

# Permutation Tests

UGARTE, M.D.(\*)

(\*)DEPARTAMENTO DE ESTADÍSTICA E I. O., UNIVERSIDAD PÚBLICA DE NAVARRA, PAMPLONA, SPAIN

E-MAIL: LOLA@UNAVARRA.ES

---

**Material from the book *Probability and Statistics with R*  
by Ugarte, Militino, and Arnholt. Chapman and Hall/CRC, 2008**

# Introduction

**Permutation tests are computationally intensive techniques that actually pre-date computers.**

Until recently, permutation tests were more of a theoretical ideal than a useful technique. With the advent of high-powered computers, permutation tests have moved out of the abstract into the world of the practical.

The permutation test is examined in only one context here: **the two-sample problem.**

## Introduction -Cont.

**The fundamental idea** behind the permutation test is that if there are no differences between two treatments, all data sets obtained by randomly assigning the data to the two treatments have an equal chance of being observed.

Permutation tests are especially advantageous when working with small samples where verification of assumptions required for tests such as the pooled *t*-test are difficult.

## Introduction -Cont.

To test a hypothesis with a permutation test:

1. Choose a test statistic  $\hat{\theta}$  that measures the effect under study. Note that certain statistics will have more power to detect the effect of interest than others.
2. Create the sampling distribution that the test statistic in Step 1 would have if the effect is not present in the population (the null hypothesis is true).

3. Find the “observed test statistic” in the sampling distribution from Step 2.

Observed values in the extremes of the sampling distribution suggest that the effect under study is “real”. In contrast, observed values in the main body of the sampling distribution imply that the effect is likely to have occurred by chance.

4. Calculate the p-value based on the observed test statistic. This may be

$$P(|\hat{\theta}| \geq |\hat{\theta}_{obs}|), \quad \text{in a two-sided test}$$

$$P(\hat{\theta} \leq \hat{\theta}_{obs}), \quad \text{in a left one-sided test}$$

$$P(\hat{\theta} \geq \hat{\theta}_{obs}), \quad \text{in a right one-sided test}$$

# The two-sample problem

Suppose two independent random samples  $\mathbf{z} = \{z_1, \dots, z_n\}$  and  $\mathbf{y} = \{y_1, \dots, y_m\}$  are drawn from possibly different probability distributions  $F$  and  $G$ .

**The question of interest is whether  $F = G$ .**

Assuming  $H_0 : F = G$  is true, it is possible to create the permutation sampling distribution for some statistic of interest,  $\hat{\theta}$ .

Let  $N$  equal the combined sample size  $n + m$ .

# The two-sample problem

- Let  $\mathbf{v} = \{v_1, v_2, \dots, v_N\}$ , be the combined and ordered vector of values, (the first  $n$  values correspond to the first sample and the rest to the second).
- Let  $\mathbf{g} = \{g_1, g_2, \dots, g_N\}$  be a vector indicating the group membership for  $\mathbf{v}$ . Since there are  $n$   $z_i$ 's and  $m$   $y_j$ 's in  $\mathbf{g}$ , there are  $\binom{N}{n}$  possible ways of partitioning  $N$  elements into two subsets of sizes  $n$  and  $m$ .
- Consequently, under the null hypothesis that  $F = G$ , the vector  $\mathbf{g}$  has probability  $1/\binom{N}{n}$  of equalling any one of its possible values. That is, all combinations of  $z_i$ 's and  $y_j$ 's are equally likely if  $F = G$ .

# The two-sample problem

Suppose  $H_A : F > G$  and  $\hat{\theta} = \bar{z} - \bar{y}$ , then **the exact p-value** is found by finding

$$\frac{\#\{\hat{\theta} \geq \hat{\theta}_{obs}\}}{\binom{N}{n}}$$

For all but relatively trivial sized samples  $n$  and  $m$ , the number  $\binom{N}{n}$  will be huge, **making the enumeration of all possible samples of the statistic of interest a monumental task.**

For example, consider that if  $n = 10$  and  $m = 10$  that complete enumeration would require listing  $\binom{20}{10} = 184,756$  possible outcomes.



# The two-sample problem

Consequently, an approximation to the exact p-value is often obtained by re-sampling **WITHOUT** replacement the original data some “large” number of times,  $B$ , which is usually at least 999, and approximating the p-value with

$$\text{p-value} \doteq \frac{[1 + \#\{\hat{\theta} \geq \hat{\theta}_{obs}\}]}{(B + 1)} = \frac{[1 + \sum_{i=1}^B \mathbf{I}\{\hat{\theta} \geq \hat{\theta}_{obs}\}]}{(B + 1)}$$

When the sampling is done with replacement, a bootstrap test is performed. Bootstrap tests are somewhat more general than permutation tests since they apply to a wider class of problems; however, they do not return “exact” p-values.

## Example (Original data: Ott and Mendenhall, 1985)

The data for this problem will be stored in a data frame called **Ratbp**. Researchers wanted to know whether a drug was able to reduce blood pressure of rats. Twelve rats were chosen and the drug was administered to six rats, the treatment group, chosen at random. The other six rats, the control group, received a placebo. The drops in blood pressure (mmHg) for the treatment group (with probability distribution  $F$ ) and the control group (with probability distribution  $G$ ) are stored in the variables **Treat** ( $z$ ) and **Cont** ( $y$ ), respectively. Note that positive numbers indicate blood pressure decreased while negative numbers indicate that it rose.

## Example

Under the null hypothesis,  $H_0 : F = G$ , the data come from a single population (i.e., we are not able to prove the drug efficacy). The question of interest is: How likely are differences as extreme as those observed between the treatment and control groups to be seen if the null hypothesis is correct? Use  $\hat{\theta} = \bar{z} - \bar{y}$  as the statistic of interest and compute

- (a) the exact permutation p-value
- (b) an estimated permutation p-value based on 499 permutation replications
- (c) an estimated bootstrap p-value based on 499 bootstrap replications

## Solution (a) Exact permutation p-value

The test statistic of interest (Step 1) has been specified to be  $\hat{\theta} = \bar{z} - \bar{y}$ . Finding the *p-value* requires creation of sampling distributions with different methods.

First, create **Ratbp** data frame.

Combine the treatment (**Treat**) and control (**Cont**) data in a single variable called **Blood**.

```
## To introduce the data in R
Treat<-c(69,24,63,87.5,77.5,40)
Cont<-c(9,12,36,77.5,-7.5,32.5)
Ratbp<-data.frame(Treat, Cont) # to create a data frame
> Ratbp
  Treat Cont
1  69.0  9.0
2  24.0 12.0
3  63.0 36.0
4  87.5 77.5
5  77.5 -7.5
6  40.0 32.5
Blood<-c(Treat,Cont)
> Blood
[1] 69.0 24.0 63.0 87.5 77.5 40.0  9.0 12.0 36.0 77.5 -7.5 32.5
```

## Solution (a) Exact permutation p-value

- We compute all of the possible combinations of the indices of **Blood**:  $\binom{12}{6} = 924$  and put them in a vector called **pdT6**.
- To do that we may install in R the package BSDA. Alternatively, we may load the function **Combinations.R**.
- Now we construct a matrix called **Comb** of dimension  $924 \times 12$  where each row contains the values of **Blood**.
- We also construct a vector **Theta** with all of the possible values of  $\hat{\theta} = \bar{z} - \bar{y}$ , i.e., a value for each possible combination.

```

library(BSDA)
pdT6<-t(Combinations(12,6))
#Let us show the first three rows of pdT6:
pdT6[1:3,]
> pdT6[1:3,]
      N N N N N
[1,] 1 2 3 4 5 6
[2,] 1 2 3 4 5 7
[3,] 1 2 3 4 6 7
## Now we construct the matrix with 924 rows -all the same-
Comb<-matrix(rep(c(Treat,Cont),924),ncol=12,byrow=T)
Comb[1:3,]          #three first rows of Comb
> Comb[1:3,]
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]   69   24   63 87.5 77.5   40    9   12   36  77.5  -7.5  32.5
[2,]   69   24   63 87.5 77.5   40    9   12   36  77.5  -7.5  32.5
[3,]   69   24   63 87.5 77.5   40    9   12   36  77.5  -7.5  32.5

```



```
# This is for understanding what we are doing

Comb[2,pdT6[2,]]

#we extract from row 2 of matrix Comb the columns
#specified by the second row of matrix pdT6

> pdT6[2,]
[1] 1 2 3 4 5 7

# In words, we take values of row 2 of Comb
#corresponding to columns 1, 2, 3, 4, 5 y 7

> Comb[2,pdT6[2,]]
[1] 69.0 24.0 63.0 87.5 77.5 9.0

#Could you imagine the result of the following command?
Comb[2, -pdT6[2,]]
```



RESULT:

```
Comb[2, -pdT6[2,]]
```

```
> Comb[2, -pdT6[2,]]
```

```
[1] 40.0 12.0 36.0 77.5 -7.5 32.5
```

```
#Note that it gives the data from row 2 that DO NOT correspond  
#to columns 1, 2, 3, 4, 5 y 7
```

Let's remember the contents of Comb[2,]

```
> Comb[2,]
```

```
[1] 69.0 24.0 63.0 87.5 77.5 40.0 9.0 12.0 36.0 77.5 -7.5 32.5
```

Let's create now the vector **Theta** that contains all of the possible values of  $\hat{\theta} = \bar{z} - \bar{y}$ .

```
Theta<-array(0,924)

for (i in 1:924) {
Theta[i]<-mean(Comb[i,pdT6[i,]])-mean(Comb[i, -pdT6[i,]])}

Theta.obs<-mean(Treat)-mean(Cont)
pval<-sum(Theta >=Theta.obs)/choose(12,6) # exact permutation p-value

> pval
[1] 0.03138528
```

## Solution (a) exact permutation p-values using the package coin

```
library(coin)
```

```
GR<-as.factor(c(rep("Treat",6), rep("Cont",6)))
```

```
oneway_test(Blood~GR, distribution="exact",alternative="less")
```

```
Exact 2-Sample Permutation Test
```

```
data: Blood by groups Cont, Treat
```

```
Z = -1.871, p-value = 0.03139
```

```
alternative hypothesis: true mu is less than 0
```

## Example (Solution (b)). Estimated permutation p-value

To compute the estimated permutation p-value we draw randomly and without replacement a certain amount of samples. From a total of 924 permutations we draw randomly  $B = 499$ .

```
set.seed(13)
theta.obs<-mean(Ratbp$Treat)-mean(Ratbp$Cont)
boot.blood<-sample(1:924,size=499, replace=F)
B<-499
theta<-array(0,B)
for (i in 1:length(boot.blood)){
  j<-boot.blood[i]
  media1<-mean(Comb[j,pdT6[j,]])
  media2<-mean(Comb[j,-pdT6[j,]])
  theta[i]<-media1-media2}
pval.aprox<-(sum(theta >=theta.obs)+1)/(B+1) ##pval.aprox

pval.aprox
0.032
```

## Example (Solution (b)). -Using package **boot-**

- For that we need to load package **boot**
- Let us construct a function called **blood.fun()** that computes the difference among the six first values and the six last values of a vector of length 12.
- To obtain samples without replacement, we need to add the argument **sim="permutation"** in the function **boot()**

Let us see the **R** solution:

```
library(boot)

set.seed(13)
B<-499

blood.fun<- function(datos,i){
  d<-datos[i]
  MD<-mean(d[1:6])-mean(d[7:12])
  MD}

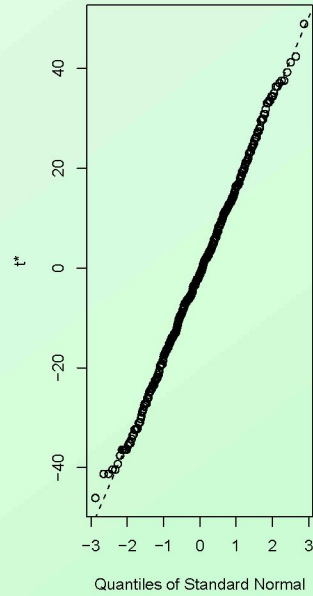
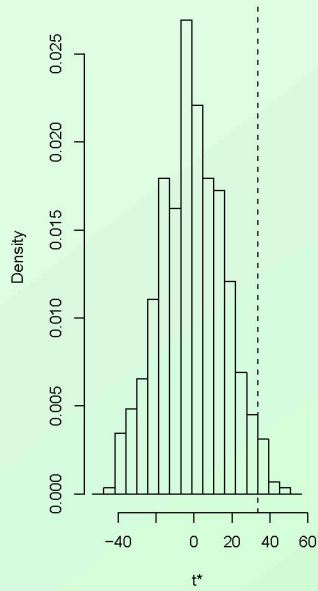
boot.blood.b<-boot(Blood,blood.fun, R=B, sim="permutation")

plot(boot.blood.b) #This graphs the sampling
#distribution of the statistic

pval.boot.b<-(sum(boot.blood.b$t >= boot.blood.b$t0)+1)/(B+1)

> pval.boot.b
[1] 0.03
```

Histogram of  $t$



## Example (Solución (c)). Estimated bootstrap p-value using boot

If we select the 499 samples with replacement, we are doing a **BOOTSTRAP TEST**

- Load package **boot**
- Let us construct a function called **blood.fun()** that computes the difference among the six first values and the six last values of a vector of length 12.
- To obtain samples with replacement, we need to add the argument **sim="ordinary"** in the function **boot()**

Let us see the **R** solution:



```
library(boot)

set.seed(13)
B<-499

blood.fun<- function(datos,i){
  d<-datos[i]
  MD<-mean(d[1:6])-mean(d[7:12])
  MD}

boot.blood.b<-boot(Blood,blood.fun, R=B, sim="ordinary")

plot(boot.blood.b) #Podemos representar la distribuci\`on
#en el muestreo del estad\`istico

pval.boot.b<-(sum(boot.blood.b$t >= boot.blood.b$t0)+1)/(B+1)

> pval.boot.b
[1] 0.038
```

## Example (Solution(c)). Estimated bootstrap p-value without using boot

```
set.seed(13)
theta.obs<-mean(Ratbp$Treat)-mean(Ratbp$Cont)

boot.blood<-sample(1:924,size=499, replace=T)

B<-499
theta<-array(0,B)
for (i in 1:length(boot.blood)){
  j<-boot.blood[i]
  media1<-mean(Comb[j,pdT6[j,]])
  media2<-mean(Comb[j,-pdT6[j,]])
  theta[i]<-media1-media2}
pval.boot<-(sum(theta >=theta.obs)+1)/(B+1) ##pval.boot
#plot(boot.blood)

pval.boot
0.04
```

## Example Conclusions

- Using a permutation test the exact p-value is 0.0314.
- The estimated permutation p-value is 0.03.
- The estimated bootstrap p-value is 0.038.
- **The three procedures lead to the same conclusion.** We reject the null hypothesis, showing an effect of the drug. It seems to decrease the blood pressure in rats.

## Exercise for the students

A company wants to know if it is efficient to teach some new tools to its workers using internet courses. The company randomly select 7 workers and randomly assign them to two groups of sizes 4 and 3. The first group attended traditional courses and the second internet courses. After the courses a test was administered to the workers whose results were:

Internet courses: 37, 49, 55, 57

Traditional courses: 23, 31, 46

Show if internet courses are preferred over traditional courses using a permutation test and a bootstrap test. Compute the p-value if the company is interested in checking if both learning methods are different.

## Exercise for the students

A Japanese company and an American company claim that they have developed new technology to increase network transmission speeds. The marketing managers of both companies simultaneously announce that they can transmit 1 terabyte per second. To substantiate their claims, each company submits trial data (in seconds) to transmit one terabyte with their new technologies.

Japanese company	0.98	0.85	0.9	1	
American company	0.95	0.94	0.8	1	0.99

Is there evidence to suggest the transmission speed using the technology developed by the American company is superior to the transmission speed using the technology developed by the Japanese company? Compute the  $p$ -value to answer the question with the following techniques:

1. Enumerate all possible combinations to find the  $p$ -value for a permutation test.

2. Use the function *onewaytest* from the **coin** package to calculate an appropriate *p* – value. Does this *p* – value match the one in part 1?
3. Obtain an estimated permutation *p* – value using the **boot** function from the **boot** package.
4. What conclusion do the *p* – value support?