55617

Statistical Methods for Environmental Pollution Monitoring

Richard O. Gilbert
Pacific Northwest Laboratory

LOS ALAMOS
NATIONAL LABORATORY
DEC 05 1991
LIBRARIES
PROPERTY

Chapter 18 : Comparing Populations

VNR VAN NOSTRAND REINHOLD
New York

3505

Dedicated to my parents, Mary Margaret and Donald I. Gilbert

Copyright © 1987 by **Van Nostrand Reinhold**
Library of Congress Catalog Card Number: 86-26758
ISBN 0-442-23050-8

Work supported by the U.S. Department of Energy under Contract No. DE-AC06-76RLO1830.

All rights reserved. No part of this work covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without written permission of the publisher.

Printed in the United States of America

Van Nostrand Reinhold
115 Fifth Avenue
New York, New York 10003

Chapman & Hall
2-6 Boundary Row
London SE1 8HN, England

Thomas Nelson Australia
102 Dodds Street
South Melbourne, Victoria 3205, Australia

Nelson Canada
1120 Birchmount Road
Scarborough, Ontario M1K 5G4, Canada

16 15 14 13 12 11 10 9 8 7 6 5 4

Library of Congress Cataloging-in-Publication Data
Gilbert, Richard O.
Statistical methods for environmental pollution monitoring.
Bibliography: p.
Includes index.
1. Pollution—Environmental aspects—Statistical methods. 2. Environmental monitoring—Statistical methods. I. Title.
TD193.G55 1987 628.5′028 86-26758
ISBN 0-442-23050-8

18 Comparing Populations

An objective of many pollution monitoring and research studies is to make comparisons between pollution levels at different times or places or collected by different measurement techniques. This chapter provides simple nonparametric tests for making such comparisons. These tests do not require that data follow the normal distribution or any other specific distribution. Moreover, many of these tests can accommodate a few missing data or concentrations at the trace or ND (not detected) levels.

We begin with procedures for comparing two populations. The procedures are of two types: those for paired data, and those for independent data sets. Examples of paired data are (i) measurements of two pollutants on each of n field samples, (ii) measurements of a pollutant on air filters collected at two adjacent locations for n time periods, and (iii) measurements of a pollutant on both leaves and roots of the same n plants. The paired test we consider is the sign test. Friedman's test, an extension of the sign test to more than two populations, is also given.

Independent data sets are those for which there is no natural way to pair the data. For example, if n soil samples are collected at each of two hazardous waste sites, there may be no rational way to pair a pollution measurement from one site with a pollution measurement from the other site. For this type of data we illustrate Wilcoxon's rank sum test (also known as the Mann-Whitney test) for the comparison of two populations and the Kruskal-Wallis test for the comparison of more than two populations. The tests discussed in this chapter can be computed by using a statistical software computer package such as Biomedical Computer Programs P Series, (1983) and Statistical Package for the Social Sciences (1985). Additional information on the tests in this chapter and on related testing, parameter estimation, and confidence interval procedures are given in Lehmann (1975), Conover (1980) and/or Hollander and Wolfe (1973).

18.1 TESTS USING PAIRED DATA

Suppose n paired measurements have been made. Denote these pairs by $(x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n})$, where x_{1i} is the ith observation from population 1 and x_{2i} is the paired ith observation from population 2. When data are paired, we could compare the two populations by looking at the sign or the magnitudes of the set of n differences $D_i = x_{2i} - x_{1i}$, $i = 1, 2, \ldots, n$. The sign test uses the signs and the Wilcoxon signed rank test uses the magnitudes. These two tests are alternatives to the commonly used paired t test described in many statistical methods books, e.g., Snedecor and Cochran (1980, p. 85). The latter test should be used if the differences are a random sample from a normal distribution.

18.1.1 Sign Test

The sign test is simple to compute and can be used no matter what the underlying distribution may be. It can also accommodate a few ND (not detected) concentrations. It is more versatile than the Wilcoxon signed rank test since the latter test requires that the underlying distribution be symmetric (though not necessarily normal) and that no NDs be present. However, the Wilcoxon test will usually have more power than the sign test to detect differences between the two populations. The sign test may be the better choice if ease of computation is an important consideration.

The sign test statistic, B, is the number of pairs (x_{1i}, x_{2i}) for which $x_{1i} < x_{2i}$, that is, the number of positive differences D_i. The magnitudes of the D_i are not considered; only their signs are. If any D_i is zero so that a + or − sign cannot be assigned, this data pair is dropped from the data set and n is reduced by 1. The statistic B is used to test the null hypothesis:

H_0: The median of the population of all possible differences is zero, that is, x_{1i} is as likely to be larger than x_{2i} as x_{2i} is likely to be larger than x_{1i} **18.1**

Clearly, if the number of + and − signs are about equal, there is little reason to reject H_0.

Two-Sided Test

If the number of paired data, n, is 75 or less, we may use Table A14 to test H_0 versus the alternative hypothesis

H_A: The median difference does not equal zero, that is, x_{1i} is more likely to exceed x_{2i} than x_{2i} is likely to exceed x_{1i}, or vice versa **18.2**

Then reject H_0 and accept H_A at the α significance level if

$$B \le l - 1 \quad \text{or} \quad B \ge u$$

where l and u are integers taken from Table A14 for the appropriate n and chosen α.

For example, suppose there are $n = 34$ differences, and we choose to test at the $\alpha = 0.05$ level. Then we see from Table A14 that we reject H_0 and accept H_A if $B \le 10$ or if $B \ge 24$.

EXAMPLE 18.1

Grivet (1980) reports average and maximum oxidant pollution concentrations at several air monitoring stations in California. The daily maximum (of hourly average) oxidant concentrations (parts per hundred million) at 2 stations for the first 20 days in July 1972 are given in Table 18.1. The data at the 2 stations are paired and perhaps correlated because they were taken on the same day. This type of correlation is permitted, but correlation between the pairs, that is, between observations taken on different days, should not be present for the test to be completely valid. If this latter type of positive correlation is present, the test would indicate more than the allowed $100\alpha\%$ of the time a significant difference between the 2 stations when none actually exists. This problem is discussed by Gastwirth and Rubin (1971) and by Albers (1978*a*).

We test the null hypothesis that the median difference in maximum concentrations between the two stations is zero, that is, there is no tendency for the oxidant concentrations at one station to be larger than at the other station. Since concentrations are tied on 3 days, n equals 17 rather than 20. The number of + signs is $B = 9$, the number of days that the maximum concentration at station 41541 exceeds that at station 28783. Suppose we use $\alpha = 0.05$. Then from Table A14 for $n = 17$ we find $l = 6$ and $u = 15$. Since B is not less than or equal to $l - 1 = 5$ nor greater than or equal to 15, we cannot reject H_0.

Table A14 gives values of l and u for $n \leq 75$. When $n > 20$, we may use the following approximate test procedure:

1. Compute

$$Z_B = \frac{B - n/2}{\sqrt{n/4}} = \frac{2B - n}{\sqrt{n}} \tag{18.3}$$

2. Reject H_0 and accept H_A (Eq. 18.2) if $Z_B \leq -Z_{1-\alpha/2}$ or if $Z_B \geq Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is obtained from Table A1.

EXAMPLE 18.2

Using the data in Example 18.1, we test H_0 versus H_A, using Z_B. We have $Z_B = (18 - 17)/\sqrt{17} = 0.243$. For $\alpha = 0.05$, Table A1 gives $Z_{0.975} = 1.96$. Since Z_B is not less than or equal to -1.96

Table 18.1 Maximum Oxidant Concentrations[a] at Two Stations in July 1972

Day	*Station 28783*	*Station 41541*	*Sign of Difference*	*Day*	*Station 28783*	*Station 41541*	*Sign of Difference*
1	8	10	+	11	11	13	+
2	5	7	+	12	12	14	+
3	6	7	+	13	13	20	+
4	7	7	[b]	14	14	28	+
5	4	6	+	15	12	6	−
6	4	6	+	16	12	7	−
7	3	3	[b]	17	13	7	−
8	5	4	−	18	14	6	−
9	5	5	[b]	19	12	4	−
10	6	4	−	20	15	5	−

[a]Data are parts per hundred million.
[b]Tied concentrations.

nor greater than or equal to 1.96, we cannot reject H_0. This conclusion is the same as that obtained by using Table A14 in Example 18.1.

One-Sided Test

Thus far we have considered only a two-sided alternative hypothesis (Eq. 18.2). One-sided tests may also be used. There are two such tests:

1. Test H_0 versus the alternative hypothesis, H_A, that the x_2 measurements tend to exceed the x_1 measurements more often than the reverse. In this case reject H_0 and accept H_A if $B \geq u$, where u is obtained from Table A14. Alternatively, if $n > 20$, reject H_0 and accept H_A if $Z_B \geq Z_{1-\alpha}$, where Z_B is computed by Eq. 18.3 and $Z_{1-\alpha}$ is from Table A1.
2. Test H_0 versus the alternative hypothesis that the x_1 measurements tend to exceed the x_2 measurements more often than the reverse. If $n \leq 75$, use Table A14 and reject H_0 and accept H_A if $B \leq l - 1$. Alternatively, if $n > 20$, reject H_0 and accept H_A if $Z_B \leq -Z_{1-\alpha}$, where Z_B is computed by Eq. 18.3.

When one-sided tests are conducted with Table A14, the α levels indicated in the table are divided by 2. Hence, Table A14 may only be used to make one-sided tests at the 0.025 and 0.005 significance levels.

Trace Concentrations

The sign test can be conducted even though some data are missing or are ND concentrations. See Table 18.2 for a summary of the types of data that can occur, whether or not the sign can be determined, and the effect on n. The effect of decreasing n is to lower the power of the test to indicate differences between the two populations.

18.1.2 Wilcoxon Signed Rank Test

The *Wilcoxon signed rank test* can be used instead of the sign test if the underlying distribution is symmetric, though it need not be a normal distribution. This Wilcoxon test (not to be confused with the Wilcoxon rank sum test discussed in Section 18.2.1) is more complicated to compute than the sign test

Table 18.2 Determination of the Sign Test

Type of Data	*Can Sign Be Computed?*	*Decreases n?*
One or both members of a pair are absent	No	Yes
$x_{2i} = x_{1i}$	No	Yes
One member of a pair is ND	Yes[a]	No
Both members of a pair are ND	No	Yes

[a]If the numerical value is greater than the detection limit of the ND value.

because it requires computing and ranking the D_i. In most situations it should have greater power to find differences in two populations than does the sign test. The null and alternative hypotheses are the same as for the sign test. The test is described by Hollander and Wolfe (1973).

18.1.3 Friedman's Test

Friedman's test is an extension of the sign test from two paired populations to k related populations. The underlying distribution need not be normal or even symmetric. Also, a moderate number of ND values can be accommodated without seriously affecting the test conclusions. However, no missing values are allowed. The null hypothesis is

H_0: There is no tendency for one population to have larger or smaller values than any other of the k populations

The usual alternative hypothesis is

H_A: At least one population tends to have larger values than one or more of the other populations

Examples of "populations" appropriate for Friedman's test are (i) measurements of $k = 3$ or more pollutants on each of n field samples, (ii) measurements of a single pollutant on air filters collected at $k = 3$ or more air monitoring stations for n time periods, or (iii) measurements obtained by $k = 3$ or more analytical laboratories on a set of n identical spiked samples. The data are laid out as follows:

	Block				
	1	2	3	....	n
Population 1	x_{11}	x_{12}	x_{13}	...	x_{1n}
Population 2	x_{21}	x_{22}	x_{23}	...	x_{2n}
⋮					
Population k	x_{k1}	x_{k2}	x_{k3}	...	x_{kn}

The steps in the testing procedure are as follows:

1. For each block, assign the rank 1 to the smallest measurement, the rank 2 to the next largest measurement, . . . , and the rank k to the largest measurement. If two or more measurements in the block are tied, then assign to each the midrank for that tied group (illustrated Example 18.3).
2. Compute R_j, the sum of the ranks for the jth population.
3. If no tied values occur within any block, compute the Friedman test statistic as follows:

$$F_r = \left[\frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2\right] - 3n(k+1) \qquad \textbf{18.4}$$

4. If tied values are present within one or more blocks, compute the Friedman statistic as follows:

$$F_r = \frac{12 \sum_{j=1}^{k} \left[R_j - \frac{n(k+1)}{2} \right]^2}{nk(k+1) - \frac{1}{k-1} \sum_{i=1}^{n} \left\{ \left(\sum_{j=1}^{g_i} t_{i,j}^3 \right) - k \right\}} \qquad \textbf{18.5}$$

where g_i is the number of tied groups in block i and $t_{i,j}$ is the number of tied data in the jth tied group in block i. Each untied value within block i is considered to be a "group" of ties of size 1. The quantity in braces { } in the denominator of Eq. 18.5 is zero for any block that contains no ties. This method of handling ties is illustrated in Example 18.3. Equation 18.5 reduces to Eq. 18.4 when there are no ties in any block.

5. For an α level test, reject H_0 and accept H_A if $F_r \geq \chi^2_{1-\alpha,k-1}$, where $\chi^2_{1-\alpha,k-1}$ is the $1 - \alpha$ quantile of the chi-square distribution with $k - 1$ df, as obtained from Table A19, where k is the number of populations. The chi-square distribution is appropriate only if n is reasonably large. Hollander and Wolfe provide exact critical values (their Table A.15) for testing F_r for the following combinations of k and n: $k = 3$, $n = 2, 3, \ldots, 13$; $k = 4$, $n = 2, 3, \ldots, 8$; $k = 5$, $n = 3, 4, 5$. Odeh et al. (1977) extend these tables to $k = 5$, $n = 6, 7, 8$; $k = 6$, $n = 2, 3, 4, 5, 6$. These tables will give only an approximate test if ties are present. The use of F_r computed by Eq. 18.5 and evaluated using the chi-square tables may be preferred in this situation. The foregoing tests are completely valid only if the observations in different blocks are not correlated.

In step 1, if there is one ND value within a block, assign it the rank 1. If there are two or more ND values within a block, treat them as tied values and assign them the midrank. For example, if three NDs are present within a block, each is assigned the rank of 2, the average of 1, 2, and 3. This method of handling NDs assumes all measurements in the block are greater than the detection limit of all the ND values in the block.

EXAMPLE 18.3

The data in Table 18.3 are daily maximum oxidant air concentrations (parts per hundred million) at $k = 5$ monitoring stations in California for the first $n = 6$ days in July 1973 (from Grivet, 1980). We shall use Friedman's procedure to test at the $\alpha = 0.025$ significance level the null hypothesis, H_0, that there is no tendency for any station to

Table 18.3 Daily Maximum Air Concentrations[a] in California During July 1973

Station Number	Day (Block) 1	2	3	4	5	6	Sum of Ranks (R_j)
28783	7 (2)[b]	5 (4)	7 (3.5)	12 (2)	4 (3)	4 (4)	18.5
41541	5 (1)	3 (1.5)	3 (1)	8 (1)	3 (1.5)	2 (1)	7
43382	11 (4)	4 (3)	7 (3.5)	17 (4)	5 (4.5)	4 (4)	23
60335	13 (5)	6 (5)	12 (5)	21 (5)	5 (4.5)	4 (4)	28.5
60336	8 (3)	3 (1.5)	4 (2)	13 (3)	3 (1.5)	3 (2)	13

[a] Data are parts per hundred millions.
[b] Rank of the measurement.

Table 18.4 Computing the Correction for Ties in Eq. 18.5 for Friedman's Test

Blocks	g_i	$t_{i,j}$	$\sum_{j=1}^{g_i} t_{i,j}^3 - k$
2	4	$t_{2,1} = 2,\ t_{2,2} = t_{2,3} = t_{2,4} = 1$	$11 - 5 = 6$
3	4	$t_{3,1} = 2,\ t_{3,2} = t_{3,3} = t_{3,4} = 1$	$11 - 5 = 6$
5	3	$t_{5,1} = t_{5,2} = 2,\ t_{5,3} = 1$	$17 - 5 = 12$
6	3	$t_{6,1} = 3,\ t_{6,2} = t_{6,3} = 1$	$29 - 5 = 24$
			Sum $= 48$

have oxidant levels greater or smaller than any other station. Also shown in Table 18.3 are the ranks of the measurements obtained as in step 1 and the sum of the ranks for each station (step 2).

Since there are ties in blocks 2, 3, 5, and 6, we must use Eq. 18.5 to compute F_r. First compute the quantity in braces { } in the denominator of Eq. 18.5 for all blocks that contain ties. This computation is done in Table 18.4.

Note that the values of all the $t_{i,j}$'s in each block must sum to k, the number of populations (stations). Since $k = 5$, $n = 6$, and Sum $= 48$, Eq. 18.5 is

$$F_r = \frac{12[(18.5 - 18)^2 + (7 - 18)^2 + \cdots + (13 - 18)^2]}{6(5)(6) - 48/4}$$

$$= 20.1$$

For $\alpha = 0.025$ we find from Table A19 that $\chi^2_{0.975,4} = 11.14$. Since $F_r > 11.14$, we reject H_0 and accept the H_A that at least 1 station tends to have daily maximum oxidant concentrations at a different level than the other stations. From Table 18.3 it appears that stations 41541 and 60336 have consistently lower concentrations than the other stations.

18.2 INDEPENDENT DATA SETS

We discuss two nonparametric tests for independent data sets: the Wilcoxon rank sum test (not to be confused with the Wilcoxon signed rank test discussed in Section 18.1.2) and the Kruskal-Wallis rank test, which generalizes the Wilcoxon rank sum test to more than two populations.

18.2.1 Wilcoxon Rank Sum Test

The Wilcoxon rank sum test may be used to test for a shift in location between two independent populations, that is, the measurements from one population tend to be consistently larger (or smaller) than those from the other population. This test is an easily computed alternative to the usual independent-sample t test discussed in most statistics methods books (see, e.g., Snedecor and Cochran, 1980, p. 83). (Do not confuse the independent-sample t test with the paired t test for paired data. The latter is discussed by Snedecor and Cochran, 1980, p. 85.)

The rank sum test has two main advantages over the independent-sample t test: (i) The two data sets need not be drawn from normal distributions, and (ii) the rank sum test can handle a moderate number of ND values by treating them as ties (illustrated in Example 18.4). However, both tests assume that the distributions of the two populations are identical in shape (variance), but the distributions need not be symmetric. Modifications to the t test to account for unequal variances can be made as described in Snedecor and Cochran (1980, p. 96). Evidently, no such modification exists for the rank sum test. Reckhow and Chapra (1983) illustrate the use of the rank sum test on chlorophyll data in two lakes.

Suppose there are n_1 and n_2 data in data sets 1 and 2, respectively (n_1 need not equal n_2). We test

H_0: The populations from which the two data sets have been drawn have the same mean **18.6**

versus the alternative hypothesis

H_A: The populations have different means **18.7**

The Wilcoxon rank sum test procedure is as follows:

1. Consider all $m = n_1 + n_2$ data as one data set. Rank the m data from 1 to m, that is, assign the rank 1 to the smallest datum, the rank 2 to the next largest datum, . . . , and the rank m to the largest datum. If several data have the same value, assign them the midrank, that is, the average of the ranks that would otherwise be assigned to those data.
2. Sum the ranks assigned to the n_1 measurements from population 1 Denote this sum by W_{rs}.
3. If $n_1 \leq 10$ and $n_2 \leq 10$, the test of H_0 may be made by referring W_{rs} to the appropriate critical value in Table A.5 in Hollander and Wolfe (1973) (see their pages 67–74 for the test method).
4. If $n_1 > 10$ and $n_2 > 10$ and no ties are present, compute the large sample statistic

$$Z_{rs} = \frac{W_{rs} - n_1(m+1)/2}{[n_1 n_2(m+1)/12]^{1/2}} \qquad \textbf{18.8}$$

5. If $n_1 > 10$ and $n_2 > 10$ and ties are present, do not compute Eq. 18.8. Instead, compute

$$Z_{rs} = \frac{W_{rs} - n_1(m+1)/2}{\left\{\dfrac{n_1 n_2}{12}\left[m + 1 - \dfrac{\sum_{j=1}^{g} t_j(t_j^2 - 1)}{m(m-1)}\right]\right\}^{1/2}} \qquad \textbf{18.9}$$

where g is the number of tied groups and t_j is the number of tied data in the jth group. Equation 18.9 reduces to Eq. 18.8 when there are no ties.
6. For an α level two-tailed test, reject H_0 (Eq. 18.6) and accept H_A (Eq. 18.7) if $Z_{rs} \leq -Z_{1-\alpha/2}$ or if $Z_{rs} \geq Z_{1-\alpha/2}$.
7. For a one-tailed α level test of H_0 versus the H_A that the measurements from population 1 tend to exceed those from population 2, reject H_0 and accept H_A if $Z_{rs} \geq Z_{1-\alpha}$.

8. For a one-tailed α level test of H_0 versus the H_A that the measurements from population 2 tend to exceed those from population 1, reject H_0 and accept H_A if $Z_{rs} \leq -Z_{1-\alpha}$.

EXAMPLE 18.4

In Table 18.5 are ^{241}Am concentrations (pCi/g) in soil crust material collected within 2 plots, one near ("onsite") and one far ("offsite") from a nuclear reprocessing facility (Price, Gilbert, and Gano, 1981). Twenty measurements were obtained in each plot. We use the Wilcoxon rank sum test to test the null hypothesis that average concentrations at the 2 plots are equal versus the alternative hypothesis that the onsite plot (population 1) has larger concentrations than in the offsite plot (population 2). That is, we perform the test in step 7. We shall use $\alpha = 0.05$. The ranks of the combined data are shown in Table 18.5 and W_{rs} is computed to be 500.

There are $g = 6$ groups of ties. Four groups have length 2, that is, $t = 2$, and 2 groups have $t = 3$. Equation 18.9 gives

$$Z_{rs} = \frac{500 - 20(41)/2}{\{[20(20)/12]\,[41 - [(4)(2)(3) + (2)(3)(8)]/40(39)]\}^{1/2}} = 2.44$$

Performing the test in step 7, since $Z_{rs} > 1.645$, we reject H_0 and accept H_A that the onsite population has larger ^{241}Am concentrations than the offsite plot.

We note that the correction for ties, that is, using Eq. 18.9 instead of Eq. 18.8, will usually have a negligible effect on the value of Z_{rs}. The correction becomes more important if the t_j are large. Also, if NDs are present but occur in only one of the populations, it is still possible to rank all the data and perform the test. For instance in Example 18.4 if the negative concentrations had been reported by the analytical laboratory as ND values, they would still have been assigned the ranks 1, 2, and 3 if NDs were treated as being less in value than the smallest numerical value (0.0056). In addition, if the three ND values had been considered to be tied, all three would have been assigned the

Table 18.5 ^{241}Am Concentrations in (pCi/g) Soil Crust Material

Population 1 (onsite)				*Population 2 (offsite)*			
0.0059	(5)[a]	0.036	(28)	−0.011[b]	(1)	0.019	(16)
0.0074	(7)	0.040	(29)	−0.0088[b]	(2)	0.020	(18.5)
0.015	(9.5)	0.042	(30)	−0.0055[b]	(3)	0.020	(18.5)
0.018	(13.5)	0.045	(31)	0.0056	(4)	0.022	(20)
0.019	(16)	0.046	(32)	0.0063	(6)	0.025	(22)
0.019	(16)	0.053	(34)	0.013	(8)	0.030	(23)
0.024	(21)	0.062	(36)	0.015	(9.5)	0.031	(25)
0.031	(25)	0.066	(37)	0.016	(11.5)	0.050	(33)
0.031	(25)	0.069	(38)	0.016	(11.5)	0.057	(35)
0.034	(27)	0.081	(40)	0.018	(13.5)	0.073	(39)

$W_{rs} = 5 + 7 + 9.5 + \cdots + 38 + 40 = 500$.

[a]Rank of the datum.

[b]Negative measurements reported by the analytical laboratory.

average rank of 2, which would not have changed the value of W_{rs}. If NDs occur in both populations, they can be treated as tied values all less than the smallest numerical value in the combined data set. Hence, they would each receive the average rank value for that group of NDs, and the Wilcoxon test could still be conducted. (See Exercise 18.4.)

18.2.2 Kruskal-Wallis Test

The Kruskal-Wallis test is an extension of the Wilcoxon rank sum test from two to k independent data sets. These data sets need not be drawn from underlying distributions that are normal or even symmetric, but the k distributions are assumed to be identical in shape. A moderate number of tied and ND values can be accommodated. The null hypothesis is

H_0: The populations from which the k data sets have been drawn have the same mean **18.10**

The alternative hypothesis is

H_A: At least one population has a mean larger or smaller than at least one other population **18.11**

The data take the form

Population

1	2	3	...	k
x_{11}	x_{21}	x_{31}	...	x_{k1}
x_{12}	x_{22}	x_{32}	...	x_{k2}
⋮	⋮	⋮	⋮	⋮
x_{1n_1}	x_{2n_2}	x_{3n_3}		x_{kn_k}

The total number of data is $m = n_1 + n_2 + \cdots + n_k$, where the n_i need not be equal. The steps in the testing procedure are as follows:

1. Rank the m data from smallest to largest, that is, assign the rank 1 to the smallest datum, the rank 2 to the next largest, and so on. If ties occur, assign the midrank (illustrated in Example 18.5). If NDs occur, treat these as a group of tied values that are less than the smallest numerical value in the data set (assuming the detection limit of the ND values is less than the smallest numerical value).
2. Compute the sum of the ranks for each data set. Denote this sum for the jth data set by R_j.
3. If there are no tied or ND values, compute the Kruskal-Wallis statistic as follows:

$$K_w = \left[\frac{12}{m(m+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j}\right] - 3(m+1) \qquad \textbf{18.12}$$

4. If there are ties or NDs treated as ties, compute a modified Kruskal-Wallis statistic by dividing K_w (Eq. 18.12) by a correction for ties, that is, compute

$$K'_w = \frac{K_w}{1 - \dfrac{1}{m(m^2 - 1)} \sum_{j=1}^{g} t_j(t_j^2 - 1)} \qquad \text{18.13}$$

where g is the number of tied groups and t_j is the number of tied data in the jth group. Equation 18.13 reduces to Eq. 18.12 when there are no ties.

5. For an α level test, reject H_0 and accept H_A if $K'_w \geq \chi^2_{1-\alpha,k-1}$, where $\chi^2_{1-\alpha,k-1}$ is the $1 - \alpha$ quantile of the chi-square distribution with $k - 1$ df, as obtained from Table A19, where k is the number of data sets. Iman, Quade, and Alexander (1975) provide exact significance levels for the following cases:

$$\begin{array}{ll} k = 3 & n_i \leq 6 \\ & n_1 = n_2 = n_3 = 7 \\ & n_1 = n_2 = n_3 = 8 \\ k = 4 & n_i \leq 4 \\ k = 5 & n_i \leq 3 \end{array}$$

Less extensive exact tables are given in Conover (1980) and Hollander and Wolfe (1973) for $k = 3$ data sets.

EXAMPLE 18.5

An aliquot-size variability study is conducted in which multiple soil aliquots of sizes 1 g, 10 g, 25 g, 50 g, and 100 g are analyzed for ^{241}Am. A portion of the data for aliquot sizes 1 g, 25 g, and 100 g is used in this example. (Two ND values are added for illustration.) The full data set is discussed by Gilbert and Doctor (1985). We test the null hypothesis that the concentrations from all 3 aliquot sizes have the same mean. The alternative hypothesis is that the concentrations for at least 1 aliquot size tend to be larger or smaller than those for at least 1 other aliquot size. We test at the $\alpha = 0.05$ level. The data, ranks, and rank sums are given in Table 18.6.

Table 18.6 Aliquot-Size Variability Study

^{241}*Am Concentrations (nCi/g)*		
1 g	*25 g*	*100 g*
1.45 (7)[a]	1.52 (8.5)	1.74 (13)
1.27 (6)	2.46 (22)	2.00 (17.5)
1.17 (4)	1.23 (5)	1.79 (14)
1.01 (3)	2.20 (20)	1.81 (15)
2.30 (21)	2.68 (23)	1.91 (16)
1.54 (10)	1.52 (8.5)	2.11 (19)
1.71 (11.5)	ND (1.5)	2.00 (17.5)
1.71 (11.5)		
ND (1.5)		
$R_1 = 75.5$	$R_2 = 88.5$	$R_3 = 112$
$n_1 = 9$	$n_2 = 7$	$n_3 = 7$

[a]Rank of the datum.
ND = not detected.

There are $g = 4$ groups of ties and $t = 2$ for each group. The modified Kruskal-Wallis statistic (Eq. 18.13) is

$$K'_w = \frac{[12/23(24)](75.5^2/9 + 88.5^2/7 + 112^2/7) - 3(24)}{1 - 4(2)(3)/23(528)} = 5.06$$

From Table A19 we find $\chi^2_{0.95,2} = 5.99$. Since $K'_w < 5.99$, we cannot reject H_0 at the $\alpha = 0.05$ level.

Note that the correction for ties made a negligible difference in the test statistic. However, a bigger correction is obtained if t is large for one or more groups of ties. This could happen if there are many NDs, where t is the number of NDs.

18.3 SUMMARY

This chapter discussed simple nonparametric tests to determine whether observed differences in two or more populations are statistically significant, that is, of a greater magnitude than would be expected to occur by chance. We emphasize the correction for ties that these nonparametric tests provide, since a moderate number of trace or ND measurements can be accommodated by assuming they are a group of tied values. Hollander and Wolfe (1973) and Conover (1980) provide other uses for these tests and discuss related estimation and confidence interval procedures.

EXERCISES

18.1 Use the first 10 days of oxidant data in Example 18.1 to conduct a one-tailed sign test at the $\alpha = 0.025$ level. Use the alternative hypothesis H_A: Maximum oxidant concentrations at station 41541 tend to exceed those at station 28781 more than the reverse.

18.2 Suppose the following paired measurements have been obtained (ND = not detected; M = missing data):

	Pair										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
x_1	ND	7	ND	M	3	M	3	7	12	10	15
x_2	ND	6	6	6	1	M	2	1	11	8	3

Conduct a one-tailed sign test of H_0 versus the H_A that x_1 measurements tend to exceed x_2 measurements more often than the reverse. Use $\alpha = 0.025$.

18.3 Compute Friedman's test, using the data in Example 18.3 and $\alpha = 0.025$. Ignore the correction for ties.

18.4 Suppose all ^{241}Am concentrations less than 0.02 pCi/g in the 2 populations in Example 18.4 were reported by the analytical laboratory as ND. Use the Wilcoxon rank sum test to test H_0: means of both populations are equal versus H_A: the offsite population has a smaller mean than the onsite

population.
have on Z_{rs}

18.5 Suppose th
18.5 were
on the resul
18.5.)

ANSWERS

18.1 Denote stat
$B = 5$. Fr
the $\alpha = 0$

18.2 Delete pair
Since $B =$

18.3 Equation 1
11.14, the

18.4 $W_{rs} = 48$
Since Z_{rs}
2.

18.5 $R_1 = 74$,
Since K'_w

population. Use $\alpha = 0.025$. What effect do the large number of NDs have on Z_{rs}? Is it more difficult to reject H_0 if NDs are present?

18.5 Suppose that all ^{241}Am measurements less than 1.5 nCi/g in Example 18.5 were reported by the laboratory as ND. Use the Kruskal-Wallis test on the resulting data set. Use $\alpha = 0.05$. (Retain the 2 NDs in Example 18.5.)

ANSWERS

18.1 Denote station 28781 data as x_1 data, and station 41541 data as x_2 data. $B = 5$. From Table A14, $u = 7$. Since $B < 7$, we cannot reject H_0 at the $\alpha = 0.025$ level.

18.2 Delete pairs 1, 4, and 6. $n = 8$, $B = 1$. From Table A14, $l - 1 = 0$. Since $B = 1$, we cannot reject H_0.

18.3 Equation 18.4 gives $F_r = 18.77$. Reject H_0 and accept H_A, since $18.77 > 11.14$, the same result as when the correction for ties was made.

18.4 $W_{rs} = 487$. $g = 3$ with $t = 17$, 2, and 3. $Z_{rs} = 77/35.5163 = 2.168$. Since $Z_{rs} > 1.645$, reject H_0 and accept H_A. The NDs reduced Z_{rs} from 2.436 in Example 18.4 to 2.168. Yes!

18.5 $R_1 = 74$, $R_2 = 90$, $R_3 = 112$, $m = 23$, $K'_w = 5.339/0.97085 = 5.50$. Since $K'_w < 5.99$, we cannot reject H_0 at the $\alpha = 0.05$ level.