

## Дескриптивна статистика

3 курс, статистика, Шкляр Ірина Володимирівна

Завдання 2, варіант 9

Порахувавши свої компанії по номерам (від 161 до 180), я отримала таблиці від компанії EMR до FFIV і зробила окремий каталог tables з ними.

З цих компаній я вибрала F (Ford Motor Company), EW (Edwards Lifesciences) і EXPE (Expedia Group).

F:

```
> setwd("/Users/irynashkliar/Downloads/tables")  
  
> x<-read.csv("table_f.csv",header=F)  
> colnames(x)<-c("dat","z","opn","mx","mn","clo","vol")  
> lr<-log(x$clo[-nrow(x)]/x$clo[-1])  
> mr<-log(x$mx/x$mn)  
> or<-log(x$opn/x$clo)
```

EW:

```
> y<-read.csv("table_ew.csv",header=F)  
> colnames(y)<-c("dat","z","opn","mx","mn","clo","vol")  
> lr1<-log(y$clo[-nrow(y)]/y$clo[-1])  
> mr1<-log(y$mx/y$mn)  
> or1<-log(y$opn/y$clo)
```

EXPE:

```
> z<-read.csv("table_expe.csv",header=F)  
> colnames(z)<-c("dat","z","opn","mx","mn","clo","vol")  
> lr2<-log(z$clo[-nrow(z)]/z$clo[-1])  
> mr2<-log(z$mx/z$mn)  
> or2<-log(z$opn/z$clo)
```

В 4 пункті я вибрала компанію Ford.

Далі наведений код як я рахувала дані для скриньок з вусами та самі скриньки:

P.S. Щоб знайти  $Q1(x)$ , я спочатку рахувала медіану по відрізьку від першого (найменшого) числа до медіани самого відрізьку ( $mr, lr$  або  $or$ ), а щоб  $Q3(x)$  -- медіану по відрізьку від медіани  $mr, lr$  або  $or$  до останнього (найбільшого) значення.

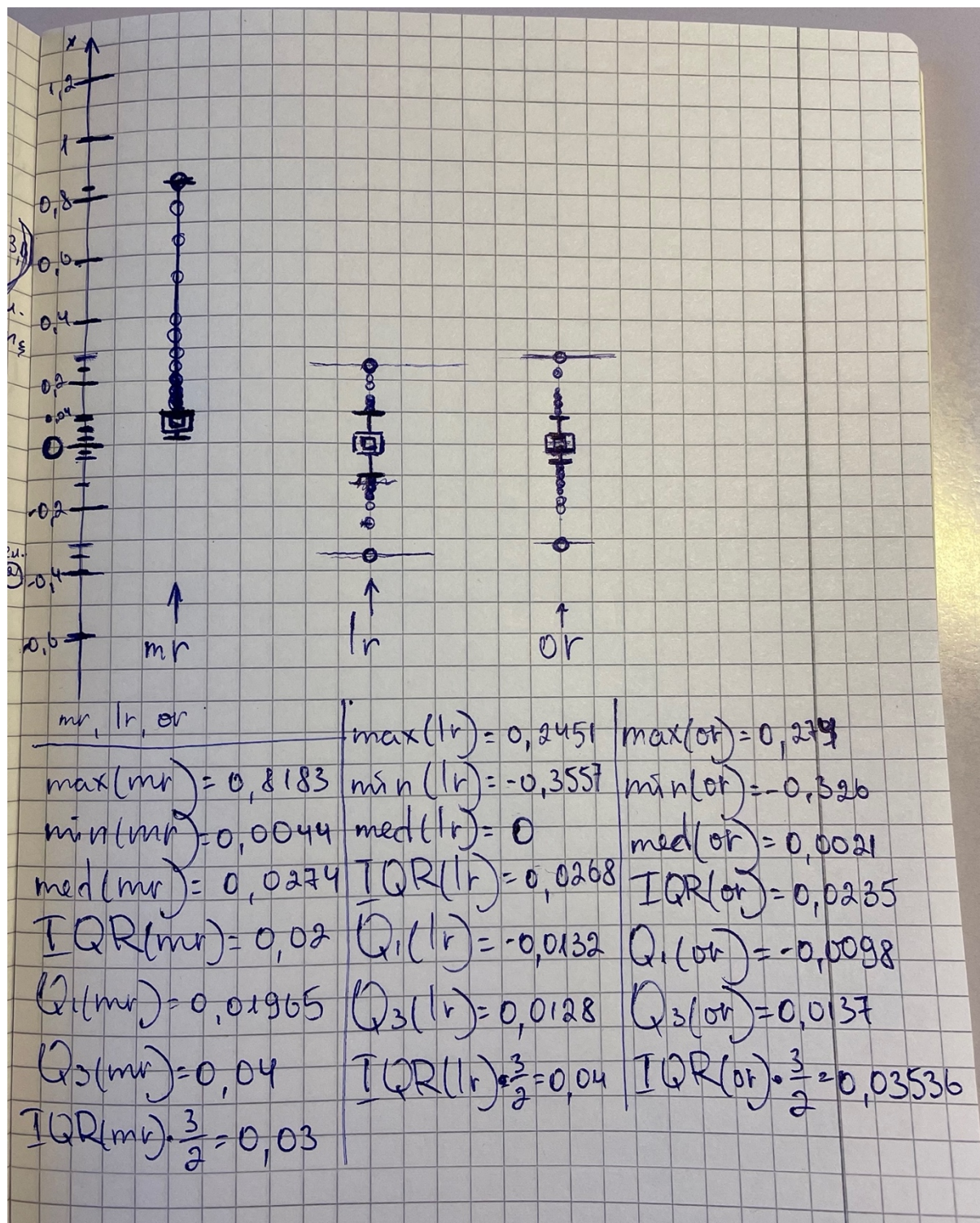
```
> max(mr)
[1] 0.8183113
> min(mr)
[1] 0.004396861
> median(mr)
[1] 0.02745334
> IQR(mr)
[1] 0.02055231
> mr0 <- sort(mr)
> mr_1 <- mr0[mr0<=0.02745334]
> median(mr_1)
[1] 0.01965062
> mr_2 <- mr0[mr0>=0.02745334]
> median(mr_2)
[1] 0.04020669
> IQR(mr0)*3/2
[1] 0.03082847
```

```
> max(lr)
[1] 0.2451207
> min(lr)
[1] -0.3557362
> lr0 <- sort(lr)
> median(lr0)
[1] 0
> IQR(lr0)
[1] 0.02681982
> lr_1 <- lr0[lr0<=0]
> median(lr_1)
[1] -0.013276
> lr_2 <- lr0[lr0>=0]
> median(lr_2)
[1] 0.01279313
> IQR(lr0)*3/2
[1] 0.04022974
```

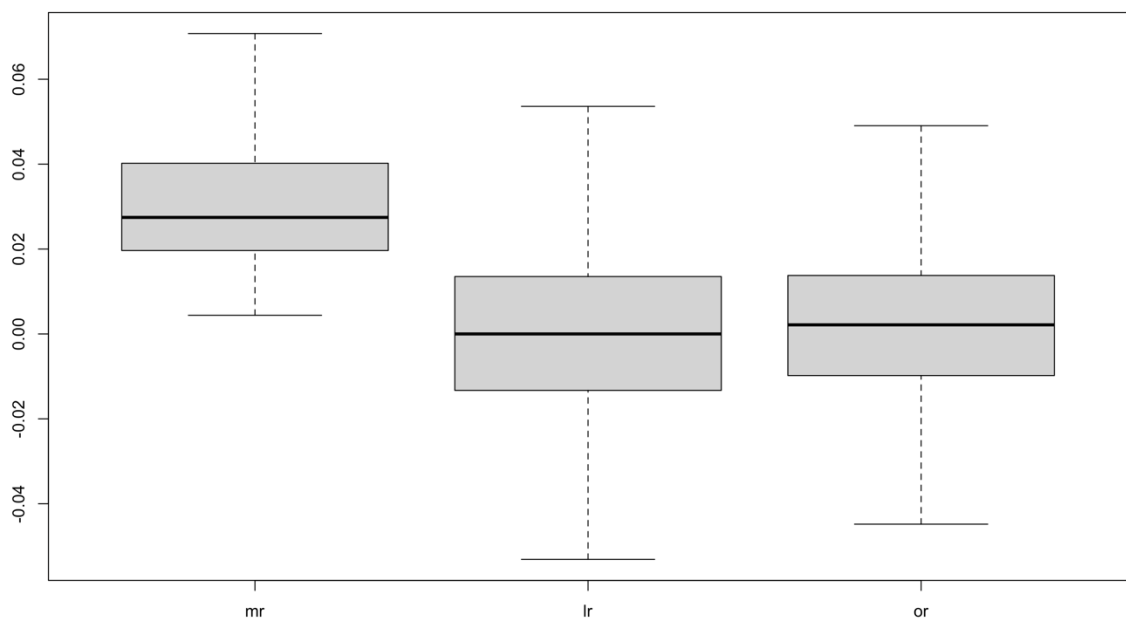
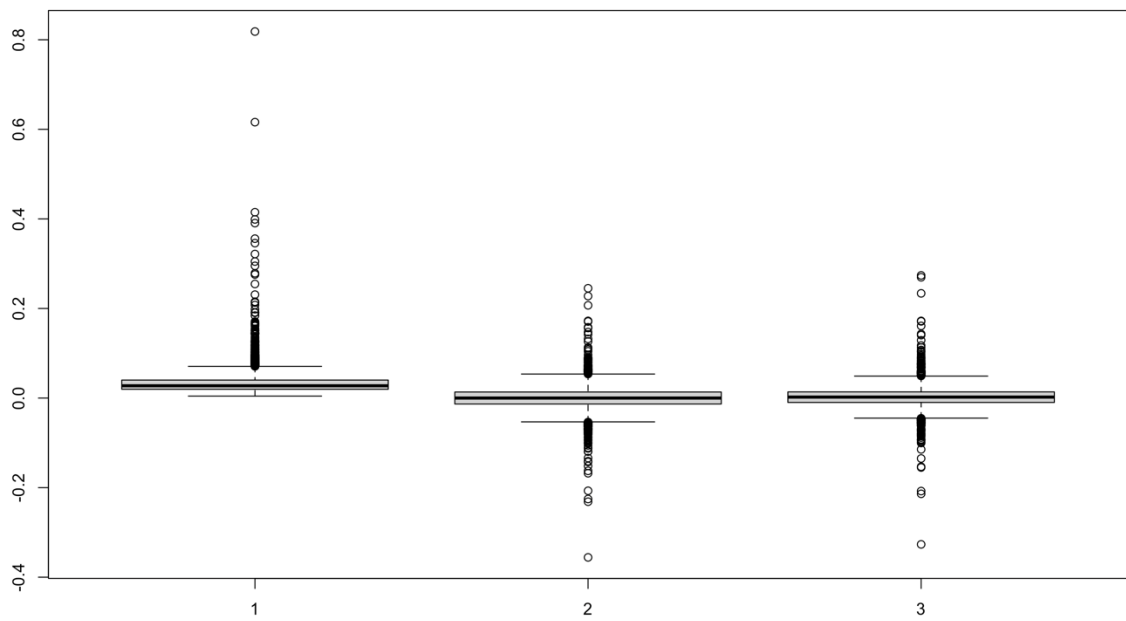
```
> max(or)
[1] 0.2738884
> min(or)
[1] -0.3266828
```

```
> or0 <- sort(or)
> median(or0)
[1] 0.002142587
> or_1 <- or0[or0<=0.00214]
> median(or_1)
[1] -0.009806133
> or_2 <- or0[or0>=0.00214]
> median(or_2)
[1] 0.01376973
> IQR(or0)
[1] 0.02357518
> IQR(or0)*3/2
[1] 0.03536277
```

Скриньки та мої підрахунки наведені на сторінці нижче:



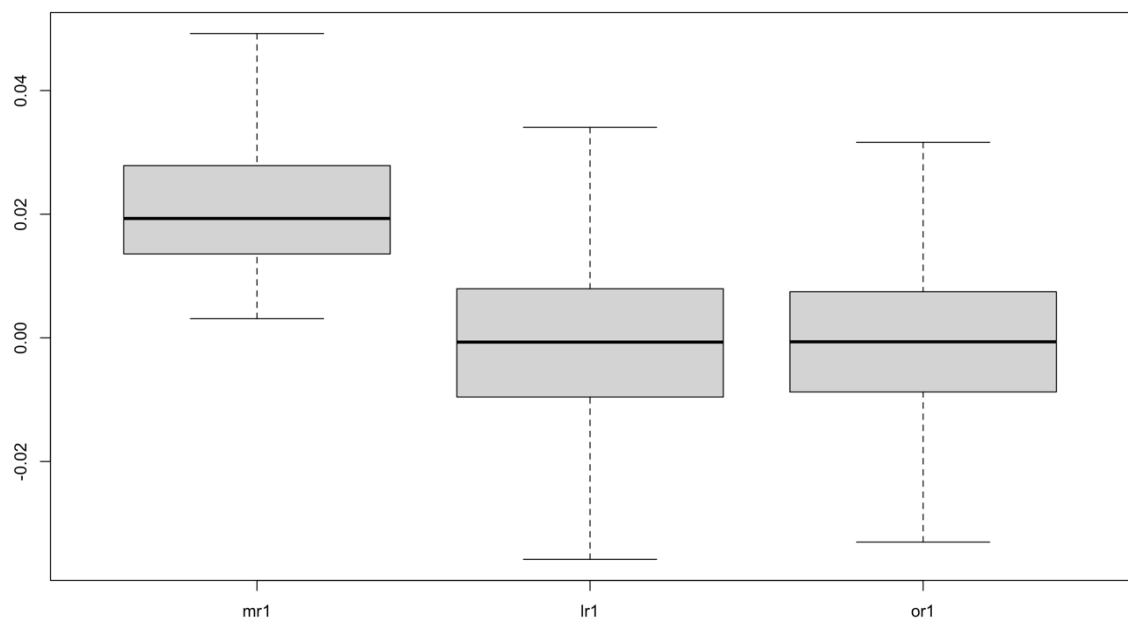
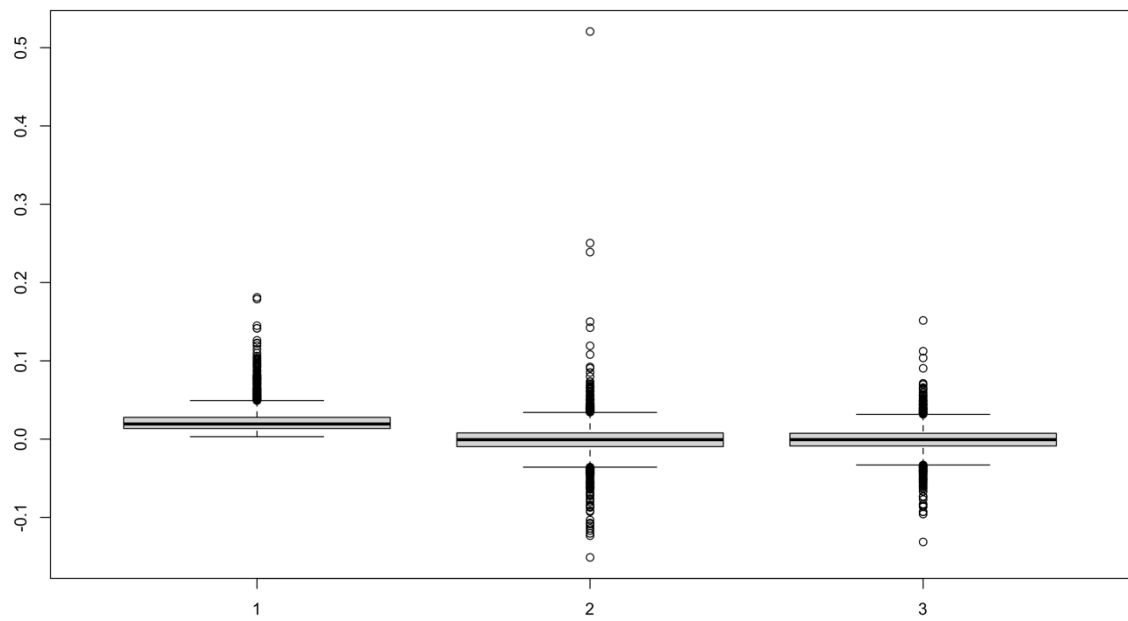
На наступній сторінці зображені скриньки з вусами, що намальовано саме в R за допомогою формул `boxplot(mr,lr,or)` та `boxplot(mr,lr,or,outline=F,names=c("mr","lr","or"))` -- скриньки без викидів :



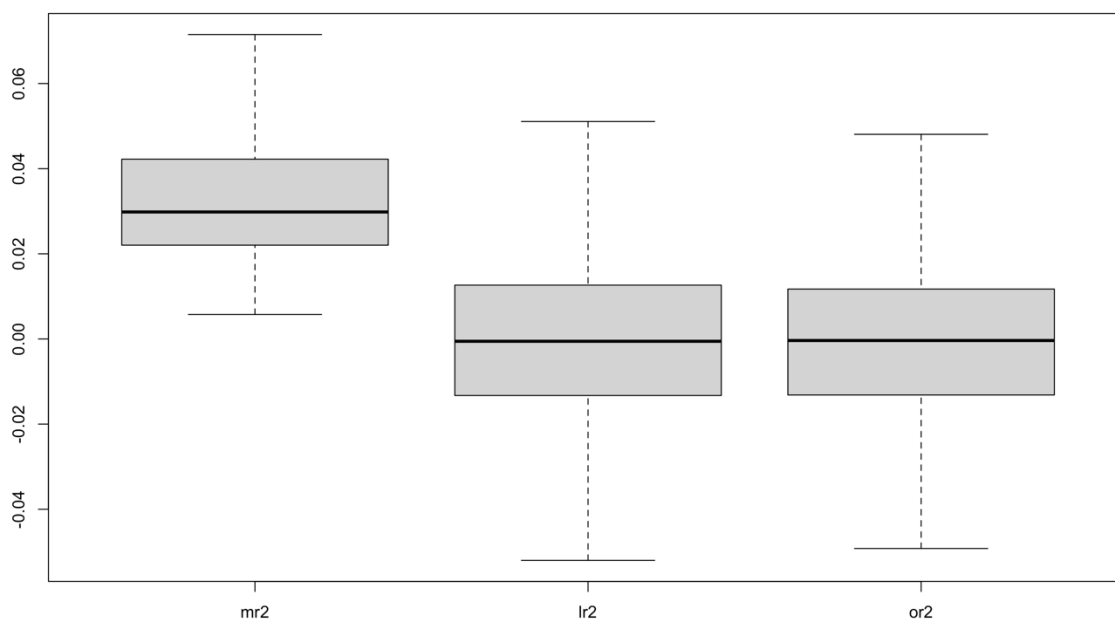
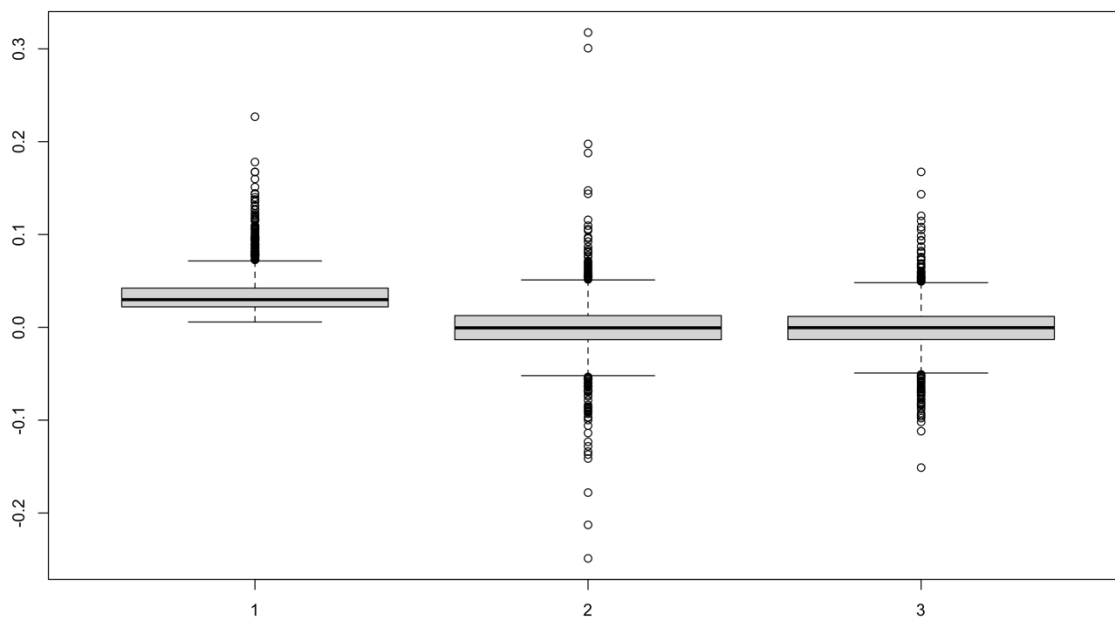
Загалом, відмінностей між моїми скриньками та скриньками, що намалював R, немає. Отже, дані збігаються, а тому підрахунок правильний.

Далі зробимо скриньки для компанії EW та EXPE:

EW:



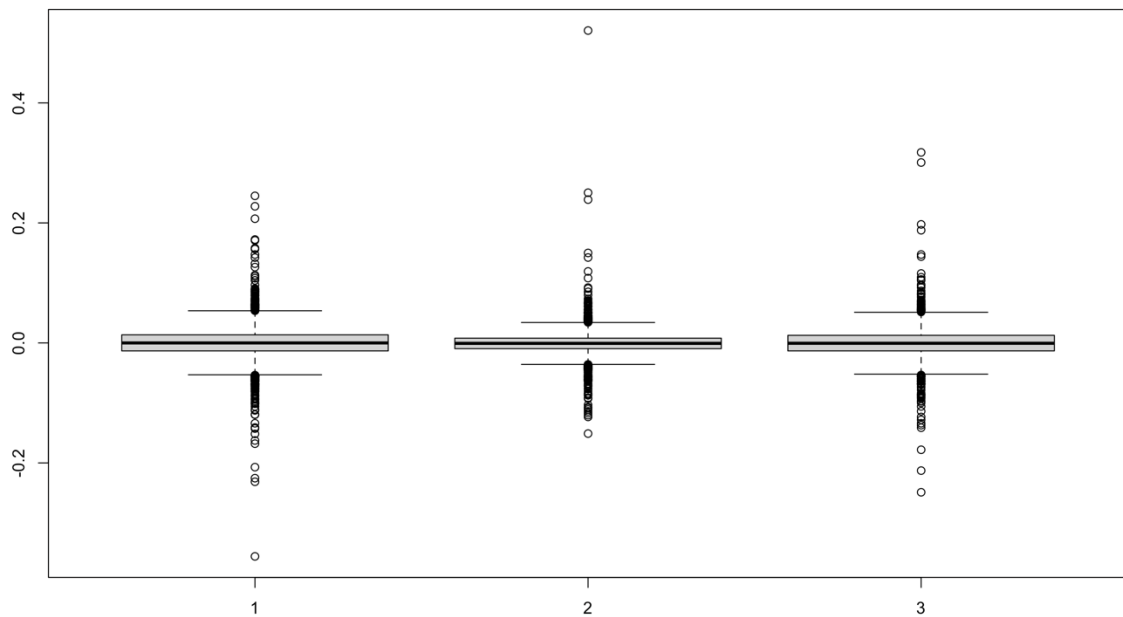
EXPE:



Бачимо, що розподіли  $mr$  та  $lr$  для двох компаній – зовсім різні функції. А  $lr$  та  $or$  дуже схожі між собою. Також видно, що  $\max(mr)$  та  $\text{med}(mr)$  у компанії EXPE вищий ніж у компанії EW. А  $\min(lr)$  та  $\max(lr)$  компанії EXPE нижчий ніж EW.

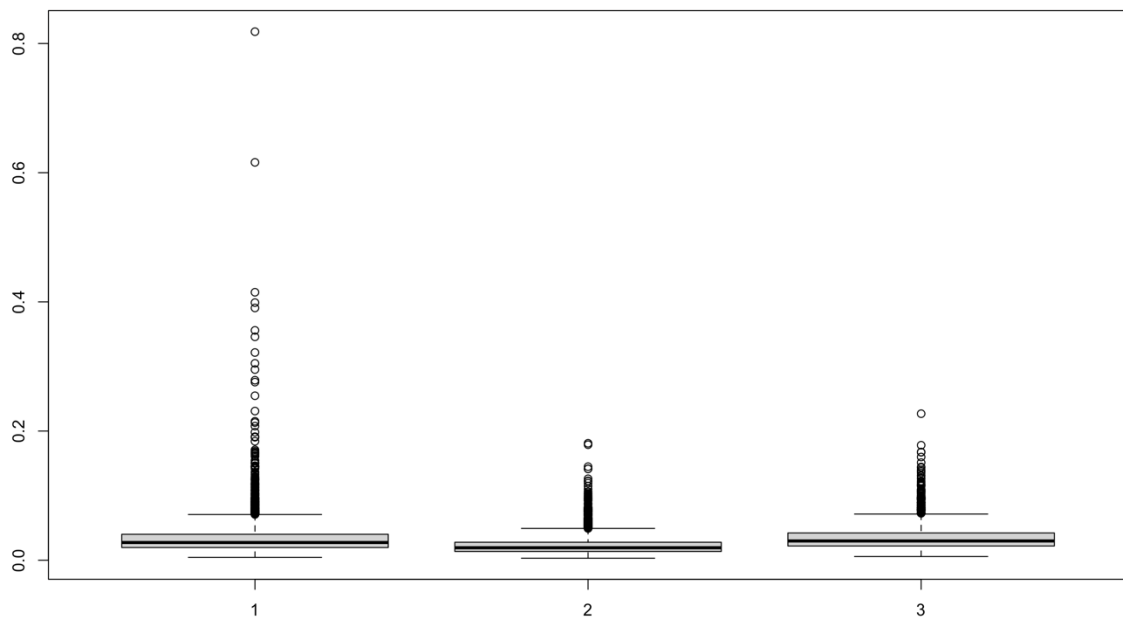
Тепер виведемо  $lr$  для всіх трьох компаній на одному рисунку. А потім аналогічно  $mr$  та  $or$ .

Lr:



З рисунку видно, що  $IQR(Lr)$  1 та 3 компаній більший ніж 2. Також у 1 компанії (Ford)  $\min(Lr)$  значно нижчий ніж у 2 та 3 компаній. А у 2 компанії (EW)  $\max(Lr)$  значно вищий ніж у 1 та 3.

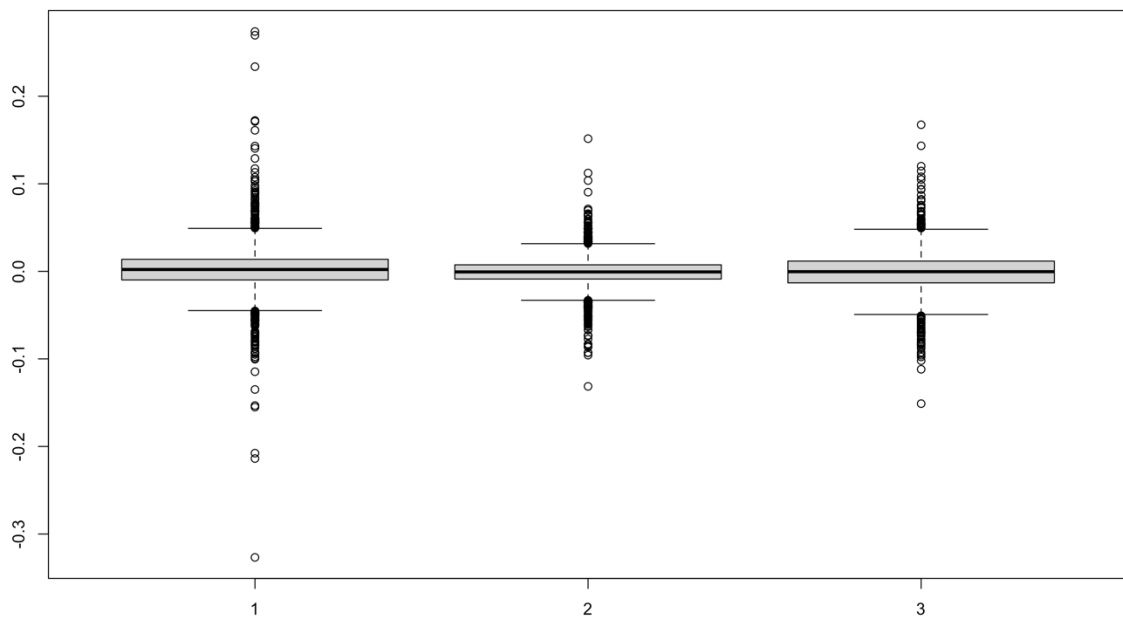
Mr:



Тут бачимо, що  $mr$  1 компанії -- Ford, має вищі значення, ніж інші.

Or:

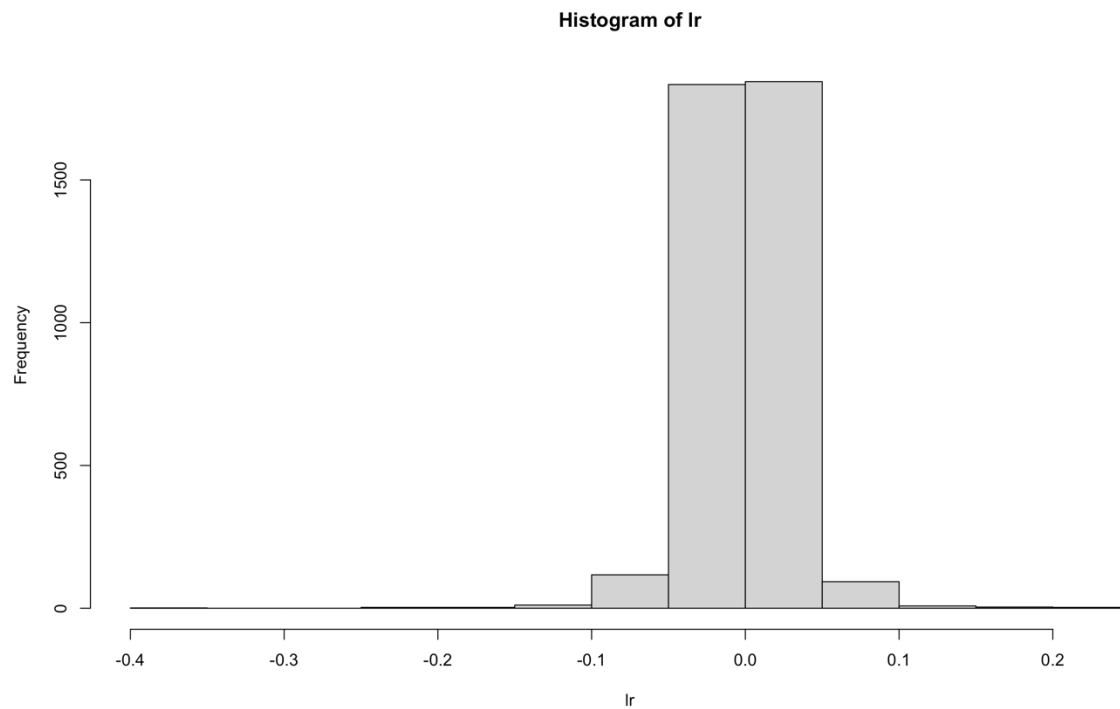




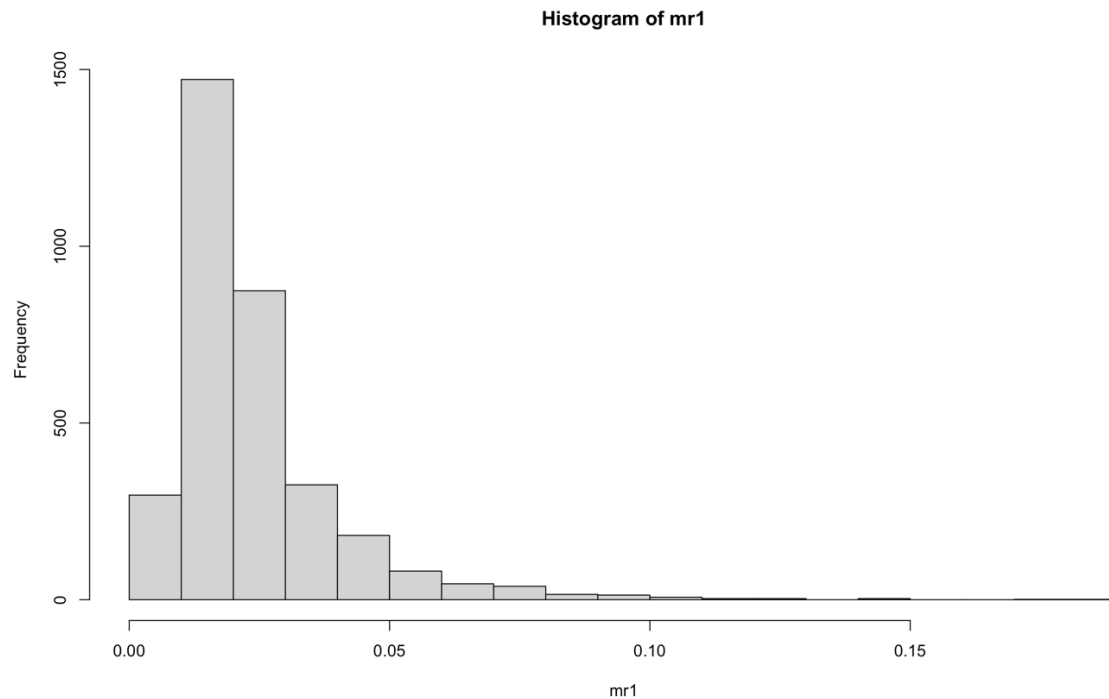
Тут  $\max(or)$  та  $\min(or)$  1 компанії дуже відрізняються від 2 та 3.

В пункті 8 я вирішила подивитись гістограми:  $lr$  для Ford,  $mr$  для EW та  $or$  для EXPE:

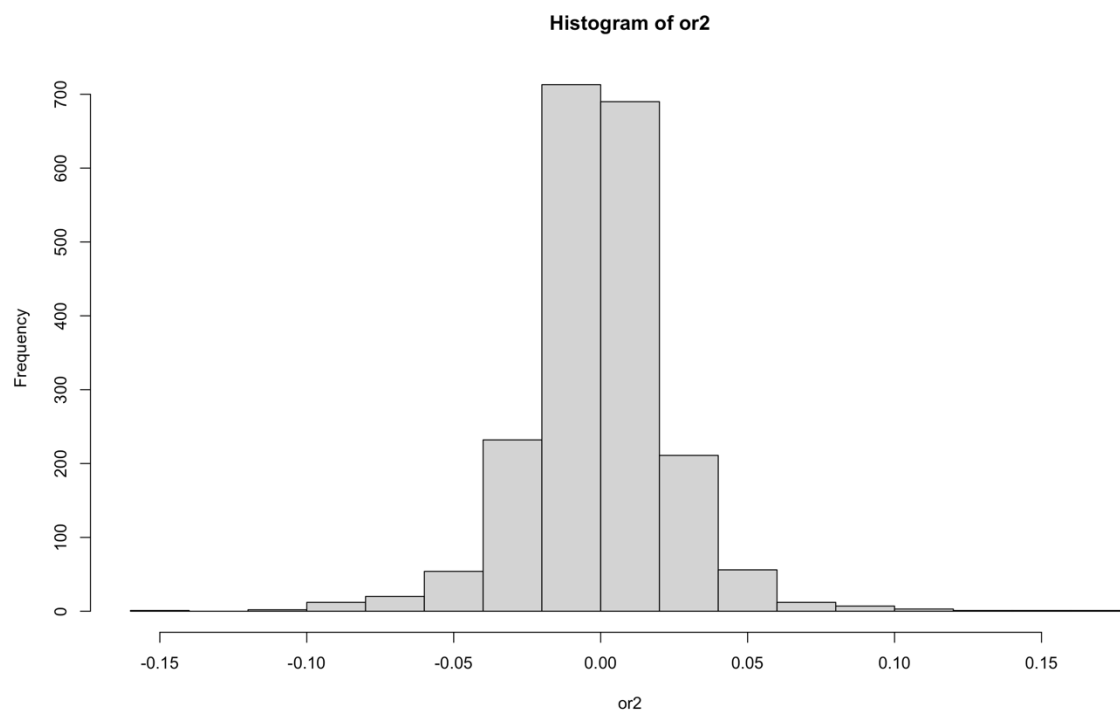
Ford:



EW:



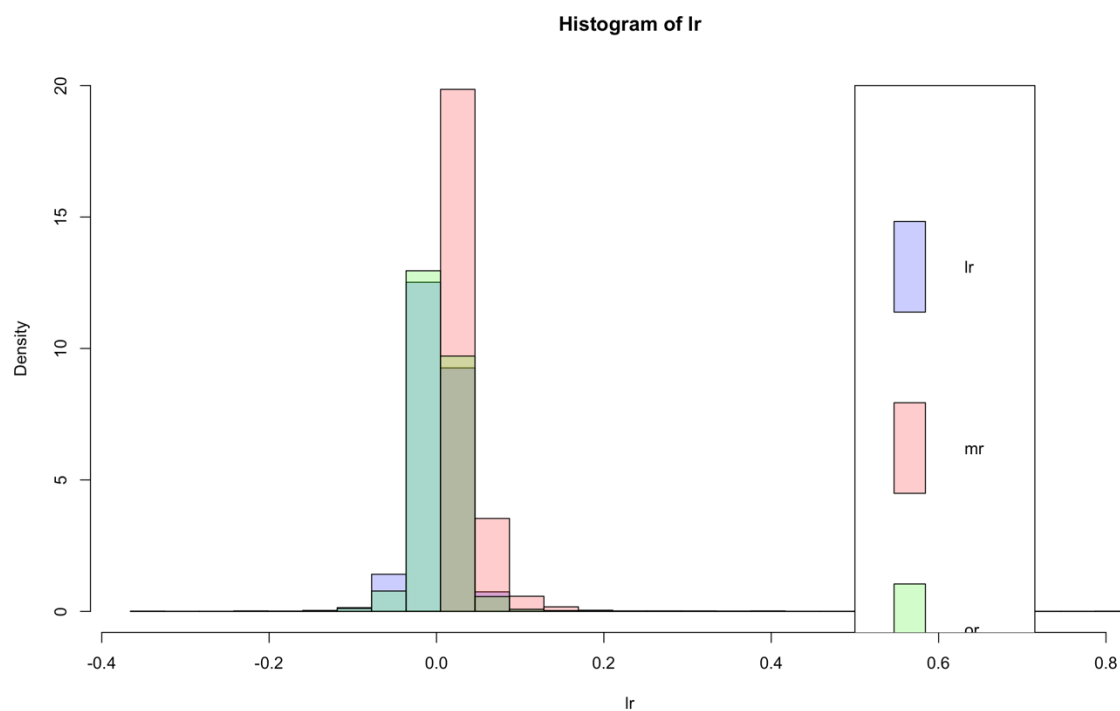
EXPE:



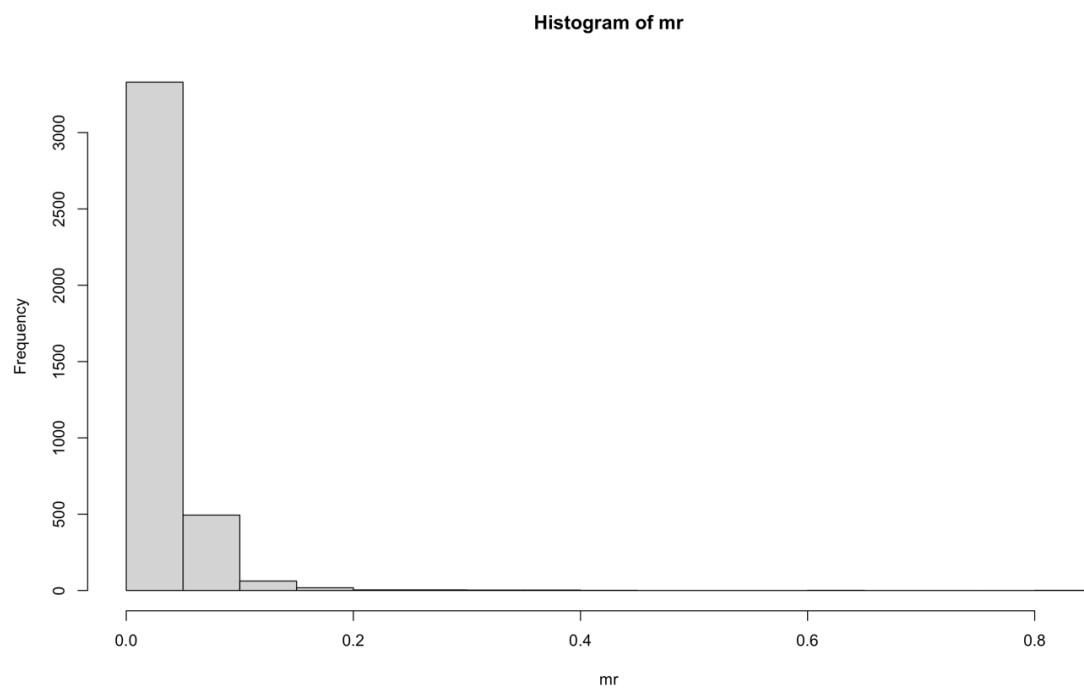
В пункті 9 я зробила гістограму для mr, lr та or для фірми 1 – Ford. Нижче наведений код з R, за допомогою якого я малювала гістограму:

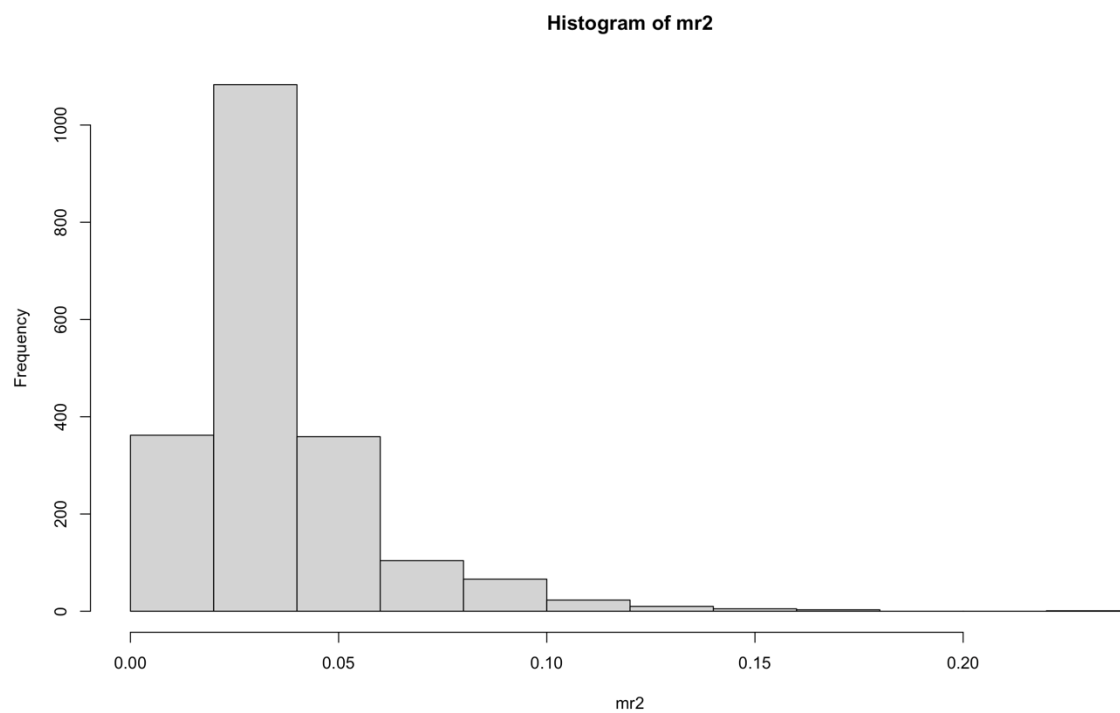
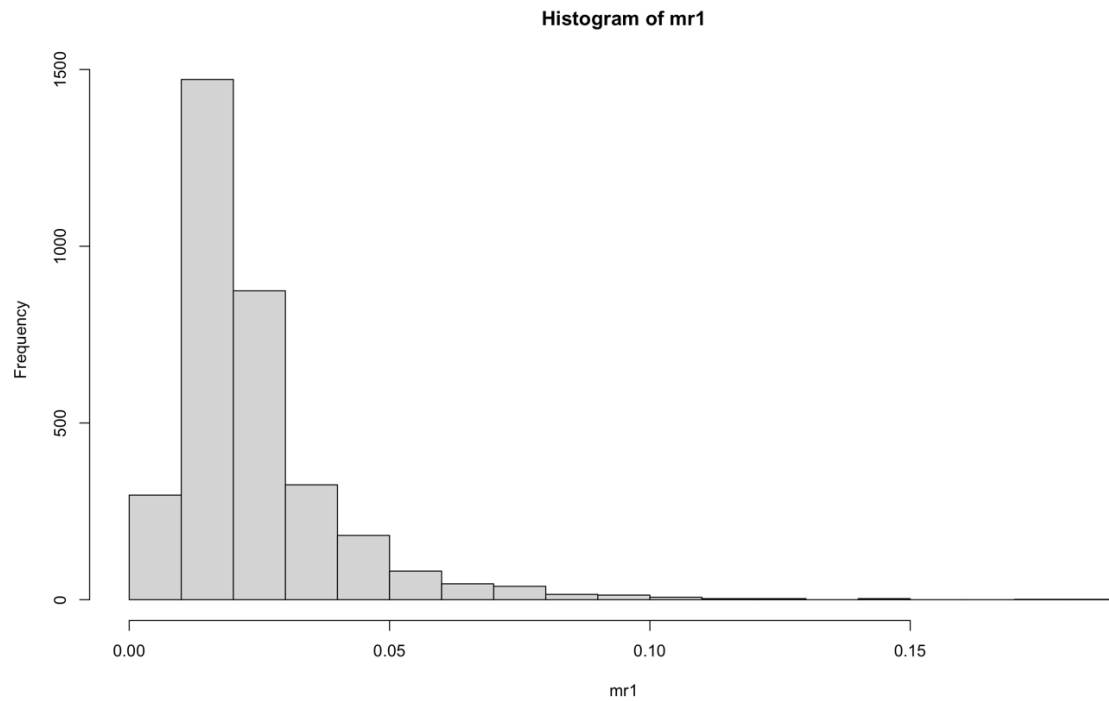
```
> brks<-seq(min(lr,mr,or)-0.01,max(lr,mr,or)+0.01,length.out=30)
> hist(lr,breaks=brks,probability=T,col=rgb(0,0,1,1/4),ylim=c(0,20))
> hist(mr,breaks=brks,probability=T,col=rgb(1,0,0,1/4), add=T)
> hist(or,breaks=brks,probability=T,col=rgb(0,1,0,1/4), add=T)
```

```
> legend(x=0.5,y=20,c("lr","mr","or"),fill=c(rgb(0,0,1,1/4),rgb(1,0,0,1/4),rgb(0,1,0,1/4)))
```



В пункті 10 зробимо гістограми *mr* для трьох компаній (Ford - *mr*, EW - *mr1*, EXPE - *mr2*):





Отже, у висновку можемо сказати, якщо розподіли – скриньки з вусами майже однакові, для якихось даних, то і гістограми також будуть дуже схожими.