

Nonparametric Test Procedures

1 Introduction to Nonparametrics

Nonparametric tests do not require that samples come from populations with normal distributions or any other specific distribution. Hence they are sometimes called distribution-free tests. They may have other more easily satisfied assumptions - such as requiring that the population distribution is symmetric. In general, when the parametric assumptions are satisfied, it is better to use parametric test procedures. However, nonparametric test procedures are a valuable tool to use when those assumptions are not satisfied, and most have pretty high efficiency relative to the corresponding parametric test when the parametric assumptions are satisfied. This means you usually don't lose much testing power if you swap to the nonparametric test even when not necessary. The exception is the sign test, discussed below. The following table discusses the parametric and corresponding nonparametric tests we have covered so far. (z-test, t-test, F-test refer to the parametric test statistics used in each procedure).

Parameter	Parametric Test	Nonparametric Test
π	One sample z-test	Exact Binomial
μ or μ_d	One sample t-test	Sign test or Wilcoxon Signed-Ranks test
$\mu_1 - \mu_2$	Two sample t-test	Wilcoxon Rank-Sum test
ANOVA	F-test	Kruskal-Wallis

As a note on p-values before we proceed, details of computations of p-values are left out of this discussion. Usually these are found by considering the distribution of the test statistic across all possible permutations of ranks, etc. in each case. It is very tedious to do by hand, which is why there are established tables for it, or why we rely on the computer to do it.

The examples throughout these notes are brief - I do not list the hypotheses or check the assumptions explicitly. Assume $\alpha = .05$. You should be able to determine the hypotheses as we discuss the examples as well as make a decision based on the reported p-value.

Finally, you are NOT expected to know how to do these tests by hand or even the R command. At most, you will be asked which nonparametric test to use (see chart above) and to make a decision based on a p-value from one of these tests (decision rule still applies). You will NOT have to list these assumptions, etc. I am providing some of the details as a reference for you if you want to employ these in your project or some other future time.

References:

Devore and Peck. Fifth edition. Has short descriptions of each test. Was removed to an optional online chapter in our Sixth edition.

Triola. Elementary Statistics. Has a chapter on nonparametrics - no discussion of binomial.

Hollander and Wolfe. Nonparametric Statistical Methods. In-depth explanation of each of these and more.

2 Exact Binomial Test for a Population Proportion π

Situation: You observe n (fixed) number of trials with only 2 possible outcomes per trial.

Hypotheses: Same forms as for a population proportion z -test

Necessary Assumptions: (Combined with n fixed these are the conditions for a binomial)

1. The trials must be independent.
2. The two outcomes can be considered successes and failures.
3. Probability of success is constant for all trials.

Test Procedure: Let B be the number of successes in your sample. B is used as the test statistic. The p -value is computed based on the Binomial distribution as the probability of observing B successes or more extreme in the direction of your alternative for a given probability of success in the null hypothesis, denoted π_0 .

Function in R: `binom.test(x, n, p = 0.5, alternative = c("two.sided", "less", "greater"))`
 x is the number of successes, n is the number of trials, p = hypothesized probability of success (can be changed) = π_0 , and pick the appropriate alternative

Benefits: This allows you to do an exact binomial test for small samples where $n\pi_0$ and $n(1 - \pi_0)$ are not both ≥ 10 . You can also do an exact binomial test for larger samples and see how close the z -test approximation is to the exact p -value.

Example: You want to test whether or not a coin is fair by flipping it. Imagine you only have time for 10 flips and 8 come up heads. Is this enough information to conclude that the coin is not fair?

```
binom.test(8,10,p=.5,alternative="two.sided")
Exact binomial test
data: 8 and 10
number of successes = 8, number of trials = 10, p-value = 0.1094
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
0.4439045 0.9747893
sample estimates:
probability of success
0.8
```

3 Sign Test for a Population Median

Can use as paired test on the *differences*

Situation: You observe n observations from a population or $2n$ observations which are paired

leaving you with n differences after the differences have been computed.

Hypotheses: Same forms as for μ (or μ_d) except μ is replaced by median - θ (or θ_d).

Necessary Assumptions:

1. The n observations or n differences are independent.
2. The observations or differences come from a continuous population which has a median θ and which is symmetric.

Test Procedure: Compute the differences as observations minus hypothesized value or as the differences (given a specific subtraction order) if paired. Count the number of positive differences - that is the test statistic. An equal number of positive and negative differences is expected if the null hypothesis is true. Compute the p-value as the probability you observe your number of positive differences or more extreme using binomial probabilities.

Function in R: No direct function, but can be done using binomial probabilities. Adjustments do need to be made for upper/lower/two-tailed tests.

Benefits/Issues: Simplistic idea. Not used often because it has low relative efficiency compared to parametric test. You lose all information about the magnitude of the differences, which can be very important. The Signed-Rank test is a better choice.

For example, for the sign test, both of the following data sets would have the same test statistic of 3. But you lose all information about spread.

Data set 1: -1, -.5, -.25, 1, 2, 3

Data set 2: -1, -.5, -.25, 2, 6, 100

4 Wilcoxon Signed-Rank Test for a Population Median

Can use as paired test on the *differences*

Situation: You observe n observations from a population or $2n$ observations which are paired leaving you with n differences after the differences have been computed.

Hypotheses: Same forms as for μ (or μ_d) except μ is replaced by median - θ (or θ_d).

Necessary Assumptions:

1. The n observations or n differences are independent.
2. The observations or differences come from a continuous population which has a median θ and which is symmetric.

Test Procedure: If dealing with observations, the test computes observations minus hypothesized θ to get differences. Starting with the differences (if the median really is θ , these differences should be symmetric around 0), compute their absolute values. Rank each difference by absolute value

(small to large) and add up the ranks of the differences which are positive as your test statistic. Example: If your observations are 4, 14, 19, 30, 59 and θ is hypothesized to be 25, the test first computes the differences 4-25, 14-25, 19-25, 30-25, 59-25. Those differences are -21, -11, -6, 5, 34. The ranked absolute values of the differences are 4,3,2,1, and 5 respectively. The sums of the ranks for the positive differences are 1+5=6. The test then finds the probability of getting that test statistic or something more extreme.

Function in R: `wilcox.test(x,y,paired=F or T,alternative="",mu=hypothesized value (default 0))`
If doing a one-sample test for μ , you would run `wilcox.test(x,mu=hyp value, alternative = "two.sided")` and fill in the value for hypothesized value and change alternative to match what you have.
If doing a one-sample test for μ_d , you would run `wilcox.test(x,y,paired=T,alternative = "two.sided")` and change the alternative to match what you have. x and y are the before/after scores or whatever that you want the differences computed between.

Benefits: Does not lose all information about magnitude of differences. Has high relative efficiency compared to parametric test.

Example: The paired example we did in class on spending amounts comparing actual and budgeted amounts did not satisfy the normal distribution assumption for the population. In that situation, you might try to do a nonparametric test due to assumption worries (even with $n=42$). Below is another example paired test where the qq-plot looked very bad, but also had a very small sample size, $n=15$. For this example, the p-value in the parametric test was .04934, but there are major concerns about assumptions being violated.

```
x=firm, y=audit
wilcox.test(x, y, paired = TRUE, alternative = "greater")
Wilcoxon signed rank test with continuity correction
data: x and y
V = 97.5, p-value = 0.01771
alternative hypothesis: true location shift is greater than 0
Warning message:
In wilcox.test.default(x, y, paired = TRUE, alternative = "greater") :
cannot compute exact p-value with ties
```

There are some ties with ranks that appear here, and the output lets you know that the p-values are therefore approximate.

5 Wilcoxon Rank-Sum Test for a difference in two Population Means

Can also be done using the Mann-Whitney U statistic

Situation: You obtain $n+m$ observations from two groups and you want to compare their medians.

Hypotheses: This is a little different than anything we have seen. In probability, there is something called a distribution function and this test tests for equality of the distribution functions

explicitly. However, it is equivalent to say that it tests for the equality of population medians instead of means (which was the parametric test), with hypotheses now in terms of medians as $\theta_1 - \theta_2$.

Necessary Assumptions:

1. The samples from each populations were random samples.
2. The samples were independent.
3. Both populations are continuous.

Test Procedure: Sort the combined observations and compute their ranks. The test statistic is the sum of ranks of observations from the second sample. Compute p-value accordingly. Note that the expected value of the test statistic here depends on the sample sizes - you would expect different values depending on how many observations were in each sample. This is less intuitive than expecting a zero difference like in the sign test.

Function in R: `wilcox.test(x,y,alternative = “”)`
x and y are the two sets of observations (do not need to be same length), and set alternative as appropriate.

Benefits: Does not require parametric assumptions. Keeps information about magnitude by using combined rank values. Has high relative efficiency compared to parametric test.

Example: Salamander data:

Assume instead of 35 male and 42 female salamanders as before we only had 4 males and 6 females with length measurements as below. We perform a test to see if females are longer than males based on medians (note, putting females first means $>$ in the alternative).

```
females=c(59.66,71.64,61.38,67.96,63.56,64.23)
```

```
males=c(62.49,59.04,58.49,62.39)
```

```
wilcox.test(females,males,alternative=“greater”)
```

Wilcoxon rank sum test

data: females and males

W = 20, p-value = 0.05714

alternative hypothesis: true location shift is greater than 0

Here is the parametric test on just these 10 combined observations, just so you can see the difference:

```
t.test(females,males,alternative=“greater”)
```

Welch Two Sample t-test

data: females and males

t = 1.9801, df = 7.579, p-value = 0.04252

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

0.2236468 Inf

sample estimates:

mean of x mean of y

64.73833 60.60250

6 Nonparametric ANOVA - Kruskal-Wallis for General Alternatives

Situation: You have n total observations from k populations (i.e. k samples) and you want to compare their medians. (It's actually a bit more general than that but that is enough detail to get the gist of it).

Hypotheses: Same forms as ANOVA, just with μ_i replaced by θ_i for medians.

Necessary Assumptions:

1. The k samples were independent.
2. Each sample was a random sample.
3. The populations are related by at most a shift in distribution function (this means the populations differ at most by their medians but other characteristics are the same).

Test Procedure: Combine all samples. Rank all observations. Compute the sum of ranks for each sample. Compute average rank of each sample. Combine that information into a test statistic and compute p-value.

Function in R: For an example with $k = 3$, let x, y, z be the list of observations in each sample. Then run

```
kruskal.test(list(x,y,z))
```

No information about alternatives is required since this is like an ANOVA.

Benefits: You don't need the requirement that all populations are normally distributed, though you still need spreads to be somewhat similar. More advanced versions for directed alternatives (called umbrella hypotheses) exist. Has pretty high relative efficiency compared to ANOVA, but less efficient than either of the Wilcoxon methods compared to their parametric counterparts.

Example: Flammability study from class

Five labs are being compared based on a clothing flammability test. Each lab tested 11 samples of the same fabric.

```
lab1=c(2.9,3.1,3.1,3.7,3.1,4.2,3.7,3.9,3.1,3.0,2.9)
```

```
lab2=c(2.7,3.4,3.6,3.2,4.0,4.1,3.8,3.8,4.3,3.4,3.3)
```

```
lab3=c(3.3,3.3,3.5,3.5,2.8,2.8,3.2,2.8,3.8,3.5,3.8)
```

```
lab4=c(3.3,3.2,3.4,2.7,2.7,3.3,2.9,3.2,2.9,2.6,2.8)
```

```
lab5=c(4.1,4.1,3.7,4.2,3.1,3.5,2.8,3.5,3.7,3.5,3.9)
```

```
kruskal.test(list(lab1,lab2,lab3,lab4,lab5))
```

Kruskal-Wallis rank sum test

```
data: list(lab1, lab2, lab3, lab4, lab5)
```

Kruskal-Wallis chi-squared = 14.2622, df = 4, p-value = 0.006504