



PROBABILIDADE E ESTATÍSTICA (IND2604)

MODELO DE REGRESSÃO MÚLTIPLA PARA A PREVISÃO DO ÍNDICE BOVESPA - IBOV

Grupo 4:
Daiane Oliveira
Fernanda Nucci
Leonardo Domingues

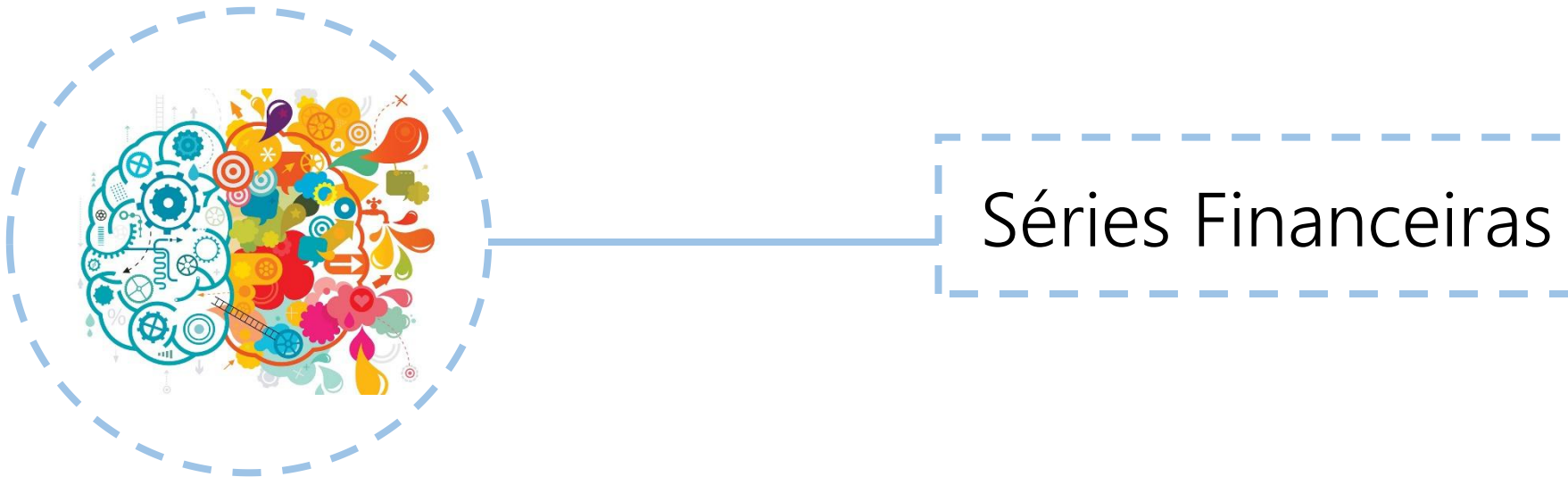
Agenda

- Motivação
- Objetivo
- Contextualização
- Variáveis
- Regressão Múltipla
- Seleção de variáveis
- Análise dos resíduos
- Modelo
- Próximos passos
- Referências



Motivação

- **Área:** Finanças e Análise de Investimentos.



Objetivo

Mensurar a influência de variáveis macroeconômicas no retorno do Índice Bovespa através da utilização de regressão múltipla, propondo um modelo que possibilite previsões, simulações e uma análise mais assertiva do índice.

Contextualização

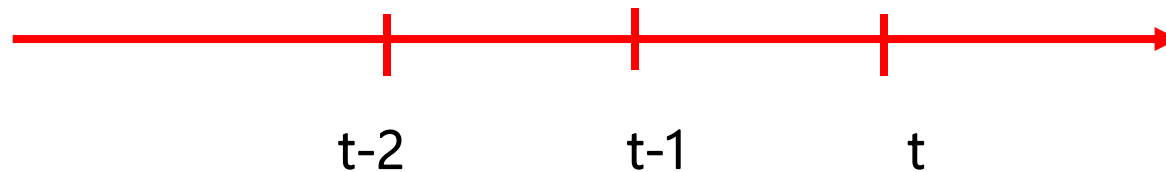
- Modelos de Regressão Linear

De acordo com Johnston e Dinardo (1998), nos modelos de regressão linear supõe-se que os erros “gerados” pelo modelo possuem algumas características como: média zero, variância constante, distribuição normal e independência, implicando, assim, na inexistência de correlação serial.

Contextualização

- Séries Temporais

Segundo Morettin e Toloi (2006), “uma série temporal é qualquer conjunto de observações ordenadas no tempo”.



Contextualização

- Conforme Barros e Souza (1995), ao se tentar modelar uma série temporal através de um modelo de regressão, a hipótese de independência dos ruídos não é realista.

- Os estimadores usuais por mínimos quadrados são ainda não tendenciosos, mas não têm variância mínima.

- Os estimadores da variância e dos erros padrões dos coeficientes da regressão são subestimados, o que levaria à conclusão de que os estimadores são mais precisos do que na realidade.

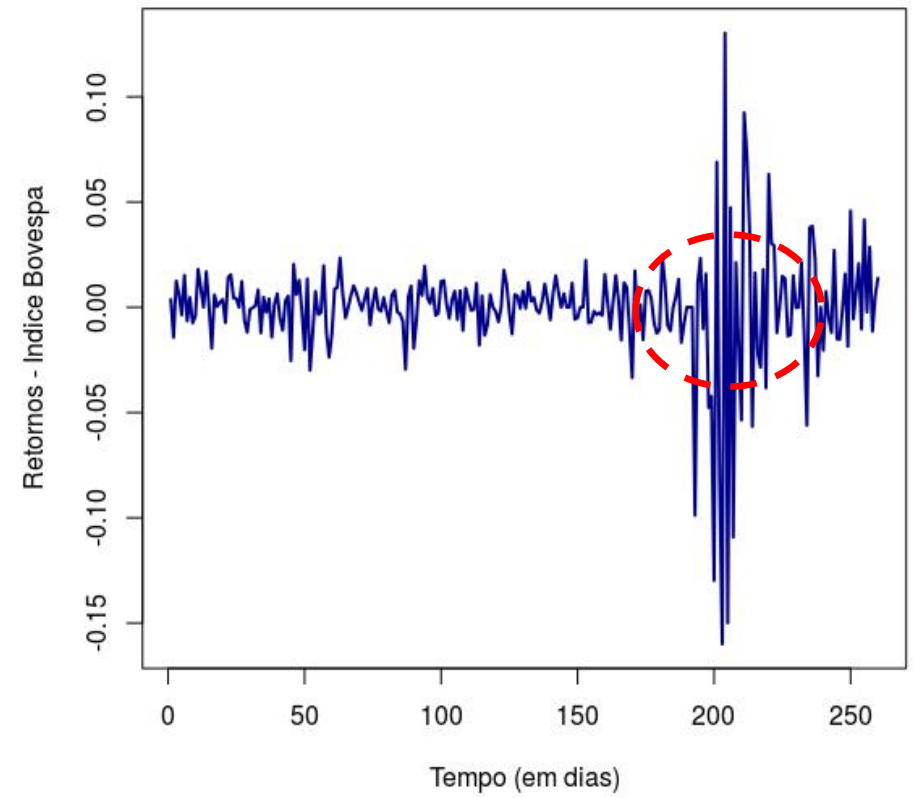
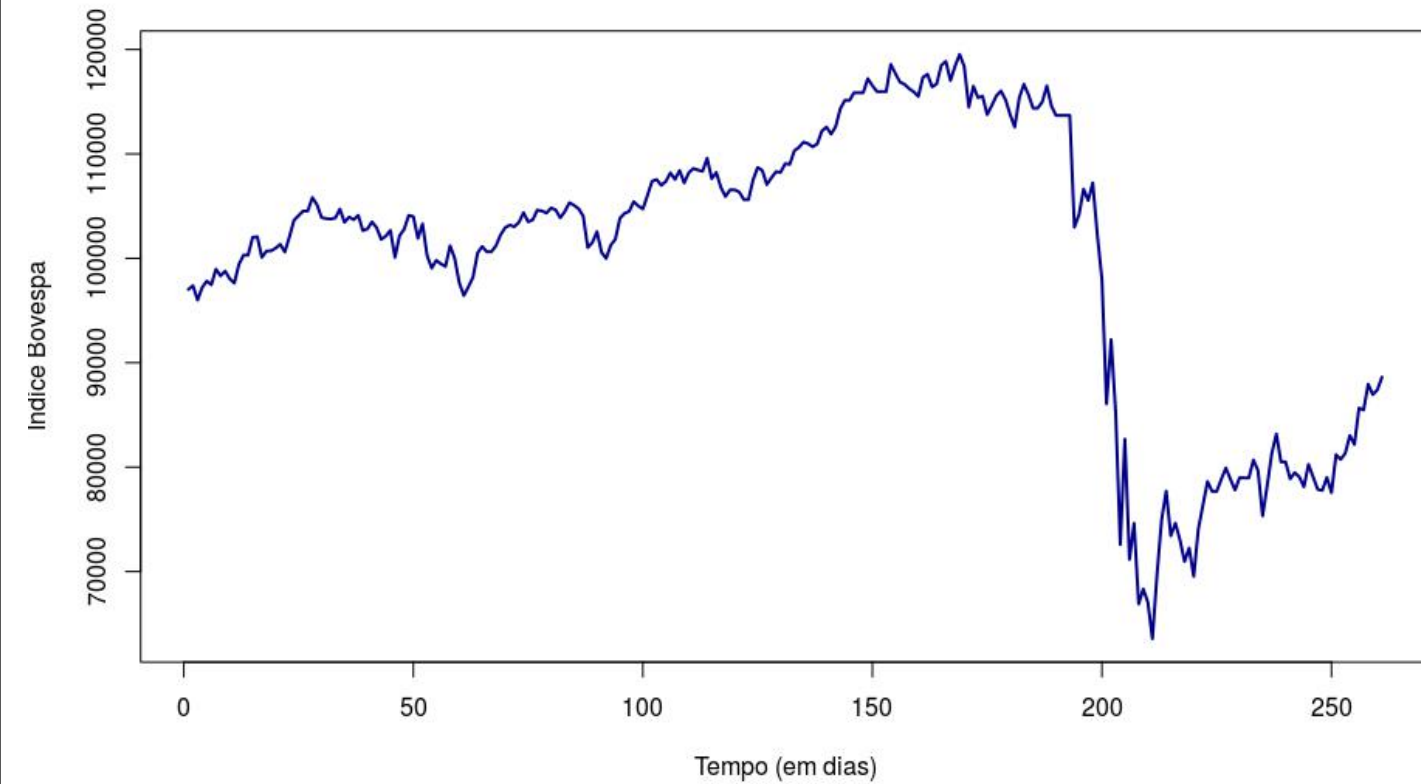
- Os intervalos de confiança para os parâmetros da regressão e os testes de hipóteses relacionados a estes intervalos perdem a validade.

Variável Resposta

- Principal indicador de desempenho das ações negociadas na B3 e reúne as empresas mais importantes do mercado de capitais brasileiro;
- Criado em 1968 e, ao longo desses 50 anos, consolidou-se como referência para investidores ao redor do mundo;
- Reavaliado a cada quatro meses, o índice é resultado de uma carteira teórica de ativos;
- Composto pelas ações e units de companhias listadas na B3 que atendem aos critérios descritos na sua metodologia, correspondendo a cerca de 80% do número de negócios e do volume financeiro do nosso mercado de capitais.

TOP 20 AÇÕES IBOV (01/07/20)		
Ação	Tipo	Part. (%)
VALE3	ON	10,7
ITUB4	PN	7,006
B3SA3	ON	6,537
PETR4	PN	5,662
BBDC4	PN	5,127
PETR3	ON	3,83
ABEV3	ON	3,579
MGLU3	ON	2,756
BBAS3	ON	2,651
ITSA4	PN	2,505
WEGE3	ON	2,027
JBSS3	ON	1,992
LREN3	ON	1,908
GNDI3	ON	1,812
NTCO3	ON	1,638
SUZB3	ON	1,552
RENT3	ON	1,411
BBDC3	ON	1,381
RAIL3	ON	1,377
EQTL3	ON	1,364

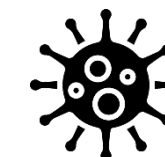
IBOV



Variáveis Explicativas

Variável Resposta: IBOV

Categoria	Variável Explicativa
Bolsas	Dow Jones Nasdaq SP&500
Índices	EWZ EWZS VIX
Risco País	Risco País
Moeda	Bitcoin Dólar americano
Commodities	Brent Milho Ouro
Juros	Juros Futuro (Jan22) Selic
Inflação	IPCA
Dummy	Covid-19



Premissas e Definições

- Dados diários de 03/06/2019 a 01/06/2020;
- Calculados os retornos de todas as variáveis;
- **Retorno:** medida de resultado financeiro;
- **Retorno e risco** são medidas importantes em Finanças pois sintetizam os ganhos e as perdas potenciais de um investimento para o qual existe uma série histórica;
- De acordo com Lima *et al.* (2017), dada a série histórica de preços de uma ação P_t , o **retorno contínuo** (ou log-retorno) dessa ação no dia t é definido por:

$$r_t = \ln \left(\frac{P_t}{P_{t-1}} \right) = \ln(P_t) - \ln(P_{t-1})$$

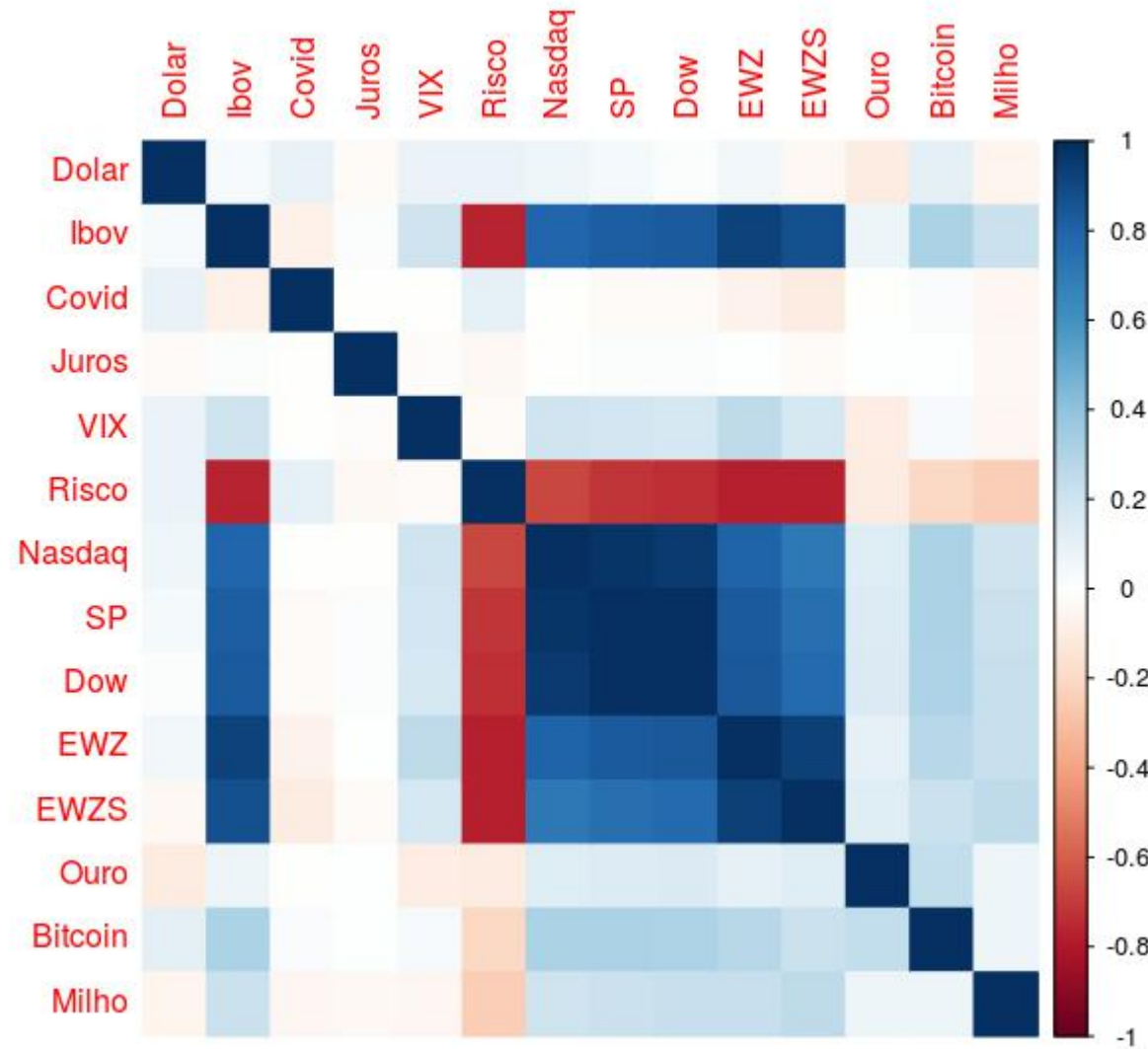
Correlação

- A correlação de X e Y é o número definido por:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}.$$

A correlação é sempre entre -1 e 1, com os valores -1 e 1 indicando uma perfeita relação linear entre X e Y .

Correlações



Passo a Passo

- 1 Seleção de variáveis
- 2 Análise de resíduos
- 3 Modelo

Seleção de Variáveis

Seleção de todos
modelos possíveis

Seleção automática

Forward

Backward

Stepwise

Seleção de Variáveis

1º Modelo: Ibov ~ Risco + SP + EWZ + EWZS + Ouro + Bitcoin

Call:

```
lm(formula = Ibov ~ Risco + SP + EWZ + EWZS + Ouro + Bitcoin,  
    data = retornos)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.066363	-0.003671	0.000132	0.004055	0.041841

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.0004454	0.0006104	0.730	0.46625
Risco	-0.0001755	0.0000808	-2.172	0.03079 *
SP	0.2323800	0.0576377	4.032	0.000073322093 ***
EWZ	0.4231707	0.0583523	7.252	0.000000000005 ***
EWZS	0.0864677	0.0493099	1.754	0.08072 .
Ouro	-0.1234268	0.0533396	-2.314	0.02147 *
Bitcoin	0.0393210	0.0135955	2.892	0.00416 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009744 on 253 degrees of freedom

Multiple R-squared: 0.8738, Adjusted R-squared: 0.8708

F-statistic: 292 on 6 and 253 DF, p-value: < 0.0000000000000022




Variáveis não significativas

Análise dos Resíduos

Normalidade	Homocedasticidade	Independência	Análise de Variância	Colinearidade
<ul style="list-style-type: none"> • Normalidade • Shapiro-Wilk • Kolmogorov-Smirnov 	<ul style="list-style-type: none"> • Homocedasticidade • Breusch-Page • Goldfeld-Quandt 	<ul style="list-style-type: none"> • Durbin-Watson 	<ul style="list-style-type: none"> • <i>Lack of fit</i> 	<ul style="list-style-type: none"> • Colinearidade • Multicolinearidade

	Ibov	EWZ	Ouro	Bitcoin	EWZS
Ibov	1.00000000	0.9238209	0.07449419	0.3144443	0.8756099
EWZ	0.92382095	1.0000000	0.10360015	0.2724310	0.9396088
Ouro	0.07449419	0.1036002	1.00000000	0.2454893	0.1332270
Bitcoin	0.31444430	0.2724310	0.24548926	1.0000000	0.2194541
EWZS	0.87560989	0.9396088	0.13322697	0.2194541	1.0000000
SP	0.82873234	0.8353153	0.14718268	0.3121092	0.7507179
Risco	-0.76120957	-0.7711884	-0.10617552	-0.2016273	-0.7710508
	SP	Risco			
Ibov	0.8287323	-0.7612096			
EWZ	0.8353153	-0.7711884			
Ouro	0.1471827	-0.1061755			
Bitcoin	0.3121092	-0.2016273			
EWZS	0.7507179	-0.7710508			
SP	1.0000000	-0.7153619			
Risco	-0.7153619	1.0000000			

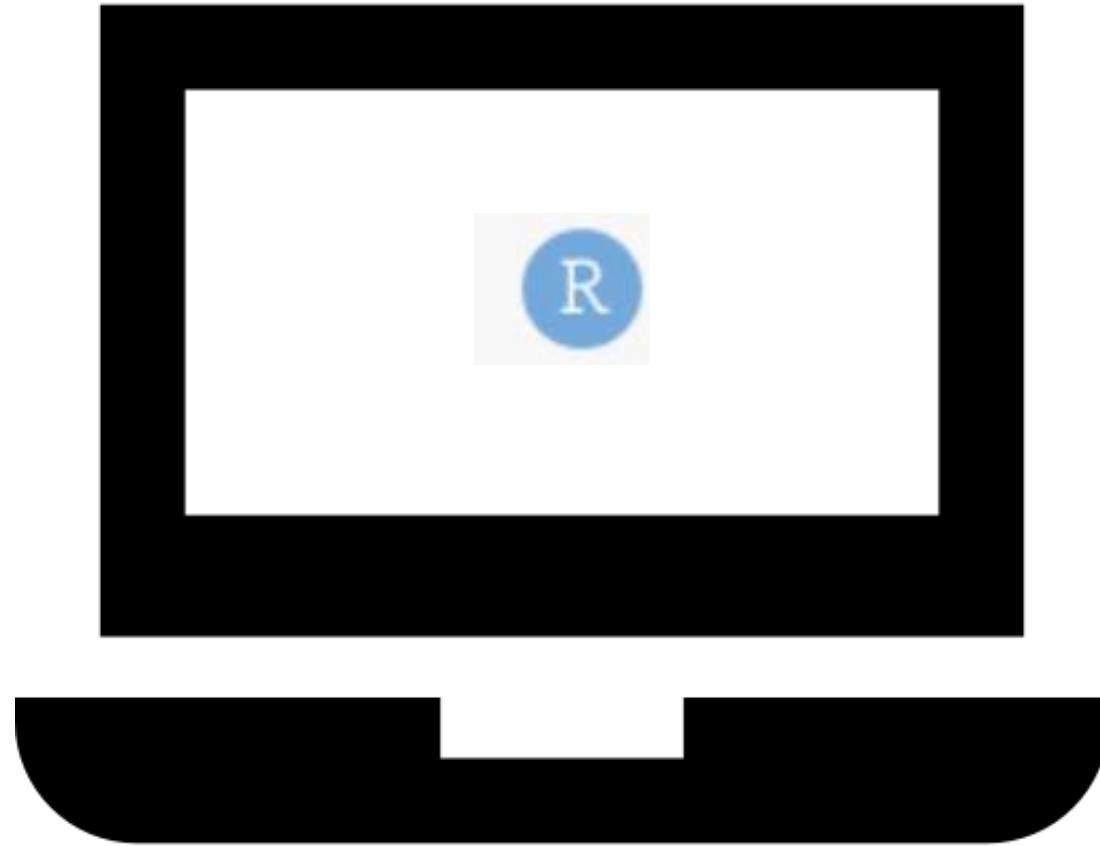
Análise de Resíduos

Normalidade	Homocedasticidade	Independência	Análise de Variância	Colinearidade
<ul style="list-style-type: none">• QQplot• Shapiro-Wilk• Kolmogorov-Smirnov	<ul style="list-style-type: none">• Homocedasticidade• Breusch-Page• Goldfeld-Quandt	<ul style="list-style-type: none">• Durbin-Watson	<ul style="list-style-type: none">• <i>Lack of fit</i>	<ul style="list-style-type: none">• Colinearidade• Multicolinearidade 

```
> fit = lm(Ibov ~ Risco + SP + EWZ + EWZ5 + Ouro + Bitcoin, data = retornos)
> vif(fit)
```

Risco	SP	EWZ	EWZ5	Ouro	Bitcoin
2.762312	3.770365	13.152775	9.731414	1.097153	1.177710

Após muitas simulações...



Seleção de Variáveis

Modelo: Ibov ~ Risco + VIX + Dow + Bitcoin

Call:

```
lm(formula = Ibov ~ Risco + VIX + Dow + Bitcoin, data = retornos)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.056266	-0.007110	0.000899	0.008153	0.044546

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.00021247	0.00082327	-0.258	0.79656
Risco	-0.00074522	0.00009804	-7.601	0.000000000000561 ***
VIX	0.02937683	0.00931592	3.153	0.00181 **
Dow	0.68860473	0.05886371	11.698	< 0.0000000000000002 ***
Bitcoin	0.04069912	0.01791073	2.272	0.02390 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01325 on 255 degrees of freedom

Multiple R-squared: 0.765, Adjusted R-squared: 0.7613

F-statistic: 207.5 on 4 and 255 DF, p-value: < 0.00000000000000022

Análise de Resíduos

Normalidade

- QQplot
- Shapiro-Wilk
- Kolmogorov-Smirnov



Homocedasticidade

- Homocedasticidade
- Breusch-Page
- Goldfeld-Quandt

Independência

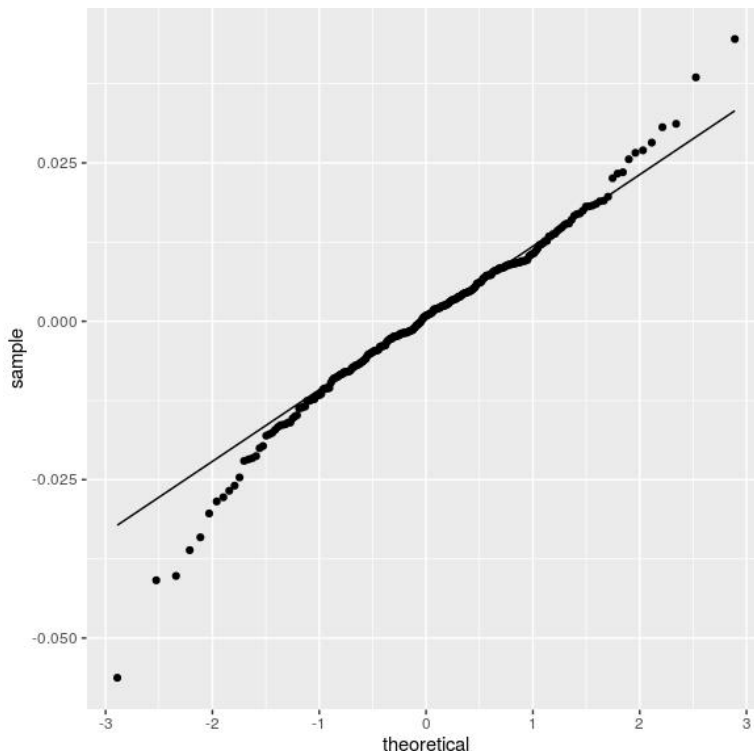
- Durbin-Watson

Análise de linearidade

- *Lack of fit*

Colinearidade

- Colinearidade
- Multicolinearidade



Shapiro-Wilk normality test

```
data: residuos  
W = 0.97615, p-value = 0.0002374
```

One-sample Kolmogorov-Smirnov test

```
data: residuos  
D = 0.061224, p-value = 0.284  
alternative hypothesis: two-sided
```

Análise de Resíduos

Normalidade

- QQplot
- Shapiro-Wilk
- Kolmogorov-Smirnov



Homocedasticidade

- Homocedasticidade
- Breusch-Pagan
- Goldfeld-Quandt



Independência

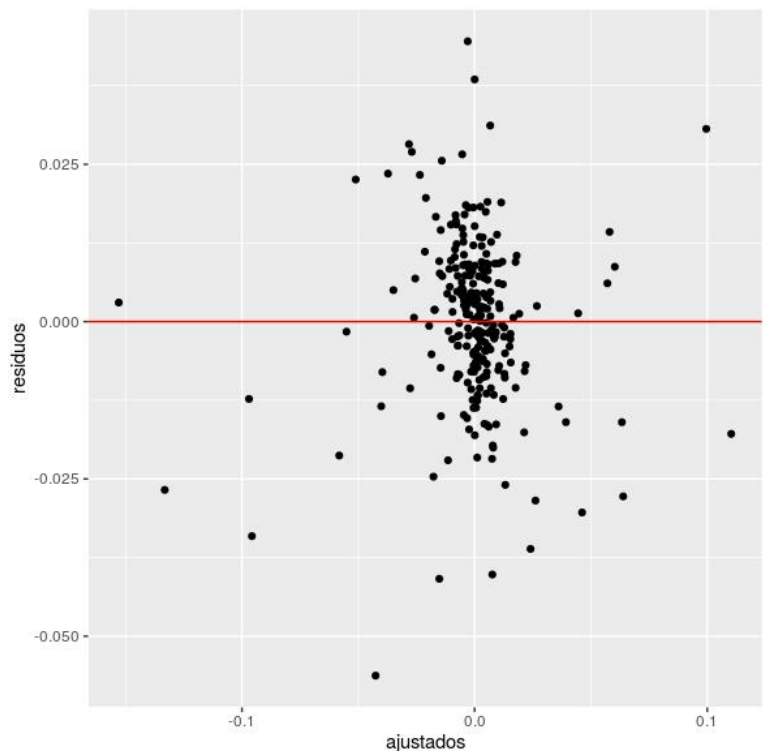
- Durbin-Watson

Análise de linearidade

- *Lack of fit*

Colinearidade

- Colinearidade
- Multicolinearidade



studentized Breusch-Pagan test

```
data: fit
BP = 4.3366, df = 4, p-value = 0.3624
```

Goldfeld-Quandt test

```
data: fit
GQ = 3.7723, df1 = 125, df2 = 125, p-value = 0.0000000000004041
alternative hypothesis: variance increases from segment 1 to 2
```

Análise de Resíduos

Normalidade

- QQplot
- Shapiro-Wilk
- Kolmogorov-Smirnov



Homocedasticidade

- Homocedasticidade
- Breusch-Page
- Goldfeld-Quandt



Independência

- Durbin-Watson

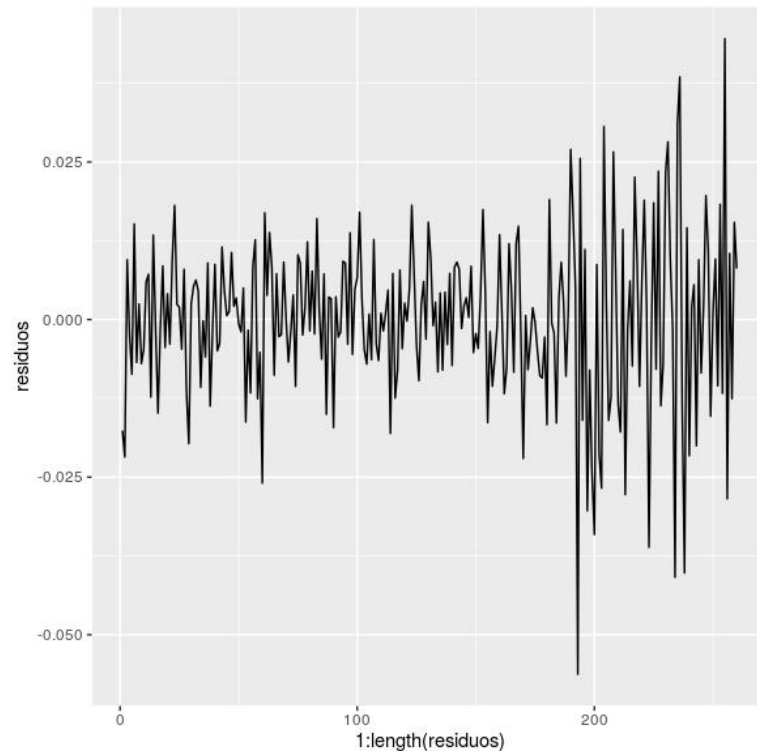


Análise de Variância

- *Lack of fit*

Colinearidade

- Colinearidade
- Multicolinearidade



Durbin-Watson test

```
data: fit
DW = 2.3017, p-value = 0.9929
alternative hypothesis: true autocorrelation is greater than 0
```

Análise de Resíduos

Normalidade	Homocedasticidade	Independência	Análise de Variância	Colinearidade
<ul style="list-style-type: none">• QQplot• Shapiro-Wilk• Kolmogorov-Smirnov	<ul style="list-style-type: none">• Homocedasticidade• Breusch-Page• Goldfeld-Quandt	<ul style="list-style-type: none">• Durbin-Watson	<ul style="list-style-type: none">• <i>Lack of fit</i>	<ul style="list-style-type: none">• Colinearidade• Multicolinearidade
✓	✗	✓	✓	






```
          rstudent unadjusted p-value Bonferroni p
193 -4.457528          0.000012448    0.0032365
```

```
Response: Ibov
      Df  Sum Sq Mean Sq F value    Pr(>F)
Risco   1  0.110308  0.110308  628.7636 < 0.00000000000000022 ***
VIX      1  0.006571  0.006571   37.4565    0.000000003505 ***
Dow      1  0.027849  0.027849  158.7392 < 0.00000000000000022 ***
Bitcoin  1  0.000906  0.000906    5.1635    0.0239 *
Residuals 255  0.044736  0.000175
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Valores
Extremos

Lack of fit

Análise de Resíduos

Normalidade	Homocedasticidade	Independência	Análise de linearidade	Colinearidade
<ul style="list-style-type: none"> • QQplot • Shapiro-Wilk • Kolmogorov-Smirnov 	<ul style="list-style-type: none"> • Homocedasticidade • Breusch-Page • Goldfeld-Quandt 	<ul style="list-style-type: none"> • Durbin-Watson 	<ul style="list-style-type: none"> • Teste da falta de ajustes 	<ul style="list-style-type: none"> • Colinearidade • Multicolinearidade 

	Ibov	VIX	Dow	Bitcoin	Risco
Ibov	1.0000000	0.20319159	0.8393370	0.31444430	-0.76120957
VIX	0.2031916	1.00000000	0.1712427	0.03910411	-0.02292343
Dow	0.8393370	0.17124268	1.0000000	0.30732740	-0.73091465
Bitcoin	0.3144443	0.03910411	0.3073274	1.00000000	-0.20162729
Risco	-0.7612096	-0.02292343	-0.7309146	-0.20162729	1.00000000

Colinearidade

Multicolinearidade

Risco	VIX	Dow	Bitcoin
2.201015	1.055017	2.398474	1.106133

Modelo Final

Modelo: Ibov ~ Risco + VIX + Dow + Bitcoin

Call:

```
lm(formula = Ibov ~ Risco + VIX + Dow + Bitcoin, data = retornos)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.056266	-0.007110	0.000899	0.008153	0.044546

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.00021247	0.00082327	-0.258	0.79656
Risco	-0.00074522	0.00009804	-7.601	0.000000000000561 ***
VIX	0.02937683	0.00931592	3.153	0.00181 **
Dow	0.68860473	0.05886371	11.698	< 0.0000000000000002 ***
Bitcoin	0.04069912	0.01791073	2.272	0.02390 *

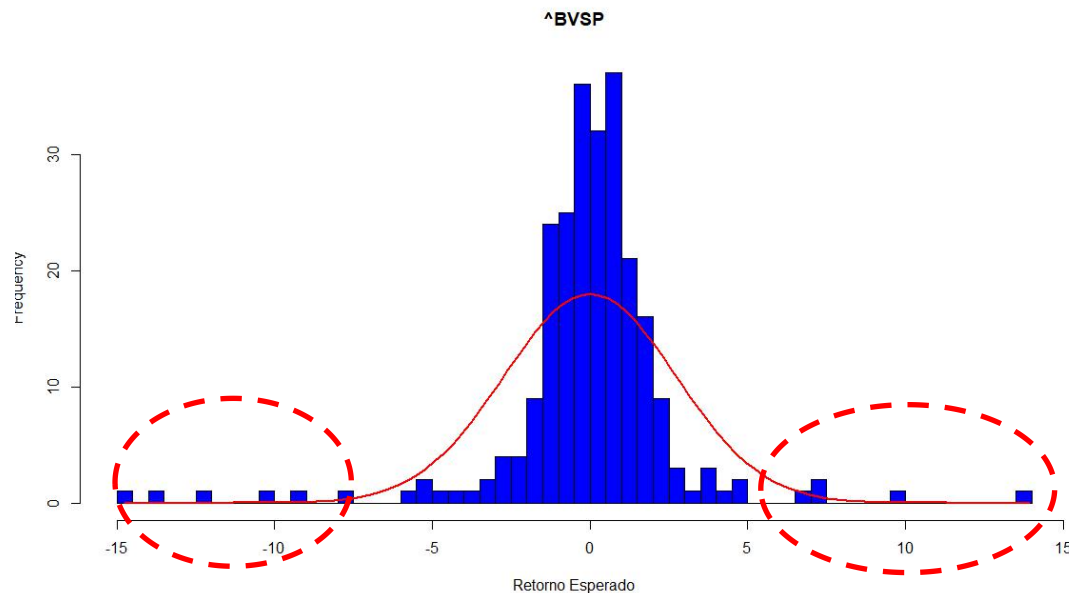
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01325 on 255 degrees of freedom

Multiple R-squared: 0.765, Adjusted R-squared: 0.7613

F-statistic: 207.5 on 4 and 255 DF, p-value: < 0.00000000000000022

Próximos Passos



Fatos estilizados: regularidades estatísticas observadas num grande número de séries de *retornos financeiros*, obtidas a partir de estudos empíricos iniciados na década de 60, utilizando séries financeiras dos diversos mercados mundiais:

- estacionariedade;
- fraca dependência linear;
- dependência não linear com aglomerados de volatilidade;
- alta curtose ou não normalidade incondicional e etc.

Necessário utilizar Modelo GARCH

Próximos Passos

- Modelos de Regressão Dinâmica
 - Os modelos de regressão dinâmica combinam a dinâmica de séries temporais e o efeito de variáveis explicativas;
 - Segundo Zanini (2000), nos modelos de regressão dinâmica, a variável dependente é explicada por seus valores defasados e pelos valores atuais e passados de variáveis causais ou exógenas.

Referências

BARROS, M.; SOUZA, R. C. Regressão Dinâmica. Núcleo de Estatística Computacional. PUC-Rio, 1995.

JOHNSTON, J.; DINARDO, J. Econometric Methods. New York, John Wiley, 1998.

LIMA, I. S.; DE LIMA, G. S. F.; PIMENTEL, R. C. *et al.* Curso de Mercado Financeiro. 2ª Edição. São Paulo: Atlas, 2017.

MORETTIN, P. A.; TOLOI, C. M. C. Análise de Séries Temporais. 2ª Edição. São Paulo: Blucher, 2006.

ZANINI, A. Redes Neurais e Regressão Dinâmica: um modelo híbrido para a previsão de curto prazo da demanda de gasolina automotiva no Brasil. Tese de Mestrado, DEE, PUC-Rio, 2000.