Dr Isabella Gollini

Probability and Statistics

Lab - Non Parametric test for the median η

- Section 1 contains a summary of non parametric test for the median.
- Section 2 [page 4] contains a guided example
- Section 3 [page 8] contains other two exercises. [Try to do them on your own, but do not hesitate to ask any questions!]

Remember to write your code into an `R` script, in order to be able to save and reuse it.

1.1 Sign test

1. We have a random sample, $x_1, x_2, \ldots, x_n$, of size n from some unknown distribution. To each observation x_i attach a $+$ sign if $x_i > \eta_0$ and a $-$ sign if $x_i < \eta_0$. Discard any values equal to η_0. Let n the number of values that remain.

2. **H_0 and H_1**

Null hypothesis H_0	Alternative hypothesis H_1				
	two-sided		one-sided greater		one-sided less
$\eta = \eta_0$	$\eta \neq \eta_0$	or	$\eta > \eta_0$	or	$\eta < \eta_0$

3. **Test Statistic**

   To test the null hypothesis H_0 we use the test statistic

$$s = \text{“the number of } - \text{ signs”} \sim Bin(n, 0.5).$$

5 **p-value**

two-sided		one-sided greater		one-sided less
$2\ \min\left(F(s\|n, 0.5), 1 - F(s-1\|n, 0.5)\right)$	or	$F(s\|n, 0.5)$	or	$1 - F(s-1\|n, 0.5)$

where F is the c.d.f. of a Binomial distribution.

R function

Two-sided H_1:

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

One-sided greater H_1:

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

One-sided less H_1:

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

1.2 Wilcoxon Signed-Rank test

1. We have a random sample, $x_1, x_2, \ldots, x_n$, of size n from some unknown symmetric distribution (so that $\eta = \mu$). To construct the Wilcoxon test statistic for the signed-rank test, carry out the following steps.
   (a) Define $x_i^* = x_i - \eta_0$ $(i = 1, 2 \ldots, n)$.
   (b) Eliminate any observations $x_i^* = 0$. Let n the number of values that remain.
   (c) Rank the absolute values $|x_i^*|$ by assigning 1 to the smallest, 2 to the second smallest, …, n to the largest. Tied values are assigned the average of the ranks that would have been assigned with no ties.
   (d) Assign signs to the ranks, a $+$ sign if $x_i^* > 0$, and a $-$ sign if $x_i^* < 0$.

2. **H_0 and H_1**

Null hypothesis H_0	Alternative hypothesis H_1				
	two-sided		one-sided greater		one-sided less
$\eta = \eta_0$	$\eta \neq \eta_0$	or	$\eta > \eta_0$	or	$\eta < \eta_0$

3. **Test Statistic**

   To test the null hypothesis H_0 we use the test statistics

$$\begin{aligned} T^- &= \text{“the sum of the ranks with } - \text{ sign”} \\ T^+ &= \text{“the sum of the ranks with } + \text{ sign”}. \end{aligned}$$

   Under H_0, T^- and T^+ are identically distributed, and $E(T^-) = E(T^+) = \frac{1}{4}n(n+1)$.

4. **Rejection region Significance-level-α test**

two-sided		one-sided greater		one-sided less
$\min(T^-, T^+) \leq x(P/2)$	or	$T^- \leq x(P)$	or	$T^+ \leq x(P)$

   where $P = 100\alpha$, and $x(P)$ are the lower percentage points of T^-, T^+ [see Table 20 of *Lindley and Scott*].

R function

Two-sided H_1:

```
wilcox.test(x, mu = eta0, correct = FALSE)
```

One-sided greater H_1:

```
wilcox.test(x, mu = eta0, alternative = "greater", correct = FALSE)
```

One-sided less H_1:

```
wilcox.test(x, mu = eta0, alternative = "less", correct = FALSE)
```

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)
```

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				
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				

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:

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

```r
d <- Caliper1 - Caliper2
d
```

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
n
```

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.

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

Draw conclusions:

(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.

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

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:

(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.

3.2 Student's Sleep Data

This dataset is more then 100 years old: the data was collected by Cushny and Peebles (1905), and was used by Gosset ("Student") to demonstrate the theoretical developments of the t-distribution (Student, 1908).

The data are the average number of hours of sleep (relative to a baseline) gained by 10 patients on two different soporific drugs, Dextro-hyoscyamine hydrobromide and Laevo-hyoscyamine hydrobromide.

Patient ID	Drug 1	Drug 2
1	0.7	1.9
2	-1.6	0.8
3	-0.2	1.1
4	-1.2	0.1
5	-0.1	-0.1
6	3.4	4.4
7	3.7	5.5
8	0.8	1.6
9	0.0	4.6
10	2.0	3.4

The question is whether one drug is better than the other.

(i) Specify any assumptions and appropriate null and alternative hypotheses, and carry out a parametric test. Draw conclusions.

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

(iii) Specify any assumptions and carry out the Wilcoxon signed-rank test. Draw conclusions.

(iv) Comment on any differences between the conclusions of the three tests performed.

NOTE: The data set `sleep` is available in `R`:

```r
library(datasets)
data(sleep)
attach(sleep)
drug1 <- extra[group == 1]
drug2 <- extra[group == 2]
```


- Cushny, A. R. and Peebles, A. R. (1905) "The action of optical isomers: II hyoscines". *The Journal of Physiology* 32, 501–510.
- Student (1908) "The probable error of a mean". *Biometrika* 6: 1–25.