

4 Paired Comparisons and Blocked Designs

4.1 Paired-Comparison Permutation Test

Previously we considered the *unpaired* t -test and the *unpaired* permutation test.

Recall (from Math 220) the paired t -test for testing the difference between two means μ_1 and μ_2 .

Suppose we want to test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ at level $\alpha = 0.05$.

For paired data (X_i, Y_i) , $i = 1, \dots, n$, define the difference $D_i = X_i - Y_i$.

Then, perform the one-sample t -test using D_i .

$$T = \frac{\bar{D} - 0}{s_D/\sqrt{n}}$$

What are the assumptions?

How is the p -value determined?

4.1.1 Steps for a Paired-Comparison Permutation Test

The *permutation* test statistic is based on \bar{D} .

The **null** hypothesis is that the population of D_i is **symmetric** about zero.

For an *upper*-tailed test, the **alternative** hypothesis is that the differences tend to fall more to the *positive* side of zero.

Special case (a more *cleanly* stated alternative hypothesis): For the alternative hypothesis, the distribution of D_i is symmetric about $\Delta \neq 0$.

In general, what are the assumptions needed for this *special case*?

Example: Sample just one (X, Y) data pair, and define the difference $D = X - Y$.

Under H_0 , is D more likely to be in the interval $(10, 20)$ or $(-20, -10)$?

□

Problem 4.1.1 (lung capacity; hypothetical data set),

problem4.1.1.txt: Total lung capacity, in milliliters, of four patients is measured *before* and *after* surgical correction of tuberculosis kyphosis of the spine. We wish to test at level $\alpha = 0.05$ whether or not the *median* change in lung capacity differs from zero.

Patient number	1	2	3	4
Preoperative (X)	1870	1980	1520	3155
Postoperative (Y)	2220	2080	2160	3040

(a) State the null and alternative hypotheses.

(b) Obtain the *permutation distribution* of \bar{D} .

```
> z = read.table2( "problem4.1.1.txt", header=TRUE )
> x = as.numeric( z[ 1, 2:5 ] )
> y = as.numeric( z[ 2, 2:5 ] )
> d = x - y
```

Permutation	Patient 1	Patient 2	Patient 3	Patient 4	\bar{D}
1	350	100	640	115	301.25
2	350	100	640	-115	243.75
3	350	100	-640	115	-18.75
4	350	100	-640	-115	-76.25
5	350	-100	640	115	251.25
6	350	-100	640	-115	193.75
7	350	-100	-640	115	-68.75
8	350	-100	-640	-115	-126.25
9	-350	100	640	115	126.25
10	-350	100	640	-115	68.75
11	-350	100	-640	115	-193.75
12	-350	100	-640	-115	-251.25
13	-350	-100	640	115	76.25
14	-350	-100	640	-115	18.75
15*	-350	-100	-640	115	-243.75
16	-350	-100	-640	-115	-301.25

\bar{D}	probability
-301.25	1/16
-251.25	1/16
-243.75*	1/16
-193.75	1/16
-126.25	1/16
-76.25	1/16
-68.75	1/16
-18.75	1/16
18.75	1/16
68.75	1/16
76.25	1/16
126.25	1/16
193.75	1/16
243.75	1/16
251.25	1/16
301.25	1/16
sum	1

(c) Perform the *paired-comparison permutation test* using hand-calculations.

(d) Perform the *paired-comparison permutation test* using the function `perm.test`.

(e) What assumptions were needed to perform this *permutation* test?

(f) Perform the *paired t*-test.

(g) What assumptions were needed to perform this *paired t*-test?

□

4.1.2 Randomly Selected Permutations

In general, how many permutations exist for a paired-comparison permutation test?

How many groupings exist for a paired-comparison permutation test for a sample of size 50?

If the number of permutations required for an *exact p*-value is too computationally intensive, how should the procedure be modified?

Problem 4.1.2 (sutured vs. taped wounds), [problem4.1.2.txt](#): In

a study of the comparative tensile strength of tape-closed and sutured wounds, the following results were obtained on 10 rats, 40 days after the incisions made on their backs had been closed by suture or by surgical tape (Ury and Forrester, 1970, *American Statistician*). Suppose it is desired to test the hypothesis of no difference versus the alternative that the tape-closed wounds are stronger, at level $\alpha = 0.05$.

	X	Y
Rat	Tape	Suture
1	659	452
2	984	587
3	397	460
4	574	787
5	447	351
6	479	277
7	676	234
8	761	516
9	647	577
10	577	513

(a) State the null and alternative hypotheses.

(b) Determine an approximate p -value based on the *paired-comparison permutation*

test using 20,000 simulated permutations.

```
> z = read.table2( "problem4.1.2.txt", header=TRUE )
```

```
> x = z[ , 2 ]
```

```
> y = z[ , 3 ]
```

(c) Determine the *exact* p -value of the *paired-comparison permutation test*.

(d) What assumptions were needed to perform this *permutation test*?

(e) Perform the *paired t-test*.

(f) What assumptions were needed to perform this *paired t-test*?

□

4.1.3 Large-Sample Approximations

The test statistic used in the *paired-comparison permutation test* is \bar{D} .

When n , the number of paired observations, is large, what is the approximate distribution of \bar{D} under H_0 for finite σ_D ?

How is \bar{D} often standardized (for *parametric* statistics)?

What is the approximate distribution of the *standardized* value of \bar{D} under H_0 for large n and finite σ_D ?

4.1.4 A Test for the Median of a Symmetric Population

In this subsection, again assume that the population is **symmetric** with median Δ , and we perform hypothesis tests on Δ , based on the data $X_i, i = 1, \dots, n$.

Problem 4.1.3: Based on the simple random sample $\{32, 45, 4, 76\}$ from a **symmetric** population, test if the population median Δ is 20 versus the alternative that the median is *greater* than 20 at level $\alpha = 0.05$.

(a) Read in the data.

```
> x = c( 32, 45, 4, 76 )
```

(b) State the null and alternative hypotheses.

(c) Obtain the *permutation distribution* of the appropriate nonparametric statistic based on means.

Permutation	Obs. 1	Obs. 2	Obs. 3	Obs. 4	\bar{D}
1	12	25	16	56	27.25
2	12	25	16	-56	-0.75
3*	12	25	-16	56	19.25
4	12	25	-16	-56	-8.75
5	12	-25	16	56	14.75
6	12	-25	16	-56	-13.25
7	12	-25	-16	56	6.75
8	12	-25	-16	-56	-21.25
9	-12	25	16	56	21.25
10	-12	25	16	-56	-6.75
11	-12	25	-16	56	13.25
12	-12	25	-16	-56	-14.75
13	-12	-25	16	56	8.75
14	-12	-25	16	-56	-19.25
15	-12	-25	-16	56	0.75
16	-12	-25	-16	-56	-27.25

\bar{D}	probability
-27.25	1/16
-21.25	1/16
-19.25	1/16
-14.75	1/16
-13.25	1/16
-8.75	1/16
-6.75	1/16
-0.75	1/16
0.75	1/16
6.75	1/16
8.75	1/16
13.25	1/16
14.75	1/16
19.25*	1/16
21.25	1/16
27.25	1/16
sum	1

(d) Perform the *permutation test* based on the sample mean difference, using hand-calculations.

(e) Perform the *permutation test* using the function `perm.test`.

- (f) What assumptions were needed to perform this *permutation* test?
- (g) Perform the one-sample *t*-test.
- (h) What assumptions were needed to perform this *t*-test?
- (i) Are the *permutation test for the median of a symmetric population* and the one-sample *t*-test heavily influenced by outliers?

□

Homework p. 141: Exercise 4.1*

Hints for homework exercise 4.1*: You do **NOT** need to state H_0 , H_a , a *p*-value, or conclusion. Show the *R*-code and output for **all** of your calculations; e.g., the mean differences. Clearly, the permutation distribution must be symmetric. Introduce the question number as a comment using “#” or in red using .html code; e.g., ` Exercise 4.1 `.

4.2 Signed-Rank Test

4.2.1 The Wilcoxon Signed-Rank Test without Ties in the Data

Paired (two-)sample case:

Assumptions: The *differences* of the paired observations are independent and are from a *symmetric* population.

Also, assume that the differences are all nonzero and contain no ties.

Computing the *Wilcoxon signed-rank* test statistic:

- ⊙ Determine the **differences** of the paired observations, and **rank** the *absolute values* of these **differences**.
- ⊙ Produce the *signed ranks* by attaching a *negative sign* to the *rank* if the difference is *negative*.

- ⊙ The *observed* test statistic is the **sum** of the *positive* ranks.
- ⊙ Obtain a *permuted* test statistic based on the sum of the *positive* ranks, by independently multiplying each *rank* by +1 or -1, each with probability 0.5. Compute this *permuted* test statistic for all possible groupings or for a large number of simulated groupings.
- ⊙ The *p*-value is based on the proportion of the *permuted* test statistics at least as extreme as the *observed* test statistic.

Revisit problem 4.1.1 (lung capacity; hypothetical data set),

problem4.1.1.txt: Total lung capacity, in milliliters, of four patients is measured *before* and *after* surgical correction of tuberculosis kyphosis of the spine. We wish to test at level $\alpha = 0.05$ whether or not the *median* change in lung capacity differs from zero.

Patient number	1	2	3	4
Preoperative (X)	1870	1980	1520	3155
Postoperative (Y)	2220	2080	2160	3040

(a) State the null and alternative hypotheses.

(b) Determine the signed ranks of the differences between X and Y .

```
> z = read.table2( "problem4.1.1.txt", header=TRUE )
> x = as.numeric( z[ 1, 2:5 ] )
> y = as.numeric( z[ 2, 2:5 ] )
> d = x - y
```

(c) Obtain the *permutation distribution* of the *sum* of the *positive* ranks.

Permutation	Obs. 1	Obs. 2	Obs. 3	Obs. 4	SR ₊
1	1	2	3	4	10
2	1	2	3	-4	6
3	1	2	-3	4	7
4	1	2	-3	-4	3
5	1	-2	3	4	8
6	1	-2	3	-4	4
7	1	-2	-3	4	5
8	1	-2	-3	-4	1
9	-1	2	3	4	9
10	-1	2	3	-4	5
11	-1	2	-3	4	6
12*	-1	2	-3	-4	2
13	-1	-2	3	4	7
14	-1	-2	3	-4	3
15	-1	-2	-3	4	4
16	-1	-2	-3	-4	0

SR ₊	probability
0	1/16
1	1/16
2*	1/16
3	2/16
4	2/16
5	2/16
6	2/16
7	2/16
8	1/16
9	1/16
10	1/16
sum	1

(d) Perform the *Wilcoxon signed-rank test* using hand-calculations.

(e) Perform the *Wilcoxon signed-rank test* using the function `wilcox.test`.

(f) Is the *Wilcoxon signed-rank test* heavily influenced by outliers?

(g) What assumptions were needed to perform this test?

□

◇

One-sample case:

Assumption: The observations are independently sampled from a **symmetric** population with median Δ .

We perform hypothesis tests on Δ , based on the data $X_i, i = 1, \dots, n$.

Define $D_i = X_i - \Delta_0$, and proceed as in the two-sample paired case.

Revisit problem 4.1.3: Based on the simple random sample $\{32, 45, 4, 76\}$ from a **symmetric** population, test if the population median Δ is 20 versus the alternative that the median is *greater* than 20 at level $\alpha = 0.05$.

(a) Read in the data.

```
> x = c( 32, 45, 4, 76 )
```

(b) State the null and alternative hypotheses.

(c) Obtain the *permutation distribution* of SR_+ .

Permutation	Obs. 1	Obs. 2	Obs. 3	Obs. 4	SR ₊
1	1	2	3	4	10
2	1	2	3	-4	6
3	1	2	-3	4	7
4	1	2	-3	-4	3
5*	1	-2	3	4	8
6	1	-2	3	-4	4
7	1	-2	-3	4	5
8	1	-2	-3	-4	1
9	-1	2	3	4	9
10	-1	2	3	-4	5
11	-1	2	-3	4	6
12	-1	2	-3	-4	2
13	-1	-2	3	4	7
14	-1	-2	3	-4	3
15	-1	-2	-3	4	4
16	-1	-2	-3	-4	0

SR ₊	probability
0	1/16
1	1/16
2	1/16
3	2/16
4	2/16
5	2/16
6	2/16
7	2/16
8*	1/16
9	1/16
10	1/16

sum	1
-----	---

(d) Perform the *Wilcoxon signed-rank test* using hand-calculations.

(e) Perform the *Wilcoxon signed-rank test* using the function `wilcox.test`.

(f) Perform the t -test.

□

4.2.2 Large Sample Approximation

Each *rank* takes on a value in $\{1, 2, \dots, n\}$.

The *signed-rank* test statistic may be written as:

$$\text{SR}_+ = \sum_{i=1}^n iV_i,$$

where each V_i is either *zero* (if the observation D_i is *negative*) or *one* (if the observation D_i is *positive*).

Under H_0 (i.e., the population D_i is symmetric about zero), what is the mean of V_i ?

What is the mean of SR_+ under H_0 ?

The variance of SR_+ can be shown to be $n(n+1)(2n+1)/24$ (need not memorize).

Should a continuity correction be used when the distribution of a Wilcoxon signed-rank test statistic is approximated by a normal distribution?

Revisit problem 4.1.2 (sutured vs. taped wounds),

problem4.1.2.txt: In a study of the comparative tensile strength of tape-closed and sutured wounds, the following results were obtained on 10 rats, 40 days after the incisions made on their backs had been closed by suture or by surgical tape (Ury and Forrester, 1970, *American Statistician*). Suppose it is desired to test the hypothesis of no difference versus the alternative that the tape-closed wounds are stronger, at level $\alpha = 0.05$.

(a) Approximate the p -value, using the normal approximation with continuity correction but withOUT using the function `wilcox.test`.

```
> z = read.table2( "problem4.1.2.txt", header=TRUE )
> x = z[ , 2 ]
> y = z[ , 3 ]
> d = x - y
```

```
> ranks = rank(abs(d))

> # Determine the value of the Wilcoxon signed-rank test statistic.

> n = length( d )

> mu = n * (n+1) / 4

> stand.dev = sqrt( n * (n+1) * (2*n+1) / 24 )
```

(b) Approximate the p -value of the *Wilcoxon signed-rank* test statistic, using the normal approximation with continuity correction and *using* the function `wilcox.test`.

(c) Determine the *exact* p -value of the *Wilcoxon signed-rank* test statistic.

(d) Determine the *exact* p -value based on the *paired-comparison permutation test* and \bar{D} .

□

Note: In parts (a) and (b) of this above example, the normal approximation with continuity correction to the Wilcoxon signed-rank test statistic was performed for **paired samples** (when the *differences* are independent and symmetrically distributed).

Is such a normal approximation also valid for a **one-sample** test on a **median** (when the observations are independent and symmetrically distributed)?

4.2.3 Adjustment for Ties

Consider the *paired* samples case.

Two types of *ties* may occur.

(1) For a particular pair, $X = Y$.

(2) $|D|$ is the same for two or more pairs of observations.

When should observations of zero be removed?

When should observations of zero NOT be removed?

Ranking with zeros:

Problem 4.2.1: Based on the observations, test if the median of $X - Y$ is 40 versus the alternative that this median difference is *less* than 40.

Observation	1	2	3	4
X	108	131	84	165
Y	47	123	44	146

(a) State the null and alternative hypotheses.

(b) Obtain the *permutation distribution* of the Wilcoxon signed-rank test statistic, ranking WITH zeros.

```
> x = c( 108, 131, 84, 165 )
```

```
> y = c( 47, 123, 44, 146 )
```

```
> d = x - y - 40
```

Permutation	Obs. 1	Obs. 2	Obs. 3	Obs. 4	SR ₊
1	2.5	4	0	2.5	9
2	2.5	4	0	-2.5	6.5
3	2.5	-4	0	2.5	5
4*	2.5	-4	0	-2.5	2.5
5	-2.5	4	0	2.5	6.5
6	-2.5	4	0	-2.5	4
7	-2.5	-4	0	2.5	2.5
8	-2.5	-4	0	-2.5	0

SR ₊	probability
0	1/8
2.5*	2/8
4	1/8
5	1/8
6.5	2/8
9	1/8
sum	1

(c) Determine the p -value of the Wilcoxon signed-rank test, ranking with zeros, withOUT using the function `wilcox.test`.

□

Ranking withOUT zeros:

Revisit problem 4.2.1: Based on the observations, test if the median of $X - Y$ is 40 versus the alternative that this median difference is *less* than 40.

Observation	1	2	3	4
X	108	131	84	165
Y	47	123	44	146

(a) State the null and alternative hypotheses.

(b) Obtain the *permutation distribution* of the Wilcoxon signed-rank test statistic, ranking withOUT zeros.

```
> d1 = d[ d != 0 ]
```

Permutation	Obs. 1	Obs. 2	Obs. 4	SR ₊
1	1.5	3	1.5	6
2	1.5	3	-1.5	4.5
3	1.5	-3	1.5	3
4*	1.5	-3	-1.5	1.5
5	-1.5	3	1.5	4.5
6	-1.5	3	-1.5	3
7	-1.5	-3	1.5	1.5
8	-1.5	-3	-1.5	0

SR ₊	probability
0	1/8
1.5*	2/8
3	2/8
4.5	2/8
6	1/8
sum	1

(c) Determine the p -value of the Wilcoxon signed-rank test, ranking withOUT zeros, withOUT using the function `wilcox.test`.

□

Large sample approximation:

Recall from section 4.2.2: The *signed-rank* test statistic may be written as

$$SR_+ = \sum_{i=1}^n iV_i,$$

where each V_i is either *zero* (if the observation D_i is *negative*) or *one* (if the observation D_i is *positive*).

Simple example (no ties): The values of D are $\{25, -30, 57, 13\}$.

□

For large n and possible *ties*, what is the approximate distribution of SR_+ ?

Problem #4.2.2 (IQ scores), [problem4.2.2.txt](#): In a study of the effect that familiarity with the examiner may have on the performance of a child in an IQ test, 48 first-graders were divided into 24 matched pairs, and one child from each pair was selected at random for the experimental group. Each child in the experimental group spent four 20-minute periods with the investigator, who then gave the child an IQ test. The corresponding test was also given to the controls, and an initial IQ score was established for all children at the beginning of the study. Marine (1929, *Contrib. to Educ.*, Columbia Univ.) reported the following differences (treated – control) in the gains achieved in the IQ scores: $-17, -15, -8, -7, -10, 4, -10, 4, 12, 1, 21, 2, -20, 2, 0, 5, -7, 19, -8, 9, -3, -7, -1, 34$.

We are interested in testing the hypothesis of no difference versus the alternative that familiarity with the examiner tends to improve the performance of the test.

(a) State the null and alternative hypotheses.

(b) Read in the data.

```
> d = scan2( "problem4.2.2.txt" )
```

(c) Are there any ties or zeros?

(d) Determine the approximate p -value of the Wilcoxon signed-rank test.

(e) What assumptions were needed to perform this test?

(f) Determine the approximate p -value of the t -test.

(g) What assumptions were needed to perform this test?

□

Homework p. 142: Exercises 4.3, 4.4 (exact test), 4.4* (by hand AND by R , use normal approximation with continuity correction), 4.5

Hints for homework exercise 4.4*: Make sure that you read in (i.e., do not retype) the data. State H_0 , H_a , and conclusion in **statistical terms** and **regular English**, and **define** any notation used. When using the normal approximation with continuity correction, determine the **approximated** p -value **twice**, once by hand (i.e., withOUT using a function containing the word **test**) and once by R (i.e., USING a function containing the word **test**). Retype the two **approximated** p -values with the remark '# approximated', or **highlight** the two **approximated** p -values in **yellow**. When computing the **exact** p -value, solve only by R (i.e., USING a function containing the word **test**). Retype the **exact** p -value with the remark '# exact', or **highlight** the **exact** p -value in **pink**. Remember to **COMPARE** the two **approximated** p -values (which should be the same) to the **exact** p -value. Introduce the question **number** as a comment using “#” or in **red** using .html code; e.g., ` Exercise 4.4 `.

4.3 Other Paired-Comparison Tests

4.3.1 Sign Test

The **sign** test is a special case of the **binomial** test (from section 1.1.1), where the null population median is set to zero.

Let SN_+ be the number of observations greater than zero.

What is the distribution of SN_+ under the null hypothesis that $\theta_H = 0$ (for a continuous population)?

How can zeros occur in the data, if the population is continuous?

What should be done with zeros when computing SN_+ ?

What is the *asymptotic* distribution of SN_+ under the null hypothesis that $\theta_H = 0$ (for a continuous population), for large n ?

Problem #4.3.1 (schizophrenia), [problem4.3.1.txt](#): Stanley and Walton (1961, *J. Mental Sci.*) reported the following behavior ratings before and after six weeks of receiving a placebo, for 12 chronic schizophrenics who were used as controls in a clinical trial. *Larger behavior ratings are more favorable than smaller behavior ratings.* Use the *sign* test to determine if behavior ratings tend to increase after using a placebo.

Before	2.4	2.2	2.1	2.9	2.2	2.3	2.4	1.5	2.7	1.9	1.8	1.3
After	2.54	3.18	2.54	3.27	2.09	2.45	3.09	1.45	3.45	3.09	1.81	1.45

(a) State the null and alternative hypotheses.

(b) Read in the data.

```
> z = scan2( "problem4.3.1.txt" )
```

(c) Determine the value of the test statistic SN_+ .

(d) Determine the *exact* p -value of the *sign* test and state the conclusion, withOUT using the function `binom.test`.

(e) Determine the *exact* p -value of the *sign* test, using the function `binom.test`.

(f) Plot the distribution of SN_+ under H_0 , and shade in the appropriate region corresponding to the p -value.

(g) Determine the *asymptotic* p -value of the *sign* test and state the conclusion.

(h) Plot the asymptotic distribution of SN_+ under H_0 , and shade in the appropriate region corresponding to the p -value.

□

4.3.3 Selecting Among Paired Comparison Tests

Compare the three paired tests:

⊙ paired t -test

- ⊙ Wilcoxon signed-rank test
- ⊙ sign test

Which test is *most* susceptible to outliers?

Which test is *least* susceptible to outliers?

Normal alternatives:

Which test is most powerful when the population is *normal* or has *light tails*?

For *normal* distributions, the asymptotic efficiency of the *Wilcoxon signed-rank* test to the *t*-test is 0.955.

For *normal* distributions, the asymptotic efficiency of the *sign* test to the *Wilcoxon signed-rank* test is $2/3$.

Laplace alternatives:

For *Laplace* distributions, the asymptotic efficiency of the *Wilcoxon signed-rank* test to the *t*-test is 1.5.

Cauchy alternatives:

For *Cauchy* distributions, the asymptotic efficiency of the *Wilcoxon signed-rank* test to the *t*-test is infinite.

For *Cauchy* distributions, the asymptotic efficiency of the *sign* test to the *Wilcoxon signed-rank* test is 1.3.

Homework p. 142: Exercise 4.6 (exact p -value by hand and by R , normal approximation with continuity correction)

p. 141

Exercises

1 For the data in the table, obtain the permutation distribution of the mean of the differences.

Pair	1	2	3	4
Treatment 1	100	250	50	80
Treatment 2	112	240	58	82

142 Chapter 4: Paired Comparisons and Blocked Designs

- 2 Students in an introductory statistics course were asked at the beginning of the course and at the end of the course the extent to which they disagreed or agreed with the statement: "Statistics is important to my major area of study." They responded on a scale of 1 to 5 with (1) strongly disagree, (2) moderately disagree, (3) neutral, (4) moderately agree, and (5) strongly agree. Data are shown in the table. Use a paired-comparison permutation test to determine whether responses to this question changed significantly from the beginning to the end of the semester.

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Before	2	3	4	4	3	1	3	4	4	5	3	4	2	2	4	3	4	2	2
After	2	4	4	4	4	4	3	5	4	4	4	5	4	2	5	5	4	1	2

- 3 Obtain the permutation distribution of the signed-rank statistic SR_+ for the data in Exercise 1.
- 4 Measurements of a blood enzyme LDH were taken on seven subjects before fasting and after fasting. Is there a significant difference between the LDH readings before and after fasting? Test using Wilcoxon's signed-rank statistic. Compare the p -value of the exact test with that of the normal approximation.

Subject	1	2	3	4	5	6	7
Before	89	90	87	98	120	85	97
After	76	101	84	86	105	84	93

- 5 Apply the Wilcoxon signed-rank test adjusted for ties to the data in Exercise 2.
- 6 Apply the sign test to the data in Exercises 2 and 4.