Finanças Quantitativas: Lista 5

Lucas Moschen

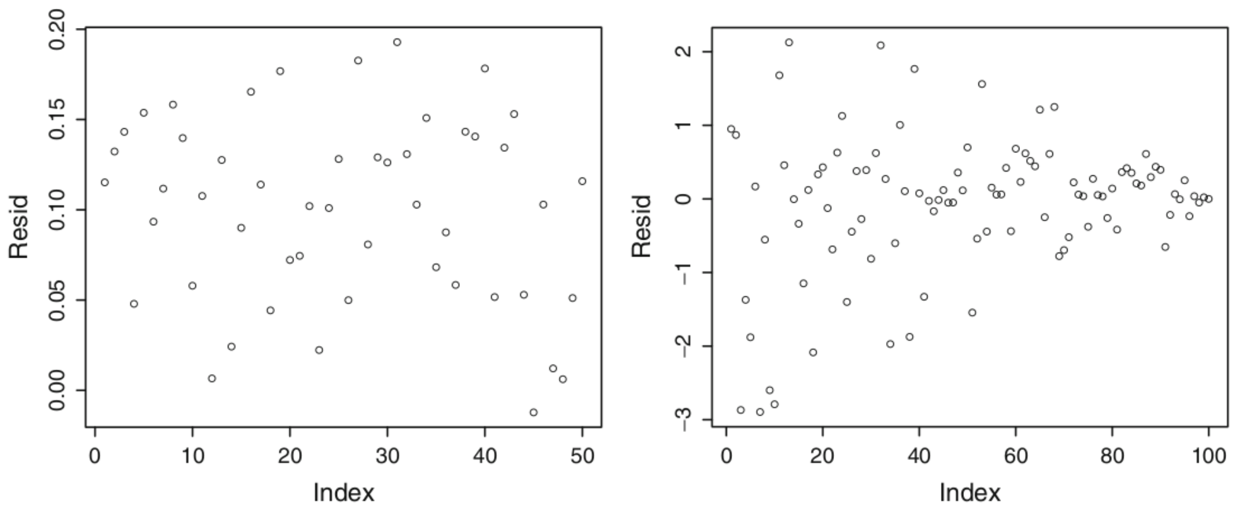
24 de maio de 2020

Exercício 1

Problema 4.1 (Carmona)

Item 1

Considere a seguinte imagem do livro.



Podemos observar que o plote da esquerda tem o eixo X marcado pelo índice de determinadas observações e o eixo Y valores, quase todos, estritamente positivos (com excessão aparente de um). Isso não é a característica de resíduos brutos, onde $\hat{\epsilon}_i = y - \hat{y}_i$, dado que a média desses resíduos é de longe 0. Como temos uma quantidade relativamente grande de observações e esperamos que os erros tenham média 0 ao modelarmos o erro quadrático, vemos que não é possível que esse plote represente isso.

Item 2

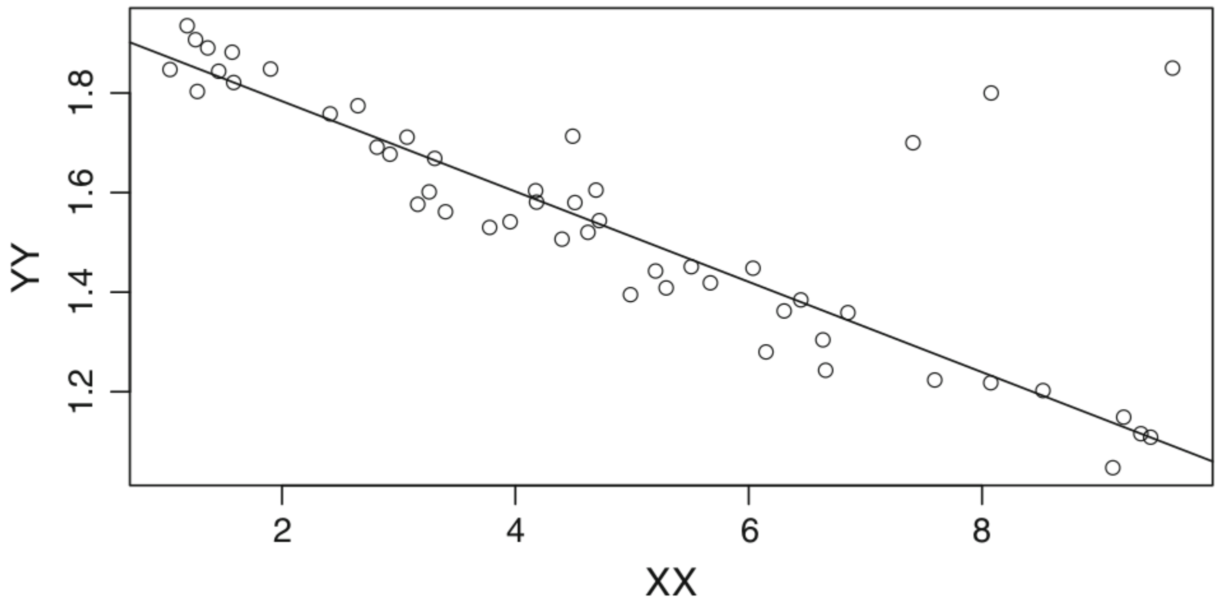
Ainda sobre a figura acima, obseve que quanto maior o índice, mais os resíduos se concentram em torno de 0. Como eles se concentram, podemos dizer, com base no gráfico que quanto maior o índice i, menor a variância em torno da média do resíduo. Conhecemos nesse capítulo que os resíduos tem desvio padrão

$$\sigma_{\epsilon_i} = \sigma\sqrt{1 - h_{i,i}}$$

Assim, $h_{i,i}$ tem um comportamento crescente em relação ao índice, isto é, a diagonal tem termos em uma sequência crescente.

Item 3

Agora considere essa imagem:



Parece-me que essa linha é uma **regressão linear com desvios absolutos**, visto que ela não se influencia tanto com os três *outliers* na parte de cima do gráfico. Isto acontece, pois a regressão $\mathcal{L}_1$ é menos sensível a *outliers*.

Problema 4.11 (Carmona)

Neste exercício, pretendemos analizar um exemplo de regressão não linear. Os dados são de uma droga *Puromycin* e possui uma tabela com três variáveis de um experimento biomédico em células, tratadas ou não. Denotamos y como a velocidade inicial da reação, enquanto x é a concentração da enzima. Pela relação de Michaelis-Menten, onde V_a é a velocidade assintótica e K uma constante:

$$y = \phi(x) = V_a \frac{x}{x + K}$$

Itens 1 e 2

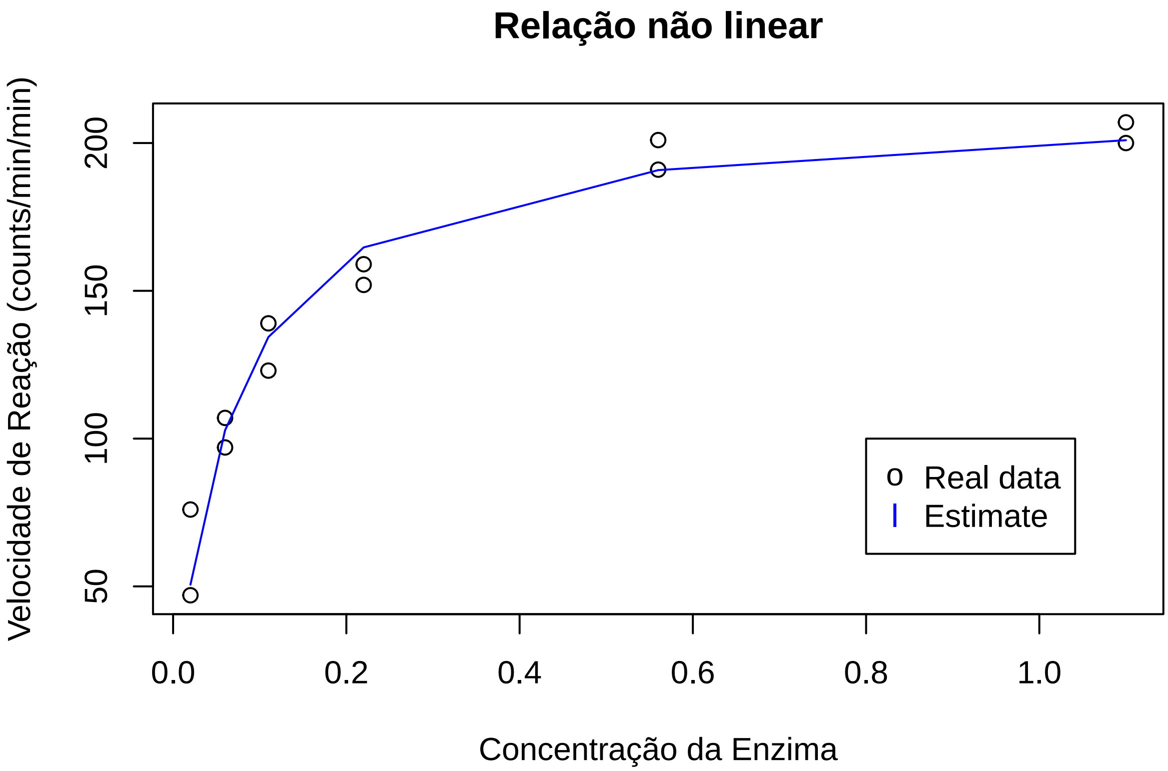
```r
data("Puromycin", Puromycin)
attach(Puromycin)
y = Puromycin[state == "treated",]$rate
x = Puromycin[state == "treated",]$conc
plot(x,y, xlab = "Concentração da Enzima", ylab = "Velocidade de Reação (counts/min/min)",
    main = "Relação não linear")

y_fit = nls(y ~ Va*(x/(x + K)), start = c(Va = 200, K = 0.1))
print("These are the estimated parameters:")
```

```
## [1] "These are the estimated parameters:"
```

```r
print(coef(y_fit))
```

```
##          Va           K
## 212.6836296   0.0641211
```

```r
lines(x, fitted(y_fit), col = "blue")

legend(0.8, 100, legend = c("Real data","Estimate"), pch = c("o","l"), col = c("black", "blue"))
```



Problema 4.12 (Carmona)

Nesse exercício, usadmos a família generalizada Vasicek para parametrizar o termo estrutural da taxa de juros. Definimos $Y_{GV}(x, \theta)$ como:

$$Y_{GC}(x, \theta) = \theta_1 - \theta_2\theta_4 \frac{1 - e^{-x/\theta_4}}{x} + \theta_3\theta_4 \frac{(1 - e^{-x/\theta_4})^2}{4x}$$

Item 1

Item 2

Item 3

Exercício 2

Considere o modelo de regressão linear múltipla:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon$$

E o custo associado a β, que queremos minimizar, é:

$$\begin{aligned}
\mathcal{L}_2(\beta) &= ||\boldsymbol{Y} - \boldsymbol{X}\beta||^2 = \langle \boldsymbol{Y} - \boldsymbol{X}\beta, \boldsymbol{Y} - \boldsymbol{X}\beta \rangle \\
&= \langle \boldsymbol{Y}, \boldsymbol{Y} \rangle - 2\langle \boldsymbol{Y}, \boldsymbol{X}\beta \rangle + \langle \boldsymbol{X}\beta, \boldsymbol{X}\beta \rangle \\
&= \langle \boldsymbol{Y}, \boldsymbol{Y} \rangle - 2\langle \boldsymbol{X}\beta, \boldsymbol{Y} \rangle + \langle \boldsymbol{X}\beta, \boldsymbol{X}\beta \rangle \\
&= \boldsymbol{Y}^T\boldsymbol{Y} - 2\beta^T\boldsymbol{X}^T\boldsymbol{Y} + \beta^T\boldsymbol{X}^T\boldsymbol{X}\beta
\end{aligned}$$

Para minimizar esse valor, primeiro procuramos os pontos críticos. Nesse caso, veja que a primeira expressão independe do vetor β e, portanto, sua derivada será 0. A segunda expressão, temos uma combinação dos elementos do vetor β, logo, ao derivar parcialmente em relação a cada valor, obtemos a expressão equivalente e, por isso, a derivação é linear. Na última expressão, veja que $\boldsymbol{X}^T\boldsymbol{X}$ é uma expressão com os valores de β quadráticos e, portanto, a expressão se segue:

$$\frac{d}{d\beta}\mathcal{L}_2(\hat{\beta}) = -2\boldsymbol{X}^T\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X}\hat{\beta} = 0$$

Desta forma, como o posto de $\boldsymbol{X}$ é completo $\boldsymbol{X}^T\boldsymbol{X}$ tem colunas linearmente independentes e é, portanto, invertível. Desta maneita, temos uma solução única e:

$$\boldsymbol{X}^T\boldsymbol{X}\hat{\beta} = \boldsymbol{X}^T\boldsymbol{Y} \implies \hat{\beta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

Nesse obtemos que $\hat{\beta}$ é um ponto crítico. Para averiguar se é argumento mínimo, façamos:

$$\frac{d^2}{d\beta^2}\mathcal{L}_2(\hat{\beta}) = 2\boldsymbol{X}^T\boldsymbol{X}$$

A matriz $\boldsymbol{X}^T\boldsymbol{X}$ é simétrica e, $\forall x \in \mathbb{R}^{p+1}, x^T\boldsymbol{X}^T\boldsymbol{X}x = \langle \boldsymbol{X}x, \boldsymbol{X}x\rangle \geq 0$ e será igual a 0 somente se $\boldsymbol{X}x = 0$. Como X tem posto completo, ele tem espaço anulado com dimensão 0 e, poranto, $anul(X) = \{0\}$ e, se $\boldsymbol{X}x = 0, x = 0$. Concluo, então que essa matriz é estritamente positiva e, desta maneira, $\hat{\beta}$ é de fato um mínimo da expressão.

Exercício 3

```r
library(BatchGetSymbols) # get financial data
```

Item a

Esses são os *trading symbols* das componentes que integram a Ibovespa, segundo a página oficial.

```r
df.ibov <- GetIbovStocks()
print(df.ibov$tickers)
```

```
##  [1] "ABEV3"  "AZUL4"  "B3SA3"  "BBAS3"  "BBDC3"  "BBDC4"  "BBSE3"  "BEEF3"
##  [9] "BPAC11" "BRAP4"  "BRDT3"  "BRFS3"  "BRKM5"  "BRML3"  "BTOW3"  "CCRO3"
## [17] "CIEL3"  "CMIG4"  "COGN3"  "CPFE3"  "CRFB3"  "CSAN3"  "CSNA3"  "CVCB3"
## [25] "CYRE3"  "ECOR3"  "EGIE3"  "ELET3"  "ELET6"  "EMBR3"  "ENBR3"  "ENGI11"
## [33] "EQTL3"  "FLRY3"  "GGBR4"  "GNDI3"  "GOAU4"  "GOLL4"  "HAPV3"  "HGTX3"
## [41] "HYPE3"  "IGTA3"  "IRBR3"  "ITSA4"  "ITUB4"  "JBSS3"  "KLBN11" "LAME4"
## [49] "LREN3"  "MGLU3"  "MRFG3"  "MRVE3"  "MULT3"  "NTCO3"  "PCAR3"  "PETR3"
## [57] "PETR4"  "QUAL3"  "RADL3"  "RAIL3"  "RENT3"  "SANB11" "SBSP3"  "SULA11"
## [65] "SUZB3"  "TAEE11" "TIMP3"  "TOTS3"  "UGPA3"  "USIM5"  "VALE3"  "VIVT4"
## [73] "VVAR3"  "WEGE3"  "YDUQ3"
```

Item b

Tomo os dados históricos do período de 01/01/2015 até o dia de 31/12/2019 cada uma das 66 ações da Ibovespa. Daquelas que printei acima e integram atualmente a bolsa, as ações de "AZUL4.SA", "BPAC11.SA", "BRDT3.SA", "CRFB3.SA", "GNDI3.SA", "HAPV3.SA", "IRBR3.SA", "NTCO3.SA" e "RAIL3.SA" (9 ao total) não possuiam atividade nesse período. Por simplicidade e para analisar um tempo mais complexo, eu as retiro das analisadas. Além disso, eu integro as ações em uma *dataframe* que conterá também a informação dos dados do índice da Ibovespa.

```
first.date <- "2015-01-01"
last.date <- "2019-12-31"

df <- BatchGetSymbols("^BVSP", first.date = first.date, last.date = last.date)
stocks <- data.frame(BVSP = df$df.tickers$price.close)

for (ticker in df.ibov$tickers) {
  tickerSA = paste(ticker, ".SA", sep = "")
  df <- BatchGetSymbols(tickerSA, first.date =   first.date, last.date = last.date)
  if(length(stocks$BVSP) != length(df$df.tickers$price.close)){next}
  stocks <- cbind(stocks, df$df.tickers$price.close)
}

# Removing NA values
stocks <- stocks[complete.cases(stocks),]
```

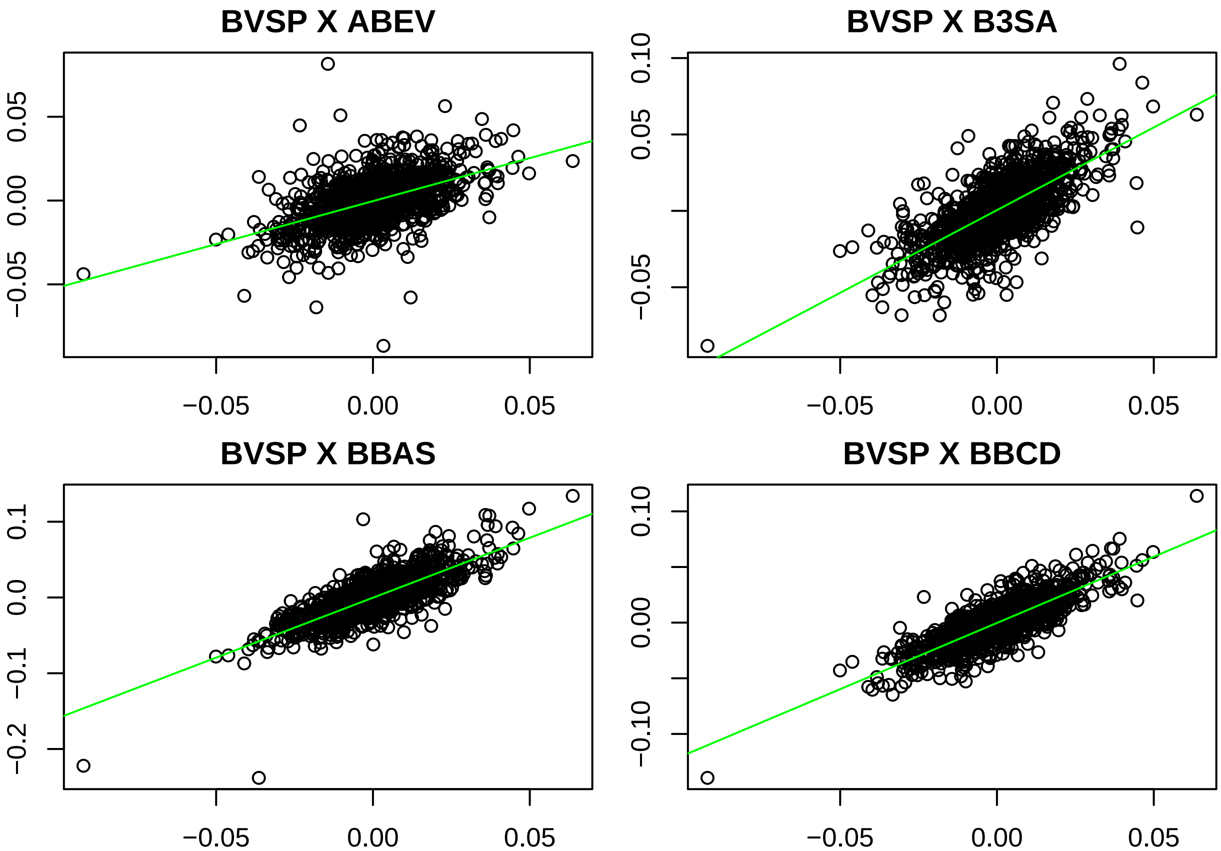
Item c

```
# Primeiro, vamos calcular os log-retornos de cada ação.
stocks <- data.frame(diff(log(as.matrix(stocks))))

# Vou supor que o portfolio de mercado é dado pelo índice da Ibovespa.
# Vou considerar a taxa de juros predominante nesse período e tomar o excess return.
r = 0.00018985
stocks <- stocks - r

# Renomear corretamente
names(stocks) <- append(c("BVSP"), df.ibov$tickers[-c(2,9,11,21,36,39,43,54,60)])

# Regressão Linear de Mínimos Quadrados
coefficients <- data.frame(BVSP = c("Intersept","Slope"))
for(i in 2:length(names(stocks))){
  model <- lsfit(stocks$BVSP, stocks[names(stocks)[i]])
  coefficients[names(stocks)[i]] = model$coefficients
}

# Destaco os coeficientes estimados para as 4 primeiras ações:
print(coefficients[,1:6])
```

```
##        BVSP         ABEV3       B3SA3         BBAS3         BBDC3         BBDC4
## 1 Intersept -0.0003308279 0.000473314 -0.0003204702 -0.0002761612 -0.0002738513
## 2     Slope  0.5132610403 1.082581334  1.5818037297  1.1906179748  1.2669861505
```

```
# Plotting dessas quatro primeiras ações, cada uma em seu respectivo eixo y contra o retorno de mercado
par(mfrow = c(2,2), mar = c(2,2,2,1))

# Plotting dos gráficos
plot(stocks$BVSP, stocks$ABEV3, main = "BVSP X ABEV")
abline(a = coefficients[1,2], b = coefficients[2,2], col = "green")
plot(stocks$BVSP, stocks$B3SA3, main = "BVSP X B3SA")
abline(a = coefficients[1,3], b = coefficients[2,3], col = "green")
plot(stocks$BVSP, stocks$BBAS3, main = "BVSP X BBAS")
abline(a = coefficients[1,4], b = coefficients[2,4], col = "green")
plot(stocks$BVSP, stocks$BBDC3, main = "BVSP X BBCD")
```

```r
abline(a = coefficients[1,5], b = coefficients[2,5], col = "green")
```



É interessante que visualmente já percebemos que os interceptos são muito próximos a 0, como esperávamos. De fato, a média do valor absoluto é mostrada na figura. Também vemos como de fato os dados tem um comportamento linear. Além disso, aprendemos que a estimativa da inclinação da reta ($\hat{\beta}_j$) para cada ativo é conhecida como investimento beta e mede a sensitividade do retorno com as variações dos retornos do portifólio de mercado. Assim, quando esse valor é maior do que 1, temos ativos com mais risco. Para visualizar, considere o seguinte gráfico, ele indica as diferenças de cada dessa estimativa para cada ativo.

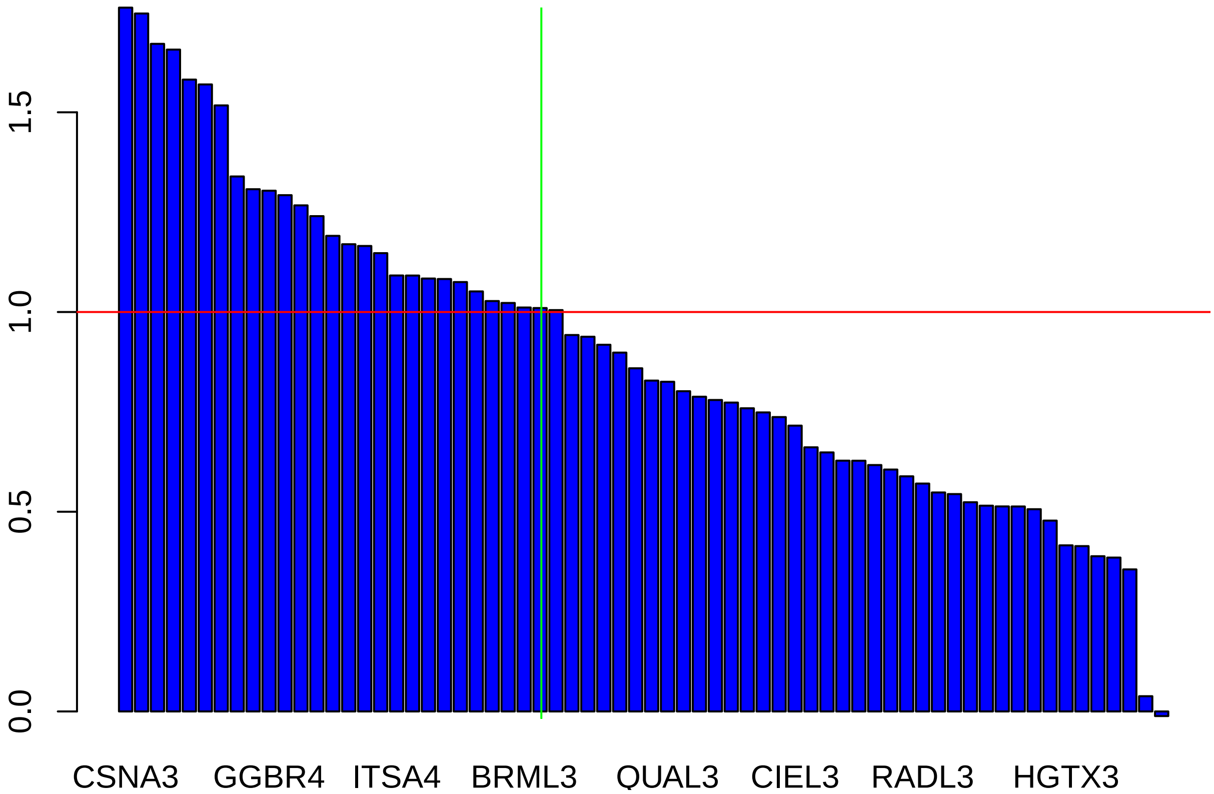
```r
print("Valor médio absoluto do intersepto")
```

```
## [1] "Valor médio absoluto do intersepto"
```

```r
print(mean(abs(as.matrix(coefficients[1,-1]))))
```

```
## [1] 0.0003920817
```

```r
par(mar = c(2,2,2,1))
barplot(as.matrix(sort(coefficients[2,-1], decreasing = TRUE)), col = "blue")
abline(h = 1.0, col = "red")
abline(v = 32.0, col = "green")
```



Nesse gráfico, podemos perceber que aproximadamente 40% dos ativos tem investimento beta maior do que 1. Além disso, dois ativos demonstraram que a estimativa da inclinação é negativo. Mas os valores são próximos de 0, o que indica possível atuação de *outliers*.

Item d

Item e