Dr Isabella Gollini

Probability and Statistics

Lab - Non Parametric test for the median η

2 Guided Example

2.1 Example 1 of 2013 Exam

In an experiment to investigate whether there was significant evidence of an underlying difference between the mean measurements given by two types of caliper, Caliper 1 and Caliper 2, the diameter of a ball bearing was measured by 12 inspectors, each using the two types of caliper. The results are given in the table below.

Diameter of ball bearing in cm.

Inspector	Caliper 1	Caliper 2
1	0.265	0.264
2	0.265	0.265
3	0.266	0.264
4	0.267	0.266
5	0.267	0.267
6	0.265	0.268
7	0.267	0.264
8	0.267	0.265
9	0.265	0.265
10	0.268	0.267
11	0.268	0.268
12	0.265	0.269

(i) Specify appropriate null and alternative hypotheses and carry out the sign test. Draw conclusions.

(ii) Specify any assumptions and appropriate null and alternative hypotheses, and carry out the Wilcoxon signed-rank test. Draw conclusions.

(iii) Using the tabulated data, show exactly how the Wilcoxon statistic has been calculated.

(iv) Comment on the essential difference between the assumptions that underlie the parametric and the non-parametric tests, and on how the conclusions that may be drawn from the two procedures compare in the present case.

Solution:

We have already used this dataset during the lab on “the testing of hypothesis and t-test”, if you saved the `R` script you can use it to load the data. Otherwise you have to type in the two vectors for the data on Caliper 1 and Caliper 2 manually:

```r
Caliper1 <- c(0.265, 0.265, 0.266, 0.267, 0.267, 0.265,
              0.267, 0.267, 0.265, 0.268, 0.268, 0.265)

Caliper2 <- c(0.264, 0.265, 0.264, 0.266, 0.267, 0.268,
              0.264, 0.265, 0.265, 0.267, 0.268, 0.269)

cbind(Caliper1, Caliper2)

##       Caliper1 Caliper2
##  [1,]    0.265    0.264
##  [2,]    0.265    0.265
##  [3,]    0.266    0.264
##  [4,]    0.267    0.266
##  [5,]    0.267    0.267
##  [6,]    0.265    0.268
##  [7,]    0.267    0.264
##  [8,]    0.267    0.265
##  [9,]    0.265    0.265
## [10,]    0.268    0.267
## [11,]    0.268    0.268
## [12,]    0.265    0.269
```

Fill in this table, and then check if you got it right:

Inspector	Caliper 1 x_i	Caliper 2 y_i	difference $d_i = x_i - y_i$	sign of d_i	rank of d_i	signed rank
1	0.265	0.264	*0.001*	$+$	*2*	*+2*
2	0.265	0.265	*0*	.		
3	0.266	0.264	*0.002*	$+$	*4.5*	*+4.5*
4	0.267	0.266	*0.001*	$+$	*2*	*+2*
5	0.267	0.267	*0*	.		
6	0.265	0.268	*-0.003*	$-$	*6.5*	$-$*6.5*
7	0.267	0.264	*0.003*	$+$	*6.5*	*+6.5*
8	0.267	0.265	*0.002*	$+$	*4.5*	*+4.5*
9	0.265	0.265	*0*	.		
10	0.268	0.267	*0.001*	$+$	*2*	*+2*
11	0.268	0.268	*0*	.		
12	0.265	0.269	*-0.004*	$-$	*8*	$-$*8*

Remember that tied values are assigned the average of the ranks that would have been assigned with no ties.

(i) Firstly we have to specify the null and alternative hypotheses, and noticing that in order to perform the sign test we are making the assumptions that the sample is random, and the observations come from the same (unknown) distribution.

> Write the null and alternative hypotheses:
>
> *We test the null hypothesis* $H_0 : \eta = 0$ *against the two-sided alternative* $H_1 : \eta \neq 0$.

In order to perform the sign test we have to calculate the vector of the differences:

```
d <- Caliper1 - Caliper2
d

##  [1]  0.001  0.000  0.002  0.001  0.000 -0.003  0.003  0.002
##  [9]  0.000  0.001  0.000 -0.004
```

We calculate the test statistic s that is the number of $-$ signs that corresponds to the number of observations with $d_i < 0$. `n` is the number of observations with $d_i \neq 0$.

```
s <- sum(d < 0)
n <- sum(d != 0)
s

## [1] 2

n

## [1] 8
```

To perform the sign test in `R` we use the function `binom.test`. You have to notice that he first argument to be specified is the number of successes, that in this context is represented by the number of positives differences (that is equal to the number of trials `n` minus the number of failures `s`). The second argument is the number of trials, that in this context is given by the number of observations with differences not equal to 0. Then we can specify the alternative hypothesis, that by default is `two.sided`. Type `?binom.test` for further details and options.

```r
binom.test(n - s, n)

##
##  Exact binomial test
##
## data:  n - s and n
## number of successes = 6, number of trials = 8,
## p-value = 0.2891
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.3491442 0.9681460
## sample estimates:
## probability of success
##                   0.75
```

Draw conclusions:

The p-value of 0.289 is not significant at the 5% significance level. There is no significant evidence that there is any difference in the median measurements given by the two types of caliper.

(ii) In the case of the non-parametric Wilcoxon signed-rank test it is assumed that the differences d_i are symmetrically distributed about their median value η. This implies that we assume that the median is equal to the mean.

```r
wilcox.test(d, correct = FALSE)

## Warning in wilcox.test.default(d, correct = FALSE): cannot
compute exact p-value with ties
## Warning in wilcox.test.default(d, correct = FALSE): cannot
compute exact p-value with zeroes

##
##  Wilcoxon signed rank test
##
## data:  d
## V = 21.5, p-value = 0.6215
## alternative hypothesis: true location is not equal to 0
```

NOTE: In presence of ties, and/or values equal to 0 the function `wilcox.test` uses the normal approximation to calculate the p-value, this is the reason of the `Warning` message. The value of the test statistic T^+ that is indicated in the output with `V` is correct.

Draw conclusions:

The p-value of 0.621 is certainly not significant at the 5% significance level, nor even remotely significant at any level that we would normally consider. There is no significant evidence that there is any difference in the mean [median] measurements given by the two types of caliper.

(iii) In order to carry out the Wilcoxon test without using `R` we have to fill in the table on the top of page 5. The Wilcoxon statistic given by the sum of the positive signed ranks is $T^+ = 2 + 4.5 + 2 + 6.5 + 4.5 + 2 = 21.5$, and the Wilcoxon statistic given by the sum of the negative signed ranks is $T^- = 6.5 + 8 = 14.5$. We can check that $T^+ + T^- = n(n+1)/2 = 8 \times 9/2 = 36$.

Given the value of the test statistic draw the conclusion to the test by using the statistical tables:

(iv) In all the tests the data are assumed to be random and coming from the same distribution. With the non-parametric sign test no assumption on the form of the distribution is made. With the non-parametric Wilcoxon test no normality assumption is made, but the data are assumed to came from a symmetric distribution. With the parametric t-test it is assumed that the differences are normally distributed.

In all the three cases the p-values are $> \alpha = 0.05$. The p-values for the Wilcoxon test and the t-test are quite similar [the p-value of the t-test was 0.674]. The conclusions to be drawn are the same in all the cases.

3 Extra Examples

3.1 Running Times

Eight athletes ran a 400 metre race at sea level and, at a later meeting, ran another 400 metre race at high altitude. Their times in seconds are presented below.

Runner	Sea level time	High altitude time
1	48.3	50.4
2	47.6	47.3
3	49.2	50.8
4	50.3	52.3
5	48.8	47.7
6	51.1	54.5
7	49.0	48.9
8	48.1	49.9

Do these data provide evidence that athletes perform better at sea level?

(i) Specify appropriate null and alternative hypotheses and carry out the sign test. Draw conclusions.

(ii) Specify any assumptions and appropriate null and alternative hypotheses, and carry out the Wilcoxon signed-rank test. Draw conclusions.

(iii) Comment on any differences between the conclusions of the sign test, the Wilcoxon signed-rank test and the paired comparisons t-test of Examples 5 – Question 3.

Solution:

(i) The time differences and their signs are

-2.1	0.3	-1.6	-2.0	1.1	-3.4	0.1	-1.8
$-$	$+$	$-$	$-$	$+$	$-$	$+$	$-$

Let η denote the median of the population from which our sample of differences is drawn. We test $H_0 : \eta = 0$ against $H_1 : \eta < 0$. The number of $+$ signs is $S = 3$. Using a one-sided test and Table 1 of the binomial distribution function with $n = 8, p = 0.5$, we obtain the p-value,

$$p = \Pr(S \leq 3) = F_3 = 0.3633 \, .$$

This is not significant even at the 25% level. According to this test, there is no evidence that athletes perform better at sea level than at high altitude.

The corresponding `R` output is:

```
SeaLevel <- c(48.3, 47.6, 49.2, 50.3, 48.8, 51.1, 49.0, 48.1)
HighAlt <- c(50.4, 47.3, 50.8, 52.3, 47.7, 54.5, 48.9, 49.9)
cbind(SeaLevel, HighAlt)
```

```
##      SeaLevel HighAlt
## [1,]     48.3    50.4
## [2,]     47.6    47.3
## [3,]     49.2    50.8
## [4,]     50.3    52.3
## [5,]     48.8    47.7
## [6,]     51.1    54.5
## [7,]     49.0    48.9
## [8,]     48.1    49.9
```

```
diff <- SeaLevel - HighAlt

s <- sum(diff < 0)
n <- sum(diff != 0)
```

```
binom.test(n - s, n,
           alternative = "less")
```

```
##
##  Exact binomial test
##
## data:  n - s and n
## number of successes = 3, number of trials = 8,
## p-value = 0.3633
## alternative hypothesis: true probability of success is less than 0.5
## 95 percent confidence interval:
##  0.0000000 0.7107592
## sample estimates:
## probability of success
##                  0.375
```

(ii) We assume that the population distribution of the differences is symmetric. The sample is random. We again test $H_0 : \eta = 0$ against $H_1 : \eta < 0$. The time differences and their signed ranks are

-2.1	0.3	-1.6	-2.0	1.1	-3.4	0.1	-1.8
-7	$+2$	-4	-6	$+3$	-8	$+1$	-5

The sum of the ranks with positive signs is $T^+ = 6$. Using a one-sided test and Table 20 with $n = 8$, we see that we would need $T^+ \leq 5$ to have a result that was significant at the 5% level. So there is no strong evidence that athletes perform better at sea level than at high altitude.

The corresponding R output is:

```r
wilcox.test(diff,
            alternative = "less",
            correct = FALSE)

##
##  Wilcoxon signed rank test
##
## data:  diff
## V = 6, p-value = 0.05469
## alternative hypothesis: true location is less than 0
```

(iii) The p-value of 0.055 for the Wilcoxon signed-ranks test shows that it comes close to giving a result that is significant at the 5% level, but the sign test gives a result that is nowhere near significant at the 5% level. The sign test does not take into account the fact that the positive differences are the smallest ones in absolute value, whereas the Wilcoxon signed-rank test does, so giving a value of T^+ that is small enough to be almost significant at the 5% level.

Under the normality assumption, the paired comparisons t-test gave a p-value of 0.032 that was significant at the 5% level.