Chapter 6: Nonparametric procedures

If you decide that the assumptions of normality are not sufficiently met for doing a t-test or ANOVA, then what do you do?

One possibility is to use *nonparametric* procedures. The word *nonparametric* is in contrast to *parametric* procedures, where it is assumed that the data come from a family of distributions (such as the normal) which is *parameterized* by a small number of parameters.

For example, normal distributions are parameterized by the mean μ and variance σ^2. The t family of distributions is parameterized by the degrees of freedom, sometimes denoted ν. (It is common, but not universal, to use Greek letters for the parameters of a distribution.)

Nonparametric procedures

Nonparametric procedures make fewer assumptions about the distribution of the data than do parameteric procedures. The *t*-tests and ANOVA procedures are examples of parametric procedures. Probability statements such as *p*-values, and the width of confidence intervals based on these procedures assume a very specific family of distributions (normal distributions) for the underlying data.

Nonparametric procedures still make assumtions about the data, usually especially that each observation is independently sampled. Nonparametric procedures often make weaker assumptions than parametric procedures.

For example, some (not all) nonparametric procedures assume that the data come from a symmetric distribution, but that distribution is not assumed to be normal. If the distribution does happen to be normal, then the procudure would still be valid.

Nonparametric procedures

If the normality assumption is reasonable for *t*-tests and ANOVA, and if the equal variances assumption is reasonable for ANOVA, there is no need to use nonparametric procedures. However, if these assumptions seem questionable, then it is reasonable to consider nonparametric alternatives.

If nonparametric procedures are used when the assumptions of *t*-tests or ANOVA are met, then it is likely that any evidence against the null hypothesis (i.e., the p-value) would be weakened. Another way of saying this is that *t*-tests and ANOVA tend to be more *powerful* (higher probability of rejecting the null when the null is false) then nonparametric procedures when the assumptions of the procedures are met. We will explore this idea using simulation after introducing some of the methods.

Nonparametric procedures: Sign test

The sign test is a test of the hypothesis that a *median* of a population is equal to a certain value. The sign test a nonparametric alternative to the one-sample t-test.

Let η (pronounced *Ay-duh*) denote the population median. The null hypothesis can be written:

$$H_0 : \eta = \eta_0$$

If the null hypothesis is true, then approximately half of the observations should be above η_0 and half should be below η_0.

The alternative hypothesis can be based on either a two-sided or one-sided test, so we could have

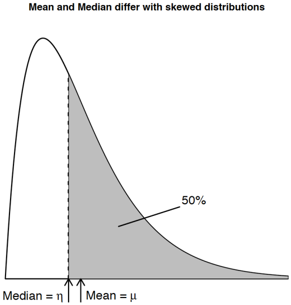
$$\begin{aligned} H_A :& \nu \neq \nu_0 \\ H_A :& \nu < \nu_0, \text{ or} \\ H_A :& \nu > \nu_0 \end{aligned}$$

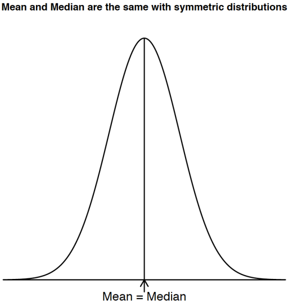
Nonparametric procedures: Sign test

If the distribution is symmetric (such as for the normal), then the population median is equal to the population median, so the statement that the population median is a certain value is equivalent to the statment that the population mean is that value as well.

However, the test also works for distributions that are skewed, so that the population median is different from the population mean.

Nonparametric procedures: Sign test


Mean and Median differ with skewed distributions


Mean and Median are the same with symmetric distributions

Nonparametric procedures: Sign test

For hypothesis testing, the usual procedure is:

- to construct a test statistic based on the data (e.g., t_{obs} or F)
- determine the distribution of the test statistic under the null hypothesis
- quantify how consistent the data are with the null hypothesis (get a p-value)
- make a decision based on the test statistic or p-value

This general approach to hypothesis testing works for *many* different cases, including t-tests, ANOVA and here the sign test.

For the sign test, the test statistic is S, the number of observations larger than η_0, the hypothesized median.

Nonparametric procedures: Sign test

Once you have determined S, you need to find a distribution that S should follow under the null hypothesis. If the null hypothesis is correct, then each observation has a 50% chance to be either above or below η_0.

The procedure is similar to flipping a coin for each observation. With probability 50%, you get heads (the value is above ν_0), and with probability 50%, you get tails (the value is below ν_0.

The right distribution for describing this is well known in probability and is called the *binomial distirbution*. This distribution describes the probability of getting k successes in n trials, where each trial is independent and has probability p of success. For this application, $p = 1/2$.

Nonparametric procedures: Sign test

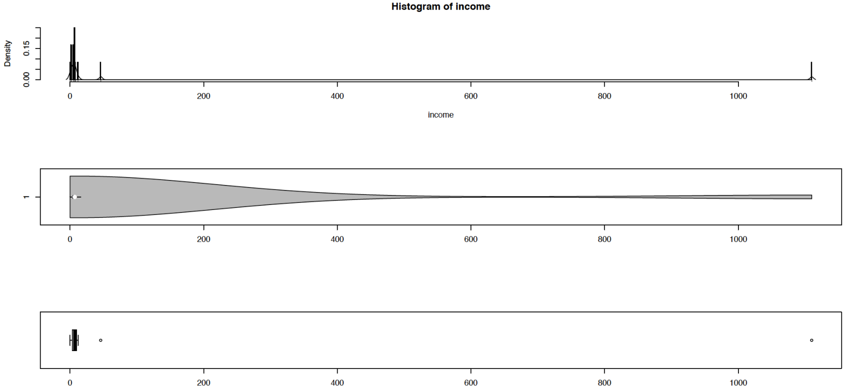
Rather than calculating the binomial probabilities yourself, you can use the function SIGN.test() in the BSDA package in R. The following is an example with an extreme outlier:

```
#### Example: Income Data
income <- c(7, 1110, 7, 5, 8, 12, 0, 5, 2, 2, 46, 7)
# sort in decreasing order
income <- sort(income, decreasing = TRUE)
income
## [1] 1110 46 12 8 7 7 7 5 5 2 2 0
summary(income)
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.00 4.25 7.00 100.90 9.00 1110.00
sd(income)
## [1] 318.0078
```

Nonparametric procedures: Sign test

```
par(mfrow=c(3,1))
# Histogram overlaid with kernel density curve
hist(income, freq = FALSE, breaks = 1000)
points(density(income), type = "l")
rug(income)
# violin plot
library(vioplot)
vioplot(income, horizontal=TRUE, col="gray")
# boxplot
boxplot(income, horizontal=TRUE)
```

Nonparametric procedures: Sign test



Nonparametric procedures: Sign test

Also try doing qqnorm() to see what the QQ-plot looks like (I'll let you do this on your own). Notice the extreme outlier.

A *t*-distribution based CI for this data is unreasonable since it includes negative values

```
income <- c(7, 1110, 7, 5, 8, 12, 0, 5, 2, 2, 46, 7)
t.test(income)

One Sample t-test

data:  income
t = 1.0993, df = 11, p-value = 0.2951
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -101.1359  302.9692
sample estimates:
mean of x
 100.9167
```

Nonparametric procedures: Sign test

Instead let's try a sign test. The sign test will automatically compute the median for you.

```
library(BSDA)
SIGN.test(income)
s = 11, p-value = 0.0009766
alternative hypothesis: true median is not equal to 0
95 percent confidence interval:
  2.319091 11.574545
sample estimates:
median of x
          7
Achieved and Interpolated Confidence Intervals:
                  Conf.Level L.E.pt  U.E.pt
Lower Achieved CI     0.8540 5.0000  8.0000
Interpolated CI       0.9500 2.3191 11.5745
Upper Achieved CI     0.9614 2.0000 12.0000
```

Nonparametric procedures: Sign test

Note that the confidence interval here are for the population median, not the population mean. This is slightly different than for the *t*-test.

Also, because the exact distribution of incomes is not known, can only be computed with a 95% confidence level if assumptions are made about the data, which R describes as interpolation. Otherwise, the CI can include values in the data, but the confidence level will not be exactly 95%. It therefore outputs a range of CIs for you to choose from. For example, you are approximately 96% confident that the population median income is between $2,000 and $12,000.

Nonparametric procedures: Sign test

It might have occurred to you that since the original data had one extreme outlier, we could have analyzed the data by removing that outlier and then analyzing the remaining data using the usual t-test approach.

The advantage for this approach is that we use the more common t statistics, which can be more powerful and (often) lead to narrower confidence intervals.

For this data, even removing the observation of 1110 leads to a second outlier of 46. Potentially you could remove this outlier as well. But remember that in inferential statistics we are making inferences about a population from which we sampled. If we remove observations that are genuine (not due to typos, incorrectly copied data, etc.), what population are making inferences about? For incomes, we seem to be making inferences about the population of incomes that are not extremely high, rather than the general population of incomes, which includes some genuinely high values.

Nonparametric procedures: Sign test

```
> shapiro.test(income)

Shapiro-Wilk normality test

data:  income
W = 0.35148, p-value = 1.718e-06

> shapiro.test(income[income<100])

data:  income[income < 100]
W = 0.59454, p-value = 2.175e-05

> shapiro.test(income[income<46])


data:  income[income < 46]
W = 0.95189, p-value = 0.6909
```

Nonparametric procedures: Sign test

To illustrate the sensitive of the *t* based confidence intervals (and p-values) to the outliers compare what happens to the *t*-tests versus signed rank tests as the extreme observations are made smaller (but still larger than other observations).

```
 income
# [1]    7 1110    7    5    8   12    0    5    2    2   46    7
income2[2] <- 110
income2[11] <- 16
income2
# [1]   7 110   7   5   8  12   0   5   2   2  16   7
income3 <- income2
income3[2] <- 17
income3
 [1]  7 17  7  5  8 12  0  5  2  2 16  7
```

Nonparametric procedures: Sign test

Sensitivity of *t*-test to outliers:

```
 t.test(income)$conf.int
#[1] -101.1359  302.9692
t.test(income2)$conf.int
#[1] -4.111024 34.277691
 t.test(income3)$conf.int
#[1]  3.945899 10.720768
t.test(income)$p.value
#[1] 0.295115
t.test(income2)$p.value
#[1] 0.1116271
t.test(income3)$p.value
#[1] 0.0005855308
```

Nonparametric procedures: Sign test

Robustness of sign test to outliers:

```
SIGN.test(income)$conf.int
#[1]  2.319091 11.574545
SIGN.test(income2)$conf.int
#[1]  2.319091 11.574545
SIGN.test(income3)$conf.int
#[1]  2.319091 11.574545
```

Nonparametric procedures: Sign test

We could also look at what happens if the outliers are removed. Again, the sign test is less sensitive than the *t*-test:

```
SIGN.test(income)
#s = 11, p-value = 0.0009766
#95 percent confidence interval: 2.319091 11.574545
SIGN.test(income[income<40])
#s = 9, p-value = 0.003906
#95 percent confidence interval: 2.000000 7.675556
t.test(income)
#t = 1.0993, df = 11, p-value = 0.2951
#95 percent confidence interval: -101.1359  302.9692
t.test(income[income<40])
t = 4.9637, df = 9, p-value = 0.0007766
#95 percent confidence interval: 2.993414 8.006586
```

Nonparametric procedures: Rank-sum test

An alternative to the sign test is the Wilcoxon signed rank test. In this nonparametric procedure, it is assumed that the underlying distirbution is symmetric, but not necessarily normal. It makes stronger assumpetions than the sign test, but not as strong as the *t*-test.

Here the null is that $\mu = \mu_0$ where μ is equivalently the mean or median. You compute both the signs of $X_i - \mu_0$ and the ranks of $|X_i - \mu_0|$ for each data point. By ranks, we mean that the largest deviation $|X_i - \mu_0|$ gets rank n, where n is the sample size, and the smallest deviation $|X_i - \mu_0|$ gets rank 1.

Nonparametric procedures: Rank-sum test

Example where we test $H_0 : \mu = 10$.

X_i	$X_i - 10$	sign	$\lvert X_i - 10 \rvert$	rank	rank × sign
20	10	$+$	10	6	6
18	8	$+$	8	4.5	4:5
23	13	$+$	13	8	8
5	-5	$-$	5	3	-3
14	4	$+$	4	2	2
8	-2	$-$	4	2	-1
18	8	$+$	8	4.5	4.5
22	12	$+$	12	7	7

Note that for the tied observations, these would have ranks 4 and 5, so we give them each the average of 4 and 5. Gernally, if k observations are tied for rank r, give them each rank $((r+0)+(r+1)+\cdots+(r+k-1))/k = r + (k-1)/2$.

Nonparametric procedures: Rank-sum test

The test statistic is $W =$ the sum of the positive signed ranks. For the above example

$$W = 6 + 4.5 + 8 + 2 + 4.5 + 7 = 32$$

Note that the sum of the unsigned ranks is

$$1 + 2 + \cdots + n = n(n+1)/2$$

where n is the sample size. For this example, $n = 8$, so

$$n(n+1)/2 = (8)(9)/2 = 36$$

If half of the observations are above μ_0, then you expect half of the observations to contribute to the W statistic, and the expected value of W is $(1/2) \times n(n+1)/2 = n(n+1)/4 = 18$ for this example. The question then is whether 32 is significantly different from 18. This depends on the distribution of W.

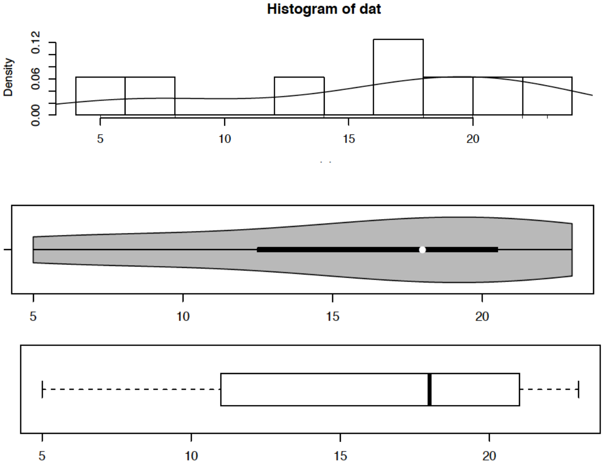
Nonparametric procedures: Rank-sum test

```
#### Example: Made-up Data
dat <- c(20, 18, 23, 5, 14, 8, 18, 22)
# sort in decreasing order
dat <- sort(dat, decreasing = TRUE)
dat
## [1] 23 22 20 18 18 14 8 5
summary(dat)
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 5.0 12.5 18.0 16.0 20.5 23.0
sd(dat)
## [1] 6.524678
```

Nonparametric procedures: Rank-sum test

```
par(mfrow=c(3,1))
# Histogram overlaid with kernel density curve
hist(dat, freq = FALSE, breaks = 10)
points(density(dat), type = "l")
rug(dat)
# violin plot
library(vioplot)
vioplot(dat, horizontal=TRUE, col="gray")
# boxplot
boxplot(dat, horizontal=TRUE)
```

Nonparametric procedures: Sign test



Nonparametric procedures: Rank-sum test

QQplot and Shapiro-Wilk test do not suggest evidence against normality. There do not appear to be outliers, the distribution is unimodal, and the there does not appear strong skew (is is slightly left-skewed). So I would be comfortable using a *t*-test for this data. Nevertheless, we illustrate using both the *t*-test and signed rank test.

Nonparametric procedures: Rank-sum test

```
t.test(dat, mu=10)
##
## One Sample t-test
##
## data: dat
## t = 2.601, df = 7, p-value = 0.03537
## alternative hypothesis: true mean is not equal to 10
## 95 percent confidence interval:
## 10.54523 21.45477
## sample estimates:
## mean of x
## 16
```

Nonparametric procedures: Rank-sum test

```
wilcox.test(dat, mu=10, conf.int=TRUE)
## Warning in wilcox.test.default(dat, mu = 10, conf.int = TRUE):
cannot compute exact p-value with ties
## Wilcoxon signed rank test with continuity correction
##
## V = 32, p-value = 0.0584
## 95 percent confidence interval:
## 9.500002 21.499942
## (pseudo)median
## 16.0056
# without continuity correction
wilcox.test(dat, mu=10, conf.int=TRUE, correct=FALSE)
## V = 32, p-value = 0.04967
## alternative hypothesis: true location is not equal to 10
## 95 percent confidence interval:
## 10.99996 21.00005
## (pseudo)median
## 16.0056
```

Nonparametric procedures: Rank-sum test

Note that the p-value is slightly different depending on whether a continuity correction is used or not. Continuity corrections are often used for discrete tests such as this one and the chi-square test (which we haven't covered), particularly when p-values are based on normal approximations.

When there are ties in the ranks or the sample size is large (50 or above), R uses normal approximations. The idea is that $W/SE(W)$ is approximately normally distributed, so this quantity acts like a z-score. The test is still considered nonparametric even when a normal approximation for the distribution of W is used.

The idea is that W is approximately normally distributed (due to the Central Limit Theorem) even if the underlying data isn't. Of course, you might argue that if W is approximately normally distributed, then so is $\overline{X}$. This is likely true as long as there are no extreme outliers.

Nonparametric procedures: Rank-sum test

You might be tempted to ask, is the correct p-value really below .05 or not? Scientifically, however, p-values of 0.0584 and 0.0497 are quite close. They indicate similar amounts of evidence against the null hypothesis.

This just happens to be a case where a slight difference in method leads to a different conclusion if you are using 0.05 as a cutoff. This is an example where people might argue that paying too much attention to p-values is a bad thing. In the end, I would prefer using the continuity correction since it is the default method and tends to lead to better performance for discrete methods.

Nonparametric procedures: Rank-sum test

If we apply the Wilcoxon test to the income data we would get the same conclusion, although the p-value is somewhat larger.

```
wilcox.test(income)
#
# Wilcoxon signed rank test with continuity correction
#
#data:  income
#V = 66, p-value = 0.003753
#alternative hypothesis: true location is not equal to 0
```

Nonparametric procedures: paired data

Just like for the *t*-test, you might have two paired samples (e.g., pre- vs post scores) or you might have two independent samples. For paired data, just like with the matched pairs *t*-test, you can analyze the differences as a single sample rather than think of it as a two-sample problem. This works for both the sign test and the Wilcoxon rank-sum test. For paired data, you will usually be interested in testing $H_0 : \nu = 0$ or $H_0 : \mu = 0$.