Data Visualization and Modelling- Task 1

Ruth Kristianingsih

October 8, 2017

EXERCISE 1

The first row is chest circumference (in inches) of five subjects. Let us call this X. The second row is the respective total volumes of air that can be breathes in and out in one minute (in litres) for the same five subjets. Let us call this Y. X : 39, 29, 60, 40, 32; Y : 11, 5, 20, 8, 6
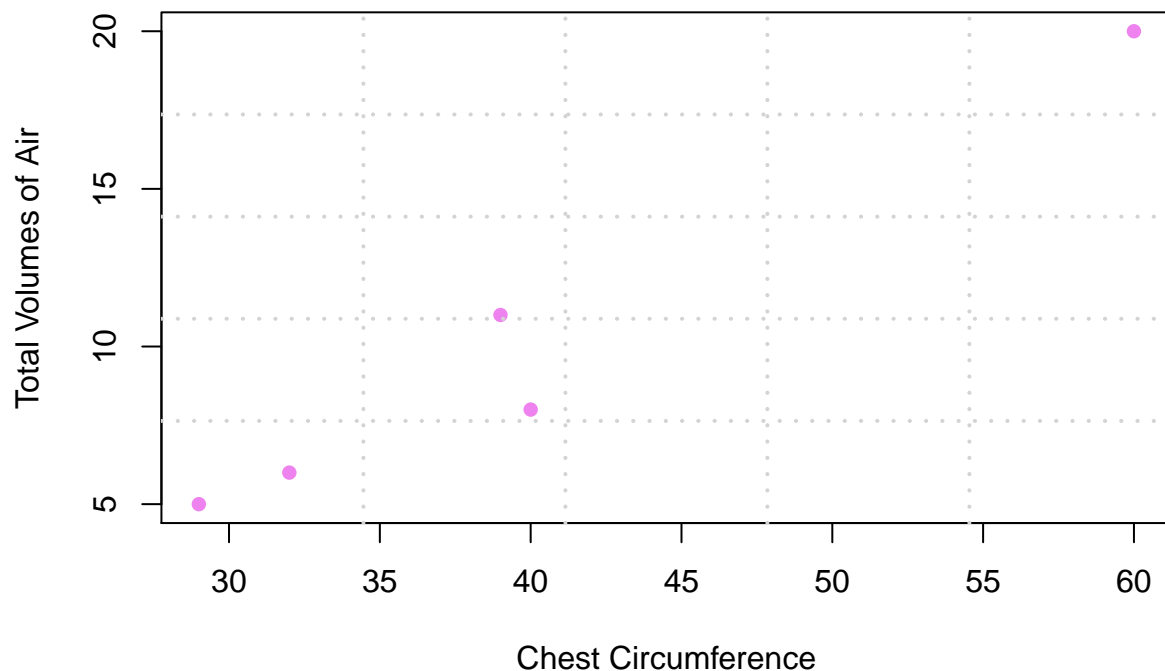
Perform an exact test for

$$Ho : \rho = 0$$

$$Ho : \rho > 0$$

Solution

The first solution will be finding correlation using Pearson method. Firstly, the data will be plotted.

```
x <- c(39, 29, 60, 40, 32)
y <- c(11, 5, 20, 8, 6)
plot(x, y, xlab="Chest Circumference", ylab="Total Volumes of Air", col="Violet", pch=16)
grid(5, 5, lwd = 2)
```

Now, the Pearson correlation between chest circumference and total volume of air will be tested using the code below.

```r
x = c(39, 29, 60, 40, 32)
y = c(11,5, 20, 8, 6)
nr= 100 #number of arrangements to be examined
st = numeric(nr)
sttrue = cor(x,y)
n = length(y)
cnt=0
for (i in 1:nr){
  d=sample(y,n)
  st[i] <- cor(d,x)
  if(st[i] > sttrue) cnt=cnt+1
}
cnt/nr
```

```
## [1] 0.01
```

the results of the test conducted with 100 sample is 0.01 and the correlation of x and y itself is 0.9778 as shown below.

```r
sttrue
```

```
## [1] 0.9777792
```

```r
cor.test(x,y, method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  x and y
## t = 8.0785, df = 3, p-value = 0.003963
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6954566 0.9985955
## sample estimates:
##       cor
## 0.9777792
```

Moreover, from the result above, it can be seen that the p-value from the sample x and y itself is 0.003963 which is less than 0.05, therefore Ho is rejected, meaning that chest circumference and volume of air is positively correlated.

The position of the number of correlation obtained in the histogram of simulation can be showed in the following graph.

```r
hist(st)
abline(v = sttrue, col="red", lwd=3, lty=2)
```

Histogram of st
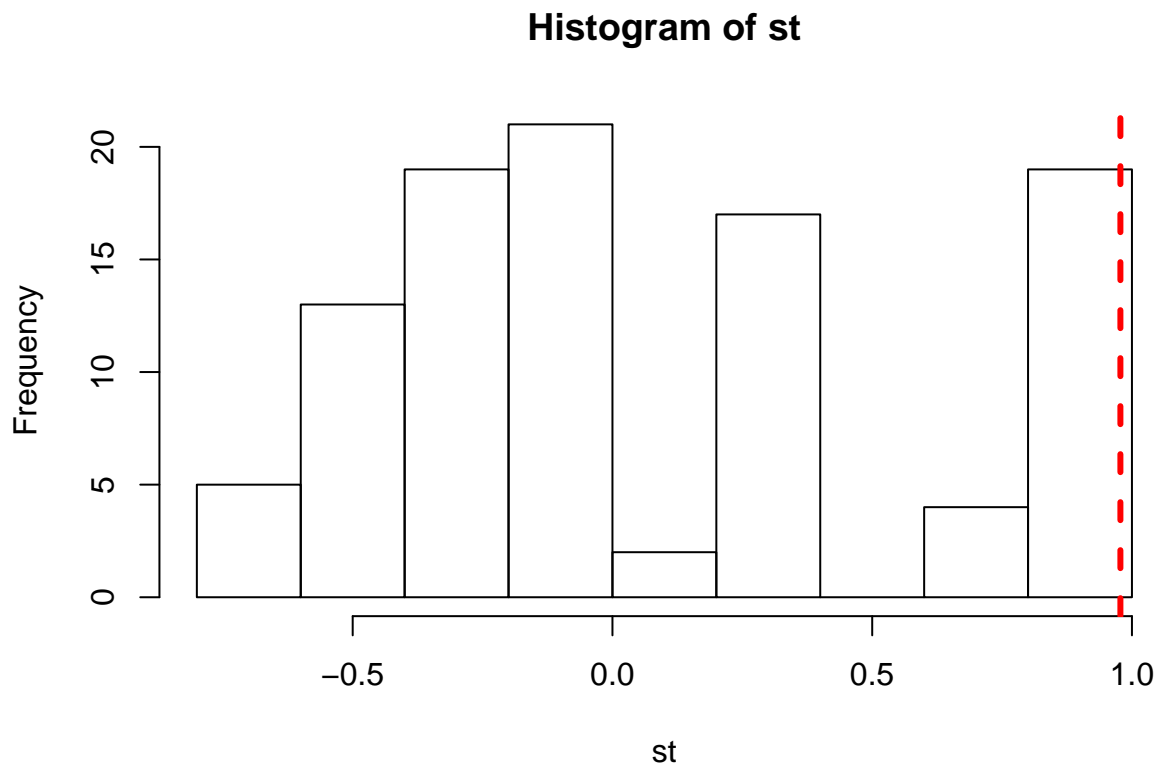
And now, the exact correlation test will be performed. Using the the code below, as it can be seen that the result is 0.0083 which is still less than 0.05, therefore Ho is rejected.

```
library(combinat)
```

```
##
## Attaching package: 'combinat'

## The following object is masked from 'package:utils':
##
##     combn
```

```
x = c(39, 29, 60, 40, 32)
y = c(11,5, 20, 8, 6)
sttrue = cor(x,y)
n = length(y)
nr= fact(n) #number of arrangements to be examined
st = numeric(nr)
cnt=0
d=permn(y)
for (i in 1:nr){
  st[i] <- cor(d[[i]],x)
  if(st[i] > sttrue) cnt=cnt+1
}
cnt/nr
```

```
## [1] 0.008333333
```

The second solution will be regarding test correlation between chest circumference and total volumes of air

using Spearman method.

Using code performed below, it can be obtained that the p-value is 0.03 and it can be seen also that the correlation using spearman shows that the value 0.9, meaning that the correlation is positive and Ho is rejected.

```r
x = c(39, 29, 60, 40, 32)
y = c(11,5, 20, 8, 6)
nr= 100 #number of arrangements to be examined
st = numeric(nr)
sttrue = cor(x,y, method="spearman")
n = length(y)
cnt=0
for (i in 1:nr){
  d=sample(y,n)
  st[i] <- cor(d,x)
  if(st[i] > sttrue) cnt=cnt+1
}
cnt/nr
```

```
## [1] 0.08
```

```r
sttrue = cor(x,y, method="spearman")
sttrue
```

```
## [1] 0.9
```

Next, correlation test using the sample x and y is done and for the sample x and y itself, it is obtained that the correlation value (rho) is equal to 0.9 and the p-value is equal 0.083 which is less than alpha=0.1. Therefore, Ho is rejected, meaning that x and y are positively corelated.

```r
cor.test(x,y, method="spearman")
```

```
##
##  Spearman's rank correlation rho
##
## data:  x and y
## S = 2, p-value = 0.08333
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.9
```

Now, exact test should be perform to see whether this value is still consistent.

```r
library(combinat)
x = c(39, 29, 60, 40, 32)
y = c(11,5, 20, 8, 6)
sttrue = cor(x,y, method="spearman")
n = length(y)
nr= fact(n) #number of arrangements to be examined
st = numeric(nr)
cnt=0
d=permn(y)
for (i in 1:nr){
  st[i] <- cor(d[[i]],x)
  if(st[i] > sttrue) cnt=cnt+1
}
```

4

```
cnt/nr
```

```
## [1] 0.06666667
```

From the results above, the correlation value between x and y is 0.9 with the p-value 0.8333. When the simulation is conducted with the bigger sample which is 100, it is obatined that p-value is 0.08, and for the exact test, it is obtained that p-value is 0.0667. This results are consistent and based on the results, meaning that Ho is rejected, since the value of the correlation is bigger than zero and p-value is less than alpha is equal to 0.1.

EXERCISE 2

These are the increments of weight recorded in an experiment where a new addictive has been added to a standard compound feed: Standard : 2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8 + Additive : 3.5, 6.3, 4.2, 4.3, 3.8, 5.7, 4.4

Number 1

Is the additive efficient? Perform a classical t-test and use corresponding T-statistic to perform a permutation test.

Before we perfom any t-test, we need to check whether the variances are homogeneous with a F-test of Fisher using the function var.test() in R.

```
std <- c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8)
add <- c(3.5, 6.3, 4.2, 4.3, 3.8, 5.7, 4.4)
var.test (std,add)
```

```
##
##  F test to compare two variances
##
## data:  std and add
## F = 1.1461, num df = 8, denom df = 6, p-value = 0.8929
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.2046746 5.3313102
## sample estimates:
## ratio of variances
##            1.1461
```

With the hypotesis: Ho: variance of standard and additive are equal (homogeneous) H1: variance of standard and additive are not equal

From the results above, it can be examined since p-value is greater than alpha=0.05, then Ho is accepted, which means that the variance of x and y are equal. (Homogenity of variance)

To make the comparison between the two groups, the information regarding variance above can be used, the function t.test with homogeneous variances (var.equal = TRUE) and independent samples (paired = FALSE).

Now, classical t-test using R will be done. From the results below, it is obtained that the t value is -2.6586 and p-value is 0.01871 which is less than 0.05. Based on this results, then it is concluded that Ho is rejected, meaning that H1 is accepted (Average value of additive data is greater than average value of standard data). Therefore, the additive is effective.

```
t.test(std, add, paired=F, var.equal=TRUE)
```

```
## 
##  Two Sample t-test
## 
## data:  std and add
## t = -2.6586, df = 14, p-value = 0.01871
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.5695886 -0.2748558
## sample estimates:
## mean of x mean of y
##  3.177778  4.600000
```

And now, T-statistics will be used to perform a permutation test. It yields the results below.

```r
nr= 10000 #number of rearrangements to be examined
st <- numeric (nr)
std = c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8)
add = c(3.5, 6.3, 4.2, 4.3, 3.8, 5.7, 4.4)
n1 <- length(add); n2 <- length(std)
total <- n1+n2
sttrue = mean(add)-mean(std)
cnt = 0 #number of counter
vect = c(add, std)
for (i in 1:nr){
  d= sample(vect,n1+n2)
  ne <- d[1 : n1]
  a <- n1 + 1; co <- d[a:total]
  st[i] <- mean (ne) - mean (co)
  if(st[i] > sttrue)cnt=cnt+1
}
cnt/nr #pvalue
```

```
## [1] 0.0098
```

It is clear that the p-value is less than 0.05, meaning that Ho is rejected.

From these two results, it can be said that the results are consistent from clasical t-test and permutation test using t-statistic.

Number 2

For the second question, permutation tests using several different test statistics will be performed.

Using sum

For this case: 1. Ho and H1 Ho: standard (Xs) and additive (Xa) are equal (belong to the same population), that is Xs and Xa are copies from the same random variable H1: sum(Xa) > sum (Xs)

  2. Test Statistics An appropriate statisitic could be the difference of sum: sum(additive) - sum (standard) = 3.6

  3. Permutation test will be held using code below:

```r
nr= 10000 #number of rearrangements to be examined
st <- numeric (nr)
std = c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8)
```

```
add = c(3.5, 6.3, 4.2, 4.3, 3.8, 5.7, 4.4)
n1 <- length(add); n2 <- length(std)
total <- n1+n2
sttrue = sum(add)-sum(std)
cnt = 0 #number of counter
vect = c(add, std)
for (i in 1:nr){
  d= sample(vect,n1+n2)
  ne <- d[1 : n1]
  a <- n1 + 1; co <- d[a:total]
  st[i] <- sum (ne) - sum (co)
  if(st[i] > sttrue)cnt=cnt+1
}
cnt/nr #pvalue
```
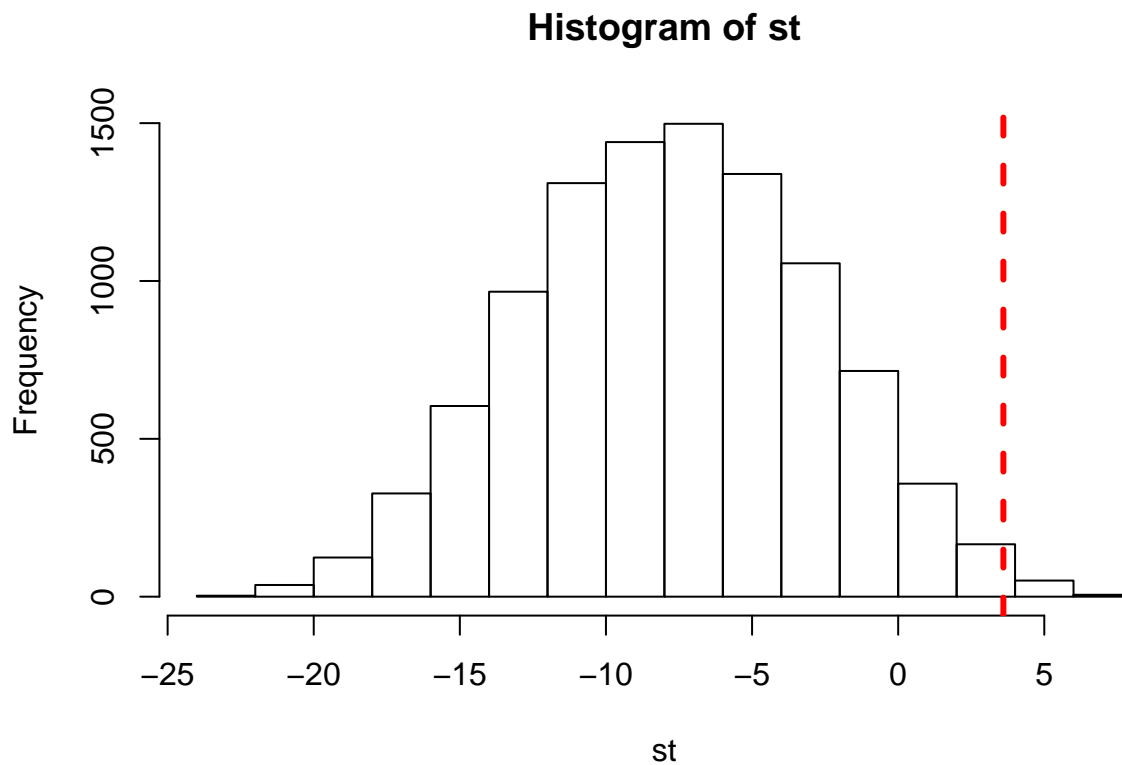
[1] 0.0075

The result of p-value above is shown that Ho is rejected and H1 is accepted, since it is less than zero.

From the following picture, the position of the difference between sum of standard and additive on the histogram of the resampling data is shown.

```
hist(st)
abline(v = sttrue, col="red", lwd=3, lty=2)
```


Histogram of st

Using median

For this case: 1. Ho and H1 Ho: standard (Xs) and additive (Xa) are equal (belong to the same population), that is Xs and Xa are copies from the same random variable H1: median(Xa) > median (Xs)

2. Test Statistics An appropriate statisitic could be the difference of sum: median(additive) - median (standard) $= 1$

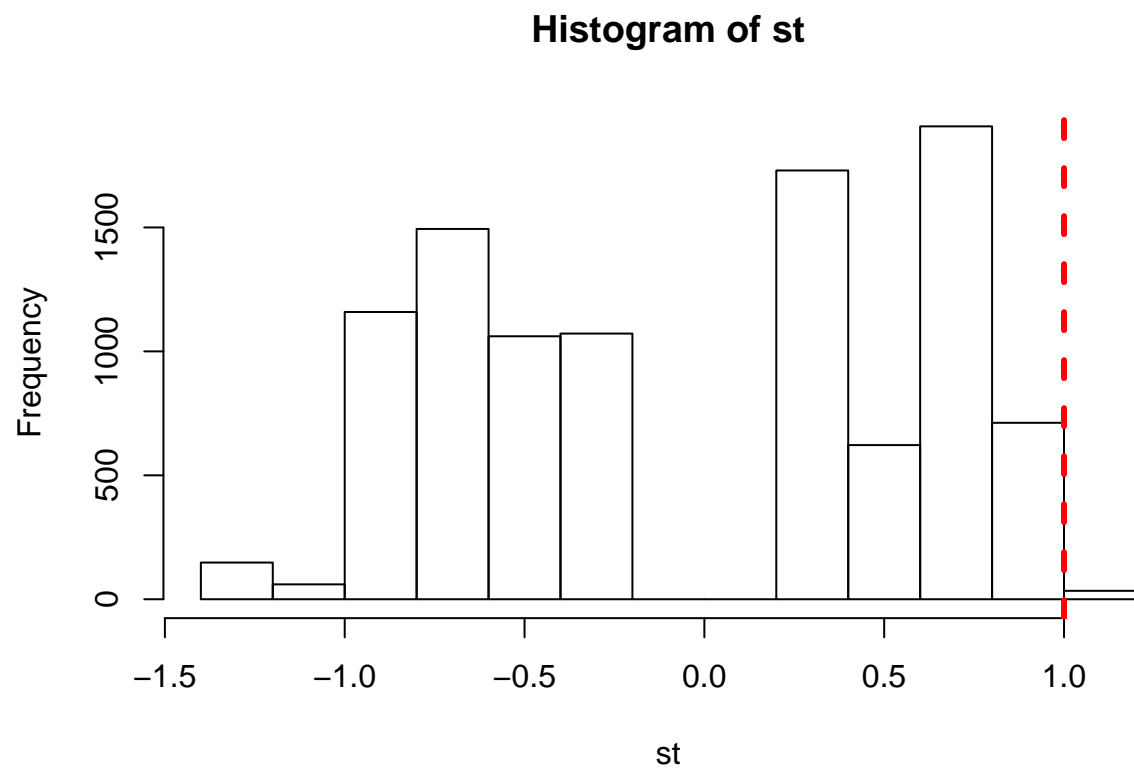3. Permutation test will be held using code below:

```r
nr= 10000 #number of rearrangements to be examined
st <- numeric (nr)
std = c(2.5, 3.4, 2.9, 4.1, 5.3, 3.4, 1.9, 3.3, 1.8)
add = c(3.5, 6.3, 4.2, 4.3, 3.8, 5.7, 4.4)
n1 <- length(add); n2 <- length(std)
total <- n1+n2
sttrue = median(add)-median(std)
cnt = 0 #number of counter
vect = c(add, std)
for (i in 1:nr){
  d= sample(vect,n1+n2)
  ne <- d[1 : n1]
  a <- n1 + 1; co <- d[a:total]
  st[i] <- median (ne) - median(co)
  if(st[i] > sttrue)cnt=cnt+1
}
cnt/nr #pvalue
```

```
## [1] 0.0164
```

The result of p-value above is shown that Ho is rejected and H1 is accepted, since it is less than zero.

From the following picture, the position of the difference between median of standard and additive on the histogram of the resampling data is shown.

```r
hist(st)
abline(v = sttrue, col="red", lwd=3, lty=2)
```

Histogram of st

Conclusion for exercise 2: The results of analysis using several methods of statistic in permutation test/resampling yielded the same results which is H1 is accepted (Additive is effective).