

# Estatística Aplicada

## Projecto intercalar

Pedro Magalhães (200202298)

2022-01-03

Declaro que o presente relatório é de minha autoria e não foi utilizado previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escurpulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referência. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

## Introdução

### Definição da questão

Investir no mercado de capitais implica tomar de decisões de compra de ativos financeiros num contexto de incerteza. De forma a tomarem decisões de investimento em linha com o seu perfil de risco e estratégias diferentes investidores recorrem a diferentes técnicas para entender e prever a evolução do preço do ativo transacionado e identificar oportunidades de entrada (compra) e saída (venda).

No caso de ações cotadas, é comum a utilização da análise técnica que consiste em modelar da evolução temporal do preço de uma ação em face da evolução de determinados indicadores técnicos.

Este projeto estuda a relação linear entre a cotação diária das ações da Microsoft no mercado de valores de Nova York e alguns indicadores técnicos normalmente utilizados neste tipo de análises.

Foram utilizados dados públicos da cotação diária de abertura, de fecho, máxima e mínima entre 01-01-2015 e 31-12-2016 no mercado de Nova York assim como informação das demonstrações financeiras anuais divulgados no site Kaggle.com.

### Critérios de selecção do modelo

O modelo seleccionado é que aquela que apresenta o maior  $R^2$  ajustado e menor  $AIC$ .

### Pressupostos e limitações

Apesar deste projeto estudar a relação linear entre a cotação diária de encerramento e vários indicadores, do ponto de vista da teoria da análise de mercados e da avaliação de carteiras o uso de regressão constitui uma limitação e simplificação. Apesar de válida do ponto de vista académico, na realidade um modelo definido desta forma não permite retirar conclusões adequadas para suportar qualquer conclusão ou decisão.

De entre as limitações e pressupostos assumidos, o mais importante é o pressuposto de **independência entre as observações e consequentemente dos resíduos do modelo**. Dada a natureza dos dados utilizados

é de esperar que as observações  $y_s$  representando a cotação de fecho do momento S e  $y_{s-h}$  representando a observação no momento S-H **não são independentes** estando portanto perante uma série temporal.

O estudo dos efeitos autorregressivos contidos na amostra da cotação está fora do âmbito deste projeto pelo que iremos desconsiderar este efeito no modelo e apenas levar em linha de conta sinais externos à série os quais serão dados pelos indicadores técnicos calculados.

## Análise exploratória dos dados

### Descrição da base de dados

```
##          date          close          eps          ema_15
## Min.      :2015-01-02   Min.      :40.29   Min.      :2.12   Min.      :41.48
## 1st Qu.:2015-07-05   1st Qu.:46.23   1st Qu.:2.12   1st Qu.:45.99
## Median :2016-01-02   Median :51.33   Median :2.12   Median :51.42
## Mean      :2016-01-01   Mean      :50.99   Mean      :2.12   Mean      :50.78
## 3rd Qu.:2016-07-02   3rd Qu.:55.72   3rd Qu.:2.12   3rd Qu.:54.87
## Max.      :2016-12-30   Max.      :63.62   Max.      :2.12   Max.      :62.65
##          ema_30          evwma_30          rsi          roc
## Min.      :41.94   Min.      :42.59   Min.      :21.82   Min.      : -0.0970986
## 1st Qu.:45.76   1st Qu.:45.45   1st Qu.:45.19   1st Qu.: -0.0066319
## Median :51.71   Median :51.46   Median :52.76   Median : 0.0001881
## Mean      :50.57   Mean      :50.07   Mean      :53.03   Mean      : 0.0005774
## 3rd Qu.:54.17   3rd Qu.:53.01   3rd Qu.:61.38   3rd Qu.: 0.0076617
## Max.      :61.82   Max.      :60.35   Max.      :84.19   Max.      : 0.0994130
##          momentum          per_30          per_30_mean          diff_close_open
## Min.      : -4.35000   Min.      :19.78   Length:504   Length:504
## 1st Qu.: -0.34500   1st Qu.:21.58   Class :character   Class :character
## Median : 0.01000   Median :24.39   Mode  :character   Mode  :character
## Mean      : 0.03113   Mean      :23.86
## 3rd Qu.: 0.36250   3rd Qu.:25.55
## Max.      : 4.84000   Max.      :29.16
## volatility_cat
## Length:504
## Class :character
## Mode  :character
##
##
##
```

A base de dados utilizada contém os seguintes dados:

**date:**

Índice diário entre 2015-01-02 e 2016-12-30

**close:**

Cotação de fecho da sessão diária ajustado por eventuais splits de ações ocorridos durante o ano. (um split de ações significa aumentar o nº de ações em circulação mantendo o mesmo market cap o que gera uma variação abrupta na cotação no momento da realização e no choque de mercado)

```
## Warning in prettyNum(.Internal(format(x, trim, digits, nsmall, width, 3L, :
## 'big.mark' and 'decimal.mark' are both '.', which could be confusing
```

## Cotação de fecho

Microsoft



Fig 1: evolução da cotação de fecho das ações da Microsoft no mercado de valores de NY em 2015 e 2016

```
## <ScaleContinuousDate>  
## Range:  
## Limits: 1.64e+04 -- 1.72e+04
```

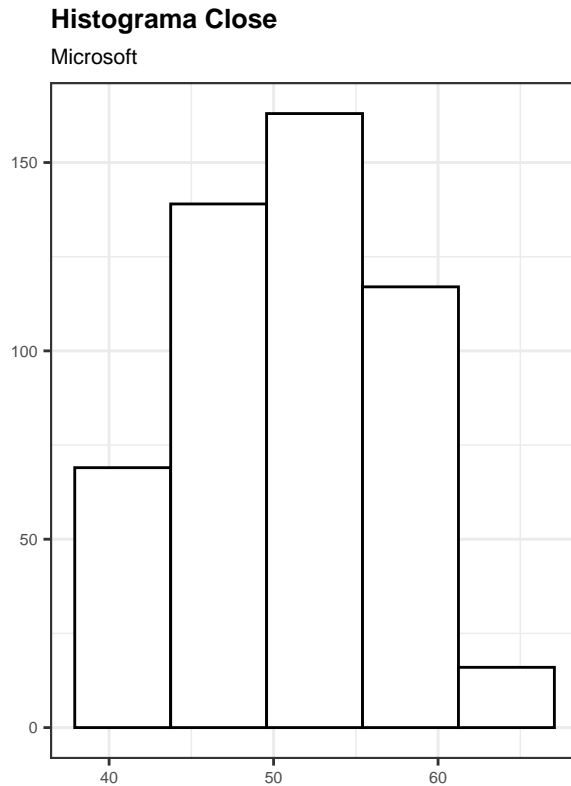


Fig 2: Histograma de cotação de fecho

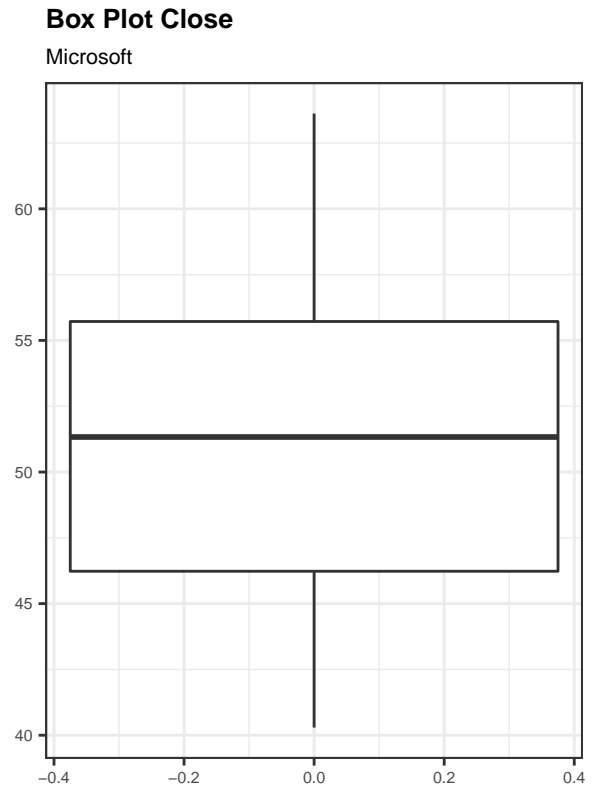


Fig 3: Box Plot de cotação de fecho

Esta é a Variável dependente que o modelo procura explicar. A figura 2 e 3 descrevem graficamente a distribuição da amostra. A análise gráfica não parece sugerir que existam outliers.

**low:** mínimo diário a que a acção foi transaccionada

**high:** máximo diário a que a acção foi transaccionada

**volume:** volume de transacções diário

**eps:** Earning per Share anual conforme reportado pelas demonstrações financeiras de final do ano e usando o seguinte rácio

$$EPS = \frac{\text{Net Income}}{\text{Outstanding Shares}}$$

Os dados de low, high, volume e eps foram utilizados para cálculo dos rácio técnicos, mas não serão utilizados para efeito do modelo.

**ema\_15 e ema\_30**

Exponential Moving Average para a janela de 15 e 30 dias respectivamente.

```
## Warning in prettyNum(.Internal(format(x, trim, digits, nsmall, width, 3L, :
## 'big.mark' and 'decimal.mark' are both '.', which could be confusing
```

## Cotação de fecho

Microsoft

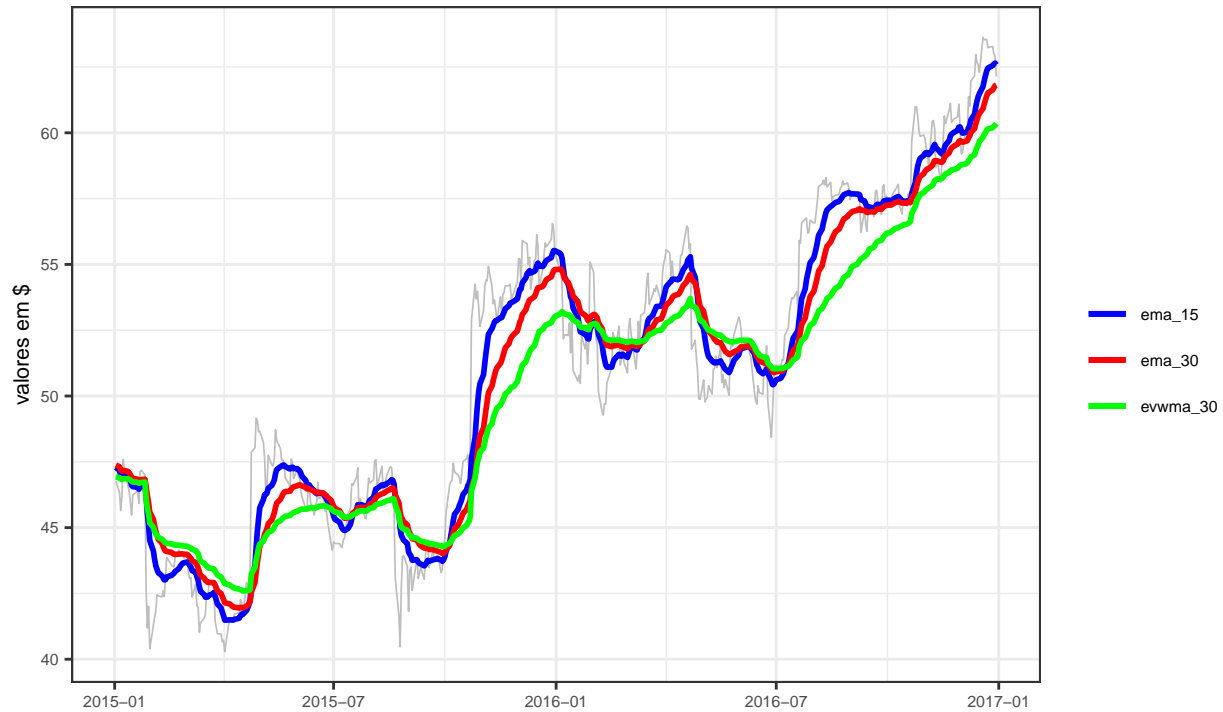


Fig 1: evolução da cotação de fecho das ações da Microsoft no mercado de valores de NY em 2015 e 2016

```
## <ScaleContinuousDate>  
## Range:  
## Limits: 1.64e+04 -- 1.72e+04
```

### ewwma\_30

Exponential Moving Average Volume Weighed. Semelhante ao indicador anterior mas corrigido pelo volume de transação diário

### rsi:

Relative Strength Index mensura a força do movimento de subida ou descida em face de um intervalo de evolução. Diferenças entre a evolução deste indicador e a cotação podem indicar uma inversão da tendência

```
## `geom_smooth()` using formula 'y ~ x'
```

### Histograma Rsi

Microsoft

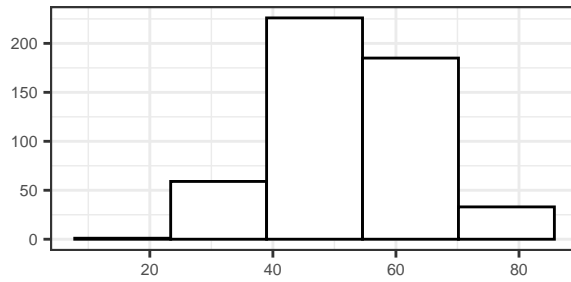


Fig 4: Histograma de rsi

### Box Plot Rsi

Microsoft

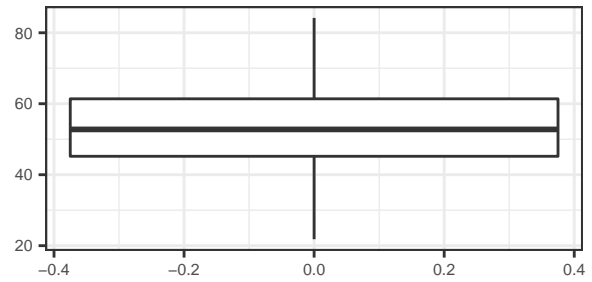


Fig 5: Box Plot de rsi

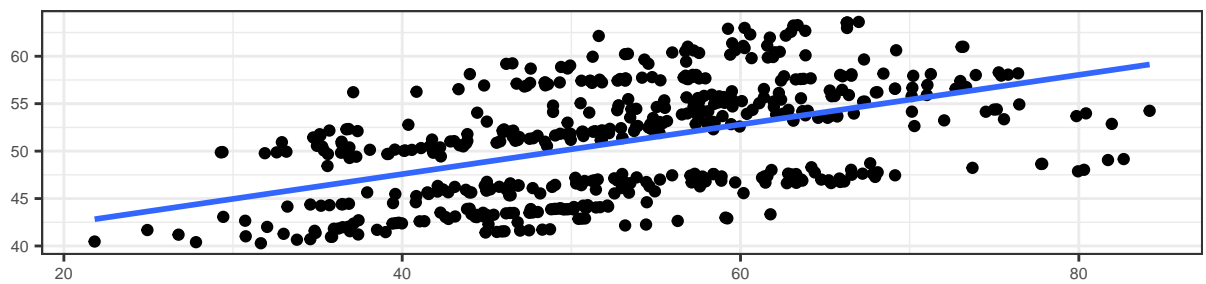


Fig 6: relação com fecho

A análise gráfica não indicia a existência de outliers e aparenta existir uma relação linear com a variável dependente.

**roc:**

Percentagem de diferença entre n observações

```
## `geom_smooth()` using formula 'y ~ x'
```

### Histograma roc

Microsoft

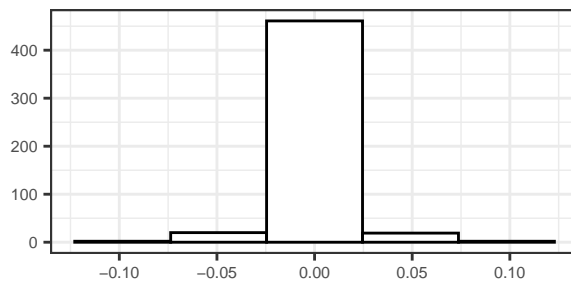


Fig 4: Histograma de roc

### Box Plot roc

Microsoft

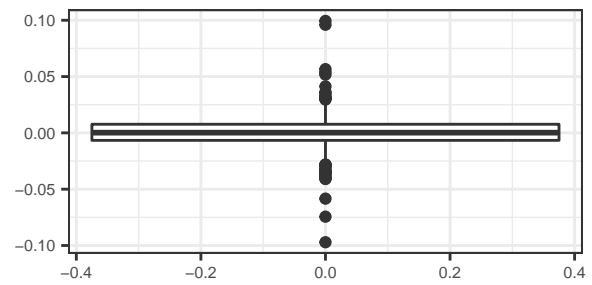


Fig 5: Box Plot de roc

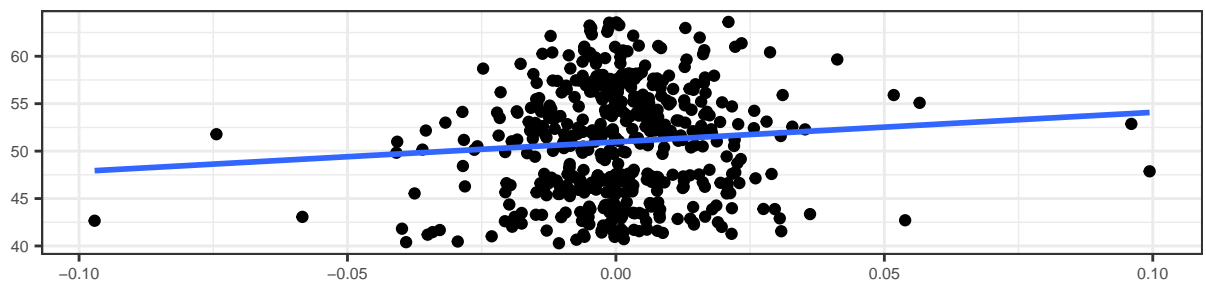


Fig 6: relação com fecho

A análise gráfica aponta para a existência de eventuais outliers que necessitariam de ser analisados mais detalhadamente. Comparada com a variável dependente verificamos que existem pontos que indiciam ter um “leverage” elevado.

### momentum:

Diferença entre n observações (lag)

```
## `geom_smooth()` using formula 'y ~ x'
```

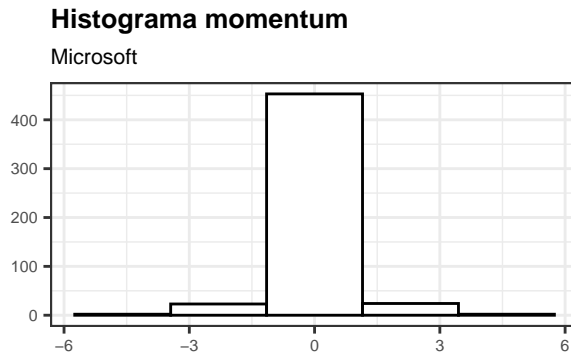


Fig 4: Histograma de momentum

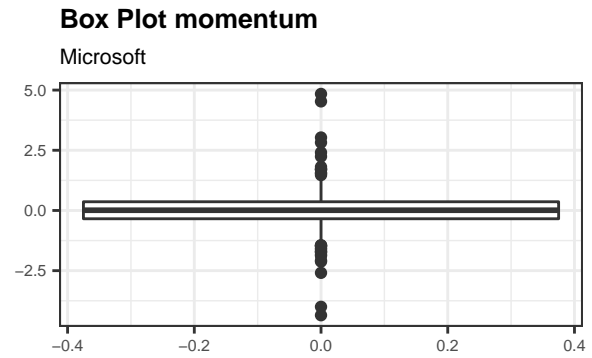


Fig 5: Box Plot de momentum

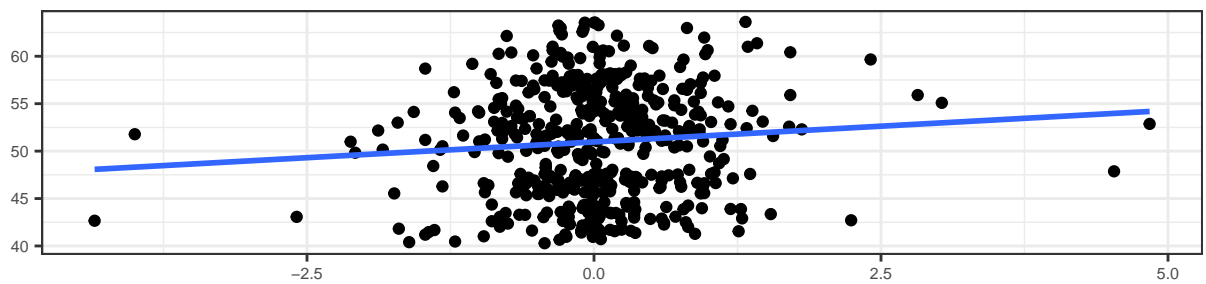


Fig 6: relação com fecho

Conclusões idênticas às dp ROC

**per\_30:**

Price Earning Ratio 30 days trailing

**per\_30\_mean:**

Variável categórica que assume valores se per\_30 for superior à média dos dois anos então “Higher than average” caso contrário “Lower than average”

**volatility:**

Variância Close-to-Close para uma janela de 15 dias.

```
## `geom_smooth()` using formula 'y ~ x'
```



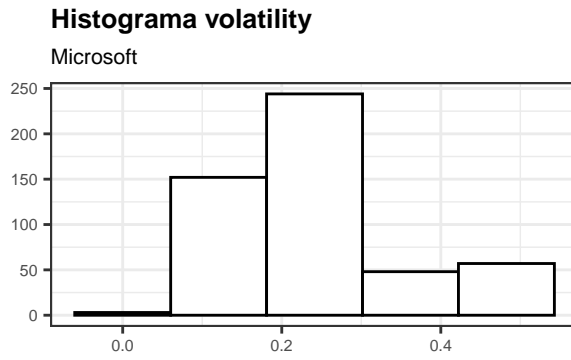


Fig 4: Histograma de volatility

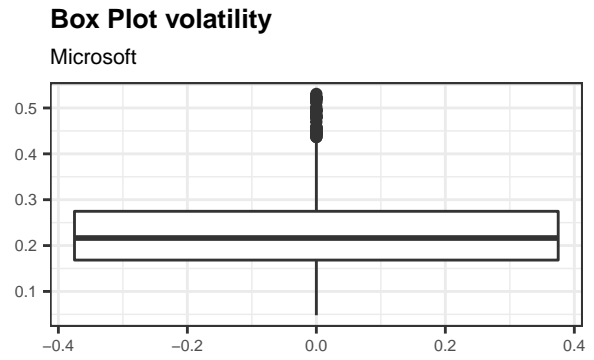


Fig 5: Box Plot de volatility

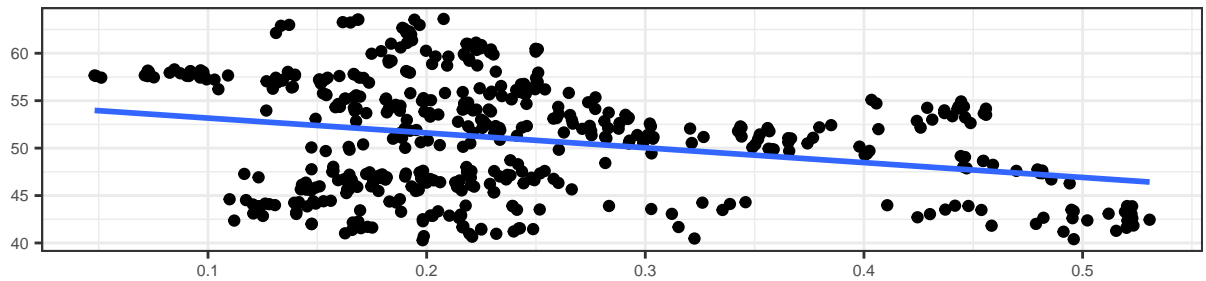


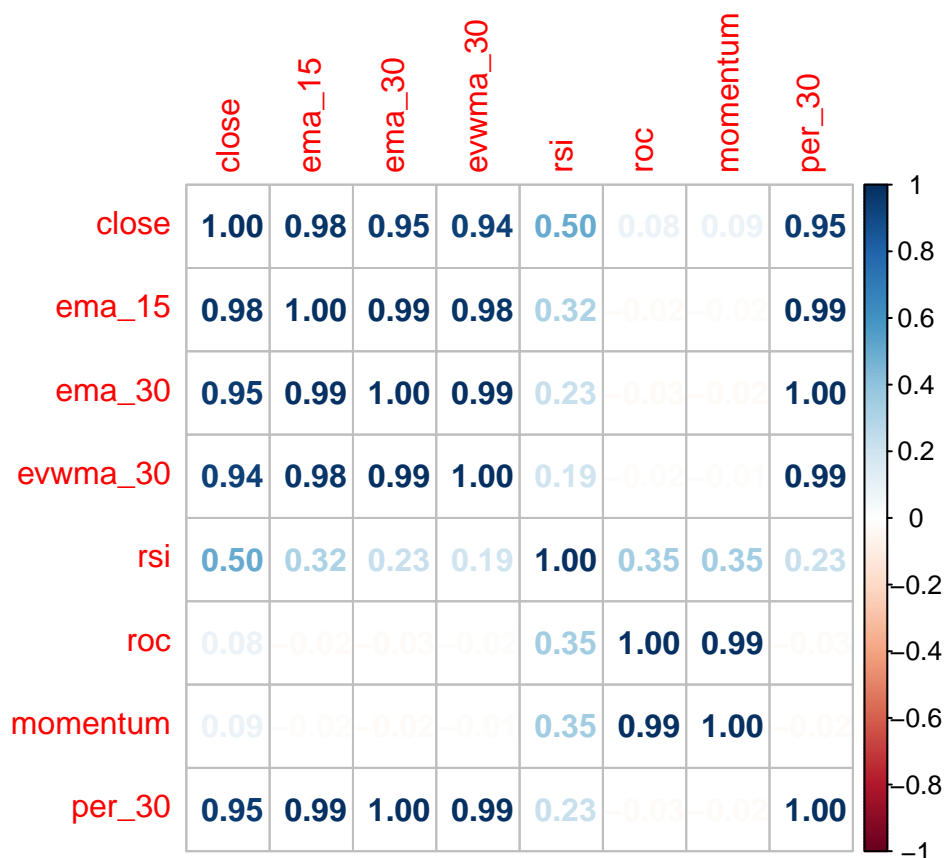
Fig 6: relação com fecho

Parece existir uma relação negativa entre as variáveis indicando quanto maior a volatilidade menor o preço.

**diff\_close\_open:**

Variável categórica para se ocorreu um ganho intradia ( $\text{Close} > \text{Open}$ ) ou uma perda ( $\text{Close} < \text{Open}$ )

## Matriz de correlação



A variável dependente em estudo apresenta uma elevada correlação positiva relativamente à média móvel exponencial a 15, 30 e 30 dias ajustado pelo volume assim como com Price Earning Ratio com 30 dias. Esta conclusão não é surpreendente dado a natureza dos indicadores calculados.

As conclusões da matriz form incorporadas no modelo tendo sido apenas considerado a média movel exponencial ajustada pelo volume por se tratar de um indicador sintetico mais completo.

## Modelos e avaliação

```
## Warning in predict.lm(model, data): prediction from a rank-deficient fit may be
## misleading
```

```
## # A tibble: 13 x 4
##   model                                model_name      r2    aic
##   <chr>                                <chr>         <dbl> <dbl>
## 1 close ~ evwma_30 * volatility_cat + rsi * volat~ relation      9.98e-1  63.1
## 2 close ~ ema_15 + ema_30 + evwma_30 + rsi + roc ~ complete      9.97e-1  365.
## 3 close ~ evwma_30 + rsi + momentum + per_30 + vo~ adjusted      9.95e-1  601.
## 4 close ~ ema_15                                ema_15      9.52e-1 1700.
## 5 close ~ ema_30                                ema_30      9.08e-1 2022.
## 6 close ~ per_30                                per_30      9.08e-1 2022.
## 7 close ~ evwma_30                              evwma_30     8.78e-1 2167.
## 8 close ~ per_30_mean                          per_30_mean   6.78e-1 2656.
```

```
## 9 close ~ rsi                rsi                2.52e-1 3080.
## 10 close ~ volatility_cat     volatility        4.93e-2 3205.
## 11 close ~ momentum          momentum          7.97e-3 3222.
## 12 close ~ roc               roc                7.19e-3 3223.
## 13 close ~ diff_close_open   diff_close_o~ 8.92e-4 3226.
```

```
##
```

```
## Call:
```

```
## lm(formula = ., data = msft_prices_2015_2016_df)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -0.80955 -0.13018 -0.00987  0.14760  0.96825
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value
## (Intercept)    -9.880514   0.532082 -18.570
## evwma_30       -0.131555   0.057815  -2.275
## volatility_catLow Volatility    5.161957   0.511693  10.088
## volatility_catMedium High volatility    3.735717   0.465725   8.021
## volatility_catMedium Low volatility    3.994074   0.484536   8.243
## rsi              0.188753   0.001683 112.180
## momentum        0.025695   0.022622   1.136
## per_30          2.414443   0.114123  21.156
## per_30_meanLower than average   -0.105449   0.057056  -1.848
## diff_close_openintraday loss    0.006103   0.028003   0.218
## evwma_30:volatility_catLow Volatility    0.058243   0.075785   0.769
## evwma_30:volatility_catMedium High volatility    0.069441   0.068923   1.008
## evwma_30:volatility_catMedium Low volatility   -0.085048   0.072918  -1.166
## volatility_catLow Volatility:rsi    -0.083029   0.003382 -24.547
## volatility_catMedium High volatility:rsi   -0.047919   0.003371 -14.217
## volatility_catMedium Low volatility:rsi   -0.058341   0.003777 -15.445
## volatility_catLow Volatility:momentum   -0.020292   0.064056  -0.317
## volatility_catMedium High volatility:momentum  0.048687   0.039911   1.220
## volatility_catMedium Low volatility:momentum -0.006696   0.045145  -0.148
## volatility_catLow Volatility:per_30    -0.160945   0.149769  -1.075
## volatility_catMedium High volatility:per_30  -0.197417   0.138575  -1.425
## volatility_catMedium Low volatility:per_30    0.139666   0.146414   0.954
```

```
##
```

```
## (Intercept)          < 2e-16 ***
## evwma_30             0.0233 *
## volatility_catLow Volatility          < 2e-16 ***
## volatility_catMedium High volatility  8.01e-15 ***
## volatility_catMedium Low volatility  1.60e-15 ***
## rsi                   < 2e-16 ***
## momentum              0.2566
## per_30                < 2e-16 ***
## per_30_meanLower than average        0.0652 .
## diff_close_openintraday loss         0.8276
## evwma_30:volatility_catLow Volatility  0.4426
## evwma_30:volatility_catMedium High volatility  0.3142
## evwma_30:volatility_catMedium Low volatility  0.2441
## volatility_catLow Volatility:rsi       < 2e-16 ***
## volatility_catMedium High volatility:rsi < 2e-16 ***
```

```

## volatility_catMedium Low volatility:rsi          < 2e-16 ***
## volatility_catLow Volatility:momentum           0.7515
## volatility_catMedium High volatility:momentum    0.2231
## volatility_catMedium Low volatility:momentum     0.8822
## volatility_catLow Volatility:per_30             0.2831
## volatility_catMedium High volatility:per_30      0.1549
## volatility_catMedium Low volatility:per_30      0.3406
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2517 on 482 degrees of freedom
## Multiple R-squared:  0.9983, Adjusted R-squared:  0.9982
## F-statistic: 1.319e+04 on 21 and 482 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = close ~ evwma_30 * volatility_cat + rsi * volatility_cat +
##     per_30, data = msft_prices_2015_2016_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82584 -0.12072 -0.01055  0.12677  0.95040
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                  -10.680199   0.343397 -31.102
## evwma_30                     -0.078522   0.022956  -3.421
## volatility_catLow Volatility    5.650334   0.422973  13.359
## volatility_catMedium High volatility  4.176813   0.425397   9.819
## volatility_catMedium Low volatility  3.896212   0.412710   9.441
## rsi                          0.189357   0.001559 121.495
## per_30                       2.333691   0.045124  51.718
## evwma_30:volatility_catLow Volatility -0.027200   0.008178  -3.326
## evwma_30:volatility_catMedium High volatility -0.034028   0.008404  -4.049
## evwma_30:volatility_catMedium Low volatility -0.020394   0.007686  -2.653
## volatility_catLow Volatility:rsi    -0.084295   0.002848 -29.600
## volatility_catMedium High volatility:rsi -0.047240   0.003006 -15.716
## volatility_catMedium Low volatility:rsi -0.054953   0.003253 -16.892
##                                Pr(>|t|)
## (Intercept)                  < 2e-16 ***
## evwma_30                     0.000677 ***
## volatility_catLow Volatility    < 2e-16 ***
## volatility_catMedium High volatility < 2e-16 ***
## volatility_catMedium Low volatility < 2e-16 ***
## rsi                          < 2e-16 ***
## per_30                       < 2e-16 ***
## evwma_30:volatility_catLow Volatility 0.000947 ***
## evwma_30:volatility_catMedium High volatility 5.98e-05 ***
## evwma_30:volatility_catMedium Low volatility 0.008230 **
## volatility_catLow Volatility:rsi    < 2e-16 ***
## volatility_catMedium High volatility:rsi < 2e-16 ***
## volatility_catMedium Low volatility:rsi < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

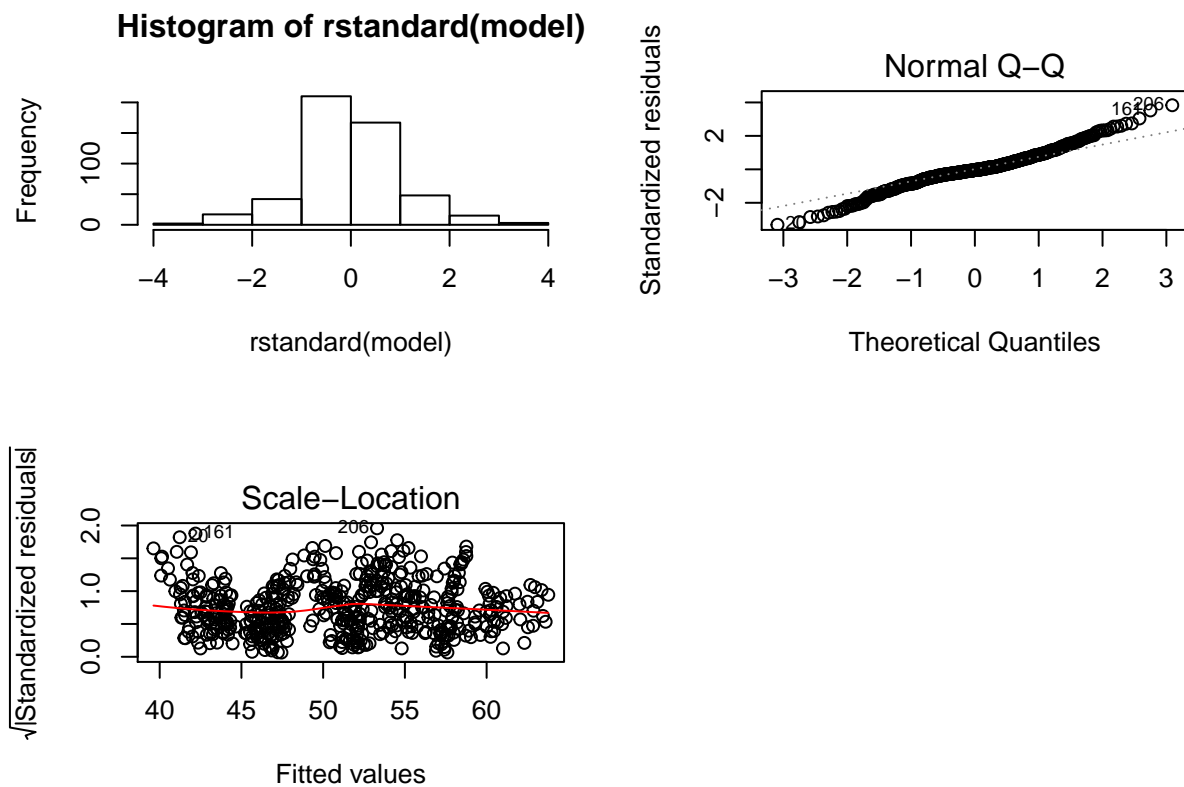
```
##
## Residual standard error: 0.2544 on 491 degrees of freedom
## Multiple R-squared:  0.9982, Adjusted R-squared:  0.9981
## F-statistic: 2.259e+04 on 12 and 491 DF,  p-value: < 2.2e-16

## Start:  AIC=-1367.01
## close ~ evwma_30 * volatility_cat + rsi * volatility_cat + per_30
##
##              Df Sum of Sq    RSS    AIC
## <none>                        31.775 -1367.01
## - evwma_30:volatility_cat    3      1.149  32.924 -1355.11
## - volatility_cat:rsi         3     63.000  94.775  -822.22
## - per_30                     1    173.092 204.866  -429.71

##
## Call:
## lm(formula = close ~ evwma_30 * volatility_cat + rsi * volatility_cat +
##     per_30, data = msft_prices_2015_2016_df)
##
## Coefficients:
##              (Intercept)
##              -10.68020
##              evwma_30
##              -0.07852
##      volatility_catLow Volatility
##              5.65033
##      volatility_catMedium High volatility
##              4.17681
##      volatility_catMedium Low volatility
##              3.89621
##              rsi
##              0.18936
##              per_30
##              2.33369
##      evwma_30:volatility_catLow Volatility
##              -0.02720
##      evwma_30:volatility_catMedium High volatility
##              -0.03403
##      evwma_30:volatility_catMedium Low volatility
##              -0.02039
##      volatility_catLow Volatility:rsi
##              -0.08429
##      volatility_catMedium High volatility:rsi
##              -0.04724
##      volatility_catMedium Low volatility:rsi
##              -0.05495
```

Não existindo qualquer fundamento teórico que obrigasse à inclusão de qualquer das variáveis utilizadas, retiramos para o modelo final aquelas cuja a Hipoteses nula não pudesse ser refutada. Modelo obtido através de Backward Selection.

## Estudo dos resíduos



Da análise gráfica não parece existir indicícos de heterostacidade nos resíduos, mas o gráfico de Normal Q-Q apresenta algum desvio que pode impactar o resultado de testes estatísticos.

## Questões sobre Regressão Linear

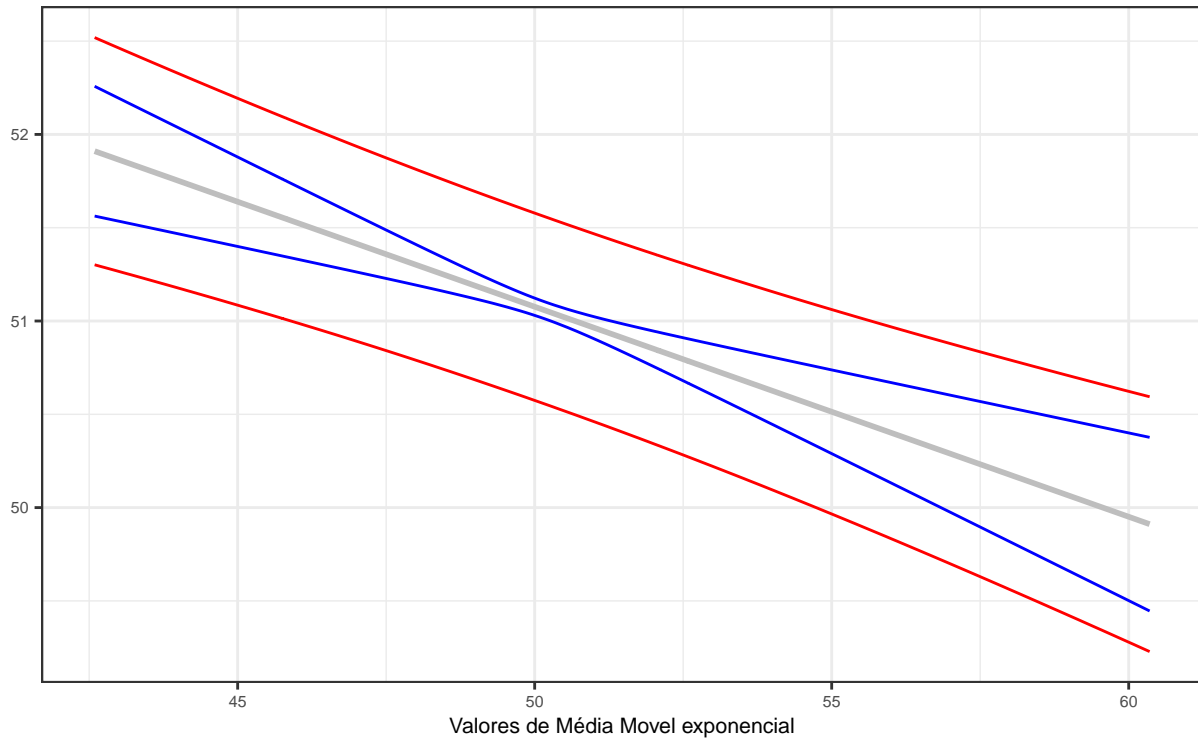
e.1) Mantendo tudo o resto constante, um aumento de 1 USD na média móvel exponencial ponderado pelo volume a 30 dias representa um aumento de -0.08 USD no valor spot. Mantendo tudo o resto constante, em média volatilidade a 30 dias baixa resultam em mais 5 USD na cotação, Média alta em 4,1 USD e Média Lpw em 3,89 USD face a se tivesse uma Volatilidade Alta.

e.2.)

```
##
##           High           Low Volatility Medium High volatility
##           126           126           126
## Medium Low volatility
##           126
```

## Previsões modelo com banda de confiança e previsão

Azul banda de confiança, vermelha banda de estimação



e3)

e4)

```
##
## Call:
## lm(formula = close ~ evwma_30 * volatility_cat + rsi * volatility_cat +
##     per_30, data = dt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.82584 -0.12072 -0.01055  0.12677  0.95040
##
## Coefficients:
##                                Estimate Std. Error t value
## (Intercept)                   -9.918030   0.473389 -20.951
## evwma_30                      -0.078522   0.022956  -3.421
## volatility_catLow Volatility     5.914345   0.498090  11.874
## volatility_catMedium High volatility  4.507098   0.501887   8.980
## volatility_catMedium Low volatility  4.094164   0.481485   8.503
## rsi                           0.189357   0.001559 121.495
## per_30                        2.333691   0.045124  51.718
## evwma_30:volatility_catLow Volatility -0.027200   0.008178  -3.326
## evwma_30:volatility_catMedium High volatility -0.034028   0.008404  -4.049
## evwma_30:volatility_catMedium Low volatility -0.020394   0.007686  -2.653
## volatility_catLow Volatility:rsi    -0.084295   0.002848 -29.600
## volatility_catMedium High volatility:rsi -0.047240   0.003006 -15.716
```

```
## volatility_catMedium Low volatility:rsi      -0.054953    0.003253 -16.892
##                                              Pr(>|t|)
## (Intercept)                                < 2e-16 ***
## evwma_30                                  0.000677 ***
## volatility_catLow Volatility                < 2e-16 ***
## volatility_catMedium High volatility        < 2e-16 ***
## volatility_catMedium Low volatility        2.23e-16 ***
## rsi                                         < 2e-16 ***
## per_30                                     < 2e-16 ***
## evwma_30:volatility_catLow Volatility      0.000947 ***
## evwma_30:volatility_catMedium High volatility 5.98e-05 ***
## evwma_30:volatility_catMedium Low volatility 0.008230 **
## volatility_catLow Volatility:rsi           < 2e-16 ***
## volatility_catMedium High volatility:rsi   < 2e-16 ***
## volatility_catMedium Low volatility:rsi    < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2544 on 491 degrees of freedom
## Multiple R-squared:  0.9982, Adjusted R-squared:  0.9981
## F-statistic: 2.259e+04 on 12 and 491 DF,  p-value: < 2.2e-16
```

Sem impacto significativo no modelo e nos coeficientes estimados.

e5) Efeito já incluído no modelo estudado