Nonparametric Statistics

Relax Normality

Radu T. Trîmbiţaş

May 19, 2016

1 Introduction

Introduction

- The term *nonparametric statistics* has no standard definition that is agreed on by all statisticians.
- *Parametric methods* – those that apply to problems where the distribution(s) from which the sample(s) is (are) taken is (are) specified except for the values of a finite number of parameters.
- *Nonparametric methods* apply in all other instances.
  - The one-sample t test applies when the population is normally distributed with unknown mean and variance. Because the distribution from which the sample is taken is specified except for the values of two parameters, μ and σ^2, the t test is a parametric procedure.
  - Suppose that independent samples are taken from two populations and we wish to test the hypothesis that the two population distributions are identical but of unspecified form. In this case, the distribution is unspecified, and the hypothesis must be tested by using nonparametric methods.
- Valid employment of some of the parametric methods presented in preceding lectures requires that certain distributional assumptions are at least approximately met. Even if all assumptions are met, research has shown that nonparametric statistical tests are almost as capable of detecting differences among populations as the applicable parametric methods. They may be, and often are, more powerful in detecting population differences when the assumptions are not satisfied.

2 A General Two-Sample Shift Model

A General Two-Sample Shift Model

- Test wether two populations have the same distribution
- Independent random samples $X_1, X_2, \ldots, X_{n_1} \sim N(\mu_X, \sigma^2)$ and $Y_1, Y_2, \ldots, Y_{n_2} \sim N(\mu_Y, \sigma^2)$; the experimenter may wish to test $H_0 : \mu_X - \mu_Y = 0$ versus $H_a : \mu_X - \mu_Y < 0$. If H_0 is true, the population distributions are identical. If H_a is true, then $\mu_Y > \mu_X$ and the distributions of X_1 and Y_1 are the same, except that the location parameter (μ_Y) for Y_1 is larger than the location parameter (μ_X) for X_1. Hence, the distribution of Y_1 is shifted to the right of the distribution of X_1 (see Figure 1).
- This is an example of a two-sample parametric shift (or location) model. The model is parametric because the distributions are specified (normal) except for the values of the parameters μ_X, μ_Y, and σ^2. The amount that the distribution of Y_1 is shifted to the right of the distribution of X_1 is $\mu_Y - \mu_X$ (see Figure 1).
- We define a shift model that applies for any distribution, normal or otherwise.
- Let $X_1, X_2, \ldots, X_{n_1}$ be a random sample from a population with distribution function $F(x)$ and let $Y_1, Y_2, \ldots, Y_{n_2}$ be a random sample from a population with distribution function $G(y)$. If we wish to test whether the two populations have the same distribution—that is, $H_0 : F(z) = G(z)$ versus $H_a : F(z) \neq G(z)$, with the actual form of $F(z)$ and $G(z)$ unspecified—a nonparametric method is required.
- Notice that H_a is a very broad hypothesis.
- we assume that $X_1, X_2, \ldots, X_{n_1}$ constitute a random sample from distribution function $F(x)$ and that $Y_1, Y_2, \ldots, Y_{n_2}$ constitute a random sample from distribution function $G(y) = F(y - \theta)$ for some unknown value θ. For the two-sample shift model, $H_0 : F(z) = G(z)$ is equivalent to $H_0 : \theta = 0$. If θ is greater (less) than 0, then the distribution of the Y-values is located to the right (left) of the distribution of the X-values. (see Figure 2)

3 Sign Test (Median Test)

Sign Test (Median Test)

- Is a nonparametric alternative to t-test for one sample
- Simple and versatile

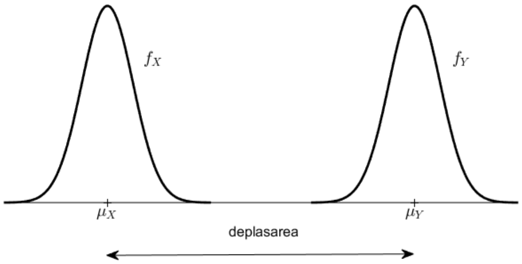

Figure 1: Two normal distributions with same variances and different means

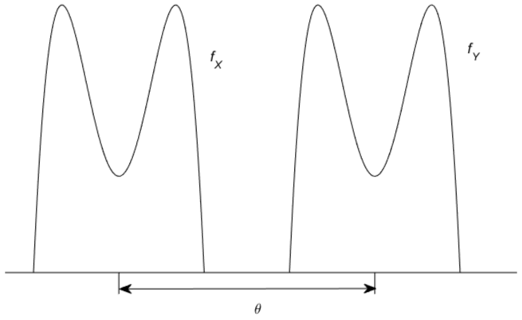

Figure 2: Two probability densities, one shifted with θ with resppect to the other

- Hypotheses $H_0 : m = m_0$ (median) with respect to usual alternatives
- The data are converted to and signs according to whether each data value is more or less than m_0. A plus sign will be assigned to each larger than m_0, a minus sign to each smaller than m_0, and a zero to those equal to m_0. The sign test uses only the plus and minus signs; the zeros are discarded.
- Test statistics $M = \min(n(+), n(-))$, where $n(+)$ is the number of + signs. Distribution of $M \sim b(n, 1/2)$.
- Normal approximation

$$Z = \frac{M - n/2}{\frac{\sqrt{n}}{2}}.$$

Example 1. The following data give temperature (in °C) for 20 days: 22 18 17 16 13 20 19 21 20 16 14 17 21 21 17 17 17 22 22 21. Test if the median is $m = 19$.

Solution. See `nonpar/ex15_0.pdf` ☐

4 The Sign Test for a Matched-Pairs Experiment

The Sign Test for a Matched-Pairs Experiment

- n pairs of observations of the form (X_i, Y_i); H_0 : X and Y have the same continuous distributions; H_a : the distributions differ in location.
- Let $D_i = X_i - Y_i$. H_0 is true $\Longrightarrow P(D_i > 0) = P(D_i < 0) = 1/2$
- Let M denote the total number of positive (or negative) differences. Then if the variables X_i and Y_i have the same distribution, M has a binomial distribution with $p = 1/2$, and the rejection region for a test based on M can be obtained by using the binomial probability distribution.
- **Problem** that may arise: observations associated with one or more pairs may be equal and therefore may result in *ties*. When this situation occurs, delete the tied pairs and reduce n, the total number of pairs.

Sign Test Summary
The Sign Test for a Matched-Pairs Experiment
Let $p = P(X > Y)$.

Null hypothesis: $H_0 : p = 1/2$.

Alternative hypothesis: $H_a : p > 1/2$ or ($p < 1/2$ or $p \neq 1/2$).

Test statistic: $M = card\{D_i = X_i - Y_i > 0 : i = 1, \ldots, n\}$.

Rejection region:
- For $Ha : p > 1/2$, reject H_0 for the largest values of M;

Day	A	B
1	172	201
2	165	179
3	206	159
4	184	192
5	174	177
6	142	170
7	190	182
8	169	179
9	161	169
10	200	210

Table 1: Data for Example 2

- for $Ha : p < 1/2$, reject H_0 for the smallest values of M;
- for $H_a : p = 1/2$, reject H_0 for very large or very small values of M.

Assumptions: The pairs (X_i, Y_i) are randomly and independently selected.

Example

Example 2. The number of defective electrical fuses produced by each of two production lines, A and B,was recorded daily for a period of 10 days, with the results shown in Table 1. Assume that both production lines produced the same daily output. Compare the number of defectives produced by A and B each day and let M equal the number of days when A exceeded B. Do the data present sufficient evidence to indicate that either production line produces more defectives than the other? State the null hypothesis to be tested and use M as a test statistic.

Solution. H_0 : identical distribution $\Longrightarrow p = P(A > B) = 0.5$ (given that there are no ties) $\Longrightarrow H_0 \Longleftrightarrow$binomial parameter $p = .5$.

Very large or very small values of M are most contradictory to the null hypothesis. Therefore, the rejection region for the test will be located by including the most extreme values of M that at the same time provide a value of α that is suitable for the test.

We see that $P(M \leq 1 \wedge M \geq 9) = 0.0215$ and $P(M \leq 2 \wedge M \geq 8) = 0.1094$. For $\alpha = 0.05$, $RR = \{0, 1, 9, 10\}$, and for $\alpha = 0.11$, $RR = \{0, 1, 2, 8, 9, 10\}$. Compute p-value (see `nonpar/ex15_1.pdf`). We code a function for sign test in R and call it: `nonpar/ex15_1b.pdf` □

Large Samples

- **Sign Test for Large Samples:** $n > 25$

Null hypothesis: $H_0 : p = .5$ (neither treatment is preferred to the other).

Alternative hypothesis: $H_a : p = .5$ for a two-tailed test (Note: We use the two-tailed test for an example. Many analyses require a one-tailed test.)

Test statistic:

$$Z = \frac{M - n/2}{\frac{1}{2}\sqrt{n}}.$$

Rejection region: Reject H_0 if $z \geq z_{\alpha/2}$ or if $z \leq -z_{\alpha/2}$, where $z_{\alpha/2}$ is the quantile of order $\alpha/2$ for standard normal distribution.

Example

The data in the file `chickenembrios.txt` 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.

File `chickenembrios.txt` gives the average number of claps per minute during the dark period (XD) and the corresponding rate during the period of illumination (YL) for 25 chick embryos. *nonpar/ chikenEmbriosls. pdf*

Comments

- Suppose that the paired differences are normally distributed with a common variance σ^2. Will the sign test detect a shift in location of the two populations as effectively as the Student's t test?
- Intuitively: no. This is correct because the Student's t test uses more information: sign +magnitude of differences for more accurate means and variances.
- Thus, we might say that the sign test is not as "efficient" as the Student's t test; but this statement is meaningful only if the differences in paired observations are normally distributed with a common variance σ_D^2.
- The sign test might be more efficient when these assumptions are not satisfied.

5 The Wilcoxon Signed-Rank Test for a Matched-Pairs Experiment

The Wilcoxon Signed-Rank Test for a Matched-Pairs Experiment

- We have n paired observations (X_i, Y_i) and $D_i = X_i - Y_i$.
- H_0: the X's and the Y's have the same distribution versus the alternative that the distributions differ in location. Under the null hypothesis of no difference in the distributions of the X's and Y's, you would expect (on the average) half of the differences in pairs to be negative and half to be positive.
- That is, the expected number of negative differences between pairs is $n/2$ (where n is the number of pairs). Further, it would follow that positive and negative differences of equal absolute magnitude should occur with equal probability.
- If we were to order the differences according to their absolute values and rank them from smallest to largest, the expected rank sums for the negative and positive differences would be equal.
- Sizable differences in the sums of the ranks assigned to the positive and negative differences would provide evidence to indicate a shift in location for the two distributions.
- To carry out the Wilcoxon test:
  1. We calculate the differences (D_i) for each of the n pairs. Differences equal to zero are eliminated, and the number of pairs, n, is reduced accordingly.
  2. We rank the absolute values of the differences, assigning a 1 to the smallest, a 2 to the second smallest, and so on. *Ties*: If two or more absolute differences are tied for the same rank, then the average of the ranks that would have been assigned to these differences is assigned to each member of the tied group. For example, if two absolute differences are tied for ranks 3 and 4, then each receives rank 3.5, and the next highest absolute difference is assigned rank 5.
  3. Then we calculate the sum of the ranks (rank sum) for the negative differences and also calculate the rank sum for the positive differences. For a two-tailed test, we use T, the smaller of these two quantities, as a test statistic to test the null hypothesis that the two population relative frequency histograms are identical. The smaller the value of T is, the greater will be the weight of evidence favoring rejection of the null hypothesis. Hence, we will reject the null hypothesis if T is less than or equal to some value, say, T_0.

4. To detect the one-sided alternative, that the distribution of the X's is shifted to the right of that of the Y's, we use the rank sum T^- of the negative differences, and we reject the null hypothesis for small values of T^-, say, $T^- \leq T_0$. If we wish to detect a shift of the distribution of the Y's to the right of the X's,we use the rank sum T^+ of the positive differences as a test statistic, and we reject small values of T^+, say, $T^+ \leq T_0$.

- The probability that T is less than or equal to some value T_0 has been calculated for a combination of sample sizes and values of T_0. These probabilities are tabulated and can be used to find the rejection region for the test based on T.

- The R functions `dsignrank`, `psignrank`, `qsignrank`, `rsignrank` are density, distribution function, quantile function and random generation, respectively, for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size n.

Summary

Wilcoxon Signed-Rank Test for a Matched-Pairs Experiment

H_0 : The population distributions for the X's and Y's are identical.

H_a : (1) The two population distributions differ in location (two-tailed), or (2) the population relative frequency distribution for the X's is shifted to the right of that for the Y 's (one-tailed).

Test statistic 1. For a two-tailed test, use $T = \min(T^+, T^-)$, where $T^+ =$ sum of the ranks of the positive differences and $T^- =$ sum of the ranks of the negative differences.

2. For a one-tailed test (to detect the one-tailed alternative just given), use the rank sum T^- of the negative differences. To detect a shift of the distribution of the Y's to the right of the distribution of the X's, use the rank sum T^+, the sum of the ranks of the positive differences, and reject H_0 if $T^+ \leq T_0$.

Rejection region: 1. For a two-tailed test, reject H_0 if $T \leq T_0$, where T_0 is the critical value for the two-sided test

2. For a one-tailed test (as described earlier), reject H_0 if $T^- \leq T_0$, where T_0 is the critical value for the one-sided test.

Example

Example 3. Due to oven-to-oven variation, a matched-pairs experiment was used to test for differences in cakes prepared using mix A and mix B. Two cakes, one prepared using each mix, were baked in each of six different ovens (a total of 12 cakes). Test the hypothesis that there is no difference in population

A	B	Difference $A - B$	Absolute Difference	Rank of Absolute Difference
.135	.129	.006	.006	3
.102	.120	$-.018$	.018	5
.108	.112	$-.004$	.004	1.5
.141	.152	$-.011$	.011	4
.131	.135	$-.004$	.004	1.5
.144	.163	$-.019$	.019	6

Table 2: Paired data and their differences for Example 3

distributions of cake densities using the two mixes. What can be said about the attained significance level? The original data and differences in densities (in ounces per cubic inch) for the six pairs of cakes are shown in Table 2.

Solution. H_0 : the two population frequency distributions of cake densities are identical. H_a : the distributions differ in location $\Longrightarrow$ a two-tailed test is required. Because the amount of data is small, we will conduct our test by using $\alpha = .10$. The critical value of T for a two-tailed test (from table), $\alpha = .10$, is $T_0 = 2$. Hence, we will reject H_0 if $T \leq 2$. There is only one positive difference, and that difference has rank 3; therefore, $T^+ = 3$. Because $T^+ + T^- = n(n+1)/2$ (why?), $T^- = 21 - 3 = 18$ and the observed value of T is $\min(3, 18) = 3$. Notice that 3 exceeds the critical value of T, implying that there is insufficient evidence to indicate a difference in the two population frequency distributions of cake densities. Because we cannot reject H_0 for $\alpha = .10$, we can only say that p-value $> .10$. See `nonpar/ex15_3.pdf` ☐

Large Samples

- Although the tables for Wilcoxon SRT contains values for the number of data pairs as large as $n = 50$, it is worth noting that T^+ (or T^-) will be approximately normally distributed when the null hypothesis is true and n is large (say, 25 or more).
- This enables us to construct a large-sample Z test, where if $T = T^+$, see [2, pages 47–49]

$$E\left(T^+\right) = \frac{n(n+1)}{4}, \qquad V\left(T^+\right) = \frac{n(n+1)(2n+1)}{24}.$$

- Then the Z statistic

$$Z = \frac{T^+ - E\left(T^+\right)}{\sqrt{V\left(T^+\right)}} = \frac{T^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

can be used as a test statistic.

Large Samples – Summary

A Large-Sample Wilcoxon Signed-Rank Test for a Matched-Pairs Experiment: $n > 25$

Null hypothesis: H_0 : The population relative frequency distributions for the X's and Y's are identical.

Alternative hypothesis:

1. H_a : The two population relative frequency distributions differ in location (a two-tailed test),
2. *or* Ha : the population relative frequency distribution for the X's is shifted to the right (or left) of the relative frequency distribution of the Ys (one-tailed tests).

Test statistic:

$$Z = \frac{T^{+} - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

Rejection region: Reject H_0 if $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ for a two-tailed test. To detect a shift in the distributions of the X's to the right of the Y 's, reject H_0 when $z \geq z_{\alpha}$. To detect a shift in the opposite direction, reject H_0 if $z \leq -z_{\alpha}$.

6 Using Ranks for Comparing Two Population Distributions: Independent Random Samples

Using Ranks for Independent Random Samples

- Wilcoxon 1947
- We assume that we are interested in testing whether the two populations have the same distribution versus the shift (or location) alternative.
- Independent random samples of size n_1 and n_2, $n_1 + n_2 = n$.
- Combine observations, rank them in order of magnitude, from 1 (the smallest) to n (the largest). Ties -> average of ranks is assigned to each member of the tied group.
- If the observations were selected from identical populations, the rank sums for the samples should be more or less proportional to the sample sizes n_1 and n_2.
- For example, if n_1 and n_2 were equal, you would expect the rank sums to be nearly equal. In contrast, if the observations in one population—say, population I—tended to be larger than those in population II, the observations in sample I would tend to receive the highest ranks and sample I would have a larger than expected rank sum.

I	II
27	32
31	29
26	35
25	28

Table 3: Data for Example 4

	I	II
	3	7
	6	5
	2	8
	1	4
Rank sum:	12	24

Table 4: Ranks for Example 4

- Thus (sample sizes being equal), if one rank sum is very large (and, correspondingly, the other is very small), it may indicate a statistically significant difference between the locations of the two populations.

Example

Example 4. The bacteria counts per unit volume are shown in Table 3 for two types of cultures, I and II. Four observations were made for each culture. Let n_1 and n_2 represent the number of observations in samples I and II, respectively. For the data given in Table 3, the corresponding ranks are as shown in Table 4. Do these data present sufficient evidence to indicate a difference in the locations of the population distributions for cultures I and II?

Finding Rejection Region

- Let W equal the rank sum for sample I (for this sample, $W = 12$).
- The minimum rank sum includes the ranks $1, 2, 3, 4$, or $W = 10$. Similarly, the maximum includes the ranks $5, 6, 7, 8$, with $W = 26$. $RR = \{10, 26\}$
- The probability α: The total number of permutations of the eight ranks is 8! The number of different arrangements of the ranks 1, 2, 3, 4 in sample I with the 5, 6, 7, 8 of sample II is $4! \cdot 4!$. Similarly, the number of arrangements that place the maximum value of W in sample I (ranks 5, 6, 7, 8) is $4! \cdot 4!$. Then, the probability that $W = 10$ or $W = 26$ is

$$p(10) + p(26) = \frac{2 \cdot 4! \cdot 4!}{8!} = \frac{1}{35} = 0.029$$

- If this value of α is too small, the rejection region can be enlarged to include the next smallest and next largest rank sums, $W = 11$ and $W = 25$. The rank sum $W = 11$ includes the ranks 1, 2, 3, 5, and

  $$p(11) = \frac{4! \cdot 4!}{8!} = \frac{1}{70}$$

  Similarly,

  $$p(25) = \frac{1}{70}.$$

  Then,

  $$\alpha = p(10) + p(11) + p(25) + p(26) = \frac{2}{35} = 0.057.$$

- Expansion of the rejection region to include 12 and 24 substantially increases the value of α. The set of sample points giving a rank of 12 includes all sample points associated with rankings of (1, 2, 3, 6) and (1, 2, 4, 5). Thus,

  $$p(12) = \frac{2 \cdot 4! \cdot 4!}{8!} = \frac{1}{35}$$

  and

  $$\begin{aligned}\alpha &= p(10) + p(11) + p(12) + p(24) + p(25) + p(26)\\ &= \frac{1}{70} + \frac{1}{70} + \frac{1}{35} + \frac{1}{35} + \frac{1}{70} + \frac{1}{70} = \frac{4}{35} = 0.114\end{aligned}$$

  This value of α might be considered too large for practical purposes. Hence, we are better satisfied with the rejection region $W = 10$, 11, 25, and 26. The rank sum for the sample, $W = 12$, does not fall in this preferred rejection region, so we do not have sufficient evidence to reject the hypothesis that the population distributions of bacteria counts for the two cultures are identical.

7 The Mann–Whitney U Test

The Mann-Whitney U Test - Logic

- The Mann–Whitney statistic U is obtained by ordering all $n_1 + n_2$ observations according to their magnitude and counting the number of observations in sample I that precede each observation in sample II. The statistic U is the sum of these counts.

- We denote the observations in sample I as $x_1, x_2, \ldots, x_{n_1}$ and the observations in sample II as $y_1, y_2, \ldots, y_{n_2}$. For example, the eight ordered observations of Example 4 are

  $$\begin{array}{cccccccc} 25 & 26 & 27 & 28 & 29 & 31 & 32 & 35 \\ x_{(1)} & x_{(2)} & x_{(3)} & y_{(1)} & y_{(2)} & x_{(4)} & y_{(3)} & y_{(4)} \end{array}$$

- The smallest y observation is $y_{(1)} = 28$, and $u_1 = 3$ x's precede it. Similarly, $u_2 = 3$ x's precede $y_{(2)} = 29$ and $u_3 = 4$, and $u_4 = 4$ x's precede $y_{(3)} = 32$ and $y_{(4)} = 35$, respectively. Then,

$$U = u_1 + u_2 + u_3 + u_4 = 3 + 3 + 4 + 4 = 14.$$

- **Formula for the Mann–Whitney U Statistic**

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W,$$

where

n_1 = number of observations in sample I,
n_2 = number of observations in sample II,
W = rank sum for sample I.

Properties

Some useful results about the distribution of U:

1. The possible values of U are $0, 1, 2, \ldots, n_1 n_2$.
2. The distribution of U is symmetric about $(n_1 n_2)/2$. That is, for any $a > 0$, $P[U \leq (n_1 n_2)/2 - a] = P[U \geq (n_1 n_2)/2 + a]$.
3. The result in (2) implies that $P(U \leq U_0) = P(U \geq n_1 n_2 - U_0)$.

Summary

The Mann–Whitney U Test

Population I is the population from which the smaller sample was taken.

Null hypothesis: H_0 : The distributions of populations I and II are identical.

Alternative hypothesis: H_a :

(1) The distributions of populations I and II have different locations (a two-tailed test),
or (2) the distribution of population I is shifted to the right of the distribution of population II,
or (3) the distribution of population I is shifted to the left of the distribution of population II.

Test Statistic:

$$U = n_1 n_2 + [n_1(n_1+1)]/2 - W$$

Rejection region: (1) For the two-tailed test and a given value of α, reject H_0 if $U \leq U_0$ or $U \geq n_1 n_2 - U_0$, where $P(U \leq U_0) = \alpha/2$.

(2) To test that population I is shifted to the right of population II with a given value of α, reject H_0 if $U \leq U_0$, where $P(U \leq U_0) = \alpha$.

(3) To test that population I is shifted to the left of population II with a given value of α, reject H_0 if $U \geq n_1 n_2 - U_0$, where $P(U \leq U_0) = \alpha$.

Assumptions: Samples have been randomly and independently selected from their respective populations. Ties in the observations can be handled by averaging the ranks that would have been assigned to the tied observations and assigning this average rank to each. Thus, if three observations are tied and are due to receive ranks 3, 4, and 5, we assign rank 4 to all three.

Examples

Example 5. Test the hypothesis that there is no difference in the locations of the population distributions for the bacteria count data of Example 4.

Solution. From tables, for $n_1 = n_2 = 4$, we find $P(U \leq 1) = .0286$. The appropriate rejection region for the two-tailed test is $U \leq 1$ or $U \geq n_1 n_2 - 1 = 16 - 1 = 15$, for which $\alpha = 2(.0286) = .0572$ or, rounding to three decimal places, $\alpha = .057$ (the same value of α obtained for Example 4).

For the bacteria data, the rank sum is $W = 12$. Then, $U = n_1 n_2 + n_1(n_1 + 1)/2 - W = (4)(4) + 4(4+1)/2 - 12 = 14$. The calculated value of U does not fall in the rejection region.

Hence, there is not sufficient evidence to show a difference in the locations of the population distributions of bacteria counts for cultures I and II. The p-value is given by $2P(U \geq 14) = 2P(U \leq 2) = 2(.0571) = .1142$. See `nonpar/ex15_4_5.pdf` □

Examples

Example 6. An experiment was conducted to compare the strengths of two types of kraft papers, one a standard kraft paper of a specified weight and the other the same standard kraft paper treated with a chemical substance. Ten pieces of each type of paper, randomly selected from production, produced the strength measurements shown in Table 5. Test the hypothesis of no difference in the distributions of strengths for the two types of paper against the alternative hypothesis that the treated paper tends to be stronger.

Solution. We may choose any population. In Table 5, the ranks are shown in parentheses, and the rank sum W is given below the first column. Because we wish to detect a shift in the distribution of population I(standard) to the left of the distribution of the population II (treated), we will reject the null hypothesis of no difference in population strength distributions when W is excessively small. Because this situation occurs when U is large, we will conduct a one-tailed statistical test and reject the null hypothesis when $U \geq n_1 n_2 - U_0$. □

	Standard, I	Treated, II
	1.21(2)	1.49(15)
	1.43(12)	1.37(7.5)
	1.35(6)	1.67(20)
	1.51(17)	1.50(16)
	1.39(9)	1.31(5)
	1.17(1)	1.29(3.5)
	1.48(14)	1.52(18)
	1.42(11)	1.37(7.5)
	1.29(3.5)	1.44(13)
	1.40(10)	1.53(19)
Rank Sum	$W = 85.5$	

Table 5: Data for Example refex15.6

Solution, continued. From tables, for $\alpha \approx .0526$, and $n_2 = 10$, we find $U_0 = 28$. Hence, we will reject if $U \geq (10)(10) - 28 = 72$. Calculate U

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W = 10 \cdot 10 + \frac{10 \cdot 11}{2} - 85.5 = 69.5$$

As you can see, U is not greater than 72. Therefore, we cannot reject the null hypothesis.

At the $\alpha = .0526$ level of significance, there is not sufficient evidence to indicate that the treated kraft paper is stronger than the standard. The p-value is given by $P(U \geq 69.5) = P(U \leq 30.5) = .0716$. See *nonpar/ex15_6.pdf* □

Large Samples

- A simplified large-sample test ($n_1 \geq 10$ and $n_2 \geq 10$) can be obtained by using the familiar Z statistic.
- When the population distributions are identical, it can be shown that the U statistic has the following expected value and variance:

$$E(U) = \frac{n_1 n_2}{2}, \qquad V(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}.$$

- Also, when n_1 and n_2 are large,

$$Z = \frac{U - E(U)}{\sigma_U}$$

  has approximately a standard normal distribution.

Comments

- The Mann–Whitney U test and the equivalent Wilcoxon rank-sum test are not very efficient because they do not appear to use all the information in the sample. Actually, theoretical studies have shown that this is not the case.
- Under the normality assumption, the two-sample t test and the Mann–Whitney U test test the same hypotheses: $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 > 0$.
- For a given α and β, the total sample size required for the t test is approximately .95 times the total sample size required for the Mann–Whitney U.
- Thus, the nonparametric procedure is almost as good as the t test for the situation in which the t test is optimal.
- For many nonnormal distributions, the nonparametric procedure requires fewer observations than a corresponding parametric procedure would require to produce the same values of α and β.

8 The Kruskal–Wallis Test for the One-Way Layout

The Kruskal–Wallis Test

- We assume that independent random samples have been drawn from k populations that differ only in location, and we let n_i, for $i = 1, 2, \ldots, k$, represent the size of the sample drawn from the ith population.
- Combine all the $n_1 + n_2 + \cdots + n_k = n$ observations and rank them from 1 (the smallest) to n (the largest).
- **Ties**: if two or more observations are tied for the same rank, then the average of the ranks that would have been assigned to these observations is assigned to each member of the tied group.
- Let R_i denote the sum of the ranks of the observations from population i and let $\overline{R}_i = R_i / n_i$ denote the corresponding average of the ranks. If $\overline{R}$ equals the overall average of all of the ranks, *consider the rank analogue of* SST, which is computed by using the ranks rather than the actual values of the measurements:

$$V = \sum_{i=1}^{k} n_i \left(\overline{R}_i - \overline{R}\right)^2 .$$

- If the null hypothesis is true and the populations do not differ in location, we would expect the $\overline{R}_i$ values to be approximately equal and the resulting value of V to be relatively small. If the alternative hypothesis is true, we would expect this to be exhibited in differences among the values of the $\overline{R}_i$ values, leading to a large value for V. Notice that $\overline{R} = (\sum_{k=1}^{n} k)/n = [n(n+1)/2]/n = (n+1)/2$ and thus that

$$V = \sum_{i=1}^{k} n_i \left(\overline{R}_i - \frac{n+1}{2} \right)^2 . \qquad (1)$$

- Instead of focusing on V, Kruskal and Wallis (1952) [3] considered the statistic

$$H = \frac{12V}{n(n+1)}, \qquad (2)$$

  which may be rewritten (homework)

$$H = \frac{12}{n\,(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3\,(n+1)\,. \qquad (3)$$

- As previously noted, the null hypothesis of equal locations is rejected in favor of the alternative that the populations differ in location if the value of H is large. Thus, the corresponding α-level test calls for rejection of the null hypothesis in favor of the alternative if $H > h(\alpha)$, where $h(\alpha)$ is such that, when H_0 is true, $P[H > h(\alpha)] = \alpha$.

- If the underlying distributions are continuous and if there are no ties among the n observations, the null distribution of H can (tediously) be found by using the methods of Probability Theory. We can find the distribution of H for any values of k and $n_1, n_2, \ldots, n_k$ by calculating the value of H for each of the $n!$ equally likely permutations of the ranks of the n observations. These calculations have been performed and tables developed for some relatively small values of k and for $n_1, n_2, \ldots, n_k$ [2, Table A.12].

Summary

Kruskal–Wallis Test Based on H for Comparing k Population Distributions

Null hypothesis: H_0: The k population distributions are identical.

Alternative hypothesis: H_a: At least two of the population distributions differ in location.

Test statistic:

$$H = \frac{12}{n\,(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3\,(n+1)\,.$$

Line 1		Line 2		Line 3	
Def	R	Def	R	Def	R
6	5	34	25	13	9.5
38	27	28	19	35	26
3	2	42	30	19	15
17	13	13	9.5	4	3
11	8	40	29	29	20
30	21	31	22	0	1
15	11	9	7	7	6
16	12	32	23	33	24
25	17	39	28	18	14
5	4	27	18	24	16
	$R_1 = 120$		$R_2 = 210.5$		$R_3 = 134.5$

Table 6: Data for Example 7

where
$n_i =$ number of measurements in the sample from population i,
$R_i =$ rank sum for sample i, where the rank of each measurement is computed according to its relative size in the overall set of $n = n_1 + n_2 + \cdots + n_k$ observations formed by combining the data from all k samples.

Rejection region: Reject H_0 if $H > \chi^2_{\alpha,k-1}$ (χ^2-quantile of order α with $k-1$ df).

Assumptions: The k samples are randomly and independently drawn. There are five or more measurements in each sample.

Example

Example 7. A quality control engineer has selected independent samples from the output of three assembly lines in an electronics plant. For each line, the output of ten randomly selected hours of production was examined for defects. Do the data in Table 6 provide evidence that the probability distributions of the number of defects per hour of output differ in location for at least two of the lines? Use $\alpha = .05$. Also give the p-value associated with the test.

Solution. In this case, $n_1 = n_2 = n_3 = 10$ and $n = 30$. Thus,

$$H = \frac{12}{30 \cdot 31}\left(\frac{120^2}{10} + \frac{210.5^2}{10} + \frac{134.5^2}{10}\right) - 3 \cdot 31 = 6.0974$$

Because all the $n_i \geq 5$, we may use the approximation for the null distribution of H and reject the null hypothesis of equal locations if $H > \chi^2_{0.05,2} = 5.99147$. Thus, we reject the null hypothesis at the $\alpha = .05$ level and conclude that at least one of the three lines tends to produce a greater number of defects than the others. See `nonpar/ex15_ 7. pdf`. □

Remarks

- It can be shown that, if we wish to compare only $k = 2$ populations, the Kruskal–Wallis test is equivalent to the Wilcoxon rank-sum two-sided test. If data are obtained from a one-way layout involving $k > 2$ populations but we wish to compare a particular pair of populations, the Wilcoxon rank-sum test (or the equivalent Mann–Whitney U test) can be used for this purpose.
- Notice that the analysis based on the Kruskal–Wallis H statistic does not require knowledge of the actual values of the observations. We need only know the ranks of the observations to complete the analysis.

9 The Friedman Test for Randomized Block Designs

The Friedman Test for Randomized Block Designs

- Milton Friedman - winner of Nobel Prize for Economy, 1937
- After the data from a randomized block design are obtained, within each block the observed values of the responses to each of the k treatments are ranked from 1 (the smallest in the block) to k (the largest in the block).
- If two or more observations in the same block are tied for the same rank, then the average of the ranks that would have been assigned to these observations is assigned to each member of the tied group. However, ties need to be dealt with in this manner only if they occur within the same block.
- R_i denote the sum of the ranks of the observations corresponding to treatment i and let $\overline{R}_i = R_i/b$ denote the corresponding average of the ranks (recall that in a randomized block design, each treatment is applied exactly once in each block, resulting in a total of b observations per treatment and hence in a total of bk total observations). Because ranks of 1 to k are assigned within each block, the sum of the ranks assigned in each block is $1 + 2 + \cdots + k = k(k+1)/2$. Thus, the sum of all the ranks assigned in the analysis is $bk(k+1)/2$. If $\overline{R}$ denotes the overall average of the ranks of all the bk observations, it follows that $\overline{R} = (k+1)/2$. Consider the *rank analog of SST for a randomized block design* given by

$$W = b\sum_{i=1}^{k}\left(\overline{R}_i - \overline{R}\right)^2 .$$

- If the null hypothesis is true and the probability distributions of the treatment responses do not differ in location, we expect the $\overline{R}_i$ -values to be approximately equal and the resulting value for W to be small. If the alternative hypothesis were true, we would expect this to lead to differences among the $\overline{R}_i$-values and corresponding large values of W. Instead of W, Friedman considered the statistic $F_r = 12W/[k(k+1)]$, which may be rewritten as

$$F_r = \frac{12}{bk(k+1)} \sum_{i=1}^{k} R_i^2 - 3b(k+1).$$

- The null hypothesis of equal locations is rejected in favor of the alternative that the treatment distributions differ in location if the value of F_r is large. That is, the corresponding α-level test rejects the null hypothesis in favor of the alternative if $F_r > f_r(\alpha)$, where $f_r(\alpha)$ is such that, when H_0 is true, $P[F_r > f_r(\alpha)] = \alpha$.
- If there are no ties among the observations within the blocks, the null distribution of F_r can (tediously) be found by using the methods of Probability Theory. For any values of b and k, the distribution of F_r is found as follows. If the null hypothesis is true, then each of the $k!$ permutations of the ranks 1, 2, . . . , k within each block is equally likely. Further, because we assume that the observations in different blocks are mutually independent, it follows that each of the $(k!)b$ possible combinations of the b sets of permutations for the within-block ranks are equally likely when H_0 is true. Consequently, we can evaluate the value of F_r for each possible case and thereby give the null distribution of F_r. Selected values for $f_r(\alpha)$ for various choices of k and b are given in Hollander and Wolfe [2, Table A.22].
- For $k = 2$, the Friedman analysis is equivalent to a two-tailed sign test.

Summary

Friedman Test Based on F_r for a Randomized Block Design

Null hypothesis: H_0: The probability distributions for the k treatments are identical.

Alternative hypothesis: H_a: At least two of the distributions differ in location.

Test statistic

$$F_r = \frac{12}{bk(k+1)} \sum_{i=1}^{k} R_i^2 - 3b(k+1),$$

where
b = number of blocks,
k = number of treatments,
R_i = sum of the ranks for the ith treatment, where the rank of each measurement is computed relative to its size within its own block.

Rejection region: $F_r > \chi^2_{\alpha,k-1}$.

Techn	Task A	Rank	Task B	Rank	Task C	Rank
1	1.21	1	1.56	3	1.48	2
2	1.63	1.5	2.01	3	1.63	1.5
3	1.42	1	1.70	2	2.06	3
4	1.16	1	1.27	2.5	1.27	2.5
5	2.43	2	2.64	3	1.98	1
6	1.94	1	2.81	3	2.44	2
Rank sum		7.5		16.5		12

Table 7: Completion times for three tasks

Assumptions: The treatments are randomly assigned to experimental units within blocks. Either the number of blocks (b) or the number of treatments (k) exceeds 5.

Example

Example 8. An experiment to compare completion times for three technical tasks was performed in the following manner. Because completion times may vary considerably from person to person, each of the six technicians was asked to perform all three tasks. The tasks were presented to each technician in a random order with suitable time lags between the tasks. Do the data in Table 7 present sufficient evidence to indicate that the distributions of completion times for the three tasks differ in location? Use $\alpha = .05$. Give bounds for the associated p-value.

Solution. RB design with $k = 3$ treatments and $b = 6$ blocks; technicians $\longrightarrow$ blocks. Because $b > 5$, we may use the Friedman test and compare the value of F_r to $\chi^2_{\alpha,k-1}$, based on $k - 1 = 2$ df. The quantile is $\chi^2_{0.05,2} = 5.99147$. For the data given in Table 7,

$$F_r = \frac{12}{6 \cdot 3 \cdot 4}\left(7.5^2 + 16.5^2 + 12^2\right) - 3 \cdot 6 \cdot 4 = 6.75.$$

Because $F_r = 6.75 > 5.99147$, we conclude at the $\alpha = .05$ level that the completion times of at least two of the three tasks possess probability distributions that differ in location. The p-value is $P(\chi^2 > 6.75) = 0.03421812$. See *nonpar/ex15_8.pdf* or *nonpar/ex15_8b.pdf* □

10 The Runs Test: A Test for Randomness

The Runs Test: A Test for Randomness

- Wald and Wolfowitz, 1940
- Why do we test randomness?

- The runs test is used to study a sequence of events with one of two outcomes, success (S) or failure (F). If we think of the sequence of items emerging from a manufacturing process as defective (F) or nondefective (S), the observation of twenty items might yield

  S S S S S F F S S S
  F F F S S S S S S S

- We notice the groupings of defectives and nondefectives and ask whether this grouping implies nonrandomness and, consequently, lack of process control.

Definition 9. A *run* is a maximal subsequence of like elements.

- The 20 elements are arranged in five runs, the first containing five S's, the second containing two F's, and so on.
- A very small or very large number of runs in a sequence indicates nonrandomness.
- Therefore, let R (the number of runs in a sequence) be the test statistic and let the rejection region be $R \le k_1$ and $R \ge k_2$, as indicated in Figure 3.
- We must then find the probability distribution for R, $P(R = r)$, to calculate α and to locate a suitable rejection region for the test.

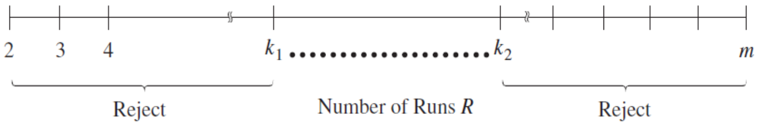

Figure 3: The rejection region for the runs test

Finding the Probabilities

- Input: sequence contains n_1 S elements and n_2 F elements, Y_1 runs of S's, Y_2 runs of F's, where $(Y_1 + Y_2) = R$.
- For a given Y_1, Y_2 can equal Y_1, $(Y_1 - 1)$, or $(Y_1 + 1)$.
- Let m denote the maximum possible number of runs. Notice that $m = 2n_1$ if $n_1 = n_2$, and that $m = (2n_1 + 1)$ if $n_1 < n_2$.

- We will suppose that every distinguishable arrangement of the $(n_1 + n_2)$ elements in the sequence constitutes a simple event for the experiment and that the sample points are equiprobable. It then remains for us to count the number of sample points that imply R runs.

- The total number of distinguishable arrangements of n_1 S elements and n_2 F elements is

  $$\binom{n_1 + n_2}{n_1},$$

  and therefore the probability per sample point is

  $$\frac{1}{\binom{n_1 + n_2}{n_1}}.$$

- The number of ways of achieving y_1 S runs is equal to the number of identifiable arrangements of n_1 indistinguishable elements in y_1 cells, none of which is empty, as represented in Figure 4. This is equal to the number of ways of distributing the $(y_1 - 1)$ inner bars in the $(n_1 - 1)$ spaces between the S elements (the outer two bars remain fixed). Consequently, it is equal to the number of ways of selecting $(y_1 - 1)$ spaces (for the bars) out of the $(n_1 - 1)$ spaces available, or

  $$\binom{n_1 - 1}{y_1 - 1}.$$

- The number of ways of observing y_1 S runs and y_2 F runs, obtained by applying the *mn* rule, is

  $$\binom{n_1 - 1}{y_1 - 1}\binom{n_2 - 1}{y_2 - 1}.$$

- This gives the number of sample points in the event "y_1 runs of S's and y_2 runs of F's." The probability of exactly y_1 runs of S's and y_2 runs of F's:

  $$p\left(y_1, y_2\right) = \frac{\binom{n_1 - 1}{y_1 - 1}\binom{n_2 - 1}{y_2 - 1}}{\binom{n_1 + n_2}{n_1}}$$

  Then, $P(R = r)$ equals the sum of $p(y_1, y_2)$ over all values of y_1 and y_2 such that $(y_1 + y_2) = r$.

- Examples: $R = 4$ could occur when $y_1 = 2$ and $y_2 = 2$ with either the S or F elements commencing the sequences. Consequently,

  $$P(R = 4) = 2P(Y_1 = 2, Y_2 = 2).$$

$$\left|S\right|SSSS\left|SS\ldots\right|SS\left|SSS\right|S\left|\right.$$

Figure 4: The distribution of n_1 S elements in y_1

On the other hand, $R = 5$ could occur when $y_1 = 2$ and $y_2 = 3$ or when $y_1 = 3$ and $y_2 = 2$, and these occurrences are mutually exclusive. Then,

$$P(R = 5) = P(Y_1 = 3, Y_2 = 2) + P(Y_1 = 2, Y_2 = 3).$$

Normal Approximations

- The probability distribution for R tends toward normality as n_1 and n_2 become large ($n_1 > 10$, $n_2 > 10$).
- Consequently, we may use the Z statistic, where

  $$\begin{aligned} Z &= \frac{R - E(R)}{\sqrt{V(R)}} \\ E(R) &= \frac{2n_1n_2}{n_1 + n_2} + 1 \\ V(R) &= \frac{2n_1n_2\left(2n_1n_2 - n_1 - n_2\right)}{\left(n_1 + n_2\right)^2\left(n_1 + n_2 - 1\right)} \end{aligned}$$

  are the expected value and variance of R, respectively.
- If α is the desired probability of a type I error, then the rejection region for a two-tailed test, is $|z| \geq z_{1-\alpha/2}$. For an upper-tail test, we reject the null hypothesis if $z > z_{1-\alpha}$ (for a lower-tail test, we reject H0 if $z < z_\alpha$).

Examples

Example 10. A true–false examination was constructed with the answers running in the following sequence:

T F F T F T F T T F T F F T F T F T T F.

Does this sequence indicate a departure from randomness in the arrangement of T and F answers?

Solution. The sequence contains $n_1 = 10$ T and $n_2 = 10$ F answers, with $y = 16$ runs. Nonrandomness can be indicated by either an unusually small or an unusually large number of runs; consequently, we will be using a two-tailed test.

Suppose that we wish to use α approximately equal to .05 with .025 or less in each tail of the rejection region. Then, from Table, with $n_1 = n_2 = 10$, we see that $P(R \leq 6) = .019$ and $P(R \leq 15) = .981$. Then, $P(R \geq 16) = 1 - P(R \leq 15) = .019$, and we would reject the hypothesis of randomness at the $\alpha = .038$ significance level if $R \leq 6$ or $R \geq 16$. Because $R = 16$ for the observed data, we conclude that evidence exists to indicate nonrandomness in the professor's arrangement of answers. The attempt to mix the answers was overdone. See *nonpar/ex15_10.pdf* □

Examples–cont.

- A second application – time series
- Departures from randomness in a series, caused either by trends or periodicities, can be detected by examining the deviations of the time series measurements from their average.
- Negative and positive deviations could be denoted by S and F, respectively, and we could then test this time sequence of deviations for nonrandomness.

Example 11. Paper is produced in a continuous process. Suppose that a brightness measurement Y is made on the paper once every hour and that the results appear as shown in Figure 5. The average y for the 15 sample measurements appears as shown. Notice the deviations about y. Do these data indicate a lack of randomness and thereby suggest periodicity and lack of control in the process?

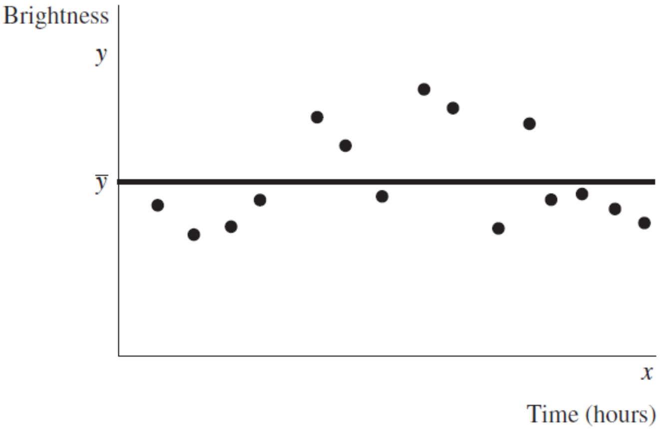

Figure 5: Paper brightness versus time

Solution. The sequence of negative (S) and positive (F) deviations as indicated in Figure 5 is

S S S S F F S F F S F S S S

Then, $n_1 = 10$, $n_2 = 5$, and $R = 7$. Consulting Table, we find $P(R \leq 7) = .455$. This value of R is not improbable, assuming the hypothesis of randomness to be true. Consequently, there is not sufficient evidence to indicate nonrandomness in the sequence of brightness measurements. See `nonpar/ex15_11.pdf` ☐

11 Rank Correlation Coefficient

Rank Correlation Coefficient

- Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a continuous bivariate population with joint distribution function $F_{X,Y}$ and marginal distribution functions F_X and F_Y. That is, the (X, Y) pairs are mutually independent and identically distributed according to some continuous bivariate population.

- The null hypothesis: X and Y are independent:

$$H_0 : [F_{X,Y}(x, y) \equiv F_X(x)F_Y(y), \text{ for all } (x, y) \text{ pairs}]. \tag{4}$$

- To compute the Spearman rank correlation coefficient r_s, we first order the n X observations from least to greatest and let R_i denote the rank of X_i, $i = 1, \ldots, n$, in this ordering. Similarly, we separately order the n Y observations from least to greatest and let S_i denote the rank of Y_i, $i = 1, \ldots, n$, in this ordering. The Spearman (1904) rank correlation coefficient is defined as the Pearson product moment sample correlation of the R_i and the S_i.

- Recall that the sample correlation coefficient for observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ is given by

$$r_P = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\left[\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2\right]^{1/2}}.$$

- When no ties within a sample are present, this is equivalent to two computationally efficient formulae:

$$r_S = \frac{12\sum_{i=1}^{n}\left(R_i - \frac{n+1}{2}\right)\left(S_i - \frac{n+1}{2}\right)}{n\left(n^2 - 1\right)} \tag{5}$$

$$= 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n\left(n^2 - 1\right)}, \tag{6}$$

where $D_i = R_i - S_i$, $i = 1, \ldots, n$.

- Rejection region

(a) Upper-tailed test: $H_a : X$ and Y are positively associated. Reject H_0 if $r_S \geq r_{S,\alpha}$, where the constant $r_{S,\alpha}$ is chosen to make the type I error probability equal to α. Values of $r_{S,\alpha}$ are found with the command `qSpearman`.

(b) Lower-tailed test: $H_a : X$ and Y are negatively associated. Reject H_0 if $r_S \leq -r_{S,\alpha}$

(c) Two-tailed test: $H_a : X$ and Y are not associated (not independent). Reject H_0 if $|r_S| \geq r_{S,\alpha/2}$.

The critical values are tabulated.

Normal Approximation

- The large-sample approximation is based on the asymptotic normality of r_S, suitably standardized. For this standardization, we need to know the expected value and variance of r_S when the null hypothesis of independence is true. Under H_0, the expected value and variance of r_S are

$$E(r_S) = 0 \tag{7}$$

$$V(r_S) = \frac{1}{n-1}. \tag{8}$$

- The standardized version of r_S is

$$r_S^* = \frac{r_S - E(r_S)}{\sqrt{V(r_S)}} = \sqrt{n-1}r_S. \tag{9}$$

- When H_0 is true, r_S^* has, as n tends to infinity, an asymptotic $N(0,1)$ distribution.

Ties

- If there are ties among the n X observations and/or separately among the n Y observations, assign each of the observations in a tied (either X or Y) group the average of the integer ranks that are associated with the tied group.

- If there are tied X's and/or tied Y's, Spearman's rank correlation coefficient calculated with Pearson's correlation does not require modification.

- If using the computationally efficient version of r_S at (6), some changes to the statistic are necessary. The statistic r_S in this case becomes

$$r_S = \frac{n(n^2-1) - 6\sum_{s=1}^{n} D_s^2 - \frac{1}{2}\left[T_1 + T_2\right]}{\left\{\left[n(n^2-1) - T_1\right]\left[n(n^2-1)\right] - T_2\right\}^{1/2}} \tag{10}$$

Teacher	Judge's Rank	Examination Score
1	7	44(1)
2	4	72(5)
3	2	69(3)
4	6	70(4)
5	1	93(8)
6	3	82(7)
7	8	67(2)
8	5	80(6)

Table 8: Data for science teachers

$$T_1 = \sum_{i=1}^{g} t_i(t_i^2 - 1), \quad T_2 = \sum_{j=1}^{h} u_j(u_j^2 - 1) \tag{11}$$

where in (10) and (11) g denotes the number of tied X groups, t_i is the size of tied X group i, h is the number of tied Y groups, and u_j is the size of tied Y group j.

- We note that an untied $X(Y)$ observation is considered to be a tied $X(Y)$ group of size 1. In particular, if neither the collection of X nor the collection of Y observations contains tied values, we have $g = h = n$, $t_j = u_j = 1$, $i = 1, \dots, n$, and $j = 1, \dots, n$. In this case of no tied X's and no tied Y's, each term involving either $(t_i^2 - 1)$ or $(u_j^2 - 1)$ reduces to zero and the "ties" expression for r_S in (10) reduces to the "no-ties" form for r_S, as given in (6).
- As a consequence of this effect that ties have on the null distribution of r_S, in order to use the large-sample approximation when there are tied X observations and/or tied Y observations, we first compute r_S^* (9) using average ranks and the ties-corrected version of r_S (10).

Example

Example 12. Suppose that eight elementary-science teachers have been ranked by a judge according to their teaching ability, and all have taken a national teachers' examination. The data are given in Table 8. Do the data suggest agreement between the judge's ranking and the examination score? Alternatively, we might express this question by asking whether a correlation exists between the judge's ranking and the ranks of examination scores. Calculate r_S for the judge's ranking and examination score data from Table 8. Test the hypothesis of no association between populations.

Solution. The computed value for r_S is $r_S = -0.7142857$. The critical value is $r_{S,0.025} = -0.6904762$. We reject the null hypothesis for $\alpha = 0.05$, left-tailed, and accept it for two-tailed. See `nonpar/ex15_12.pdf` ☐

References

[1] J. D. Gibbons, and S. Chakraborti. 2003. *Nonparametric Statistical Inference*, 4th. ed. New York, Basel: Marcel Dekker

[2] Hollander, M., and D. A. Wolfe and E. Chicken. 2014. *Nonparametric Statistical Methods*, 3rd ed. New York: Wiley.

[3] Kruskal, W. H., and W. A. Wallis. 1952. "Use of Ranks in One-Criterion Variance Analysis," *Journal of the American Statistical Association* **47**: 583–621.