

Permutation Test

Snehadrita Das

Introduction

One of the main interests of Statistics is performing inferential analysis and interpretation of the data on hand. In statistics, a hypothesis test calculates some quantity under a given assumption. The result of the test allows us to interpret whether the assumption holds or whether the assumption has been violated.

The whole procedure would have been presented, more or less, as follows:

- We want to test the validity of statement (“null hypothesis”) about a parameter associated with a well-defined underlying population against another hypothesis, the “alternative hypothesis”.
- Then we sample from the population of interest. For example, we might use simple random sampling, so that all sample values are mutually independent of each other.
- Thence we define our test statistic and its value under the null hypothesis. Usually, this test statistics would be function of the random samples which is free from the parameter under consideration. Typically, this estimator would have to be transformed to make it “pivotal” quantity, that is, having a sampling distribution that does not depend on any other unknown parameters.
- Then we calculate a quantity called, the p-value which is basically the evidence in favour of the null hypothesis and conclude our results accordingly.

Now the aforesaid is what you call a typical parametric setup which takes greater length in obtaining our interests. For example we need to take some assumptions under consideration.

- we might assume that the population values follow a Normal distribution, with a known variance, but a mean that is unknown and whose value we’re testing.
- The type of sampling method we are considering.

To put it simply, parametric procedure is quite a stretch.

Permutation Test - A Non-Parametric setup

A permutation test (also called re-randomization test), a non-parametric method, is an increasingly common statistical tool for *constructing sampling distributions* and *comparing treatment groups in Design of Experiment by means of hypothesis testing*. The parametric analogs could be seen as t-test.

why should you go for Permutation tests

- Permutation tests are non-parametric tests that require very few assumptions. A very handy tool when you do not know your population or your data generating mechanism.
- Permutation tests are non-parametric tests that solely rely on the assumption of *Exchangeability*.

What is Exchangeability

In statistics, an exchangeable sequence of random variables is a sequence whose joint probability distribution does not change when the positions in the sequence in which finitely many of them appear are altered.

- Permutation tests exist for any test statistic, regardless of whether or not its distribution is known.
- Permutation tests can be used for *analyzing unbalanced designs, paired sample problems* and *qualitative data*.

Methodology

- The goal is to test the hypothesis that the groups under study have identical distributions. Compute the difference between the sample means, where the sample sizes are, say, n_1 and n_2 .
- Pool the data and *sample without replacement*.
- Consider any permutation of the pooled data, compute the sample mean of the first n_1 observations, compute the sample mean using the remaining n_2 observations, and compute the difference between these sample means.
- Repeat the previous step for all possible permutations of the data.
- The one-sided p-value of the test is calculated as the proportion of sampled permutations where the difference in means was greater than the observed value of the test statistic. The two-sided p-value of the test is calculated as the proportion of sampled permutations where the absolute difference was greater than observed value of the test statistic.
- Since p-value could be misleading sometimes, hence we go one step above, the *Confidence interval*. We reject the null hypothesis if the observed value of the test statistic under permutation test is not contained in the sorted middle $(1 - \alpha) \times 100\%$ values.

Treatment Group Comparison Example

- We shall consider the *Toothgrowth* data from the *datasets* package, which is basically a data frame of 60 observations and 3 variables. 1. tooth length 2. supplementary type and 3. dose. We shall test if the supplementary type have any effect over tooth growth. And we shall use permutation test. Now for 60 observations we can have 60! permutations which is quite an overwhelming number.
- *Null Hypothesis* : Supplementary types have no effect at all.
- *Alternative Hypothesis* : The mean effect of supp type OJ is greater than the mean effect of supp type VC

Calculating the observed value of statistic

```
library(datasets)
obs<-ToothGrowth$len ## considering the ToothGrowth dataset
obs.mean.diff<-(mean(ToothGrowth$len[ToothGrowth$supp=="OJ"])-mean(ToothGrowth$len[ToothGrowth$supp=="VC"]))
obs.mean.diff
```

```
[1] 3.7
```

Permutations and values of the statistics

```

set.seed(17)

no.perm<-10000 ## number of permutations

n=length(ToothGrowth$len) ## number of samples to be sampled

permute.mat<-matrix(rep(0,n*no.perm),no.perm,n) ## empty matrix to store the permutations

for (i in 1:no.perm){
  permute.mat[i,]=sample(ToothGrowth$len,n)
}

mean.diff=function(data){
  (mean(data[31:60])- mean(data[1:30]))
}

list.mean.diff<-apply(permute.mat,1,mean.diff) ## 10000 mean differences
head(list.mean.diff)

```

```
[1] 2.406667 5.606667 -0.100000 1.066667 -4.100000 0.160000
```

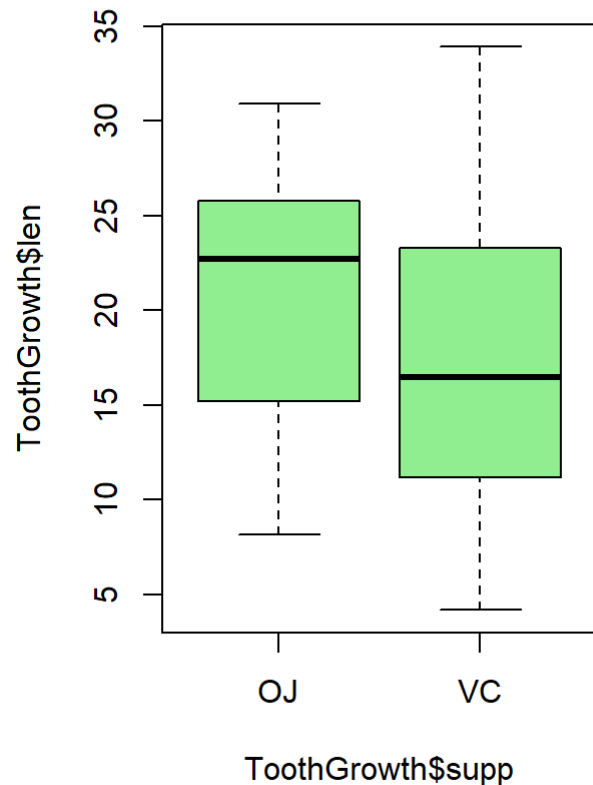
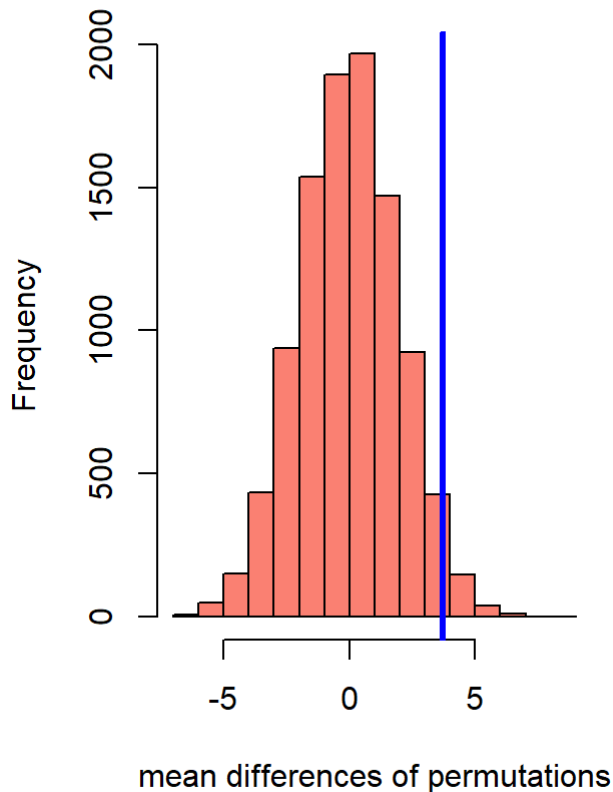
Visualization and EDA

```

par(mfrow=c(1,2))
hist(list.mean.diff,col="salmon",xlab="mean differences of permutations",
     main="Histogram of Mean differences")
abline(v=obs.mean.diff,col="blue",lwd=3)
boxplot(ToothGrowth$len~ToothGrowth$supp,col="lightgreen",
       main="Boxplot comparing tooth growth for different supp type")

```

Histogram of Mean differences t comparing tooth growth for differen



P-value and confidence interval

```
p.value<-(sum(list.mean.diff>obs.mean.diff))/no.perm
p.value ## p value
```

```
[1] 0.0281
```

```
## Confidence interval
```

```
ord.mean.diff<-sort(list.mean.diff)
quantile(ord.mean.diff,probs = c(0.025,0.975))
```

```
2.5%    97.5%
-3.840167 3.773333
```

- According to the p-value which is 0.0281(<0.05) we should reject our null hypothesis of treatment groups being equal but we can manipulate the p-value by increasing or decreasing the sample or number of permutations and hence we shall stick to our confidence interval. Since our CI contains 0, we conclude that supplementary type has no effect on the tooth growth.

A Non-Normal case study

- Consider the life time of bulbs, manufactured in factory 1 and factory 2 and we are interested in seeing the distribution of difference in mean life time.

```

set.seed(17)
life.time1<-rexp(10,(1/30))
life.time2<-rexp(10,(1/20))
pooled.data<-c(life.time1,life.time2)

## Difference in means
no.perm<-10000
n=length(pooled.data)
permute.mat.mean<-matrix(rep(0,n*no.perm),no.perm,n)
for (i in 1:no.perm){
  permute.mat.mean[i,]=sample(pooled.data,n)
}

mean.differ=function(data){
  mean(data[1:10])-mean(data[11:20])
}

list.mean<-apply(permute.mat,1,mean.differ)
head(list.mean)

```

```
[1] 6.75 -1.53 2.77 -1.03 4.83 4.52
```

```

## difference in median

permute.mat.med<-matrix(rep(0,n*no.perm),no.perm,n)
for (i in 1:no.perm){
  permute.mat.med[i,]=sample(pooled.data,n)
}
med.differ=function(data){
  median(data[1:10])-median(data[11:20])
}

list.med<-apply(permute.mat,1,med.differ)
head(list.med)

```

```
[1] 9.35 0.00 5.85 0.00 7.95 7.35
```

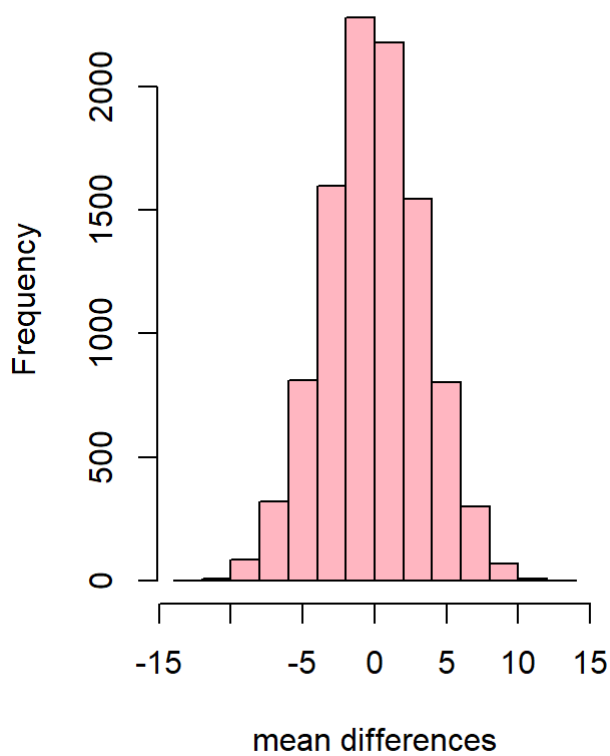
Visualization

```

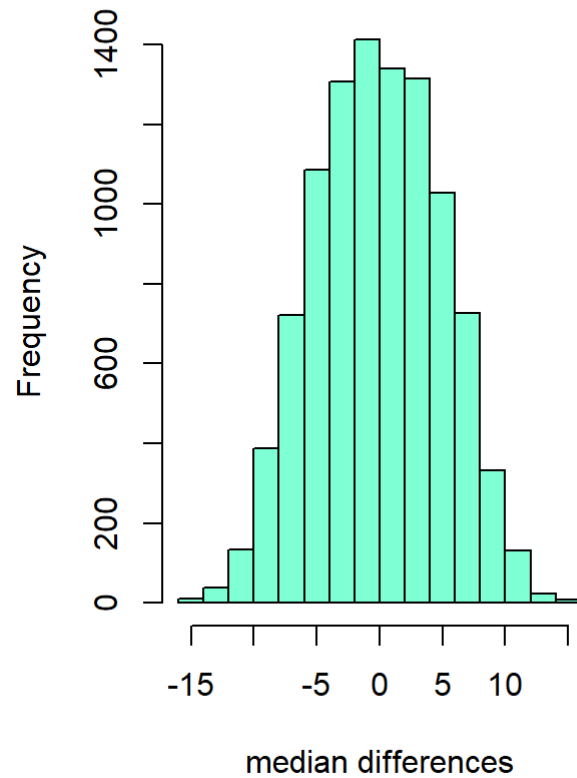
par(mfrow=c(1,2))
hist(list.mean,col="lightpink",xlab="mean differences",main="Distributions Mean Differences")
hist(list.med,col="aquamarine",xlab="median differences",main="Distributions Median Differences")

```

Distributions Mean Differences



Distributions Median Differences



KS-Test

```
ks.test(list.mean, rnorm(10000, 0, 10))
```

```
Warning in ks.test.default(list.mean, rnorm(10000, 0, 10)): p-value will be  
approximate in the presence of ties
```

Asymptotic two-sample Kolmogorov-Smirnov test

```
data: list.mean and rnorm(10000, 0, 10)  
D = 0.2401, p-value < 2.2e-16  
alternative hypothesis: two-sided
```