Assignment-2

Manuel Bottino

2023-05-24

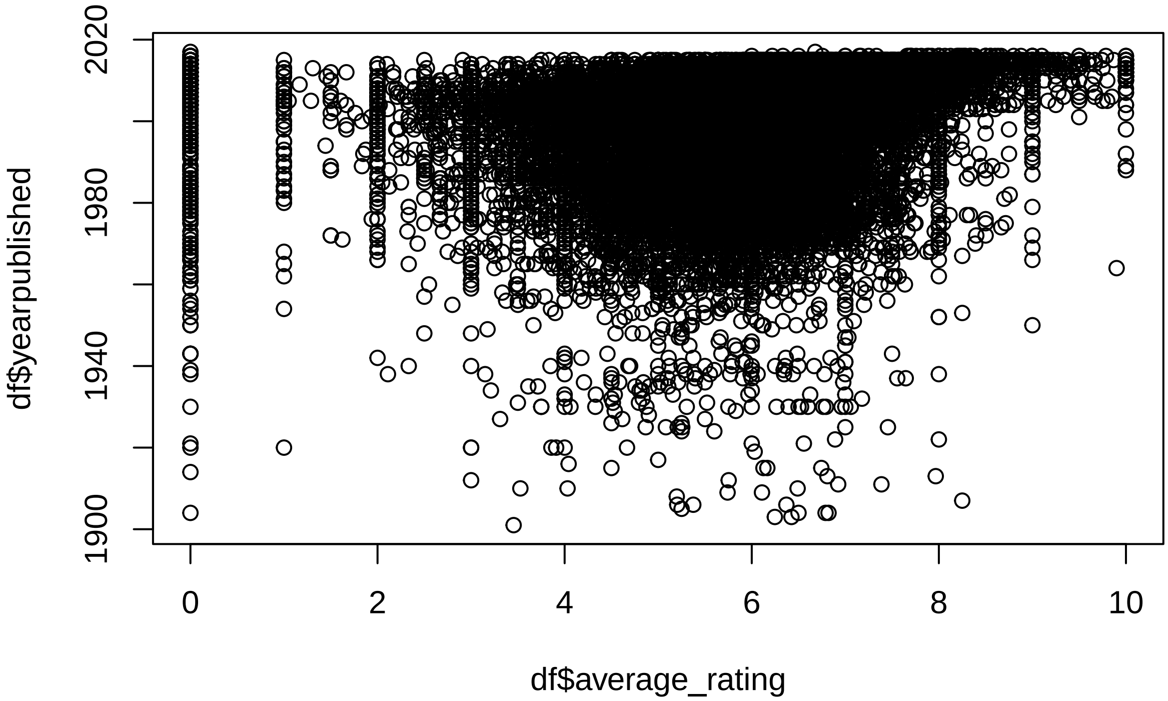
Contents

1 Excercise 1 - Risiko! is back, with friends

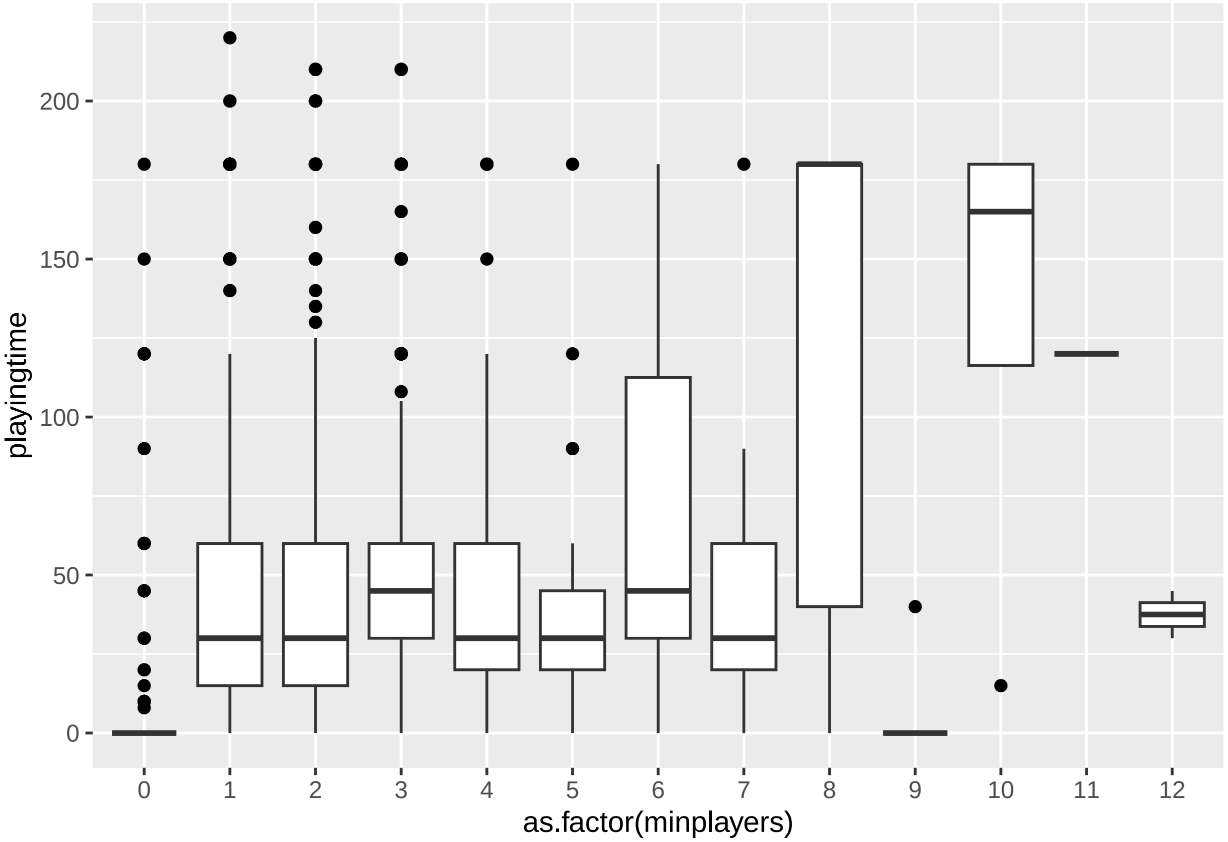
1.1 Point a)

```
df <- readRDS("~/Documents/GitHub/Computational-stat/Assignment 2/games_preprocessed.RDS")

plot(df$average_rating, df$yearpublished) # to see whether old board_games tend to be preferred more be
```



```r
ggplot(data=df, aes(y=playingtime, x=as.factor(minplayers)))+ # As more player play are needed to start
  geom_boxplot(outlier.colour="black", outlier.shape=16,
             outlier.size=2, notch=FALSE)
```



## Point b)

```
df= df[,-1:-2]
df=data.frame(scale(df))
```

```
cov(df)
```

```
##                 yearpublished  minplayers  maxplayers playingtime     minage
## yearpublished     1.00000000  0.03585626  0.03112545 -0.06572597 0.09314918
## minplayers        0.03585626  1.00000000  0.19166864  0.04429612 0.09851689
## maxplayers        0.03112545  0.19166864  1.00000000 -0.01204945 0.05626623
## playingtime      -0.06572597  0.04429612 -0.01204945  1.00000000 0.28407403
## minage            0.09314918  0.09851689  0.05626623  0.28407403 1.00000000
## average_rating    0.21657330 -0.01705543 -0.02509443  0.18694736 0.16096705
## total_owners      0.04441719  0.02724071  0.01574204  0.13281225 0.15929911
## average_weight    0.04899808 -0.06353175 -0.09918355  0.43645196 0.21685937
##                 average_rating total_owners average_weight
## yearpublished       0.21657330   0.04441719     0.04899808
## minplayers         -0.01705543   0.02724071    -0.06353175
## maxplayers         -0.02509443   0.01574204    -0.09918355
## playingtime         0.18694736   0.13281225     0.43645196
## minage              0.16096705   0.15929911     0.21685937
## average_rating      1.00000000   0.16819425     0.33090040
## total_owners        0.16819425   1.00000000     0.13106392
## average_weight      0.33090040   0.13106392     1.00000000
```

```
cor(df)
```

```
##                 yearpublished  minplayers  maxplayers playingtime     minage
## yearpublished     1.00000000  0.03585626  0.03112545 -0.06572597 0.09314918
## minplayers        0.03585626  1.00000000  0.19166864  0.04429612 0.09851689
## maxplayers        0.03112545  0.19166864  1.00000000 -0.01204945 0.05626623
## playingtime      -0.06572597  0.04429612 -0.01204945  1.00000000 0.28407403
## minage            0.09314918  0.09851689  0.05626623  0.28407403 1.00000000
## average_rating    0.21657330 -0.01705543 -0.02509443  0.18694736 0.16096705
## total_owners      0.04441719  0.02724071  0.01574204  0.13281225 0.15929911
## average_weight    0.04899808 -0.06353175 -0.09918355  0.43645196 0.21685937
##                 average_rating total_owners average_weight
## yearpublished       0.21657330   0.04441719     0.04899808
## minplayers         -0.01705543   0.02724071    -0.06353175
## maxplayers         -0.02509443   0.01574204    -0.09918355
## playingtime         0.18694736   0.13281225     0.43645196
## minage              0.16096705   0.15929911     0.21685937
## average_rating      1.00000000   0.16819425     0.33090040
## total_owners        0.16819425   1.00000000     0.13106392
## average_weight      0.33090040   0.13106392     1.00000000
```

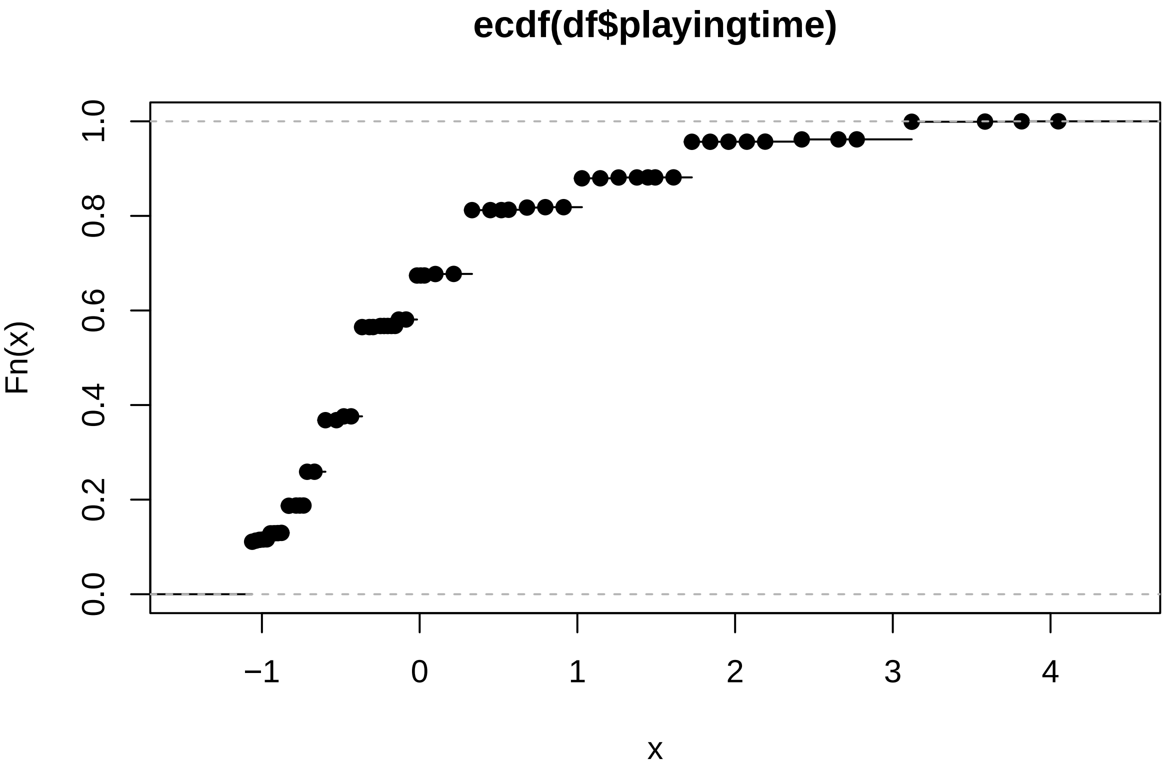
They are equal. Fonte: https://math.stackexchange.com/questions/3780344/under-what-conditions-will-the-covariance-matrix-be-identical-to-the-correlation

1.2 Point c)

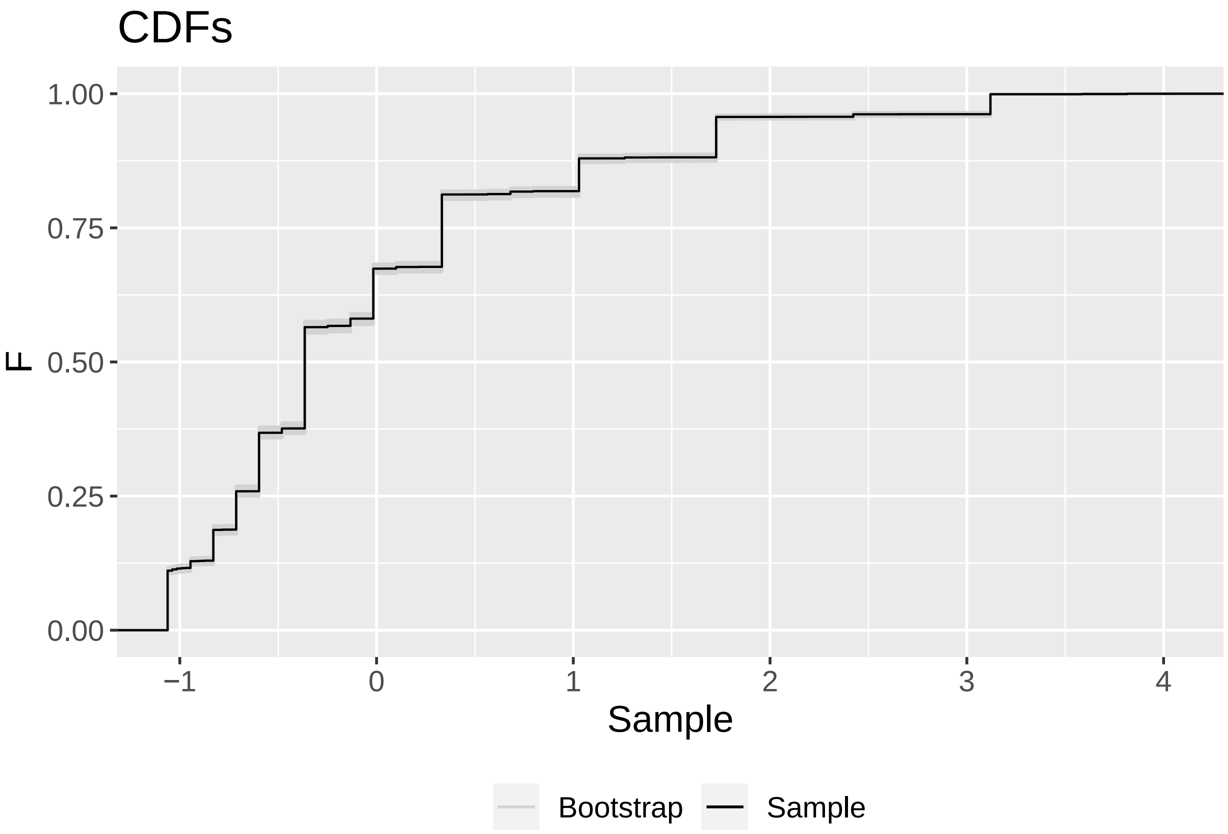
```r
bootstrap <- function(B,
                      x,
                      seed=Sys.time()){
  set.seed(seed)
  bootsamp <- replicate(B, sample(x,size = length(x),replace = T))
  return(bootsamp)
}

bootstrap_df_maxplayer =bootstrap(100, df$maxplayers)
bootstrap_df_playingtime =bootstrap(1000, df$playingtime)
bootstrap_df_minage =bootstrap(1000, df$minage)
bootstrap_df_average_rating =bootstrap(1000, df$average_rating)
bootstrap_df_average_weight =bootstrap(1000, df$average_weight)
bootstrap_df_total_owners =bootstrap(1000, df$total_owners)


plot(ecdf(df$playingtime))
```



```r
D2SUB    <- reshape2::melt(bootstrap_df_playingtime)
G2 = ggplot()+
  stat_ecdf(aes(x=D2SUB$value,group = D2SUB$Var2, col = "Bootstrap"))+
    #stat_function(aes(col="Population"),fun = function(w) cdf_distrib(w),lwd=1)+
    stat_ecdf(aes(x=df$playingtime,col="Sample"),lwd=0.4)+
    xlab("Sample")+ggtitle("CDFs")+ylab("F")+
      scale_color_manual("",values = c("lightgray",1,4))+
  theme(text = element_text(size=14),legend.position = "bottom")
G2
```


CDFs
1.00
0.75
0.50
0.25
0.00
F
−1
0
1
2
3
4
Sample
Bootstrap
Sample

1.3 Point d)

1.4 Point e)

1.5 Point f)

1.6 Point g)

1.7 Point h)

2 Exercise 2 - We need some music!

2.1 Point a)

2.2 Point b)

2.3 Point c)

2.4 Point d)

2.5 Point e)

2.6 Point f)

2.7 Point g)