$ under H_0). The critical values are $z_\alpha = z_{.04} = 1.75$ and $z_\beta = Z_{.025} - 1.96$, and we find that the required sample size for the alternative $p = .7$ is

$$n \doteq \frac{(1.75 + 1.96)^2}{4(.7 - .5)^2} = 86.03.$$

To be conservative, we take $n = 87$.

37. *Consistency of the B Test.* Under the assumption that $Z_1, \ldots, Z_n$ are a random sample from a single continuous population F, the consistency of the tests based on B depends on the parameter

$$p^* = P(Z_1 > 0) - \frac{1}{2}. \tag{3.57}$$

The test procedures defined by (3.40), (3.41), and (3.42) are consistent against the classes of alternatives corresponding to $p^* >, <$, and $\neq 0$, respectively.

Properties

1. *Consistency.* For our consistency statement, we strengthen Assumption B2 to require that each Z has the same continuous population with median θ. Then the test procedures defined by (3.40), (3.41), and (3.42) are consistent against the alternatives $\theta >, <$, and $\neq 0$, respectively. (See also Comment 37.)

2. *Asymptotic Normality.* Under a strengthened Assumption B2 that requires that each Z has the same continuous population with median θ, the asymptotic normality of the standardized form of the B statistic follows from the standard central limit theorem for sums of mutually independent and identically distributed random variables. (See also Comment 35.)
3. *Efficiency.* See Section 3.11.

Problems

43. The data in Table 3.6 are a portion of the data obtained by Cooper et al. (1967). The purpose of their investigation was to determine whether hypnotic susceptibility as measured on objective scales can be changed with practice and training. The objective measures used were the Stanford Profile Scales of Hypnotic Susceptibility, forms I and II (Hilgard, Lauer, and Morgan (1963)). The subjects were administered these Profile Scales, both forms I and II, by a hypnotist other than the experimenter. Each subject was then seen by one of the authors for an extensive period of "hypnotic training." After these sessions were concluded, each subject was retested by a different hypnotist (again not the experimenter) using equivalent forms of the Profile Scales, forms I′ and II′. Table 3.6 gives the average score obtained on forms I and II prior to hypnotic training (X) and the corresponding average score obtained on forms I′ and II′ after the training (Y) for the six subjects. Note that a high (or low) score on the Profile Scales indicates a high (or low) degree of hypnotic susceptibility.

Test the hypothesis of no change in hypnotic susceptibility versus the alternative that hypnotic susceptibility (as measured by the Profile Scales) can be increased with practice and training.

44. Change the value of Y_3 in Table 3.5 from 73 to 173. What effect does this outlying observation have on the calculations performed in Example 3.5? What does this suggest about the relative insensitivity of the sign tests to outliers? Construct an example in which changing one observation has an effect on the final decision regarding rejection or acceptance of H_0.

45. Suppose $n = 25$. Compare the exact P-value of test of $H_0 : \theta = 0$ versus $H_1 : \theta < 0$ based on $B = 8$, with the P-value found using the large-sample approximation.

46. In an investigation to determine the effect of aspirin on bleeding time and platelet adhesion, Bick, Adams, and Schmalhorst (1976) studied the reactions of normal subjects to aspirin. A subset of their data is presented in Table 3.7, where the X observation for each subject is the bleeding time (in seconds) before ingestion of 600 mg of aspirin and the Y observation is the bleeding time (again in seconds) 2 h after administration of the aspirin.

Table 3.6 Average Scores on the Stanford Profile Scales of Hypnotic Susceptibility

Subject i	X_i	Y_i
1	10.5	18.5
2	19.5	24.5
3	7.5	11.0
4	4.0	2.5
5	4.5	5.5
6	2.0	3.5

Source: L. M. Cooper, E. Schubot, S. A. Banford, and C. T. Tart (1967).

Perform the appropriate test of the hypothesis that a 600-mg dose of aspirin has no effect on bleeding time versus the alternative that it typically leads to an increase in bleeding time.

47. Assume that we have a common underlying distribution $F_1 \equiv F_2 \equiv \cdots \equiv F_n \equiv F$ (in Assumption B2). If we have 20 observations, what is the exact power of the level $\alpha = .05$ test of $H_0 : \theta = 0$ versus the alternative $\theta > 0$ when $F(0) = .3$?

48. Assume that we have a common underlying distribution $F_1 \equiv F_2 \equiv \cdots \equiv F_n \equiv F$ (in Assumption B2). If we have 18 observations and F is normal with variance 4, what is the exact power of the level $\alpha = .01$ test of $H_0 : \theta = 0$ versus the alternative $\theta < 0$ when the treatment effect is $\theta = -2$?

49. Consider a level $\alpha = .025$ test of $H_0 : \theta = 0$ versus the alternative $\theta > 0$ based on B. If our data $Z_1, \ldots, Z_n$ are a random sample from a single, continuous distribution $F(\cdot)$, how many n observations will we need to collect in order to have an approximate power of at least .75 against an alternative for which $F(0) = .20$?

50. Apply the appropriate form of the test based on B to the Hamilton depression scale factor IV data in Table 3.1.

51. Assume that we have a common underlying distribution $F_1 \equiv F_2 \equiv \cdots \equiv F_n \equiv F$ (in Assumption B2). If we have 20 observations, what is the approximate power of the level $\alpha = .05$ test of $H_0 : \theta = 0$ versus the alternative $\theta > 0$ when $F(0) = .3$? Compare this approximate power with the exact power from Problem 47.

52. Apply the large-sample approximation test of $H_0 : \theta = 1000$ versus $H_1 : \theta > 1000$ based on B to the salary data in Table 3.2.

53. For the case of $n = 5$ nonzero Z values, use the approach discussed in Comment 30 to obtain the form of the exact null (H_0) distribution of B. Verify numerically that this null distribution is, indeed, the binomial distribution with parameters $n = 5$ and $p_0 = .5$.

54. Consider the test of $H_0 : \theta = 0$ versus $H_1 : \theta > 0$ based on B for the following $n = 15$ Z observations: $Z_1 = 2.5, Z_2 = 0, Z_3 = 3.7, Z_4 = -0.6, Z_5 = 1.7, Z_6 = 0, Z_7 = 5.9, Z_8 = 4.6, Z_9 = 0, Z_{10} = -1.4, Z_{11} = 5.4, Z_{12} = 4.6, Z_{13} = 3.1, Z_{14} = -2.0$, and $Z_{15} = 6.3$. Compute the P-values for the competing B procedures based on either (i) discarding the zero Z values and reducing n accordingly, as recommended in the Ties portion of this section, or (ii) treating

Table 3.7 Bleeding Time (in seconds)

Subject i	X_i	Y_i
1	270	525
2	150	570
3	270	190
4	420	395
5	202	370
6	255	210
7	165	490
8	220	250
9	305	360
10	210	285
11	240	630
12	300	385
13	300	195
14	70	295

Source: R. L. Bick, T. Adams, and W. R. Schmalhorst (1976).

the zero Z values in a conservative manner, as presented in Comment 34. Discuss the results.

55. Consider the same setting as in Problem 54. Suppose that you had decided to use randomization to deal with the three zero Z values in the data (see Comment 34). Consider the various possible outcomes for this randomization process and compute the associated P-value for each of these outcomes. Discuss the implication of these findings in conjunction with the results of Problem 54.

56. Obtain the exact P-value for the test based on B for the bleeding time data in Table 3.7. Compare this to the P-value obtained using the large sample approximation.

57. Obtain the exact P-value for the test of $H_0 : \theta = 1000$ versus $H_1 : \theta > 1000$ based on B for the salary data in Table 3.2.

3.5 AN ESTIMATOR ASSOCIATED WITH THE SIGN STATISTIC (HODGES–LEHMANN)

Procedure

To estimate the treatment effect θ, order the sample observations and let $Z^{(1)} \leq \cdots \leq Z^{(n)}$ denote these ordered items. The estimator of θ associated with the sign statistic (see Comment 38) is

$$\tilde{\theta} = \text{median}\{Z_i,\ 1 \leq i \leq n\}. \tag{3.58}$$

Thus, if n is odd, say $n = 2k + 1$, we have $k = (n - 1)/2$ and

$$\tilde{\theta} = Z^{(k+1)}, \tag{3.59}$$

the value that occupies position $k + 1$ in the list of the ordered Z_i values. If n is even, say $n = 2k$, then $k = n/2$ and

$$\tilde{\theta} = \frac{Z^{(k)} + Z^{(k+1)}}{2}; \tag{3.60}$$

that is, when n is even, $\tilde{\theta}$ is the average of the two Z_i values that occupy positions k and $k + 1$ in the ordered list of the n data values.

EXAMPLE 3.6 ***Continuation of Example 3.5.***

To estimate θ for the beak-clapping data in Table 3.5, we first form the $n = 25$ ordered Z values, namely, $Z^{(1)} \leq \cdots \leq Z^{(25)}$: $-8.5, -4.6, -1.8, -0.8$, 1.9, 3.9, 4.7, 7.1, 7.5, 8.5, 14.8, 16.7, 17.6, 19.7, 20.6, 21.9, 23.8, 24.7, 24.7, 25.0, 40.7, 46.9, 48.3, 52.8, and 54.0. The sample size $n = 25$ is odd, so we use (3.59) with $k = (25 - 1)/2 = 12$ to obtain the estimate $\tilde{\theta} = Z^{(13)} = 17.6$ for the treatment effect θ. Thus, we estimate that a typical chick embryo of the type included in this study will produce 17.6 more beak-claps per minute during periods of illumination than during periods of darkness.

The `SIGN.test` command will provide this value automatically as seen in the `R` output in Example 3.5. Alternatively, one may use the command `median(z)` directly on the difference data $Z_i = Y_i - X_i$.

Comments

38. *Motivation for the Hodges–Lehmann Estimator.* The estimator $\tilde{\theta}$ defined by (3.58) is associated with the sign test in the same way as the estimator $\tilde{\theta}$ (3.23) is associated with the signed rank test (see Comment 15). When $\theta = 0$, the distribution of the statistic B (3.39) is symmetric about its mean $n/2$ (see Comment 33). A natural estimator of θ is the amount $\tilde{\theta}$ (say) that should be subtracted from each Z_i so that the value of B, when applied to the shifted sample $Z_1 - \tilde{\theta}, \ldots, Z_n - \tilde{\theta}$, is as close to $n/2$ as possible. Intuitively, we estimate θ by the amount ($\tilde{\theta}$) that the Z sample should be shifted in order that $Z_1 - \tilde{\theta}, \ldots, Z_n - \tilde{\theta}$ appears (when "viewed" by the sign statistic B) as a sample from a population with median 0. (Under Assumptions B1 and B2, each of the $Z_1 - \theta, \ldots, Z_n - \theta$ variables is from a population with median 0.)

    The Hodges–Lehmann method can be applied to a large class of statistics containing both B and T^+ (3.3). However, the forms of the resulting estimator for other members of this class are not always as convenient for calculation as are $\tilde{\theta}$ (3.58) or $\tilde{\theta}$ (3.23). See Hodges and Lehmann (1983) for an expository article on their method of estimation.

39. *Simplicity.* One of the virtues of $\tilde{\theta}$ (3.58) is its simplicity. While many estimators associated with distribution-free test statistics are tedious to compute (e.g., $\tilde{\theta}$ (3.23) requires computing the median of $n(n+1)/2$ values), $\tilde{\theta}$ requires only that we find the median of the n Z observations. However, although the signs of the Z differences provide sufficient information to conduct a sign test, the magnitudes of these differences are needed to obtain the value of the estimator $\tilde{\theta}$.

40. *Sensitivity to Gross Errors.* The estimator $\tilde{\theta}$ (3.58) is even less sensitive to outliers than the estimator $\hat{\theta}$ (3.23) associated with the signed rank statistic T^+ (3.3). (See Comment 16 and Problems 20 and 60.) As a result, $\tilde{\theta}$ protects well against gross errors. However, all the information contained in the collected sample is not utilized in computing $\tilde{\theta}$. Consequently, $\tilde{\theta}$ is rather inefficient for many populations.

41. *Zero Z Values.* Note that in calculating the estimator $\tilde{\theta}$, we use *all* the Z differences. Although we recommend (see Ties in Section 3.4) discarding the zero Z values (and reducing n accordingly) prior to applying the sign test to the data, it is not necessary to do so when calculating $\tilde{\theta}$. In fact, the zero Z values contain important information about the magnitude of the treatment effect. This is also the case when we consider (Section 3.6) confidence intervals and bounds for θ.

42. *Historical Perspective.* The use of the estimator $\tilde{\theta}$ predates most of the recent unified developments in the field of nonparametric statistics. A. T. Craig (1932) first found the sampling distribution of $\tilde{\theta}$, and its asymptotic properties were developed shortly thereafter by Smirnov (1935).

43. *Quasimedians.* Let $Z^{(1)} \leq \cdots \leq Z^{(n)}$ be the ordered sample observations, as in step 1 of the Procedure. Hodges and Lehmann (1967) defined the sample

quasimedians by

$$\tilde{\theta}_i = \begin{cases} \dfrac{Z^{(k+1-i)} + Z^{(k+1+i)}}{2}, & \text{if } n = 2k+1 \\ \dfrac{Z^{(k-i)} + Z^{(k+1+i)}}{2}, & \text{if } n = 2k, \end{cases}$$

for $i = 0, 1, \ldots, k$ if $n = 2k + 1$, or $i = 0, 1, \ldots, k - 1$ if $n = 2k$; that is, each quasimedian $\tilde{\theta}_i$ is an average of two symmetrically situated, ordered Z observations. (Note that this definition of a quasimedian generalizes the concept of a sample median, as the sample median $\tilde{\theta}$ (3.58) is equal to $\tilde{\theta}_0$.) These quasimedians are natural estimators for the parameter θ (see Comment 52) and were considered by Hodges and Lehmann (1967), who investigated some of the asymptotic properties of this class of statistics.

44. *Linear Combinations of Order Statistics.* Let $Z^{(1)} \leq \cdots \leq Z^{(n)}$ be the ordered sample observations, as in step 1 of the Procedure. Under the additional assumption that we have a common underlying distribution $F_1 \equiv F_2 \equiv \cdots \equiv F_n \equiv F$ (in Assumption B2), the n variables $Z^{(1)}, \ldots, Z^{(n)}$ are called the *order statistics* for the random sample $Z_1, \ldots, Z_n$. The estimator $\tilde{\theta}$ (3.58) is a special case of a general class of estimators of θ based on linear combinations of these sample order statistics, corresponding to estimators of the form

$$\tilde{\theta}_{\mathbf{b}} = \sum_{i=1}^{n} b_i Z^{(i)}, \tag{3.61}$$

where $\mathbf{b} = (b_1, \ldots, b_n)$ is a vector of n nonnegative constants such that $\sum_{i=1}^{n} b_i = 1$. For a more detailed discussion about estimators of the form $\tilde{\theta}_{\mathbf{b}}$ (3.61), see, for example, David and Nagaraja (2003) or Arnold, Balakrishnan, and Nagaraja (1992).

45. *Variance Approximation.* Hodges and Lehmann (1967) obtained an approximation for the variance of the estimator $\tilde{\theta}$ (3.58) under the additional assumption that we have a common underlying distribution $F_1 \equiv F_2 \equiv \cdots \equiv F_n \equiv F$ (in Assumption B2). (See equation (1.4) of their paper.) They point out that, up to the accuracy of their approximation, it is not wise to compute the sample median $\tilde{\theta}$ using an odd number of observations, say $n = 2k + 1$. The next smaller even number, $n = 2k$, yields a sample median that is just as accurate. This conclusion does not depend on the shape of the underlying population except that it be symmetric, although the degree of accuracy of the approximation is affected by the shape.

46. *Estimating the Asymptotic Standard Deviation of* $\tilde{\theta}$. Assume that we have a common underlying distribution $F_1 \equiv F_2 \equiv \cdots \equiv F_n \equiv F$ (in Assumption B2) and set

$$D = \sum_{i=1}^{n} a_i,$$

where

$$a_i = \begin{cases} 1, & \text{if } \left[\tilde{\theta} - (n)^{-1/5}\right] \leq Z_i \leq \left[\tilde{\theta} + (n)^{-1/5}\right] \\ 0, & \text{otherwise,} \end{cases}$$

for $i = 1, \ldots, n$. Let $A =$ maximum $\{1, D\}$. Under the additional assumption on the common distribution F that the probability of obtaining a Z observation in any (sufficiently) small interval I centered at the median θ is greater than or equal to some fixed constant (not depending on I) times the length of I, the statistic $C = n^{3/10}A^{-1}$ is a consistent estimator of the asymptotic standard deviation of the point estimator $\tilde{\theta}$ (3.58). The statistic C is related to general classes of estimators of probability density functions considered by Rosenblatt (1956), Parzen (1962), and Gupta (1967). The consistency of C follows directly from the results in Korwar (1971).

47. *Relative Merits of $\hat{\theta}$ and $\tilde{\theta}$.* The point estimator $\tilde{\theta}$ (3.58) associated with the sign test statistic B is to be preferred to the point estimator $\tilde{\theta}$ (3.23) associated with the signed rank test statistic T^+ when ease of computation is a consideration (see Comment 39). Generally (but not always), $\hat{\theta}$ is more efficient than $\tilde{\theta}$. (See Comment 40 and Section 3.11.)

Properties

1. *Standard Deviation of $\tilde{\theta}$.* For the asymptotic standard deviation of $\tilde{\theta}$ (3.58), see Fisz (1963, p. 383) and Comment 46.
2. *Asymptotic Normality.* See Fisz (1963, p. 383).
3. *Efficiency.* See Hodges and Lehman (1963) and Section 3.11.

Problems

58. Using the designated X and Y associations, estimate θ for the average Profile Scales data of Table 3.6.

59. Estimate θ for the bleeding time data of Table 3.7.

60. Change the value of Y_3 in Table 3.5 from 73 to 173. What effect does this have on the value of $\bar{Z} = \sum_{i=1}^{25} Z_i/25$? What is the new value of $\tilde{\theta}$ (3.58)? Interpret these calculations. (See Comment 40.)

61. Calculate $\tilde{\theta}$ for the heat-insoluble hydroxyproline data of Table 3.4. Compare with the value of $\hat{\theta}$ obtained in Problem 21.

62. (a) What happens to $\tilde{\theta}$ when we add a number b to each of the sample values $Z_1, \ldots, Z_n$?

(b) What happens to $\tilde{\theta}$ when we multiply each sample value by the number textitd?

(c) What happens to $\tilde{\theta}$ when we discard the k largest and k smallest values from the sample (assume $n > 2k$)? Compare your answers with the corresponding answers to Problem 23.

63. Calculate $\tilde{\theta}$ for the blood-level data of Table 3.3. Compare with the value of $\hat{\theta}$ obtained in Problem 19.

64. Calculate $\tilde{\theta}$ for the salary data in Table 3.2. Compare with the value of $\hat{\theta}$ obtained in Problem 18.

65. Calculate $\tilde{\theta}$ for the Hamilton depression scale factor IV data in Table 3.1. Compare with the value of $\tilde{\theta}$ obtained in Example 3.3.

66. Find the vector $\mathbf{b} = (b_1, \ldots, b_n)$ to show that $\tilde{\theta}$ can be written as a linear combination of the sample order statistics $Z^{(1)} \leq \cdots \leq Z^{(n)}$, as discussed in Comment 44.

67. Show that the class of quasimedian estimators of θ (see Comment 43) is a subset of the class of estimators of θ based on linear combinations of the sample order statistics $Z^{(1)} \leq \cdots \leq Z^{(n)}$, as discussed in Comment 44.

68. Find the sample quasimedians (see Comment 43) for the data in Table 3.2. How do these values compare with $\tilde{\theta}$?

69. Find the sample quasimedians (see Comment 43) for the data in Table 3.7. How do these values compare with $\tilde{\theta}$?

3.6 A DISTRIBUTION-FREE CONFIDENCE INTERVAL BASED ON THE SIGN TEST (THOMPSON, SAVUR)

Procedure

For a symmetric two-sided confidence interval for θ, with confidence coefficient $1 - \alpha$, first obtain the upper $(\alpha/2)$nd percentile point $b_{\alpha/2,1/2}$ of the null distribution of B from `qbinom`. Set

$$C_\alpha = n + 1 - b_{\alpha/2,1/2}. \tag{3.62}$$

The $100(1 - \alpha)\%$ confidence interval $(\theta_{\mathrm{L}}, \theta_{\mathrm{U}})$ for θ that is associated with the two-sided sign test (see Comment 48) of $H_0 : \theta = 0$ is then given by

$$\theta_{\mathrm{L}} = Z^{(C_\alpha)}, \theta_{\mathrm{U}} = Z^{(n+1-C_\alpha)} = Z^{(b_{\alpha/2,1/2})}, \tag{3.63}$$

where $Z^{(1)} \leq \cdots \leq Z^{(n)}$ are the ordered sample observations; that is, θ_L is the sample observation that occupies position C_α in the list of ordered sample data. The upper end point θ_U is the sample observation that occupies position $n+1 - C_\alpha = b_{\alpha/2,1/2}$ in this ordered list. With θ_L and θ_{U} given by display (3.63), we have

$$P_\theta(\theta_{\mathrm{L}} < \theta < \theta_{\mathrm{U}}) = 1 - \alpha \text{ for all } \theta. \tag{3.64}$$

(For upper or lower confidence bounds for θ associated with appropriate one-sided sign tests of H_0: $\theta = 0$, see Comment 49.)

Large-Sample Approximation

For large n, the integer C_α may be approximated by

$$C_\alpha \approx \frac{n}{2} - z_{\alpha/2}\left(\frac{n}{4}\right)^{1/2}. \tag{3.65}$$

In general, the value of the right-hand side of (3.65) is not an integer. To be conservative, take C_α to be the largest integer that is less than or equal to the right-hand side of (3.65).

EXAMPLE 3.7 ***Continuation of Examples 3.5 and 3.6.***

Consider the beak-clapping data in Table 3.5. We illustrate how to obtain the 95% confidence interval for θ. With $1-\alpha = .95$, $n = 25$, and $p = 1/2$ we see that $b_{\alpha,1/2} = b_{0.025,1/2} = 18$. From (3.62), it follows that

$$C_\alpha = 25 + 1 - 18 = 8.$$

Using C_α and $b_{\alpha,1/2}$ in (3.63), we see that

$$\theta_{\mathrm{L}} = Z^{(8)} = 7.1 \text{ and } \theta_{\mathrm{U}} = Z^{(18)} = 24.7$$

so that the 95% confidence interval for θ is

$$(\theta_{\mathrm{L}}, \theta_{\mathrm{U}}) = \left(Z^{(8)}, Z^{(18)}\right) = (7.1, 24.7).$$

The size of this confidence interval is the same as the probability a binomial random variable with parameters $n = 25$ and $p = 1/2$ is in the interval (8,18). Using `pbinom`, this is .9567. Thus, the actual confidence level is not $\alpha = .05$, but $\alpha = .0433$. This is due to the discrete nature of the statistic. `SIGN.test` provides this confidence interval. The value of $1-\alpha$ is specified through the argument `conf.level`. In the output below, three confidence intervals are provided. As it is not possible to get an interval with exactly .05 in the tails, intervals bracketing this α are given. The lower achieved interval is the exact intervals with $\alpha = 1 - .8922$, the upper achieved interval is the exact interval with $\alpha = 1 - .9567$ found above. The interpolated interval is found by linearly interpolating the lower and upper end points on $1-\alpha$. Two-sided confidence intervals are provided when the alternative hypothesis is two-sided. This is done with the argument `alternative="two.sided"` in `SIGN.test`.

```
                  Conf.Level   L.E.pt     U.E.pt
Lower Achieved CI       .8922   7.5000    23.8000
Interpolated CI         .9500   7.1417    24.6063
Upper Achieved CI       .9567   7.1000    24.7000
```

Comments

48. *Relationship of Confidence Interval to Two-Sided Test.* The $100(1-\alpha)\%$ confidence interval for θ given by display (3.63) can be obtained from the two-sided sign test as follows. The confidence interval $(\theta_{\mathrm{L}}, \theta_{\mathrm{U}})$ consists of those θ_0 values for which the two-sided α-level test of $\theta = \theta_0$ (see Comment 28) does not reject the hypothesis $\theta = \theta_0$.
49. *Confidence Bounds.* Often we are interested only in making one-sided confidence statements about the parameter θ; that is, we wish to assert with specified confidence that θ is no larger (or, in other settings, no smaller) than some upper (lower) confidence bound based on the sample data. To obtain such one-sided confidence bounds for θ, we proceed as follows. For the specified confidence coefficient $1-\alpha$, find the upper αth (not $(\alpha/2)$nd, as for the confidence interval) percentile point $b_{\alpha,1/2}$ of the null distribution of B. Set

$$C_\alpha^* = n+1 - b_{\alpha,1/2}. \tag{3.66}$$

The $100(1-\alpha)\%$ lower confidence bound θ_{L}^* for θ that is associated with the one-sided sign test of $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta > \theta_0$ is then given by

$$(\theta_{\mathrm{L}}^*, \infty) = (Z^{(C_\alpha^*)}, \infty), \tag{3.67}$$

where, as before, $Z^{(1)} \leq \cdots \leq Z^{(n)}$ are the ordered sample observations. With θ_L^* given by display (3.67), we have

$$P_\theta(\theta_{\mathrm{L}}^* < \theta < \infty) = 1 - \alpha \text{ for all } \theta. \tag{3.68}$$

The corresponding $100(1-\alpha)\%$ *upper* confidence bound $\theta_{\mathrm{U}}^\star$ for θ that is associated with the one-sided sign test of $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta < \theta_0$ is given by

$$(-\infty, \theta_{\mathrm{U}}^*) = (-\infty, Z^{(n+1-C_\alpha^*)}) = \left(-\infty, Z^{(b_{\alpha,1/2})}\right), \tag{3.69}$$

where C_α^* is given in (3.66). It follows that

$$P_\theta(-\infty < \theta < \theta_{\mathrm{U}}^*) = 1 - \alpha \text{ for all } \theta. \tag{3.70}$$

For large n, the integer C_α^* may be approximated by

$$C_\alpha^* \approx \frac{n}{2} - z_\alpha \left(\frac{n}{4}\right)^{1/2}. \tag{3.71}$$

As with C_α (3.65) and the confidence interval for θ, the value of the right-hand side of (3.71) is not an integer. To be conservative, take C_α^* to be the largest integer that is less than or equal to the right-hand side of (3.71).

The $100(1-\alpha)\%$ lower and upper confidence bounds θ_L^* (3.67) and θ_{U}^* (3.69) are related to the acceptance regions of the one-sided sign tests of H_0: $\theta = \theta_0$ against the alternatives $\theta > \theta_0$ and $\theta < \theta_0$, respectively, in the same way that the confidence interval $(\theta_{\mathrm{L}}, \theta_{\mathrm{U}})$ is related to the acceptance region of the two-sided sign test of H_0: $\theta = \theta_0$ (see Comment 48). When using `SIGN.test`, one-sided confidence intervals are produced by specifying a one-sided alternative.

50. *Zero Z Values.* Note that in calculating the confidence interval $(\theta_{\mathrm{L}}, \theta_{\mathrm{U}})$ from display (3.63) or the confidence bounds θ_{L}^* (3.67) or $\theta_{\mathrm{U}}^\star$ (3.69) for θ, we use *all* the Z differences. This is in common with our recommendation (see Comment 41) for computing the point estimator $\tilde{\theta}$ (3.58), but differs from the recommended policy (see Ties in Section 3.4) of discarding the zero Z values (and reducing n accordingly) prior to applying the sign test to the data. However, if there are zero Z's in the data, the equivalence (discussed in Comments 48 and 49) between the acceptance regions of the one-sided and two-sided sign tests and the appropriate confidence bound and confidence interval, respectively, are no longer valid.

51. *Necessity of Magnitudes.* The confidence interval and bounds (see Comment 49) for θ based on the sign tests are simple to compute, as the end points depend only on the ordered sample Z observations. However, for such a computation, knowledge of the signs of the Z differences is no longer sufficient as it was for the computation of B (3.39) for the various sign tests. We need the observation magnitudes to obtain $\theta_{\mathrm{L}}, \theta_{\mathrm{U}}, \theta_{\mathrm{L}}^*$, or θ_{U}^*.

52. *Midpoint of the Confidence Interval as an Estimator.* The midpoint of the interval (3.63), namely, $[Z^{(C_\alpha)} + Z^{(n+1-C_\alpha)}]/2$, is also a natural estimator of θ. (Note that this actually yields a class of estimators, depending on the value of α.) In general, this midpoint is not the same as $\tilde{\theta}$ (3.58). (See Hodges and Lehmann (1967) and Comment 43 for additional discussion of this midpoint class of estimators.)
53. *Comparison of Sign and Signed Rank Confidence Intervals for* θ. The confidence interval (3.63) for θ associated with the sign test and based on the n ordered Z differences is easier to compute than the confidence interval (3.27) for θ associated with the signed rank test and based on the $n(n+1)/2$ ordered Walsh averages (see Comment 17). However, the signed rank confidence interval (3.27) is generally (but not always) more efficient than the sign confidence interval (3.63). (See Section 3.11.)
54. *Extension to Discrete Distributions.* Consider the closed version $[\theta_L, \theta_U] = [Z^{(C_\alpha)}, Z^{(n+1-C_\alpha)}]$ of the $100(1-\alpha)\%$ confidence interval for θ given in display (3.63) under the alternative (to Assumptions B1 and B2) assumption that $Z_1, \ldots, Z_n$ are a random sample from an underlying distribution $F(\cdot)$ with a *unique* median θ. Suppose that this common distribution $F(\cdot)$ is such that in any bounded interval of the real line there are at most a finite number (could be zero) of values having positive probability. (If $F(\cdot)$ is continuous, this is trivially satisfied because in that case no real number has positive probability. However, the large majority of discrete probability distributions also satisfy this mild assumption.) Under these weakened conditions on the common $F(\cdot)$, the closed interval $[\theta_L, \theta_U]$ remains a conservative $100(1-\alpha)\%$ confidence interval for θ in the sense that

$$P_\theta(\theta_L \leq \theta \leq \theta_U) \geq 1 - \alpha \quad \text{for all} \quad \theta$$

is guaranteed for every such $F(\cdot)$. The closed versions of the upper and lower confidence bounds (see Comment 49), namely, $(-\infty, \theta_U^\star] = (-\infty, Z^{(b_{\alpha,1/2})}]$ and $[\theta_L^\star, \infty) = [Z^{(C_\alpha^\star)}, \infty)$, respectively, also remain conservative $100(1-\alpha)\%$ bounds over this expanded class of common distributions $F(\cdot)$. (For more details on the extension of these confidence intervals and bounds to common discrete distributions, see Scheffé and Tukey (1945) and Noether (1967a).)

Properties

1. *Distribution-Freeness.* For populations satisfying Assumptions B1 and B2, (3.64) holds. Hence, we can control the coverage probability to be $1-\alpha$ without having more specific knowledge about the forms of the underlying Z distributions. Thus, (θ_L, θ_U) is a distribution-free confidence interval for θ over a very large class of populations. (See also Comment 54.)
2. *Efficiency.* See Section 3.11.

Problems

70. For the Profile Scales data of Table 3.6, obtain a confidence interval for θ with the exact confidence coefficient .9688.

71. For the bleeding time data in Table 3.7, obtain a confidence interval for θ with the exact confidence coefficient .9426.

72. For the beak-clapping data of Table 3.5, obtain an estimate for the asymptotic standard deviation of $\tilde{\theta}$. (See Comment 46.)

73. For the beak-clapping data of Table 3.5 and $\alpha = .1078$, calculate the point estimator of θ defined in Comment 52. Compare with the value of $\tilde{\theta}$ obtained in Example 3.6.

74. For the Hamilton depression scale factor IV data of Table 3.1, find a confidence interval for θ with the exact confidence coefficient .9610.

75. For the bleeding time data in Table 3.7, obtain an approximate 94.26% confidence interval for θ using the large-sample approximation of this section. Compare this approximate confidence interval with the exact 94.26% confidence interval obtained in Problem 71.

76. How does varying α affect the length of the confidence interval defined by display (3.63)? How does it affect the point estimator of θ defined in Comment 52?

77. For the beak-clapping data of Table 3.5, find a lower confidence bound for θ with the exact confidence coefficient .9461. (See Comment 49.)

78. Consider the Stanford Profile Scores data of Table 3.6. Obtain an upper confidence bound for θ with the exact confidence coefficient .8906. (See Comment 49.)

79. For the salary data in Table 3.2, find a lower confidence bound for θ with the exact confidence coefficient .9270. How does this compare with the approximate 93.6% lower confidence bound for θ obtained in Problem 32?

80. Consider the beak-clapping data of Table 3.5. Use the large-sample approximation to obtain an approximate 95% lower confidence bound for θ (see Comment 49). Compare this approximate bound with the exact 94.61% lower confidence bound obtained in Problem 77.

81. Consider the bleeding time data of Table 3.7. Use the large-sample approximation to find an approximate 92% upper confidence bound for θ. (See Comment 49.)

82. Consider the case $n = 15$ and compare the length of the exact 96.48% confidence interval for θ given by display (3.63), with the length of the approximate 96.48% confidence interval for θ obtained using the large-sample approximation of this section.

83. Consider the case $n = 25$ and compare the exact 94.61% lower confidence bound for θ given by (3.67), with the approximate 94.61% lower confidence bound for θ obtained from the large-sample approximation in Comment 49.

84. For the bleeding time data in Table 3.7 and $\alpha = 0.05$, find the estimate of θ as described in Comment 52. Compare this with the estimate found in Problem 59.

85. Consider the two-sided confidence interval found in Problem 74. What range of α values results in the same upper and lower bounds?

86. Consider the one-sided confidence interval found in Problem 77. What range of α values results in the same lower bound?

ONE-SAMPLE DATA*

3.7 PROCEDURES BASED ON THE SIGNED RANK STATISTIC

Data. We obtain n observations $Z_1, \ldots, Z_n$.

*Sections 3.7–3.10 are optional. The contents of these sections are not used in the sequel.

Table 3.8 Estimated Values of θ from the Mariner and the Pioneer Spacecraft

Spacecraft	θ
Mariner 2 (Venus)	81.3001
Mariner 4 (Mars)	81.3015
Mariner 5 (Venus)	81.3006
Mariner 6 (Mars)	81.3011
Mariner 7 (Mars)	81.2997
Pioneer 6	81.3005
Pioneer 7	81.3021

Source: J. D. Anderson, L. Efron, and S. K. Wong (1970).

Assumptions

C1. The Z's are mutually independent.

C2. Each Z comes from a population (not necessarily the same) that is continuous and symmetric about θ.

Procedures

To test $H_0 : \theta = \theta_0$, where θ_0 is some specified number, we create the modified observations $Z_i' = Z_i - \theta_0$, for $i = 1, \ldots, n$. Then we apply any of the test procedures of Section 3.1 to these modified Z' observations.

To obtain a point estimator of θ or a confidence interval for θ, we apply the procedures of Sections 3.2 and 3.3 directly to the Z observations without modification.

EXAMPLE 3.8 ***The Mariner and the Pioneer Spacecraft Data.***

The data in Table 3.8 were reported by Anderson, Efron, and Wong (1970). The seven observations represent average measurements of θ, the ratio of the mass of the Earth to that of the moon, obtained from seven different spacecraft.

On the basis of the previous (2–3 years earlier) Ranger spacecraft findings, scientists had considered the value of θ to be approximately 81.3035. Thus, with the data of Table 3.8, we are interested in testing $H_0 : \theta = 81.3035$ versus the alternative $\theta \neq 81.3035$, and we perform test procedure (3.6). With $\alpha = .078$, we see that $t_{.078/2} = 26$.

Now, we form the modified Z' observations as follows.

i	Z_i	$Z_i' = Z_i - 81.3035$
1	81.3001	−.0034
2	81.3015	−.0020
3	81.3006	−.0029
4	81.3011	−.0024
5	81.2997	−.0038
6	81.3005	−.0030
7	81.3021	−.0014

Using the computational setup of Section 3.1 on the Z' observations, we calculate $T^+ = 0$. Thus, we reject $H_0 : \theta = 81.3035$ at the $\alpha = .078$ level, since $T^+ = 0 \leq [28 - t_{.039}] = 2$. The P-value for this symmetric test based on T^+ is `2 * psignrank(0, n=7, lower.tail=T)=.0156`. This test is implemented with `wilcox.test`. For this example,

```
wilcox.test(z, mu=81.3035)
```

results in the output

```
Wilcoxon signed rank test

data: z
V = 0, p-value = .01563
alternative hypothesis: true location is not equal to
    81.3035
```

Note the use of the symbol V in place of T^+

For the large-sample approximation, we see from (3.9) that

$$T^* = \frac{0 - [7(8)/4]}{[7(8)(15)/24]^{1/2}} = -2.366.$$

Thus, the smallest significance level at which we could reject H_0 by using a symmetric test based on the normal approximation is .018. This means that both the exact test and the large-sample approximation indicate the existence of strong evidence to reject the findings of the earlier Ranger spacecraft that $\theta = 81.3035$.

The ordered values of $(Z_i + Z_j)/2$ are $W^{(1)} \leq \cdots \leq W^{(28)}$: 81.2997, 81.2999, 81.3001, 81.3001, 81.30015, 81.3003, 81.30035, 81.3004, 81.3005, 81.30055, 81.3006, 81.3006, 81.3006, 81.3008, 81.3008, 81.30085, 81.3009, 81.3010, 81.30105, 81.3011, 81.3011, 81.3013, 81.3013, 81.30135, 81.3015, 81.3016, 81.3018, and 81.3021. If $M = 7(8)/2 = 28$, we see that $M = 2k$ with $k = 14$. Thus, from (3.25), we have

$$\hat{\theta} = \frac{W^{(14)} + W^{(15)}}{2} = \frac{81.3008 + 81.3008}{2} = 81.3008.$$

With $n = 7$ and $\alpha = .046$, we find that $t_{\alpha/2} = t_{.023} = 27$. Thus, $C_{.046} = \{7(8)/2\} + 1 - t_{.023} = 28 + 1 - 27 = 2$.

From (3.27), it follows that

$$\theta_{\mathrm{L}} = W^{(2)} = 81.2999 \text{ and } \theta_{\mathrm{U}} = W^{(27)} = 81.3018$$

so that our 95.4% confidence interval for θ is

$$(\theta_{\mathrm{L}}, \theta_{\mathrm{U}}) = (81.2999, 81.3018).$$

The above results may be produced in R through the function call

```
wilcox,test(z, mu=81.3035, exact=T, conf.int=T,
    conf.level=l-.046)
```

where z is a vector containing the seven measurements from Table 3.8. The above function call extends the R output given in Example 3.8:

```
  Wilcoxon signed rank test

data: z
V = 0, p-value = .01563
alternative hypothesis: true location is not equal to
    81.3035
95.4 percent confidence interval:
81.2999 81.3018
sample estimates:
(pseudo)median
81.3008
```

Applying the large-sample approximation, we find from (3.29) that

$$C_{.046} \approx [7(8)/4] - 1.996[7(8)(15)/24]^{1/2} \approx 2,$$

resulting in the same interval.

It is important to comment that in applying the procedures based on the signed rank statistic T^+ (3.3), we made the assumption that the population of average θ measurements for each of the satellites was symmetric about θ. (For a test of this basic assumption, see Section 3.9.) We also note that this set of data provides an example in which the populations of the Z observations are probably not the same (see Assumption C2).

Comments

55. *Assumptions.* Note that Assumption Al for the paired replicates procedures based on the signed rank statistic is not necessary for the one-sample data because these data need not consist of differences for paired observations.

Properties

1. The properties of the one-sample procedures based on the signed rank statistic are essentially the same as those of the corresponding paired replicates procedures. An exception occurs in the efficiencies of the procedures and is due to the difference in the type of data for the two problems. See Section 3.11 for a discussion of the difference in efficiencies of the procedures of Sections 3.1–3.3 when they are applied to single-sample problems.

Problems

87. The data in Table 3.9 are a subset of the data reported by Ijzermans (1970) from an investigation on the susceptibility to corrosion of 18Cr_10Ni_2Mo stainless steel (i.e., stainless steel containing 18% chromium, 10% nickel, and 2% molybdenum by weight).

Twelve specimens of steel were selected for use in the corrosion study. Although Ijzermans' experiment was directed toward corrosion, we are concerned here with the quality of the steel from which the stainless steel samples were chosen. Table 3.9 gives the percentage of chromium in the 12 samples used by Ijzermans.

Test the hypothesis that the median percentage of chromium content (θ) of the steel is 18% against the alternative that it is not 18%. Obtain a point estimate of θ and find a confidence interval for θ with the confidence coefficient .936.

88. For the percentage of chromium data in Table 3.9, obtain a point estimate of θ from the midpoint of the confidence interval calculated in Problem 87 (see Comment 23). Compare with the point estimate obtained in Problem 87.

89. Compute $\hat{\theta}$ for the settling velocity data of Table 3.12 and compare with the value of $\tilde{\theta}$ obtained in Example 3.9.

90. Lamp (1976) studied the age distribution of a common mayfly species, *Stenacron interpunctatum*, among various habitats in Big Darby Creek, Ohio. One of the measurements considered was head width (in micrometer divisions, 1 division = .0345 mm); a subset of Lamp's data from the mayflies in habitat A is presented in Table 3.10.

Test the hypothesis that the median head width for mayflies from habitat A (θ) is 22 μm divisions against the alternative that it is greater than 22 μm. Obtain a point estimate of θ and find a lower confidence bound (see Comment 21) for θ with the confidence coefficient .976.

91. The data in Table 3.11 are a subset of the data obtained by Poland et al. (1970) in an experiment concerned with the effect of occupational exposure to DDT on human drug and steroid metabolism. The DDT-exposed subjects were employees of the Montrose Chemical Corporation, who had been working in the DDT plant at Torrance, California, for more than 5 years.

Table 3.9 Percentage of Chromium in the Stainless Steel Samples

Steel sample	% of Cr
1	17.4
2	17.9
3	17.6
4	18.1
5	17.6
6	18.9
7	16.9
8	17.5
9	17.8
10	17.4
11	24.6
12	26.0

Source: A. B. Ijzermans (1970).

Table 3.10 Mayfly Head Width, Habitat A (Micrometer Divisions)

Mayfly i	Z_i
1	36
2	31
3	30
4	27
5	20
6	33
7	27
8	18
9	19
10	28

Source: W. O. Lamp (1976).

Table 3.11 6β-Hydroxycortisol Excretion (μg/24 h)

Worker i	Z_i
1	254
2	171
3	345
4	134
5	190
6	447
7	106
8	173
9	449
10	198

Source: A. Poland, D. Smith, R. Kuntzman, M. Jacobson, and A. H. Conney (1970).

Table 3.12 Settling Velocities at 22 °C

Sample i	Z_i, cm/s
1	12.9
2	13.7
3	14.5
4	13.3
5	12.8
6	13.8
7	13.4

Source: J. D. Smith (1969).

Table 3.13 Oxidant Content of Dew Water, Port Burwell, 1960

Sample i	Z_i, ppm ozone
1	.32
2	.21
3	.28
4	.15
5	.08
6	.22
7	.17
8	.35
9	.20
10	.31
11	.17
12	.11

Source: A. F. W. Cole and M. Katz (1966).

All these individuals had received moderate to intense occupational exposure to DDT, and all were in good health. One of the measures used in the study was the 24-h urinary excretion of 6β-hydroxycortisol.

Test the hypothesis that the median 6β-hydroxycortisol excretion rate for subjects with occupational exposure to DDT similar to the workers in this study (θ) is 175 μg/24 h against

the alternative that it is greater than 175. Obtain a point estimate of θ and find a confidence interval for θ with the confidence coefficient .916.

92. Consider the oxidant content of dew water data in Table 3.13. Use the computer software R to test the hypothesis that the median oxidant content of dew water (θ) was .25 against the alternative that it was less than .25. Also use R to obtain a point estimate of θ and find an upper confidence bound (see Comment 21) for θ with the confidence coefficient .961. Compare with the answers to Problem 94.

93. Consider the settling velocity data of Table 3.12. Use the computer software R to test the hypothesis that the median settling velocity for the Middle Ground sand ridge (θ) was 14 cm/s against the alternative that it was not equal to 14 cm/s. Also use R to obtain a point estimate of θ and find a confidence interval for θ with the confidence coefficient .890. Compare with the results obtained in Example 3.9.

3.8 PROCEDURES BASED ON THE SIGN STATISTIC

Data. We obtain n observations $Z_1, \ldots, Z_n$.

Assumptions

D1. The Z's are mutually independent.

D2. Each Z comes from the same continuous population with median θ, so that $P(Z_i > \theta) = P(Z_i > \theta) = \frac{1}{2}, i = 1, \ldots, n$.

Procedures

To test $H_0 : \theta = \theta_0$, where θ_0 is some specified number, we form the modified observations $Z_i' = Z_i - \theta_0$, for $i = 1, \ldots, n$. Then we can apply any of the test procedures of Section 3.4 to these modified Z' observations. (In the test of H_0, we can weaken Assumption D2 to D2′, namely, that each Z comes from a population, not necessarily the same population, such that $P(Z_i < \theta_0) = P(Z_i > \theta_0) = \frac{1}{2}, i = 1, \ldots, n$, when θ is equal to the hypothesized value θ_0.)

To obtain a point estimator of θ or a confidence interval for θ, we apply the procedures of Sections 3.5 and 3.6 directly to the Z observations without modification.

EXAMPLE 3.9 ***Sediment Settling Velocities.***

The data in Table 3.12 are a subset of the data obtained by Smith (1969) in an experiment investigating the geomorphology of the Middle Ground sand ridge, which is located in Vineyard Sound, Massachusetts.

Seven samples were obtained from a particular portion of the ridge using a Van Veen grab. One of the objective measurements reported by Smith was the settling velocity of the sediment at 22 °C. For sediment from a sand-wave crest section of a sand ridge, the settling velocity has a typical value of 14 cm/s. Table 3.12 gives the settling velocities for the seven sediment samples collected from a particular portion of the Middle Ground sand ridge.

We would like to detect whether the seven sediment samples came from a sand-wave crest section of the Middle Ground sand ridge. Let θ denote the median settling velocity for the population of sediment samples from this portion of Middle Ground. Then we are interested in testing $H_0 : \theta = 14$ cm/s versus the alternative $\theta \neq 14$ cm/s, and we perform test procedure (3.42). With $\alpha = .02$, we see that $b_{.02/2,1/2} = 7$.

Now, we create the modified Z' observations using the following setup.

i	Z_i	$Z_i' = Z_i - 14$
1	12.9	−1.1
2	13.7	−0.3
3	14.5	0.5
4	13.3	−0.7
5	12.8	−1.2
6	13.8	−0.2
7	13.4	−0.6

Using the computational setup of Section 3.4 on the Z' observations, we calculate $B = 1$. Thus, we accept $H_0 : \theta = 14$ cm/s at the $\alpha = .02$ level, since $[7 - b_{.02/2,1/2}] = 0 < B < 7 = b_{.02/2,1/2}$. The above results may be reproduced in R through the function call

```
SIGN.test(z, md=14)
```

where z is a vector containing the seven measurements from Table 3.12. This command also provides the P-value and, optionally, confidence intervals. For the current data and test, the P-value is .125. The partial R output of the above command is

```
    One-sample Sign-Test

data: z
s = 1, p-value = .125
alternative hypothesis: true median is not equal to 14
sample estimates:
median of x
13.4
```

For the large-sample approximation, we see from (3.45) that

$$B^* = \frac{1 - \left(\frac{7}{2}\right)}{\left(\frac{7}{4}\right)^{1/2}} \approx -1.89.$$

Thus the smallest significance level at which we could reject H_0 using a symmetric test based on the normal approximation is .0588.

The ordered Z observations are $Z^{(1)} \leq \cdots \leq Z^{(7)}$:12.8, 12.9, 13.3, 13.4, 13.7, 13.8, and 14.5. The sample size n is $(2k + 1)$ with $k = 3$, therefore (3.59) implies that

$$\tilde{\theta} = Z^{(4)} = 13.4.$$

With $n = 7$ and $\alpha = .1250$, we find that $b_{\alpha/2,1/2} = b_{.0625,1/2} = 6$. Thus, $C_{.1250} = 7 + 1 - 6 = 2$. From (3.63), it follows that

$$\theta_{\mathrm{L}} = Z^{(2)} = 12.9 \text{ and } \theta_{\mathrm{U}} = Z^{(6)} = 13.8,$$

so that our 87.50% confidence interval for θ is

$$(\theta_{\mathrm{L}}, \theta_{\mathrm{U}}) = (12.9, 13.8).$$

The confidence interval for θ is found with

```
SIGN.test(z, md=14, conf.level=1-.125)
```

This results in the following output being appended to the output give above:

```
                  Conf.Level     L.E.pt     U.E.pt
Lower Achieved CI      .8750       12.9       13.8
Interpolated CI        .8750       12.9       13.8
Upper Achieved CI      .9844       12.8       14.5
```

Note that the lower achieved interval is the desired interval for this α.

Applying the large-sample approximation, we find from (3.65) that

$$C_{.1250} \approx \left(\frac{7}{2}\right) - 1.534\left(\frac{7}{4}\right)^{1/2} \approx 1,$$

and as $Z^{(1)} = 12.8$ and $Z^{(n+1-1)} = Z^{(7)} = 14.5$, the approximate 87.50% confidence interval for θ is $(12.8, 14.5)$.

Comments

56. *Assumptions.* Note that Assumption B1 for the paired replicates procedures based on the sign statistic is not necessary for the one-sample data because these data do not consist of differences for paired observations.

57. *Procedures for Population Quantiles Other than the Median.* For one-sample data, the theory underlying the sign statistic can also be used to construct distribution-free test procedures for population quantiles other than the median. Such test procedures are similar to procedures (3.40), (3.41), and (3.42), but they have different P-values in the null hypothesis binomial distribution. For example, let $Z_1, \ldots, Z_n$ be a random sample from a population II. Define μ_ξ to be the unknown ξ quantile of the population. (For convenience, let us assume that μ_ξ is unique.) Consider the problem of testing $H_0 : \mu_\xi = \mu_0$ (specified) versus the one-sided alternative $\mu_\xi > \mu_0$. Define B to be the number of Z's that are greater than μ_0. Under H_0, B has the binomial distribution with parameters n and $p = 1 - \xi$. Large values of B indicate that $\mu_\xi > \mu_0$, so an appropriate one-sided α-level test is to reject H_0 in favor of $\mu_\xi > \mu_0$ if $B \geq b_{\alpha,1-\xi}$ and accept H_0 if $B < b_{\alpha,1-\xi}$. One-sided tests against $\mu_\xi < \mu_0$ and two-sided tests for alternatives $\mu_\xi \neq \mu_0$ are constructed in a similar manner. The natural point estimator of the parameter $P(Z > \mu_0)$ is the statistic B/n. Approximate confidence intervals for μ_ξ can also be obtained (cf. Conover (1999)).

Let $Z_1, Z_2, \ldots, Z_n$ be a random sample of size n from an unknown distribution. Let z_p denote the pth quantile of the distribution. Hayter (2013) constructs simultaneous confidence intervals for $z_{p_i}, 1 \leq i \leq k, 0 < p_1 < p_2 < \cdots < p_k < 1$. The intervals are of the form

$$z_{p_i} \in [Z_{(l_i)}, Z_{(u_i+1)}], \quad 1 \leq i \leq k,$$

where $Z_{(1)}, Z_{(2)} \ldots, Z_{(n)}$ are the order statistics. The integers l_i and u_i are suitably chosen as described by Hayter to provide an overall simultaneous confidence level of at least $1 - \alpha$, with the lower limits being $-\infty$ if $l_i = 0$ and the upper limits being ∞ if $u_i = n$. See Hayter (2013) for his methodology and for specific examples with $\alpha = .05$ and $n = 20, 50$ and 80.

Properties

1. The properties of the one-sample procedures based on the sign statistic are essentially the same as those of the corresponding paired replicates procedures. An exception occurs in the efficiencies of the procedures and is due to the difference in the type of data for the two problems. See Section 3.11 for a discussion of the difference in efficiencies for the procedures of Sections 3.4–3.6 when they are applied to single-sample problems.

Problems

94. The data in Table 3.13 are a subset of the data obtained by Cole and Katz (1966). They were investigating the relation between ozone concentrations and weather fleck damage to tobacco crops in southern Ontario, Canada. One of the objective measurements reported was oxidant content of dew water in parts per million (ppm) ozone. Twelve samples of dew were collected during the period August 25–30, 1960, at Port Burwell, Ontario; the resulting oxidant contents are given in Table 3.13.

Test the hypothesis that the median oxidant content (θ) of dew water was .25 against the alternative that it was less than .25. Obtain a point estimate of θ and find a confidence interval for θ with the confidence coefficient .9614.

95. For the oxidant content data of Table 3.13, obtain a point estimate of θ from the midpoint of the confidence interval calculated in Problem 94 (see Comment 52). Compare with $\tilde{\theta}$ obtained in Problem 94.

96. Compute $\tilde{\theta}$ for the mass ratio data of Table 3.8 and compare with the value of $\hat{\theta}$ obtained in Example 3.8.

97. Maxson (1977) studied the activity patterns of female ruffed grouse with broods. Using surveillance techniques, he recorded the movements of seven female ruffed grouse with broods over a fixed period. The percentage of time that these grouse were in active movement is recorded in Table 3.14.

Test the hypothesis that the median percentage time active for female ruffed grouse with broods (θ) is 50% against the alternative that it is greater than 50%. Obtain a point estimate of θ and find a lower confidence bound (see Comment 49) for θ with the confidence coefficient .99.

98. The data in Table 3.15 are a subset of the data obtained by Flores and Zohman (1970) in an experiment investigating the effect of the method of bed-making on the oxygen consumption for patients assigned to complete or modified bed rest. The subjects were inpatients of the Rehabilitation Medicine Service, Montefiore Hospital and Medical Center, Bronx, New York.

Table 3.14 Ruffed Grouse, Percentage Time in Active Movement

Grouse i	Z_i (% time active)
1	52.7
2	51.5
3	58.4
4	56.9
5	58.5
6	54.4
7	47.1

Source: S. J. Maxson (1977).

Table 3.15 Net Oxygen Consumption (cc)

Patient i	Z_i
1	339
2	349
3	387
4	159
5	579
6	586
7	519
8	275

Source: A. M. Flores and L. R. Zohman (1970).

The measure used was net oxygen consumption for the patients during bed-making. The data in Table 3.15 are the net oxygen consumptions (in cc) for the eight patients in the study during a cardiac top-to-bottom bed-making procedure, consisting of moving the patient to a sitting position and changing the sheets from the top to the bottom of the bed.

Test the hypothesis that the median oxygen consumption rate during cardiac bed-making for patients assigned to complete or modified bed rest (θ) is 350 cc against the alternative that it is not 350 cc. Obtain a point estimate of θ and find a confidence interval for θ with the confidence coefficient .95.

99. Consider the 6β-hydroxycortisol excretion data in Table 3.11. Use the computer software R to test the hypothesis that the median 6β-hydroxycortisol excretion rate for subjects with occupational exposure to DDT similar to the workers in the Poland et al. (1970) study (θ) is 175 μg/24 h against the alternative that it is greater than 175 μg/24 h. Obtain a point estimate of θ and find a confidence interval for θ with the confidence coefficient .925. Compare with the answers to Problem 91.

100. Consider the mayfly head width data in Table 3.10. Let $\mu_{.75}$ be the 75th percentile for the distribution of mayfly head widths in habitat A studied by Lamp (1976). Test the hypothesis that $\mu_{.75} = 25$ against the alternative that $\mu_{.75}$ is greater than 25. (See Comment 57.)

3.9 AN ASYMPTOTICALLY DISTRIBUTION-FREE TEST OF SYMMETRY (RANDLES–FLIGNER– POLICELLO–WOLFE, DAVIS–QUADE)

Data. We obtain n observations $Z_1, \ldots, Z_n$.

Assumptions

E1. The Z's are mutually independent.

E2. Each Z comes from the same continuous population having distribution function F and unknown median θ. This assumption requires that $F(\theta) = \frac{1}{2}$.

Hypothesis

The null hypothesis of interest here is that the common underlying distribution for the Z observations is symmetric about θ. This hypothesis of symmetry can be written as

$$H_0 : [F(\theta + b) + F(\theta - b) = 1, \text{ for every } b], \tag{3.72}$$

and it is equivalent to the statement that $P(0 < Z - \theta < b) = P(-b < Z - \theta < 0)$ for all $b > 0$.

Procedure

For each triple of observations $(Z_i, Z_j, Z_k), 1 \le i < j < k \le n$, obtain the value of

$$f^*(Z_i, Z_j, Z_k) = [\text{sign } (Z_i + Z_j - 2Z_k)] + \text{sign } (Z_i + Z_k - 2Z_j) + \text{sign } (Z_j + Z_k - 2Z_i)], \tag{3.73}$$

where sign $(t) = -1, 0, 1$ as $t <, =, > 0$. (Note that there are $n(n-1)(n-2)/6$ distinct triples in the sample.) We say that (Z_i, Z_j, Z_k) forms a **right triple** (looks skewed to the right) if $f^*(Z_i, Z_j, Z_k) = 1$. (Note that being a right triple is equivalent to the middle *ordered* observation in (Z_i, Z_j, Z_k) being closer to the smallest of the three observations than it is to the largest of them.) Conversely, (Z_i, Z_j, Z_k) is said to be a **left triple** (looks skewed to the left) if $f^*(Z_i, Z_j, Z_k) = -1$ (i.e., the middle *ordered* observation in (Z_i, Z_j, Z_k) is closer to the largest than to the smallest of the three observations). Finally, when $f^*(Z_i, Z_j, Z_k) = 0$, the triple (Z_i, Z_j, Z_k) is neither right nor left.

For the data $Z_1, \ldots, Z_n$, set

$$T = \sum_{1 \le i < j < k \le n} \sum \sum f^*(Z_i, Z_j, Z_k) = \{[\text{number of right triples}] - [\text{number of left triples}]\}. \tag{3.74}$$

For each fixed $t = 1, \ldots, n$, let

$$B_t = \{[\text{number of right triples involving } Z_t] - [\text{number of left triples involving } Z_t]\} = \left[\sum_{j=t+1}^{n-1} \sum_{k=j+1}^{n} f^*(Z_t, Z_j, Z_k) + \sum_{j=1}^{t-1} \sum_{k=t+1}^{n} f^*(Z_j, Z_t, Z_k) + \sum_{j=1}^{t-2} \sum_{k=j+1}^{t-1} f^*(Z_j, Z_k, Z_t) \right]. \tag{3.75}$$

For each fixed integer pair (s, t) such that $1 \le s < t \le n$, define

$$B_{s,t} = \{[\text{number of right triples involving } Z_s \text{ and } Z_t] - [\text{number of left triples involving } Z_s \text{ and } Z_t]\} = \left[\sum_{j=1}^{s-1} f^*(Z_j, Z_s, Z_t) + \sum_{j=s+1}^{t-1} f^*(Z_s, Z_j, Z_t) + \sum_{j=t+1}^{n} f^*(Z_s, Z_t, Z_j) \right]. \tag{3.76}$$

Using the expressions for B_t (3.75), $B_{s,t}$ (3.76), and the triple statistic T (3.74), set

$$V = \frac{T}{\hat{\sigma}}, \tag{3.77}$$

where

$$\hat{\sigma}^2 = \left[\frac{(n-3)(n-4)}{(n-1)(n-2)} \sum_{t=1}^{n} B_t^2 + \frac{(n-3)}{(n-4)} \sum_{s=1}^{n-1} \sum_{t=s+1}^{n} B_{s,t}^2 \right.$$

$$\left. + \frac{n(n-1)(n-2)}{6} - \left\{ 1 - \frac{(n-3)(n-4)(n-5)}{n(n-1)(n-2)} \right\} T^2 \right]. \tag{3.78}$$

When H_0 is true and the underlying distribution is symmetric, V has, as n tends to infinity, an asymptotic $N(0,1)$ distribution. (In order for this normal approximation to be reasonably effective, the sample size n should be at least 10. For further discussion along these lines, see Comment 60.)

To test H_0 (3.72), corresponding to symmetry of the underlying distribution, versus the general alternative of asymmetry, corresponding to

$$H_1 : [P(Z \leq \theta + b) + P(Z \leq \theta - b) \neq 1 \text{ for at least one } b], \tag{3.79}$$

at the approximate (n large) α level of significance,

$$\text{Reject } H_0 \text{ if } |V| \geq z_{\alpha/2}; \text{ otherwise do not reject.} \tag{3.80}$$

Ties

The test procedure in (3.80) is well-defined when zeros occur in the $(Z_i + Z_j - 2Z_k)$ variables and further adjustments are not necessary.

EXAMPLE 3.10 ***Percentage Chromium in Stainless Steel.***

In order to clearly illustrate the details of the rather involved calculations necessary to obtain the value of the test statistic V (3.77), we consider the application of the test for symmetry to the first five (i.e., $n = 5$) percentage chromium data values in Table 3.9, namely, $Z_1 = 17.4, Z_2 = 17.9, Z_3 = 17.6, Z_4 = 18.1$, and $Z_5 = 17.6$. (We emphasize that this application is for illustrative purposes only. The test for symmetry is totally ineffective at detecting asymmetry for sample sizes as small as $n = 5$. See Comment 60 for related discussion.) We must calculate $n(n-1)(n-2)/6 = 5(4)(3)/6 = 10$ values of the triple indicator $f^*(Z_i, Z_j, Z_k)$ given by (3.73). We have that

$$\begin{aligned} f^*(Z_1, Z_2, Z_3) &= [\text{sign } (17.4 + 17.9 - 2(17.6)) + \text{sign } (17.4 + 17.6 \\ &\quad - 2(17.9)) + \text{sign } (17.9 + 17.6 - 2(17.4))] \\ &= [\text{sign } (.1) + \text{sign } (-.8) + \text{sign } (.7)] = 1 - 1 + 1 = 1. \end{aligned} \tag{3.81}$$

Similarly, we obtain

$$\begin{aligned} f^*(Z_1, Z_2, Z_5) &= f^*(Z_1, Z_3, Z_4) = f^*(Z_1, Z_4, Z_5) \\ &= f^*(Z_2, Z_3, Z_5) = f^*(Z_3, Z_4, Z_5) = 1 \end{aligned} \tag{3.82}$$

and

$$f^*(Z_1, Z_2, Z_4) = f^*(Z_1, Z_3, Z_5) = f^*(Z_2, Z_3, Z_4)$$
$$= f^*(Z_2, Z_4, Z_5) = -1. \tag{3.83}$$

Hence, from (3.74) we have that

$$T = \sum_{1 \le i < j < k \le 5} \sum \sum f^*(Z_i, Z_j, Z_k) = 6 - 4 = 2. \tag{3.84}$$

For the calculation of $\hat{\sigma}^2$, we first need to obtain the values of $B_1, \ldots, B_5$ and $B_{s,t}$ for $1 \le s < t \le 5$. From (3.75), (3.81), (3.82), and (3.83), we have that

$$B_1 = \sum_{j=2}^{4} \sum_{k=j+1}^{5} f^*(Z_1, Z_j, Z_k) = [1 - 1 + 1 + 1 - 1 + 1] = 2,$$

$$B_2 = \left[\sum_{j=3}^{4} \sum_{k=j+1}^{5} f^*(Z_2, Z_j, Z_k) + \sum_{k=3}^{5} f^*(Z_1, Z_2, Z_k) \right]$$
$$= [(-1 + 1 - 1) + (1 - 1 + 1)] = 0,$$

$$B_3 = \left[f^*(Z_3, Z_4, Z_5) + \sum_{j=1}^{2} \sum_{k=4}^{5} f^*(Z_j, Z_3, Z_k) + f^*(Z_1, Z_2, Z_3) \right]$$
$$= [1 + (1 - 1 - 1 + 1) + 1] = 2,$$

$$B_4 = \left[\sum_{j=1}^{3} f^*(Z_j, Z_4, Z_5) + \sum_{j=1}^{2} \sum_{k=j+1}^{3} f^*(Z_j, Z_k, Z_4) \right]$$
$$= [(1 - 1 + 1) + (-1 + 1 - 1)] = 0,$$

and

$$B_5 = \sum_{j=1}^{3} \sum_{k=j+1}^{4} f^*(Z_j, Z_k, Z_5) = [1 - 1 + 1 + 1 - 1 + 1] = 2.$$

It follows that

$$\sum_{t=1}^{5} B_t^2 = [2^2 + 0^2 + 2^2 + 0^2 + 2^2] = 12. \tag{3.85}$$

Furthermore, using (3.76), (3.81), (3.82), and (3.83), we obtain

$$B_{1,2} = \sum_{j=3}^{5} f^*(Z_1, Z_2, Z_j) = [1 - 1 + 1] = 1,$$

$$B_{1,3} = f^*(Z_1, Z_2, Z_3) + \sum_{j=4}^{5} f^*(Z_1, Z_3, Z_j) = [1 + (1 - 1)] = 1,$$

$$B_{1,4} = \sum_{j=2}^{3} f^*(Z_1, Z_j, Z_4) + f^*(Z_1, Z_4, Z_5) = [(-1 + 1) + 1] = 1,$$

$$B_{1,5} = \sum_{j=2}^{4} f^*(Z_1, Z_j, Z_5) = [1 - 1 + 1] = 1,$$

$$B_{2,3} = f^*(Z_1, Z_2, Z_3) + \sum_{j=4}^{5} f^*(Z_2, Z_3, Z_j) = [1 + (-1 + 1)] = 1,$$

$$\begin{aligned} B_{2,4} &= f^*(Z_1, Z_2, Z_4) + f^*(Z_2, Z_3, Z_4) + f^*(Z_2, Z_4, Z_5) \\ &= [-1 - 1 - 1] = -3, \end{aligned}$$

$$B_{2,5} = f^*(Z_1, Z_2, Z_5) + \sum_{j=3}^{4} f^*(Z_2, Z_j, Z_5) = [1 + (1 - 1)] = 1,$$

$$B_{3,4} = \sum_{j=1}^{2} f^*(Z_j, Z_3, Z_4) + f^*(Z_3, Z_4, Z_5) = [(1 - 1) + 1] = 1,$$

$$B_{3,5} = \sum_{j=1}^{2} f^*(Z_j, Z_3, Z_5) + f^*(Z_3, Z_4, Z_5) = [(-1 + 1) + 1] = 1,$$

and

$$B_{4,5} = \sum_{j=1}^{3} f^*(Z_j, Z_4, Z_5) = [1 - 1 + 1] = 1.$$

These $B_{s,t}$ values yield

$$\begin{aligned} \sum_{s=1}^{4} \sum_{t=s+1}^{5} B_{s,t}^2 &= [1^2 + 1^2 + 1^2 + 1^2 + 1^2 + (-3)^2 + 1^2 + 1^2 + 1^2 + 1^2] \\ &= 18. \end{aligned} \tag{3.86}$$

Using the computational results from (3.84), (3.85), and (3.86) in the formula for $\hat{\sigma}^2$ (3.78), we obtain

$$\begin{aligned} \hat{\sigma}^2 &= \left[\frac{2(1)}{4(3)}(12) + \frac{2}{1}(18) + \frac{5(4)(3)}{6} - \left\{ 1 - \frac{2(1)(0)}{5(4)(3)} \right\} (2)^2 \right] \\ &= [2 + 36 + 10 - 4] = 44. \end{aligned}$$

Finally, from (3.77), we have

$$V = \frac{T}{\hat{\sigma}} = \frac{2}{(44)^{1/2}} = .30.$$

(We note that the R command RFPW(**z**) can also be used to obtain the value of the test statistic $V = \frac{T}{\hat{\sigma}}$ for the data **z**. For this example, we have $n = 5$, $\mathbf{z} = (17.4, 17.9, 17.6, 18.1, 17.6)$, and RFPW(**z**) =.30.)

With significance level $\alpha = .05$, we use the R command `qnorm(.)` to obtain the critical value $z_{.025} = 1.96$ from the fact that `qnorm(1-.025) = qnorm(.975)=1.96`. Since $|V| = .30$ is less than 1.96, we cannot reject the null hypothesis of symmetry for the underlying distribution. In fact, using the R command `pnorm (.)`, we see that the smallest significance level at which we could reject this distributional symmetry (i.e., the two-sided P-value for these data) is

$$\begin{aligned} \underline{\alpha} &= 2P(\text{standard normal variable exceeds } .30) \\ &= 2(1 - \texttt{pnorm}(.30)) \\ &= 2(.3821) = .7642, \end{aligned}$$

clearly indicating that there is virtually no evidence in this subset of the percentage chromium data to indicate asymmetry in the underlying probability distribution. (Remember, however, that this subset was a sample of only five observations. These are simply not sufficient data to detect asymmetry even if it were present. See Comment 60.)

Comments

58. *Motivation.* A right triple is indicative of skewness to the right and a left triple is indicative of skewness to the left. The absolute value of the statistic T (3.74) is the difference between the numbers of right and left triples among the $n(n-1)(n-2)/6$ triples in the sample. When the null hypothesis H_0 (3.72) of symmetry is true, we would expect half of the sample triples to be right triples and the other half to be left triples. Thus, when H_0 is true, we would expect T to be near zero. A substantial deviation in either direction from zero for T is therefore indicative of asymmetry in the population and serves as a partial motivation for the procedure defined in (3.80).
59. *Asymptotic Distribution-Freeness.* Asymptotically (i.e., for infinitely large samples), the true level of the test defined by (3.80) will agree with the nominal level. Subject to Assumptions E1 and E2, this asymptotic result does not depend on the underlying population of the Z's. More precisely, subject to Assumptions E1 and E2, V has an asymptotic $N(0, 1)$ distribution when H_0 is true. Since this asymptotic distribution does not depend on the underlying population of the Z's, we say that the test based on V is asymptotically distribution-free. Of course, in practice, we do not have the luxury of infinite samples. Thus in any particular case, with n large, we hope the level of a test based on V is close to the nominal level α but it may not be exactly equal to α. The closeness of the approximation depends on n and α and, for fixed α, the closeness generally improves as n increases. In the case of the V test, the reader is warned that the question of how large n should be, in order for the approximation to be good, is unanswered. Exact null distribution critical values for V cannot be provided because, for a specified value of n, the exact null distribution of V depends on the underlying Z population; thus exact critical values would vary with the form of the Z population. The procedure in (3.80) based on V, therefore, is not (strictly) distribution-free.

60. *Sample Size Requirement.* As noted in Comment 59, the test of symmetry described in (3.80) is not an exact distribution-free procedure. The nominal significance level α is guaranteed only asymptotically, as the number of observations, n, becomes infinite. In addition, symmetry is a rather complex property of a probability distribution. It is, therefore, virtually impossible to deny its presence without at least a moderate sample size. It is simply difficult to "see" asymmetry in a small number of sample observations. Both Randles et al. (1980) and Davis and Quade (1978) found this to be the case. They concluded that the symmetry test (3.80) is not effective at detecting asymmetry in the underlying population unless the sample size (n) is at least 20.

61. *One-Sided Tests for Right-Skewness or Left-Skewness.* The test procedure in (3.80) is a two-sided test of symmetry against a very general class of asymmetric alternatives. However, one-sided tests of symmetry versus specific classes of right-skewed (or left-skewed) asymmetric alternatives can also be based on the statistic V (3.77). In particular, a one-sided (approximate) level α test of H_0 (3.72) (symmetry) versus the specific class of right-skewed alternatives satisfying

$$F(\theta + b) \leq [1 - F(\theta - b)], \text{ for every } b > 0,$$
$$\text{with strict inequality for at least one positive } b, \tag{3.87}$$

is given by

$$\text{Reject } H_0 \text{ if } V \geq z_\alpha; \text{ otherwise do not reject.} \tag{3.88}$$

Similarly, a one-sided (approximate) level α test of H_0 (3.72) versus the specific class of left-skewed alternatives satisfying

$$F(\theta + b) \geq [1 - F(\theta - b)], \text{ for every } b > 0,$$
$$\text{with strict inequality for at least one positive } b, \tag{3.89}$$

is given by

$$\text{Reject } H_0 \text{ if } V \leq -z_\alpha; \text{ otherwise do not reject.} \tag{3.90}$$

These one-sided hypothesis tests in (3.88) and (3.90) are asymptotically distribution-free in the same sense as the two-sided test given by (3.80). See Comment 59 for further discussion of this property.

62. *Signed Rank Procedures.* One of the critical assumptions permitting the application of signed rank procedures to one-sample data is that of underlying distributional symmetry (see Assumption C2 in Section 3.7). Under this symmetry *assumption,* procedures based on the signed rank statistic T^+ (3.3) for one-sample data are used to make inferences about the median of a population. Procedure (3.80), on the other hand, is used to test for the symmetry of a population and is not directly concerned with the numerical value of the median of the population. Therefore, in an appropriate one-sample location problem we might wish to apply procedure (3.80) (to check the symmetry assumption) prior to using the signed rank procedures of Section 3.7 for making inferences about the actual value of the unknown median of the population. Procedure (3.80) is appropriate, but the *known* median test mentioned in Comment 63 is inappropriate as a pretest in this situation. (For the paired-replicates data in Sections 3.1–3.3, we remind the reader that the symmetry assumption

is most often inherently satisfied through the nature of the pairing. See Comment 2.)

63. *Case of Known Median.* For the situation when the median of the underlying population is *known* to be a specified value θ_0 (say), Gupta (1967) proposed a procedure for testing the hypothesis of symmetry about θ_0. However, situations in which the median of the underlying population is known but the symmetry of the distribution is not known are encountered considerably less frequently than situations in which both the median and the symmetry are not known (see Comment 62). (Gupta (1967) also proposed a test for symmetry when the underlying median is not known. His procedure in this case is a competitor to the test given by (3.80). He investigated the loss of efficiency that results from using his test for symmetry with unknown median when his known median procedure is applicable.)

64. *Alternative Determination of Right and Left Triples.* The original definitions of right and left triples in this section involve the sign function $f^*(Z_i, Z_j, Z_k)$ in (3.73). A more intuitive interpretation is associated with the comparison of two common sample measures of location. For a triple (Z_i, Z_j, Z_k), let $\bar{Z} = (Z_i + Z_j + Z_k)/3$ and $\tilde{Z} = \text{median } \{Z_i, Z_j, Z_k\}$ be the average and median, respectively, for the observations in the triple. Then the triple (Z_i, Z_j, Z_k) is a right triple if $\bar{Z} > \tilde{Z}$ and it is a left triple if $\bar{Z} < \tilde{Z}$. (It is neither right nor left if $\bar{Z} = \tilde{Z}$.) This formulation provides a very natural interpretation of what it means to be a right or left triple, as we know that the population mean is greater than or less than the population median according to whether the population is skewed to the right or left, respectively. If the population is symmetric, its mean and median are equal and it would be a toss up as to which of $\bar{Z}$ or $\tilde{Z}$ would be greater. This should lead to about an equal number of right and left triples in the sample.

65. *Consistent Estimator of the Asymptotic Variance of $\sqrt{n}T$.* In order to insure that V (3.77) is asymptotically distribution-free, $n\hat{\sigma}^2$ (3.78) is taken to be a consistent estimator of the asymptotic null variance of $n^{1/2}T$. The consistency of this estimator $n\hat{\sigma}^2$ follows from a standard body of theory about a class of statistics introduced by Hoeffding (1948a) and referred to as U-statistics. (For more details about U-statistics and their application in the triples test, see Randles and Wolfe (1979).) The asymptotic normality (and, thereby, the asymptotic distribution-freeness) for V (3.77) follows from standard U-statistics theory and Slutsky's theorem (see, for example, Theorem A.3.13 in Randles and Wolfe (1979)).

66. *Consistency of the V Test.* Under Assumptions El and E2, the consistency of the tests based on V depend on the parameter

$$p^* = P(Z_1 + Z_2 - 2Z_3 > 0) - \frac{1}{2}. \tag{3.91}$$

The two-sided test defined by (3.80) is consistent against the class of asymmetric alternatives corresponding to $p^* \neq 0$. We point out that while asymmetry of a probability distribution implies that $p^* \neq 0$ for that distribution, the converse is not necessarily true; that is, there are asymmetric probability distributions for which $p^* = 0$ and against which, therefore, the two-sided test (3.80) based on V will not be consistent. Randles et al. (1980) note, however, that the class of distributions with this property is quite small. (The one-sided tests discussed in Comment 61 and defined by (3.88) and (3.90) are consistent against the classes of asymmetric alternatives corresponding to $p^* > 0$ and < 0, respectively.)

Properties

1. *Consistency.* See Comment 66 and Randles et al. (1980).
2. *Asymptotic Normality.* See Randles and Wolfe (1979, pp. 99–101).

Problems

101. Consider the percentage chromium data in Table 3.9. Test the hypothesis of symmetry versus general asymmetry. (Note that some of the necessary calculations for this test have been completed in Example 3.10.)

102. Show that a triple (Z_1, Z_2, Z_3) is a right triple if and only if $\bar{Z} = (Z_1 + Z_2 + Z_3)/3$ is greater than $\tilde{Z} = \text{median } (Z_1, Z_2, Z_3)$. (See also Comment 64.)

103. Consider the oxidant content data of Table 3.13. Test the hypothesis of symmetry versus general asymmetry.

104. What effect does the addition of a number b to each of the Z observations have on the value of the V (3.77) statistic? Comment on this as a desirable property for a test of population symmetry.

105. What effect does the multiplication of each of the Z observations by a number b have on the absolute value of the V (3.77) statistic? Comment on this as a desirable property for a test of population symmetry versus general asymmetry.

106. Consider the settling velocity data in Table 3.12. Test the hypothesis of symmetry against the alternative that the population of settling velocities is skewed to the right. (See Comment 61.)

107. Consider the Z differences for the beak-clapping data in Table 3.5. Test the hypothesis of symmetry against the alternative that the population of beak-clapping differences is skewed to the left. (See Comment 61.)

108. For n observations $Z_1, \ldots, Z_n$, what is the maximum possible value for T (3.74)? What is the minimum possible value for T? For $n = 4$, construct examples where these extreme values for T are achieved.

109. Consider the four observations $Z_1 = 2, Z_2 = 2.4, Z_3 = 3$, and $Z_4 = 3.5$. Compute the value of T (3.74) for these data. Indicate how to change only one of the sample observations in such a way that T achieves its maximum value (see Problem 108) on the altered data. Similarly, indicate how to change only one of the sample observations in such a way that T achieves its minimum value (see Problem 108) on the altered data.

BIVARIATE DATA

3.10 A DISTRIBUTION-FREE TEST FOR BIVARIATE SYMMETRY (HOLLANDER)

Data. We obtain $2n$ observations, two observations on each of n subjects.

Subject i	X_i	Y_i
1	X_1	Y_1
2	X_2	Y_2
$\vdots$	$\vdots$	$\vdots$
n	X_n	Y_n

Assumptions

F1. The n bivariate observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ are mutually independent.

F2. Each $(X_i, Y_i), i = 1, \ldots, n$, comes from the same bivariate population with joint distribution function $F(x, y)$.

Hypothesis

The null hypothesis of interest here is that the X and Y variables are exchangeable or, equivalently, that there is no treatment effect (see Comment 67). This hypothesis of exchangeability can be written as

$$H_0 : [F(x, y) = F(y, x), \text{for all } (x, y)]. \tag{3.92}$$

(Another way to state this exchangeability property is that the pairs (X, Y) and (Y, X) have the same joint bivariate distribution.)

Procedure

For each observation pair $(X_i, Y_i), i = 1, \ldots, n$, let

$$a_i = \min(X_i, Y_i), \quad b_i = \max(X_i, Y_i), \tag{3.93}$$

where, without loss of generality, we take $a_1 \leq a_2 \leq \cdots \leq a_n$. (We may simply relabel the n (X, Y) pairs so that the a's defined by (3.93) are increasing.) Define the n (observed) r values $r_1, r_2, \ldots, r_n$ by

$$r_i = \begin{cases} 1, & \text{if } X_i = a_i < b_i = Y_i \\ 0, & \text{if } X_i = b_i \geq a_i = Y_i. \end{cases} \tag{3.94}$$

That is, r_i is defined to be 1 if $X_i < Y_i$ and 0 if $X_i \geq Y_i$. (Note that the designation $r_i = 0$ for those cases where $X_i = Y_i$ is purely arbitrary, because such a tied situation makes no contribution to the overall test statistic to be defined by (3.98).)

Define the n^2 values d_{ij}, for $i, j = 1, \ldots, n$, by

$$d_{ij} = \begin{cases} 1, & \text{if } a_j < b_i \leq b_j \text{ and } a_i \leq a_j \\ 0, & \text{otherwise.} \end{cases} \tag{3.95}$$

For $j = 1, \ldots, n$, set

$$T_j = \sum_{i=1}^{n} s_i d_{ij}, \tag{3.96}$$

where d_{ij} is given by (3.95) and

$$s_i = 2r_i - 1. \tag{3.97}$$

Let A_{obs}, to be read as "A observed," be defined as

$$A_{\text{obs}} = \sum_{j=1}^{n} \frac{T_j^2}{n^2}. \tag{3.98}$$

Now, in addition to our observed r configuration $(r_1, \ldots, r_n)$ defined by (3.94), there are $2^n - 1$ other possible r configurations, corresponding to the cases in which each r_i can be either 0 or 1 and excluding the observed configuration (see Comment 68). For each of these $2^n - 1$ additional r configurations, calculate the corresponding value of A using (3.96) to (3.98). It is important to note that the d's defined by (3.95) remain the same for each of these additional calculations of A_{obs}.

Let

$$A^{(1)} \leq A^{(2)} \leq \cdots \leq A^{(2^n)} \tag{3.99}$$

denote the 2^n ordered values of the A's. (Note that A_{obs} will be one of these ordered A's.) Set

$$m = 2^n - [[2^n \alpha]], \tag{3.100}$$

where $[[2^n \alpha]]$ is the greatest integer less than or equal to $2^n \alpha$. Define M_1 to be the number of ordered values $A^{(1)} \leq \cdots \leq A^{(2^n)}$ that are greater than $A^{(m)}$ and take M_2 to be the number of $A^{(1)} \leq \cdots \leq A^{(2^n)}$ values that are equal to $A^{(m)}$, where $A^{(m)}$ is determined by (3.99) and (3.100).

To test H_0 (3.92), corresponding to the exchangeability of the X and Y variables, versus the general (two-sided) alternative that they are not exchangeable, corresponding to

$$H_1 : [F(x, y) \neq F(y, x) \text{ for at least one} (x, y)], \tag{3.101}$$

at the exact α level of significance,

$$\text{Reject } H_0 \text{ if } A_{\text{obs}} > A^{(m)}; \text{ do not reject } H_0 \text{ if } A_{\text{obs}} < A^{(m)}; \tag{3.102}$$

and

$$\text{If } A_{\text{obs}} = A^{(m)}, \text{ make a randomized decision to reject } H_0$$
$$\text{with probability } q \text{ and to not reject } H_0 \text{ with probability } 1 - q,$$

where

$$q = \frac{2^n \alpha - M_1}{M_2}. \tag{3.103}$$

The R program `HollBivSym` computes the statistic A given by (3.98). The `R` program `pHollBivSym` returns A and the exact P-value for $n \leq 20$. By default, if $n > 20$, the program uses a Monte Carlo approximation with 100,000 samples. The user can change both the number of Monte Carlo samples and the largest number of pairs for which he or she is willing to wait for the exact calculation (see Example 3.11).

Koziol (1979) developed a large-sample approximation but Kepner and Randles (1984) and Hilton and Gee (1997a) found that the large-sample approximation does not perform well. Thus one can use `pHollBivSym` understanding that it is exact for $n \leq 20$ but only approximate when $n \geq 21$.

Ties

No adjustment for ties is necessary. The calculation of A_{obs} (3.98) is well defined when ties occur. As a result, the procedure (3.102) handles ties automatically.

EXAMPLE 3.11 ***Inulin Clearance in Kidney Transplants.***

The data in Table 3.16 are a subset of the data obtained by Shelp et al. (1970) in a study of renal transplants. Part of their study dealt with living related donor kidneys and pertained to a comparison of clearance capacity of the donor and recipient after the transplant was done. Table 3.16 gives inulin clearance values for seven recipients and their corresponding donors. (We note that Assumption F2 may not be satisfied because the subjects are not homogeneous. They differ in various factors that may be pertinent to clearance, such as the basic disease, age when the transplant was performed, age of the donor, and sex of the donor. In order to illustrate the bivariate symmetry test, we neglect this heterogeneity of subjects.)

From (3.93) and Table 3.16, we find

$$a_1 = 61.4, a_2 = 63.3, a_3 = 63.7, a_4 = 67.1, a_5 = 77.3,$$
$$a_6 = 84.0, a_7 = 88.1 \tag{3.104}$$

$$b_1 = 70.8, b_2 = 89.2, b_3 = 65.8, b_4 = 88.0, b_5 = 87.3,$$
$$b_6 = 85.1, b_7 = 105.0. \tag{3.105}$$

Our observed r configuration is, from (3.94),

$$r_1 = 1, r_2 = 1, r_3 = 1, r_4 = 0, r_5 = 1, r_6 = 1, r_7 = 0. \tag{3.106}$$

Table 3.16 Inulin Clearance of Living Donors and Recipients of Their Kidneys

Patient[a]	Insulin clearance, ml/min	
	Recipient, X_i	Donor, Y_i
$1'$	61.4	70.8
$2'$	63.3	89.2
$3'$	63.7	65.8
$4'$	80.0	67.1
$5'$	77.3	87.3
$6'$	84.0	85.1
$7'$	105.0	88.1

Source: W. D. Shelp, F. H. Bach, W. A. Kisken, M. Newton, R. E. Rieselbach, and A. B. Weinstein (1970).

[a] The primes on the patient numbers indicate that our numbering is different from that in the study. We have renumbered so that the a's defined by (3.93) are in the order $a_1 < a_2 < a_3 < a_4 < a_5 < a_6 < a_7$.

We next calculate the $7^2 = 49$ d values using (3.95). We have

$$\begin{aligned}
&d_{11} = 1, d_{12} = 1, d_{13} = 0, d_{14} = 1, d_{15} = 0, d_{16} = 0, d_{17} = 0,\\
&d_{21} = 0, d_{22} = 1, d_{23} = 0, d_{24} = 0, d_{25} = 0, d_{26} = 0, d_{27} = 1,\\
&d_{31} = 0, d_{32} = 0, d_{33} = 1, d_{34} = 0, d_{35} = 0, d_{36} = 0, d_{37} = 0,\\
&d_{41} = 0, d_{42} = 0, d_{43} = 0, d_{44} = 1, d_{45} = 1, d_{46} = 0, d_{47} = 0,\\
&d_{51} = 0, d_{52} = 0, d_{53} = 0, d_{54} = 0, d_{55} = 1, d_{56} = 0, d_{57} = 0,\\
&d_{61} = 0, d_{62} = 0, d_{63} = 0, d_{64} = 0, d_{65} = 0, d_{66} = 1, d_{67} = 0,\\
&d_{71} = 0, d_{72} = 0, d_{73} = 0, d_{74} = 0, d_{75} = 0, d_{76} = 0, d_{77} = 1.
\end{aligned} \tag{3.107}$$

From (3.97) and (3.106), we have

$$s_1 = 1, s_2 = 1, s_3 = 1, s_4 = -1, s_5 = 1, s_6 = 1, s_7 = -1. \tag{3.108}$$

From (3.96), (3.107), and (3.108) we obtain

$$\begin{aligned}
T_1 &= d_{11}s_1 + d_{21}s_2 + d_{31}s_3 + d_{41}s_4 + d_{51}s_5 + d_{61}s_6 + d_{71}s_7\\
&= 1(1) + 0(1) + 0(1) + 0(-1) + 0(1) + 0(1) + 0(-1) = 1,\\
T_2 &= d_{12}s_1 + d_{22}s_2 + d_{32}s_3 + d_{42}s_4 + d_{52}s_5 + d_{62}s_6 + d_{72}s_7\\
&= 1(1) + 1(1) + 0(1) + 0(-1) + 0(1) + 0(1) + 0(-1) = 2,\\
T_3 &= d_{13}s_1 + d_{23}s_2 + d_{33}s_3 + d_{43}s_4 + d_{53}s_5 + d_{63}s_6 + d_{73}s_7\\
&= 0(1) + 0(1) + 1(1) + 0(-1) + 0(1) + 0(1) + 0(-1) = 1,\\
T_4 &= d_{14}s_1 + d_{24}s_2 + d_{34}s_3 + d_{44}s_4 + d_{54}s_5 + d_{64}s_6 + d_{74}s_7\\
&= 1(1) + 0(1) + 0(1) + 1(-1) + 0(1) + 0(1) + 0(-1) = 0,\\
T_5 &= d_{15}s_1 + d_{25}s_2 + d_{35}s_3 + d_{45}s_4 + d_{55}s_5 + d_{65}s_6 + d_{75}s_7\\
&= 0(1) + 0(1) + 0(1) + 1(-1) + 1(1) + 0(1) + 0(-1) = 0,\\
T_6 &= d_{16}s_1 + d_{26}s_2 + d_{36}s_3 + d_{46}s_4 + d_{56}s_5 + d_{66}s_6 + d_{76}s_7\\
&= 0(1) + 0(1) + 0(1) + 0(-1) + 0(1) + 1(1) + 0(-1) = 1,\\
T_7 &= d_{17}s_1 + d_{27}s_2 + d_{37}s_3 + d_{47}s_4 + d_{57}s_5 + d_{67}s_6 + d_{77}s_7\\
&= 0(1) + 1(1) + 0(1) + 0(-1) + 0(1) + 0(1) + 1(-1) = 0.
\end{aligned} \tag{3.109}$$

Equation (3.98) then yields

$$\begin{aligned}
A_{\text{obs}} &= \frac{T_1^2 + T_2^2 + T_3^2 + T_4^2 + T_5^2 + T_6^2 + T_7^2}{49}\\
&= \frac{1 + 4 + 1 + 0 + 0 + 1 + 0}{49} = \frac{7}{49}.
\end{aligned} \tag{3.110}$$

Now, to apply the exact procedure given by (3.102), we need to obtain the additional $2^7 - 1 = 127$ A values, corresponding to the other 127 possible r configurations. The 128 possible r configurations, including r observed, which is given by (3.106), are displayed in Table 3.17.

The parenthetical values to the right of each r configuration in Table 3.17 are the corresponding values of $49A$. These values are calculated in the same way we calculated A_{obs} in (3.107) to (3.110). The s's corresponding to (3.108) must be recalculated for each r configuration for use in the T_j equations, but the d's remain the same for each calculation. The ordered A's defined by (3.99) are $A^{(1)} \leq \cdots \leq A^{(128)}$.

We now list the ordered values of $49A : 49A^{(1)} = \cdots = 49A^{(8)} = 3, 49A^{(9)} = \cdots = 49A^{(40)} = 7, 49A^{(41)} = \cdots = 49A^{(88)} = 11, 49A^{(89)} = \cdots = 49A^{(120)} = 15, 49A^{(121)} = \cdots = 49A^{(128)} = 19$.

Let us illustrate the $\alpha = \frac{8}{128} = .0625$ test. The value of m (3.100) is

$$m = 2^7 - \left[\left[2^7\left(\frac{8}{128}\right)\right]\right] = 128 - 8 = 120,$$

and thus $A^{(m)} = A^{(120)} = \left(\frac{15}{49}\right)$. We then have

$$M_1 = \text{number of } A \text{ values greater than } \frac{15}{49} = 8,$$

$$M_2 = \text{number of } A \text{ values equal to } \frac{15}{49} = 32,$$

and

$$q_1 = \left(128\left(\frac{8}{128}\right) - 8\right) = 0.$$

Hence, procedure (3.102) reduces to, at the $\alpha = .0625$ level,

$$\text{Reject } H_0 \text{ if } A_{\text{obs}} > \frac{15}{49}. \tag{3.111}$$

As $A_{\text{obs}} = \frac{7}{49}$, we do not reject the hypothesis of bivariate symmetry at the .0625 level. Furthermore, because there are 120 configurations (including the one corresponding to A_{obs}) that yield a value greater than or equal to A_{obs}, the lowest level at which we can reject using a nonrandomized test based on A is $\frac{120}{128} = .9375$.

To perform the A test using R let

$$x < \text{-}c(61.4, 63.3, 63.7, 80, 77.3, 84, 105)$$

$$y < \text{-}c(70.8, 89.2, 65.8, 67.1, 87.3, 85.1, 88.1)$$

Then `HollBivSym(x, y)` returns $A = .1429$ and `pHollBivSym(x, y)` returns $A = .1429$ and the P-value .9375.

The command `pHollBivSym(x, y, approx=7)` signifies that the user is willing to use the approximate P-value when $n \geq 7$. Furthermore, the command `pHollBivSym(x, y, approx=7, n.mc=200,000)` changes the default from 100,000 to 200,000 Monte Carlo samples.

Table 3.17 The 128 Possible r Configurations and Corresponding Values of 49A

r_1	r_2	r_3	r_4	r_5	r_6	r_7	(49A)
1	1	1	1	1	1	1	(19)
1	1	1	1	1	1	0	(15)
1	1	1	1	1	0	1	(19)
1	1	1	1	1	0	0	(15)
1	1	1	1	0	1	1	(15)
1	1	1	1	0	1	0	(11)
1	1	1	1	0	0	1	(15)
1	1	1	1	0	0	0	(11)
1	1	1	0	1	1	1	(11)
1	1	1	0	1	0	1	(11)
1	1	1	0	1	1	0[a]	(7)
1	1	1	0	1	0	0	(7)
1	1	1	0	0	1	1	(15)
1	1	1	0	0	0	1	(15)
1	1	1	0	0	1	0	(11)
1	1	1	0	0	0	0	(11)
1	1	0	1	1	1	1	(19)
1	1	0	1	1	1	0	(15)
1	1	0	1	1	0	1	(19)
1	1	0	1	1	0	0	(15)
1	1	0	1	0	1	1	(15)
1	1	0	1	0	1	0	(11)
1	1	0	1	0	0	1	(15)
1	1	0	1	0	0	0	(11)
1	1	0	0	1	1	1	(11)
1	1	0	0	1	0	1	(11)
1	1	0	0	1	1	0	(7)
1	1	0	0	1	0	0	(7)
1	1	0	0	0	1	1	(15)
1	1	0	0	0	0	1	(15)
1	1	0	0	0	1	0	(11)
1	1	0	0	0	0	0	(11)
1	0	1	1	1	1	1	(11)
1	0	1	1	1	1	0	(15)
1	0	1	1	1	0	1	(11)
1	0	1	1	1	0	0	(15)
1	0	1	1	0	1	1	(7)
1	0	1	1	0	1	0	(11)
1	0	1	1	0	0	1	(7)
1	0	1	1	0	0	0	(11)
1	0	1	0	1	1	1	(3)
1	0	1	0	1	0	1	(3)
1	0	1	0	1	1	0	(7)
1	0	1	0	1	0	0	(7)
1	0	1	0	0	1	1	(7)
1	0	1	0	0	0	1	(7)
1	0	1	0	0	1	0	(11)
1	0	1	0	0	0	0	(11)
1	0	0	1	1	1	1	(11)
1	0	0	1	1	1	0	(15)
1	0	0	1	1	0	1	(11)
1	0	0	1	1	0	0	(15)
1	0	0	1	0	1	1	(7)
1	0	0	1	0	1	0	(11)

Table 3.17 (*Continued*)

r_1	r_2	r_3	r_4	r_5	r_6	r_7	(49*A*)
1	0	0	1	0	0	1	(7)
1	0	0	1	0	0	0	(11)
1	0	0	0	1	1	1	(3)
1	0	0	0	1	0	1	(3)
1	0	0	0	1	1	0	(7)
1	0	0	0	1	0	0	(7)
1	0	0	0	0	1	1	(7)
1	0	0	0	0	0	1	(7)
1	0	0	0	0	1	0	(11)
1	0	0	0	0	0	0	(11)
0	1	1	1	1	1	1	(11)
0	1	1	1	1	1	0	(7)
0	1	1	1	1	0	1	(11)
0	1	1	1	1	0	0	(7)
0	1	1	1	0	1	1	(7)
0	1	1	1	0	1	0	(3)
0	1	1	1	0	0	1	(7)
0	1	1	1	0	0	0	(3)
0	1	1	0	1	1	1	(11)
0	1	1	0	1	0	1	(11)
0	1	1	0	1	1	0	(7)
0	1	1	0	1	0	0	(7)
0	1	1	0	0	1	1	(15)
0	1	1	0	0	0	1	(15)
0	1	1	0	0	1	0	(11)
0	1	1	0	0	0	0	(11)
0	1	0	1	1	1	1	(11)
0	1	0	1	1	1	0	(7)
0	1	0	1	1	0	1	(11)
0	1	0	1	1	0	0	(7)
0	1	0	1	0	1	1	(7)
0	1	0	1	0	1	0	(3)
0	1	0	1	0	0	1	(7)
0	1	0	1	0	0	0	(3)
0	1	0	0	1	1	1	(11)
0	1	0	0	1	0	1	(11)
0	1	0	0	1	1	0	(7)
0	1	0	0	1	0	0	(7)
0	1	0	0	0	1	1	(15)
0	1	0	0	0	0	1	(15)
0	1	0	0	0	1	0	(11)
0	1	0	0	0	0	0	(11)
0	0	1	1	1	1	1	(11)
0	0	1	1	1	1	0	(15)
0	0	1	1	1	0	1	(11)
0	0	1	1	1	0	0	(15)
0	0	1	1	0	1	1	(7)
0	0	1	1	0	1	0	(11)
0	0	1	1	0	0	1	(7)
0	0	1	1	0	0	0	(11)
0	0	1	0	1	1	1	(11)
0	0	1	0	1	0	1	(11)
0	0	1	0	1	1	0	(15)
0	0	1	0	1	0	0	(15)
0	0	1	0	0	1	1	(15)

(*continued*)

Table 3.17 (*Continued*)

r_1	r_2	r_3	r_4	r_5	r_6	r_7	(49*A*)
0	0	1	0	0	0	1	(15)
0	0	1	0	0	1	0	(19)
0	0	1	0	0	0	0	(19)
0	0	0	1	1	1	1	(11)
0	0	0	1	1	1	0	(15)
0	0	0	1	1	0	1	(11)
0	0	0	1	1	0	0	(15)
0	0	0	1	0	1	1	(7)
0	0	0	1	0	1	0	(11)
0	0	0	1	0	0	1	(7)
0	0	0	1	0	0	0	(11)
0	0	0	0	1	1	1	(11)
0	0	0	0	1	0	1	(11)
0	0	0	0	1	1	0	(15)
0	0	0	0	1	0	0	(15)
0	0	0	0	0	1	1	(15)
0	0	0	0	0	0	1	(15)
0	0	0	0	0	1	0	(19)
0	0	0	0	0	0	0	(19)

[a] Note that (1,1,1,0,1,1,0) was our observed configuration (see (3.106)).

Comments

67. *Motivation.* The hypothesis H_0 (3.92) is a natural one when an experimenter is testing for a treatment effect and finds it convenient (or necessary) to have the same subjects receive the treatment and also act as controls. Since (X_i, Y_i) then represent two observations on the same subject, it is unrealistic to assume that X_i and Y_i are independent. The hypothesis of no treatment effect is precisely H_0. Terms used by various workers to describe H_0 include exchangeability, interchangeability, and bivariate symmetry. (See Hollander (1971).)

68. *Conditional Nature of the Test.* The hypothesis H_0 implies that the r's defined by (3.94) are independent and identically distributed, each r_i assuming values 1 and 0 with probabilities $\frac{1}{2}$ and $\frac{1}{2}$, respectively. This leads to a conditional distribution P_c that assigns probability $\left(\frac{1}{2}\right)^n$ to each of the A-values associated with each of the possible 2^n r configurations. (In the foregoing statement, we implicitly distinguish between all A-values, although, as we see in Example 3.11, two different r's may yield the same value of A.) The test defined by (3.102) investigates how large A_{obs} is with respect to this conditional distribution. For further information on conditional tests of this nature (which are known as *permutation tests*), see Hoeffding (1952), Box and Andersen (1955), Lehmann (1959), and Scheffé (1959).

69. *Alternative Computation of the d's.* In computing the d's defined by (3.95), life can be made easier by observing:

    (i) $d_{ij} = 0$ for all j, if $a_i = b_i$.
    (ii) $d_{ii} = 1$ if $a_i \neq b_i$, and $d_{ii} = 0$ if $a_i = b_i$.
    (iii) When $i > j$, if $a_i \neq a_j$, then $d_{ij} = 0$.

70. *Parametric Representation of the Null Hypothesis* (H_0). Consider (3.92) and define

$$A^*(x,y) = P(X \le x \text{ and } Y \le y) - P(X \le y \text{ and } Y \le x). \tag{3.112}$$

The hypothesis H_0 (3.92) is true if and only if $A^*(x,y) = 0$ for all (x,y). The statistic (A/n) estimates the parameter

$$\Delta(F) = E_F\{A^*(X',Y')\}^2, \tag{3.113}$$

where (X',Y') is a random member from the underlying bivariate population with distribution F. We may view $A^*(x,y)$ as a measure of the deviation from H_0 at the point (x,y) and $\Delta(F)$ (3.113) as the average value of the square of this deviation.

71. *Consistency: Comparison of A Test and Signed Rank Test.* The A test was designed by Hollander (1971) to detect a broad class of alternatives to the hypothesis of no treatment effect. Thus, although the A test will detect alternatives of the form associated with nonzero ($\theta \neq 0$) treatment effects as discussed for paired replicates data in Section 3.1, it will also be sensitive to differences in dispersion in the (marginal) X and Y populations, as well as to more general deviations from H_0. Of course, a price must be paid for this more general type of protection; namely, we cannot expect the A test to have power as good as that of, say, the Wilcoxon signed rank test (3.6) when the location model of Section 3.1 is true, because the signed rank test is directed to location changes. On the other hand, there are many alternatives to H_0 for which the signed rank test will have power remaining at α (for any sample size), whereas the A test will have power tending to 1 (as n tends to infinity). In fact, under mild conditions on the nature of the underlying bivariate population F, the A test is consistent when H_0 is false.

72. *Other Nonparametric Tests.* Other nonparametric tests for bivariate symmetry are proposed in Kepner and Randles (1984). See Randles and Kepner (1984) and Hilton and Gee (1997a) for power comparisons of A versus competitors when F is bivariate normal and when F is bivariate exponential. See Hilton and Gee (1997b) for an efficient algorithm for conducting the exact test based on A.

Properties

1. *Consistency.* The test defined by (3.102) is consistent against populations for which the parameter $\Delta(F)$ defined by (3.113) is positive. For conditions on F insuring that $\Delta(F)$ will be positive, see Hollander (1971).
2. *Asymptotic Distribution.* See Koziol (1979).

Problems

110. Cain, Mayer, and Jones (1970) have studied albumin and fibrinogen metabolism using the carbonate-^{14}C method to measure the synthetic rate of liver-produced plasma proteins before and after a 13-day course of prednisolone. The eight subjects were patients with hepatocellular

Table 3.18 Intravascular Albumin Pool Before and After Prednisolone

Patient	Intravascular albumin pool (g)	
	Before, X_i	After, Y_i
1	74.4	83.8
2	100.0	97.5
3	82.5	77.4
4	84.3	87.2
5	91.4	116.2
6	92.8	88.2
7	104.2	115.1
8	58.3	50.5

Source: G. D. Cain, G. Mayer, and E. A. Jones (1970).

disease as established by needle biopsy. Part of the study was related to changes in the intravascular albumin pool. Table 3.18 is based on a subset of the Cain–Mayer–Jones data.

Use R to find the exact conditional P-value for these data achieved by the test based on A.

111. Consider the intravascular albumin data in Table 3.18. Use R to determine an approximate P-value for these data based on the A test. Compare with the exact conditional P-value for these data as found in Problem 110.

112. Verify directly, or illustrate with a numerical example, remarks (i)–(iii) of Comment 69.

113. Consider the immunoreactive insulin blood-level data of Table 3.3. Use R to find the exact conditional P-value obtained by the A test for those data.

114. Consider the immunoreactive insulin blood-level data of Table 3.3. Use R to find an approximate P-value and compare it with the exact conditional P-value for these data as found in Problem 113.

115. Calculate $\Delta(F)$ for the bivariate population having joint distribution function $F(x, y) = 0$ for $x < 0, y < 0$; $= xy^2$ for $0 \le x \le 1, 0 \le y \le 1$; $= 1$ for $x > 1, y > 1$.

3.11 EFFICIENCIES OF PAIRED REPLICATES AND ONE-SAMPLE LOCATION PROCEDURES

Recall the normal theory one-sample t-test based on the statistic

$$V = \frac{\sqrt{n}\bar{Z}}{S_z}, \tag{3.114}$$

where $\bar{Z} = \sum_{i=1}^{n} Z_i/n$ and $S_z^2 = \sum_{i=1}^{n} (Z_i - \bar{Z})^2/(n-1)$. The Pitman asymptotic relative efficiency of the one-sample test procedure (one- or two-sided) based on the signed rank statistic T^+ (3.3) with respect to the corresponding normal theory test based on V is

$$e(T^+, V) = 12\sigma_F^2 \left\{ \int_{-\infty}^{\infty} f^2(u)du \right\}^2, \tag{3.115}$$

where σ_F^2 is the variance of the common (continuous and symmetric) distribution $F(\cdot)$ of $Z_1, \ldots, Z_n$ and $f(\cdot)$ is the probability density function corresponding to $F(\cdot)$. The

parameter $\int_{-\infty}^{\infty} f^2(u)du$ is the area under the curve associated with $f^2(\cdot)$, the square of the common probability density function.

The expression in (3.115) was first obtained by Pitman (1948) in the context of hypothesis testing. Hodges and Lehmann (1963) showed that the same expression, $e(T^+, V)$, also pertains to the asymptotic relative efficiency of the point estimator $\hat{\theta}$ (see (3.23)) with respect to $\bar{\theta} = \bar{Z}$. Finally, Lehmann (1963c) established that (3.115) also provides the asymptotic relative efficiency of the confidence interval (or bound) for θ derived from T^+ (see Section 3.3) relative to the corresponding confidence interval (or bound) for θ associated with the one-sample t-test based on V (3.114).

Hodges and Lehmann (1956) demonstrated that within the class of continuous and symmetric $F(\cdot), e(T^+, V)$ is always at least .864. Thus, in this class of distributions, the most efficiency that can be lost when employing a procedure (test, point estimator, or confidence interval/bound) based on T^+ instead of the corresponding normal theory procedure associated with V (3.114) is about 14%. Even when $F(\cdot)$ is normal (the proper setting for procedures based on V), $e(T^+, V) = .955$ and there is only a minor loss (4.5%) in efficiency from using a T^+-based procedure rather than the optimal procedure based on V. On the other hand, $e(T^+, V)$ exceeds 1 for many populations and it can be infinite (e.g., when $F(\cdot)$ is Cauchy). Some values of $e(T^+, V)$ for selected $F(\cdot)$ are

F :	Normal	Uniform	Logistic	Double Exponential	Cauchy
$e(T^+, V)$:	.955	1.000	1.097	1.500	∞

(3.116)

The Pitman asymptotic relative efficiency of the one-sample test procedure (one- or two-sided) based on the sign statistic B (3.39) with respect to the corresponding normal theory test based on V (3.114) is

$$e(B, V) = 4\sigma_F^2 f^2(0), \tag{3.117}$$

where σ_F^2 is the variance and $f(\cdot)$ is the probability density function for the common (continuous and symmetric) distribution $F(\cdot)$ of the Z observations.

Pitman (1948) established the general efficiency expression in (3.117) for the hypothesis tests based on B and V, although Cochran (1937) had previously obtained the particular efficiency value of .637 for the case of an underlying (F) normal distribution. Hodges and Lehmann (1963) showed that the expression $e(B, V)$ also holds for the asymptotic relative efficiency of the point estimator $\tilde{\theta}$ (see (3.58)) with respect to $\bar{\theta} = \bar{Z}$, and the results in Lehmann (1963c) lead to the same conclusion for the confidence interval (or bound) for θ based on B (see Section 3.6) relative to the corresponding confidence interval (or bound) for θ associated with the one-sample t-test based on V (3.114).

Hodges and Lehmann (1956) found that within a certain class of populations, $e(B, V)$ is always at least $\frac{1}{3}$ and it can be infinite. Some values of $e(B, V)$ for selected $F(\cdot)$ are

F :	Normal	Uniform	Logistic	Double Exponential	Cauchy
$e(B, V)$:	.637	.333	.822	2.000	∞

(3.118)

We note that for the paired replicates problem, each Z is actually a difference of two observations. For the efficiency calculation, the common F in the parameters $e(T^+, V)$

and $e(B, V)$ is a distribution for a difference of two independent and identically distributed random variables. Since neither all continuous distributions nor all continuous and unimodal distributions can be distributions for such a difference, the lower bounds for $e(T^+, V)$ and $e(B, V)$ for paired replicates data are obtained over smaller classes of distributions than for the one-sample data. In particular, in the paired case, Hollander (1967a) proved that the lower bound of .864 for $e(T^+, V)$ is no longer attainable. Similarly, Puri and Sen (1968) demonstrated that the lower bound of $\frac{1}{3}$ for $e(B, V)$ is not attainable in the paired case.

For the paired replicates data, the values of $e(T^+, V)$ and $e(B, V)$ remain the same as given in expressions (3.116) and (3.118), respectively, for an underlying (F) normal, logistic, double exponential, or Cauchy distribution. However, the uniform distribution cannot be a distribution for a difference of two independent and identically distributed random variables (see Puri and Sen (1968)).

We do not know of any results for the asymptotic efficiencies of the Randles et al. test for distributional symmetry (Section 3.9) or Hollander's bivariate symmetry test (Section 3.10).