**Dr Isabella Gollini**

**Probability and Statistics**

# 15 Nonparametric methods

## 15.1 Introduction

Many statistical procedures are based on the assumption that the data are from some known family of probability distributions, most commonly from the family of normal distributions, and the hypotheses that we test are expressed in terms of the unknown parameter values.

What are known as *nonparametric* methods are tests of hypotheses (or other procedures) that do not rely on an assumption that the data come from a known family of distributions with associated parameter values. In this section, we look at a couple of simple examples of nonparametric techniques.

With nonparametric methods, it is commonly more natural to frame hypotheses not in terms of the mean $\mu$ of the distribution from which a random sample is drawn but in terms of its median $\eta$.

If a continuous random variable $X$ has a p.d.f. $f(x)$ then its *median* $\eta$ is the number such that

$$\int_{-\infty}^{\eta} f(x)dx = \frac{1}{2} .$$

Equivalently, $F(\eta) = \frac{1}{2}$, where $F$ is the c.d.f. It follows that

$$\Pr(X < \eta) = \Pr(X > \eta) = \frac{1}{2} .$$

If the p.d.f. is a symmetric function then $\eta = \mu$.

## 15.2 The sign test

Suppose that we have a random sample, $x_1, x_2, \ldots, x_n$, of size $n$ from some unknown continuous distribution and we wish to test the null hypothesis that the median $\eta$ of the distribution takes some particular value $\eta_0$. As previously, the alternative hypothesis might be two-sided, $\eta \neq \eta_0$, or one-sided, $\eta > \eta_0$ or $\eta < \eta_0$. So we test the null hypothesis

$$H_0 : \eta = \eta_0$$

against, for example, the alternative hypothesis

$$H_1 : \eta > \eta_0,$$

if we adopt a one-sided alternative.

Under $H_0$, the $x_i$ are independently and identically distributed with

$$\Pr(x_i < \eta_0) = \Pr(x_i > \eta_0) = \frac{1}{2} \qquad (i = 1, 2, \ldots, n).$$

To each observation $x_i$ attach a $+$ sign if $x_i > \eta_0$ and a $-$ sign if $x_i < \eta_0$. (Ignore observations where, to the accuracy of the observations, $x_i = \eta_0$.) Let $S$ denote the total number of $-$ signs. If $n$ now represents the sample size after removal of any observations with $x_i = \eta_0$, under $H_0$, $S \sim B(n, 1/2)$, so that $E(S) = n/2$. We use $S$ as the test statistic and reject $H_0$ if the observed value of $S$ is too extreme. To calculate the $p$-value that corresponds to an observed value of $S$ we can use Table 1 of the binomial distribution function in *Lindley and Scott*, with $p = 0.5$.

An important special case where this approach might be used is in data arising from a matched pairs design, where we have a random sample of differences, $d_1, d_2, \ldots, d_n$, where $d_i$ represents the difference in response for the $i$th pair between Treatment 1 and Treatment 2, say, $d_i = x_i - y_i$ in an obvious notation. We would be interested in testing whether there is any evidence that one of the treatments tends to give higher responses than the other. In this case $\eta_0 = 0$ and to each difference $d_i$ attach a $+$ sign if $d_i > 0$ and a $-$ sign if $d_i < 0$.

**Example from Section 10.4 — the effect of alcohol**

In a study using identical twins to test the effect of alcohol on intelligence, for each pair of twins, one of them, selected at random, was given an intelligence test while under the influence of a given dose of alcohol. The other twin was given the same test under alcohol-free conditions. The test scores and the pairwise differences $d_i$ are listed below in Table 1. Do these data provide evidence that alcohol lowers performance on intelligence tests? This question was addressed in Section 10.4 using a paired comparisons $t$-test, assuming that the differences were a random sample from a normal distribution.

| Pair | No Alcohol | Alcohol | difference | sign |
|---|---|---|---|---|
| 1 | 83 | 78 | 5 | $+$ |
| 2 | 74 | 74 | 0 | . |
| 3 | 67 | 63 | 4 | $+$ |
| 4 | 64 | 66 | -2 | $-$ |
| 5 | 70 | 68 | 2 | $+$ |
| 6 | 67 | 63 | 4 | $+$ |
| 7 | 81 | 77 | 4 | $+$ |
| 8 | 64 | 65 | -1 | $-$ |
| 9 | 72 | 70 | 2 | $+$ |

Table 1: Intelligence test scores

If we feel that the normality assumption is unjustified then we may carry out a non-parametric test, for example, the sign test as applied to the $d_i$. We test the null hypothesis

$$H_0 : \eta = 0$$

against the one-sided alternative hypothesis

$$H_1 : \eta > 0.$$

If $H_0$ is true then we would expect about half of the signs to be $-$, but if $H_1$ is true then we would expect fewer of the signs to be $-$. We use a one-tail test and reject $H_0$ if the number of minus signs $S$ is small enough. Here there is one zero difference, so we take

$n = 9 - 1 = 8$, and $S = 2$. Under $H_0$, $S \sim B(8, 1/2)$. Using Table 1 of *Lindley and Scott* with $n = 8, p = 0.5, r = 2$ we find the corresponding $p$-value,

$$p = \Pr(S \leq 2) = F_2 = 0.1445.$$

Since $p > 0.1$, we do not reject $H_0$ at the 5% level or even at the 10% level. There is no strong evidence that alcohol lowers performance. (But if with $n = 8$ we had obtained $S = 1$ then $p = F_1 = 0.0352$, significant at the 5% level.)

Thinking of the signs as a random sample from some population, the sign test may be viewed as testing whether the population proportion of minus signs is 1/2. The sign test is equivalent to the test for a single proportion as described in Section 11.3, taking $p_0 = 1/2$, though there a normal approximation to the binomial distribution is used. Here we are using the binomial distribution itself to calculate the $p$-value, though for $n$ large we might well use the normal approximation.

To perform the sign test in `R`, firstly we load the data and calculate the vector of the differences.

```
NoAlcohol <- c(83, 74, 67, 64, 70, 67, 81, 64, 72)
Alcohol <- c(78, 74, 63, 66, 68, 63, 77, 65, 70)
Diff <- NoAlcohol - Alcohol
cbind(NoAlcohol, Alcohol, Diff)

##       NoAlcohol Alcohol Diff
##  [1,]        83      78    5
##  [2,]        74      74    0
##  [3,]        67      63    4
##  [4,]        64      66   -2
##  [5,]        70      68    2
##  [6,]        67      63    4
##  [7,]        81      77    4
##  [8,]        64      65   -1
##  [9,]        72      70    2
```

Then, we use the function `binom.test`. The first argument to be specified is the number of successes, that in this context is represented by the number of positives differences. 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 have to specify the alternative hypothesis. Type `?binom.test` for further details and options.

```
binom.test(sum(Diff>0),
           sum(Diff!=0),
           alternative = "greater")

##
##  Exact binomial test
##
## data:  sum(Diff > 0) and sum(Diff != 0)
## number of successes = 6, number of trials = 8, p-value = 0.1445
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.4003106 1.0000000
## sample estimates:
## probability of success
##                   0.75
```

The same $p$-value is obtained as from the table of the binomial distribution function.

- The sign test may be used for pairwise comparisons even when the experimental variables cannot be measured on an exact numerical scale, as might be the case in some medical or psychological testing. All that is required is that we are able to compare the responses of pairs of individuals so that we can say which is greater/better. This would be the case if the responses were on an *ordinal* scale, i.e., we could rank the responses even if we could not attach numerical values to them. In terms of our example, all that we need to know is whether in each pair the twin given alcohol or the twin not given alcohol did better on the test. We then assign to the pair a $-$ or $+$ sign, respectively.

- Where the experimental variables can be measured on a numerical scale, the sign test does not make use of all the information available. It uses the signs of the $d_i$ but not their sizes. Because of this, the sign test is not as powerful a procedure as ones that make use of the information on sizes. In our example, the $t$-test of Section 10.4 gave a significant result with $p = 0.022$, whereas the sign test gives a non-significant result with $p = 0.1445$. We turn next to a nonparametric test that does to some extent make use of the sizes of the $d_i$, the Wilcoxon signed-rank test.

## 15.3 The Wilcoxon signed-rank test

As in the case of the sign test, suppose that we have a random sample, $x_1, x_2, \ldots, x_n$, of size $n$ from some unknown continuous distribution and we wish to test the null hypothesis that the median $\eta$ of the distribution takes some particular value $\eta_0$, i.e.,

$$H_0 : \eta = \eta_0.$$

The distribution of the $x_i$ need not be normal, but it is assumed to be symmetric about its median value (so that $\eta = \mu$, the median is equal to the mean). To construct the Wilcoxon test statistic for the signed-rank test, carry out the following steps.

1. Define
$$x_i^* = x_i - \eta_0 \qquad (i = 1, 2 \ldots, n).$$
2. Eliminate any observations for which $x_i^* = 0$, reducing the value of $n$ accordingly.
3. Rank the absolute values $|x_i^*|$ by assigning 1 to the smallest, 2 to the second smallest, $\ldots$, $n$ to the largest. Tied values are assigned the average of the ranks that would have been assigned with no ties.
4. Assign signs to the ranks, a $+$ sign if $x_i^* > 0$, and a $-$ sign if $x_i^* < 0$, to obtain the "signed ranks."
5. Calculate the sum $T^-$ of the ranks with $-$ sign and/or the sum $T^+$ of the ranks with $+$ sign.

$T^-$ and $T^+$ are the *Wilcoxon test statistics*. Note that the sum of the ranks of all the $x_i^*$ is
$$\sum_{i=1}^{n} i = \frac{1}{2}n(n+1).$$
It follows that
$$T^- + T^+ = \frac{1}{2}n(n+1), \tag{1}$$
so that if we have calculated one of $T^-$, $T^+$ then the value of the other follows from Equation (1). Alternatively, if we have calculated both $T^-$ and $T^+$ then Equation (1) provides a useful check of our calculations.

- In the commonly seen special case where we have a random sample of differences, $d_1, d_2, \ldots, d_n$, from a matched pairs design and $\eta_0 = 0$, we apply the above steps 1-5 directly to the differences $d_i$.

Under $H_0$, $T^-$ and $T^+$ are identically distributed, and from Equation (1) it follows that
$$E(T^-) = E(T^+) = \frac{1}{4}n(n+1). \tag{2}$$
Although there is no simple way of specifying their distributions under $H_0$, $T^-$ and $T^+$ are symmetrically distributed about the expected value of Equation (2). If the observed values of $T^-$, $T^+$ differ too much from this expected value then this will provide evidence against the null hypothesis. Lower percentage points $x(P)$ of $T^-$, $T^+$ are given in Table 20 of *Lindley and Scott*.

- In the case of a two-sided alternative hypothesis, $H_1 : \eta \neq \eta_0$, we may take $T = \min(T^-, T^+)$. Using a two-tail test, we reject $H_0 : \eta = \eta_0$ at the $P\%$ significance level if, for the given value of $n$, $T \leq x(P/2)$.
- In the case of the one-sided alternative hypothesis, $H_1 : \eta > \eta_0$, using a one-tail test, we reject $H_0$ at the $P\%$ significance level if $T^- \leq x(P)$.
- In the case of the one-sided alternative hypothesis, $H_1 : \eta < \eta_0$, using a one-tail test, we reject $H_0$ at the $P\%$ significance level if $T^+ \leq x(P)$.

**Example — the effect of alcohol (continued)**

For the matched pairs study using identical twins to test the effect of alcohol on intelligence, in Table 2 we have ranked the $|d_i|$ and attached signs to the ranks to create a column of signed ranks.

| Pair | difference $d_i$ | sign | rank of $\lvert d_i \rvert$ | signed rank |
|---|---|---|---|---|
| 1 | 5 | $+$ | 8 | +8 |
| 2 | 0 | . | . | . |
| 3 | 4 | $+$ | 6 | +6 |
| 4 | -2 | $-$ | 3 | $-3$ |
| 5 | 2 | $+$ | 3 | +3 |
| 6 | 4 | $+$ | 6 | +6 |
| 7 | 4 | $+$ | 6 | +6 |
| 8 | -1 | $-$ | 1 | $-1$ |
| 9 | 2 | $+$ | 3 | +3 |

Table 2: Intelligence test scores — signed ranks of differences

Recall that in this case we are testing $H_0 : \eta = 0$ against the one-sided alternative, $H_1 : \eta > 0$, so all that we need to calculate is $T^-$. From Table 2, $T^- = 4$. From Table 20 of *Lindley and Scott*, we find that, for $n = 8$, $x(5) = 5$ so that our test statistic is significant at the 5% level. We conclude that there is strong evidence that the use of alcohol lowers performance.

Using the Wilcoxon signed-rank test, we have found a significant result, whereas using the sign test we did not find a significant result. The information that the positive differences are greater in absolute value than the negative differences has been incorporated into the calculation of the Wilcoxon test statistic, whereas this information was not made use of in the sign test.

In `R` there is the function `wilcox.test`:

```r
wilcox.test(Diff,
            alternative = "greater",
            correct = FALSE)

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

##
##  Wilcoxon signed rank test
##
## data:  Diff
## V = 32, p-value = 0.02386
## alternative hypothesis: true location is greater than 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.

The Wilcoxon test statistic is the value of $T^+ = 32$, and the $p$-value of 0.024 (calculated using a normal approximation) is consistent with our earlier conclusion that the test statistic is significant at the 5% level. This $p$-value is also not very different from the $p$-value of 0.022 given by the $t$-test of Section 10.4.