



TRABALHO

Método Quantitativos Aplicados (às Finanças)

MESTRADO EM ANÁLISE FINANCEIRA

MESTRADO EM ANÁLISE DE DADOS E SISTEMAS DE APOIO À DECISÃO

João Campos 2021103606

Luís Santos 2021109914

Oleksandra Kukharska 2020151174

2021/2022

Índice

1. INTRODUÇÃO.....	1
2. RETORNOS E PORTFÓLIOS.....	2
2.1. Estatística Descritiva	2
2.1.1. Média	3
2.1.2. Variância Amostral.....	3
2.1.3. Desvio Padrão.....	4
2.1.4. Coeficiente de Assimetria	4
2.1.5. Coeficiente de Achatamento ou Curtose de Fisher.....	5
2.2. ANOVA ou Teste de Variância com um fator	5
2.3. Retorno Aritmético	7
2.4. Retorno Logarítmico	7
2.5. Testes de Normalidade	8
2.5.1. Teste de Enviesamento.....	8
2.5.2. Teste de Excesso de Curtose	9
2.5.3. Teste de Jarque-Bera	10
2.6. Constituição de um Portefólio.....	11
3. MODELO DE REGRESSÃO LINEAR MÚLTIPLO	13
3.1. Resultados dos três modelos	14
3.2. Teste de significância individual (TSI)	16
3.3. Teste de significância geral (TSG)	17
3.4. Teste de White.....	17
3.5. Teste de Chow	19
3.6. Multicolinearidade.....	20
3.7. Valores dos coeficientes estimados.....	21
4. CONCLUSÃO	22
REFERÊNCIAS BIBLIOGRÁFICAS.....	23

ÍNDICE DE ILUSTRAÇÕES

TABELAS

Tabela 1 Estatística Descritiva	2
Tabela 2 Média	3
Tabela 3 Variância Amostral.....	3
Tabela 4 Desvio-Padrão.....	4
Tabela 5 Coeficiente de Assimetria.....	4
Tabela 6 Coeficiente de Achatamento ou Curtose	5
Tabela 7 Teste de Variância (ANOVA)	6
Tabela 8 Retorno Aritmético	7
Tabela 9 Retorno Logarítmico	8
Tabela 10 Retorno Logarítmico da Assimetria	8
Tabela 11 Curtose do Retorno Logarítmico.....	9
Tabela 12 Dados do Portefólio	11
Tabela 13 Variáveis dependentes e variáveis independentes	13
Tabela 14 Quadro resumo dos três modelos	15

FIGURAS

Figura 1 Modelo 1	14
Figura 2 Modelo 2	14
Figura 3 Modelo 3	15
Figura 4 Teste de White do Modelo 1.....	18
Figura 5 Teste de Chow do Modelo 1	19
Figura 6 Colinearidade do modelo 1.....	20

1. INTRODUÇÃO

O presente trabalho, foi-nos proposto na unidade curricular de Métodos Quantitativos Aplicados (às Finanças), do 1º Trimestre, do Mestrado em Análise Financeira e do Mestrado em Análise de Dados e Sistemas de Apoio à Decisão. O trabalho será composto por duas partes.

A primeira parte consiste na escolha de três ativos, neste caso foram selecionadas três criptomoedas, e consequente criação de uma base de dados das suas cotações diárias, num determinado período de tempo.

Primeiramente será feita a estatística descritiva, como forma de sumarizar o conjunto de dados extraídos, *one-way* ANOVA, os Retornos Logarítmicos e Aritméticos, os Testes de Normalidade e por último a construção do portefólio dos três ativos. O objetivo final da primeira parte é o estudo do retorno e da volatilidade do portefólio, de forma a conseguir tomar as decisões de investir no portefólio ou investir apenas num ativo.

Na segunda parte, será feito o estudo e aplicação do modelo de regressão linear múltiplo, com base no artigo selecionado, neste caso foi selecionado o artigo *FINANCIAL PERFORMANCE OF PRIVATE COMMERCIAL BANKS IN INDIA: MULTIPLE REGRESSION ANALYSIS*, publicado na *Academy of Accounting and Financial Studies Journal* em 2018, com o auxílio de uma base de dados da SABI, e de empresas selecionadas por nós, foram selecionadas variáveis dependentes e independentes e consequentemente criados 3 modelos, cujo objetivo é verificar qual se ajusta melhor à estrutura de dados.

Após a seleção do modelo, neste caso, foi selecionado o modelo cuja variável dependente é o Retorno sobre o Capital Próprio (ROE), serão analisadas as variáveis independentes que contribuem mais significativamente para a performance das empresas selecionadas. O objetivo final da segunda parte é identificar se o conjunto de variáveis selecionadas explica ou não a nossa variável dependente, ROE.

2. RETORNOS E PORTFÓLIOS

Os três ativos selecionados são a Ethereum (ETH), Binance Coin (BNB) e Bitcoin (BTC), os três ativos dizem respeito a criptomoedas, que são moedas tradicionais, não controladas por nenhum governo e totalmente digitais.

A escolha destes ativos deveu-se à revolução da era digital que tem provocado uma tendência de uso das criptomoedas promissora para o futuro. Optou-se pelas três criptomoedas acima mencionadas por terem sido classificadas pela [Bankrate](#), como as mais populares moedas digitais no mercado atual (Royal, 2021).

2.1. Estatística Descritiva

Tendo em consideração a base de dados recolhida, com cotações diárias das ações de 5 de novembro de 2020 a 3 de novembro de 2021, procedeu-se ao cálculo, em Excel, da estatística descritiva, que de seguida se apresenta e analisa.

ETH-EUR		BNB-EUR		BTC-EUR	
Média	2201,875915	Média	292,0404837	Média	41976,75486
Erro-padrão	55,94520431	Erro-padrão	9,601417841	Erro-padrão	691,8926575
Mediana	2141,445191	Mediana	316,2341615	Mediana	42876,60156
Moda	#N/D	Moda	#N/D	Moda	#N/D
Desvio-padrão	1067,36647	Desvio-padrão	183,1833774	Desvio-padrão	13200,47058
Variância da amostra	1139271,182	Variância da amostra	33556,14974	Variância da amostra	174252423,6
Curtose	-0,855947229	Curtose	-1,052677475	Curtose	-0,785488004
Assimetria	0,074899763	Assimetria	-0,124902247	Assimetria	-0,318824538
Intervalo	4193,126496	Intervalo	648,323078	Intervalo	51159,08203
Mínimo	414,067352	Mínimo	27,361004	Mínimo	14833,75391
Máximo	4607,193848	Máximo	675,684082	Máximo	65992,83594
Soma	801482,8331	Soma	106302,7361	Soma	15279538,77
Contagem	364	Contagem	364	Contagem	364

Tabela 1 Estatística Descritiva

2.1.1. Média

A média da amostra, enquanto medida de tendência central, permite-nos observar o valor médio das observações da variável e é calculado através da seguinte fórmula:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

A média calculada de cada uma das variáveis utilizadas, apresenta-se na tabela seguinte:

ETH-EUR		BNB-EUR		BTC-EUR	
Média	2201,875915	Média	292,0404837	Média	41976,75486

Tabela 2 Média

De entre as três variáveis escolhidas, a média do valor por ação, corresponde a 41.976,75 euros para a Bitcoin, seguindo-se da Ethereum, com 2.201,87 euros e por último, a Binance Coin com 292,04 euros.

2.1.2. Variância Amostral

A variância é uma medida de dispersão que representa a variabilidade dos elementos da amostra em relação à média. É o desvio padrão ao quadrado, daí ser uma medida difícil de usar na prática pois a unidade difere em relação aos dados. É calculada através da seguinte fórmula:

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \bar{x}^2$$

ETH-EUR		BNB-EUR		BTC-EUR	
Variância da amostra	1139271,182	Variância da amostra	33556,14974	Variância da amostra	174252423,6

Tabela 3 Variância Amostral

Quanto maior a variância, maior a dispersão nos dados. Neste caso, das três criptomoedas, a Bitcoin é a que possui maior dispersão dos elementos da amostra em relação à média, seguindo-se a Ethereum e por último a Binance Coin.

Como exemplo, usando a Bitcoin, uma variância de 17252423,6² é equivalente a um desvio padrão de 13200,47.

2.1.3. Desvio Padrão

O desvio padrão, enquanto medida de variação estatística, permite-nos perceber se os valores analisados estão, ou não agrupados em torno da média, isto é, um grande desvio padrão indica que os valores em estudo estão espalhados longe da média e um pequeno desvio padrão indica que os valores em estudo estão agrupados perto da média, neste último caso, quanto menor é o desvio padrão, mais homogênea é a amostra. É calculado através da seguinte fórmula:

$$\sigma = +\sqrt{s^2}$$

Assim, calculando o desvio padrão para a amostra em estudo temos:

ETH-EUR		BNB-EUR		BTC-EUR	
Desvio-padrão	1067,36647	Desvio-padrão	183,1833774	Desvio-padrão	13200,47058

Tabela 4 Desvio-Padrão

O desvio padrão assume a mesma medida da unidade amostral, neste caso, euros. Sendo assim, o desvio-padrão da Ethereum é de 1.067,36 euros, da Binance Coin é 183,18 euros e para a Bitcoin é 13.200,47 euros.

2.1.4. Coeficiente de Assimetria

O coeficiente de assimetria é dado por:

$$g_1 = \frac{m_3}{\sqrt{m_2^3}}$$

ETH-EUR		BNB-EUR		BTC-EUR	
Assimetria	0,074899763	Assimetria	-0,124902247	Assimetria	-0,318824538

Tabela 5 Coeficiente de Assimetria

Podemos observar que os valores dos coeficientes são todos diferentes de 0, o que nos permite concluir que as distribuições são assimétricas para os três ativos.

Sendo a distribuição da Ethereum assimétrica positiva e as distribuições da Bitcoin e da Binance Coin são ambas assimétricas negativas.

2.1.5. Coeficiente de Achatamento ou Curtose de Fisher

A curtose é uma medida que caracteriza o achatamento da curva da função de distribuição de probabilidade.

O coeficiente de achatamento de Fisher é dado por:

$$g_3 = \frac{m_4}{m_2^2} - 3$$

ETH-EUR	BNB-EUR	BTC-EUR
Curtose	Curtose	Curtose
-0,855947229	-1,052677475	-0,785488004

Tabela 6 Coeficiente de Achatamento ou Curtose

Através dos valores dos coeficientes de achatamento observados, podemos concluir que se trata de uma distribuição platicúrtica, uma vez que o valor g_3 para os três ativos corresponde a valores negativos, inferiores a zero (< 0).

2.2. ANOVA ou Teste de Variância com um fator

O teste de variância (*one-way ANOVA*), é uma fórmula estatística usada para comparar médias entre diferentes amostras e concluir acerca da igualdade das médias teóricas das distribuições adjacentes. É sempre um teste unilateral à direita.

SUMÁRIO

Grupos	Contagem	Soma	Média	Variância
Close ETH	364	801482,8	2201,875915	1139271,182
Close BNB	364	106302,7	292,0404837	33556,14974
Close BTC	364	15279539	41976,75486	174252423,6

ANOVA

Fonte de variação	SQ	gl	MQ	F	valor P	F crítico
Entre grupos	4,03228E+11	2	2,01614E+11	3447,857721	0	3,003988
Dentro de grupos	63679366091	1089	58475083,65			
Total	4,66907E+11	1091				

Tabela 7 Teste de Variância (ANOVA)

Estabelecer as hipóteses:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_k \quad \text{VS} \quad H_1: \text{nem todos os } \mu_k \text{ são iguais}$$

(Teste unilateral à direita)

Valor observado:

Para o nosso exemplo, temos a distribuição F de Fisher- Snedecor que possui (3-1) e (1093-4) graus de liberdade e um nível de significância de 5%. Através da tabela podemos observar que a região crítica é dada por $F^* > 3,003988$. E que a estatística F observada é dada por 3.447,857721.

Tomada de decisão:

Temos duas formas de tomar a decisão, sendo a primeira através da análise da região crítica, em que, para valores maiores que 3,003988 rejeitaremos a hipótese nula. Assim, uma vez que a estatística F observada é de $3.447,857721 > 3,003988$, pertencendo a região crítica, rejeitaremos a hipótese nula, ao nível de significância de 5%, concluímos que há evidências de que existe diferença entre as médias.

A segunda forma é através da interpretação do valor-P, que neste caso é 0, sendo menor que o nível de significância, que neste caso é de 5%. Sendo assim, rejeitamos a hipótese nula e concluímos que há evidências de que existe diferença entre as médias.

2.3. Retorno Aritmético

Visto que só se sabe o valor das cotações, as fórmulas são dadas por:

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

<i>Rt ETH</i>		<i>Rt BNB</i>		<i>Rt BTC</i>	
Média	0,008277	Média	0,011126	Média	0,004747
Erro-padrão	0,002985	Erro-padrão	0,004067	Erro-padrão	0,002222
Mediana	0,00837	Mediana	0,006279	Mediana	0,003237
Moda	#N/D	Moda	#N/D	Moda	#N/D
Desvio-padrão	0,056863	Desvio-padrão	0,077493	Desvio-padrão	0,042338
Variância da amostra	0,003233	Variância da amostra	0,006005	Variância da amostra	0,001792
Curtose	3,189597	Curtose	19,25427	Curtose	1,48125
Assimetria	-0,06194	Assimetria	2,210526	Assimetria	0,045571
Intervalo	0,531479	Intervalo	1,03026	Intervalo	0,325126
Mínimo	-0,272	Mínimo	-0,33266	Mínimo	-0,13766
Máximo	0,259475	Máximo	0,697604	Máximo	0,187465
Soma	3,004563	Soma	4,038687	Soma	1,722997
Contagem	363	Contagem	363	Contagem	363

Tabela 8 Retorno Aritmético

2.4. Retorno Logarítmico

Visto que só se sabe o valor das cotações, as fórmulas são dadas por:

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

<i>rt ETH</i>		<i>rt BNB</i>		<i>rt BTC</i>	
Média	0,006637316	Média	0,008326665	Média	0,003847606
Erro-padrão	0,002991618	Erro-padrão	0,003845224	Erro-padrão	0,002217045
Mediana	0,008335432	Mediana	0,006259632	Mediana	0,003232251

Moda	#N/D	Moda	#N/D	Moda	#N/D
Desvio-padrão	0,056997982	Desvio-padrão	0,073261365	Desvio-padrão	0,042240375
Variância da amostra	0,00324877	Variância da amostra	0,005367228	Variância da amostra	0,001784249
Curtose	3,942229597	Curtose	10,85628803	Curtose	1,505392678
Assimetria	-0,51957905	Assimetria	0,741225154	Assimetria	-0,172387519
Intervalo	0,548154259	Intervalo	0,933667957	Intervalo	0,319927638
Mínimo	-0,31745903	Mínimo	-0,40445009	Mínimo	-0,148107076
Máximo	0,230695229	Máximo	0,52921787	Máximo	0,171820562
Soma	2,409345594	Soma	3,02257939	Soma	1,396680842
Contagem	363	Contagem	363	Contagem	363

Tabela 9 Retorno Logarítmico

2.5. Testes de Normalidade

2.5.1. Teste de Enviesamento

RL ETH		RL BNB		RL BTC	
Assimetria	-0,51957905	Assimetria	0,741225154	Assimetria	-0,172387519

Tabela 10 Retorno Logarítmico da Assimetria

Estabelecer as hipóteses:

$$H_0: S = 0 \quad \text{VS} \quad H_1: S \neq 0 \quad (\text{Teste Bilateral})$$

Estatística-Teste:

$$T = \frac{\hat{S}}{\sqrt{6/n}}, N(0,1)$$

Determinar a Região de Rejeição:

$$\alpha = \frac{5\%}{2} = 0,025$$

$$P(T \leq t_c) = 0,975$$

$$\Leftrightarrow t_c = 1,96$$

$$\therefore RR =]-\infty, -1,96[\cup]1,96, +\infty [$$

Valor observado:

$$t_{\text{obs ETH}} = \frac{-0,51957905}{\sqrt{\frac{6}{363}}} = -4,041376566$$

$$t_{\text{obs BNB}} = \frac{0,741225154}{\sqrt{\frac{6}{363}}} = 5,765378661$$

$$t_{\text{obs BTC}} = \frac{-0,172387519}{\sqrt{\frac{6}{363}}} = -1,340857661$$

Tomada de decisão:

O $t_{\text{obs BTC}}$ não pertence à região de rejeição, então não rejeitamos a H_0 , logo, as ações da Bitcoin são simétricas. Quanto ao $t_{\text{obs ETH}}$ e $t_{\text{obs BNB}}$ estes pertencem à região de rejeição, então rejeitamos a H_0 e concluímos que não existe simetria nas suas ações.

2.5.2. Teste de Excesso de Curtose

<i>RL ETH</i>		<i>RL BNB</i>		<i>RL BTC</i>	
Curtose	3,942229597	Curtose	10,85628803	Curtose	1,505392678

Tabela 11 Curtose do Retorno Logarítmico

Estabelecer as hipóteses:

$$H_0: K - 3 = 0 \quad \text{VS} \quad H_1: K - 3 \neq 0 \quad (\text{Teste Bilateral})$$

Estatística-Teste:

$$T = \frac{\hat{k} - 3}{\sqrt{24/n}}$$

Determinar a Região de Rejeição:

$$\alpha = \frac{5\%}{2} = 0,025$$

$$P(T \leq t_c) = 0,975$$

$$\Leftrightarrow t_c = 1,96$$

$$\therefore RR =]-\infty; -1,96[\cup]1,96; +\infty [$$

Valor observado:

$$t_{\text{obs ETH}} = \frac{3,942229597}{\sqrt{\frac{24}{363}}} = 15,33167505$$

$$t_{\text{obs BNB}} = \frac{10,85628803}{\sqrt{\frac{24}{363}}} = 42,22105185$$

$$t_{\text{obs BTC}} = \frac{1,505392678}{\sqrt{\frac{24}{363}}} = 5,854603539$$

Tomada de decisão:

Como os t_{obs} dos 3 ativos pertencem à região de rejeição, então rejeita-se a H_0 nos 3 ativos para um nível de significância de 5% e conclui-se que há um excesso de curtose nos três ativos.

2.5.3. Teste de Jarque-Bera

Estabelecer as hipóteses:

H_0 : Os retornos são normais VS H_1 : Os retornos não são normais
(Teste unilateral à direita)

Estatística-Teste:

$$JB = \frac{n}{6} \left(\hat{S}^2 + \frac{(\hat{K} - 3)^2}{4} \right), \sim \chi^2_{(2)}$$

Determinar a Região de Rejeição:

$$P(\chi^2_{(2)} \leq qc) = 0,05$$

$$\Leftrightarrow qc = 5,991465$$

$$\therefore RR =]5,991465; +\infty [$$

Valor observado:

$$q_{\text{obs ETH}} = \frac{363}{6} \times \left((-0,519579048)^2 + \frac{3,942229597^2}{4} \right) = 251,3929842$$

$$q_{\text{obs BNB}} = \frac{363}{6} \times \left((0,741225154)^2 + \frac{10,85628803^2}{4} \right) = 1815,85681$$

$$q_{\text{obs BTC}} = \frac{363}{6} \times \left((-0,172387519)^2 + \frac{1,505392678^2}{4} \right) = 36,07428873$$

Tomada de decisão:

Como tanto no $q_{\text{obs ETH}}$, como no $q_{\text{obs BNB}}$ como no $q_{\text{obs BTC}}$ pertencem à região de rejeição, rejeita-se H_0 para os 3 ativos, para um nível de significância de 5%. Conclui-se assim que os retornos dos 3 ativos não são normais.

2.6. Constituição de um Portefólio

	ETH	BNC	BTC
Constituição	25%	25%	50%
Retorno esperado	2.201,87	292,04	41.976,75
Volatilidade	1.067,36	183,18	13,200,47

Tabela 12 Dados do Portefólio

x: "Retorno do Ativo Ethereum em 364 dias", $\sim (\mu_x = 2.201,87; \sigma_x = 1.067,36)$ $\omega_x = 0,25$

y: "Retorno do Ativo Binance Coin em 364 dias", $\sim (\mu_y = 292,04; \sigma_y = 183,18)$ $\omega_y = 0,25$

z: "Retorno do Ativo Bitcoin em 364 dias", $\sim (\mu_z = 41.976,75; \sigma_z = 13.200,47)$ $\omega_z = 0,5$

$p=0,5$

Seja, R_p : "Portefólio constituído pelos ativos x, y e z"

$$R_p = 0,25x + 0,25y + 0,5z \sim N(\mu_{R_p}, \sigma_{R_p})$$

$$\mu_{R_p} = E[R_p]$$

$$\Leftrightarrow E[0,25x + 0,25y + 0,5z] = 0,25E[x] + 0,25E[y] + 0,5E[z]$$

$$\Leftrightarrow 0,25 \times 2.201,87 + 0,25 \times 292,04 + 0,5 \times 41.976,75$$

$$\Leftrightarrow 21.611,86$$

$$\sigma_{Rp}^2 = V[Rp]$$

$$\Leftrightarrow V[0,25x + 0,25y + 0,5z]$$

$$\Leftrightarrow 0,25^2V[x] + 0,25^2V[y] + 0,5^2V[z] + 2 \times \omega_x\omega_y\text{Cov}(x,y) + 2 \times \omega_x\omega_z\text{Cov}(x,z) + 2 \times \omega_y\omega_z\text{Cov}(y,z)$$

$$\Leftrightarrow 0,0625 \times 1.067,36^2 + 0,0625 \times 183,18^2 + 0,25 \times 13.200,47^2 + 2 \times 0,25 \times 0,25 \times 1.067,36 \times 183,18 + 2 \times 0,25 \times 0,5 \times 1.067,36 \times 13.200,47 + 2 \times 0,25 \times 0,5 \times 183,18 \times 13.200,47$$

$$\Leftrightarrow 44.612.375,64$$

$$\sigma_{Rp} = \sqrt{44.612.375,64} = 6.679,25$$

$$Rp = 0,25x + 0,25y + 0,5z \sim N(21.611,86 ; 6.679,25)$$

Decisão:

Analisando os retornos esperados de cada ativo e o retorno do portefólio, podemos concluir que é mais vantajoso investir nas ações da Bitcoin, cujo retorno esperado é de 41.976,75 euros do que no portefólio dos três ativos, cujo retorno esperado é de 21.611,86 euros.

3. MODELO DE REGRESSÃO LINEAR MÚLTIPLO

Através da base de dados da SABI, os dados extraídos pertencem às entidades com a CAE - Atividades e serviços financeiros, exceto seguros e fundos de pensões, da região Centro de Portugal, cujo número de observações é 497, das quais 23 foram ignoradas, por estarem omissas ou incompletas. Visa-se selecionar o modelo que se ajusta melhor à estrutura de dados, através da estimação OLS.

O presente estudo envolve três diferentes variáveis dependentes para cinco variáveis independentes comuns. As variáveis selecionadas estão apresentadas a seguir:

Variáveis:	
Dependentes:	
ROE	Retorno sobre o capital próprio (%)
ROA	Retorno sobre o total do ativo (%)
ROIC	Retorno sobre o capital investido (%)
Independentes:	
TA	Total Ativo (em milhares de euros)
CP	Capital Próprio (em milhares de euros)
VN	Volume de negócios (em milhares de euros)
CPF	Custos e Perdas Financeiras (em milhares de euros)
RLP	Resultado Líquido do Período (em milhares de euros)

Tabela 13 Variáveis dependentes e variáveis independentes

A forma geral da equação de regressão múltipla é dada por:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

$$\Leftrightarrow Y = \beta_0 + \beta_1 TA + \beta_2 CP + \beta_3 VN + \beta_4 CPF + \beta_5 RLP + \varepsilon$$

Sendo:

- Y = variável dependente
- X_i = variáveis independentes, $i = 1, 2, \dots, 5$
- $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ e β_5 são coeficientes parciais das variáveis independentes, reflete a forma como Y varia consoante a variação de 1 unidade de X_i , mantendo-se as restantes constantes.
- ε = erro aleatório

3.1. Resultados dos três modelos

Modelo 1:

Modelo 1: Mínimos Quadrados (OLS), usando as observações 1-497 (n = 474)
Observações omissas ou incompletas foram ignoradas: 23
Variável dependente: ROE

	coeficiente	erro padrão	rácio-t	valor p
const	25,7353	14,9176	1,725	0,0852 *
TA	-0,00223645	0,00147341	-1,518	0,1297
CP	0,00198465	0,00161382	1,230	0,2194
VN	0,0499878	0,0853107	0,5859	0,5582
CPF	0,0489666	0,0609010	0,8040	0,4218
RLP	0,00989891	0,00526219	1,881	0,0606 *
Média var. dependente	20,93916	D.P. var. dependente	299,4342	
Soma resid. quadrados	41747103	E.P. da regressão	298,6691	
R-quadrado	0,015621	R-quadrado ajustado	0,005104	
F(5, 468)	1,485286	valor P(F)	0,193102	
Log. da verosimilhança	-3371,043	Critério de Akaike	6754,086	
Critério de Schwarz	6779,053	Critério Hannan-Quinn	6763,905	

Excluindo a constante, o valor p foi o maior para a variável 6 (VN)

Figura 1 Modelo 1

Quando o ROE é utilizado como variável dependente, a regressão apresenta um R^2 de 1,56% e um $R^2_{ajustado}$ de 0,51%. Este resultado indica que o modelo especificado explica 1,56% do ROE. É também de referir que a variável RLP, foi considerada significativa ao nível de 10%.

Modelo 2:

Modelo 2: Mínimos Quadrados (OLS), usando as observações 1-497 (n = 474)
Observações omissas ou incompletas foram ignoradas: 23
Variável dependente: ROA

	coeficiente	erro padrão	rácio-t	valor p
const	-99,8989	52,9274	-1,887	0,0597 *
TA	0,00256709	0,00522762	0,4911	0,6236
CP	-0,00119176	0,00572580	-0,2081	0,8352
VN	0,0842005	0,302680	0,2782	0,7810
CPF	-0,0950949	0,216075	-0,4401	0,6601
RLP	0,00592111	0,0186701	0,3171	0,7513
Média var. dependente	-83,71513	D.P. var. dependente	1055,016	
Soma resid. quadrados	5,26e+08	E.P. da regressão	1059,670	
R-quadrado	0,001823	R-quadrado ajustado	-0,008841	
F(5, 468)	0,170944	valor P(F)	0,973278	
Log. da verosimilhança	-3971,306	Critério de Akaike	7954,611	
Critério de Schwarz	7979,578	Critério Hannan-Quinn	7964,430	

Excluindo a constante, o valor p foi o maior para a variável 5 (CP)

Figura 2 Modelo 2

Quando o ROA é utilizado como variável dependente, a regressão apresenta um R^2 de 0,18% e um $R^2_{ajustado}$ de -0,88%. Este resultado indica que o modelo explica 0,18%.

Modelo 3:

Modelo 3: Mínimos Quadrados (OLS), usando as observações 1-497 (n = 474)
Observações omissas ou incompletas foram ignoradas: 23
Variável dependente: ROIC

	coeficiente	erro padrão	rácio-t	valor p
const	-71,4050	71,7968	-0,9945	0,3205
TA	-0,000995597	0,00709135	-0,1404	0,8884
CP	0,00200920	0,00776713	0,2587	0,7960
VN	-0,0483256	0,410590	-0,1177	0,9064
CPF	0,0860003	0,293109	0,2934	0,7693
RLP	0,0275118	0,0253263	1,086	0,2779
Média var. dependente	-63,17399	D.P. var. dependente	1432,013	
Soma resid. quadrados	9,67e+08	E.P. da regressão	1437,458	
R-quadrado	0,003031	R-quadrado ajustado	-0,007620	
F(5, 468)	0,284601	valor P(F)	0,921516	
Log. da verosimilhança	-4115,837	Critério de Akaike	8243,674	
Critério de Schwarz	8268,642	Critério Hannan-Quinn	8253,494	

Excluindo a constante, o valor p foi o maior para a variável 6 (VN)

Figura 3 Modelo 3

Quando o ROIC é utilizado como variável dependente, a regressão apresenta um R^2 de 0,30% e um $R^2_{ajustado}$ de -0,76%. Este resultado indica que o modelo explica 0,30%.

Para escolher o modelo, é necessário ter em conta o R^2 e $R^2_{ajustado}$, a tabela seguinte mostra os valores dos modelos em estudo:

Modelo	Variáveis Independentes	R^2	$R^2_{ajustado}$
1	5	0,015621	0,005104
2	5	0,001823	-0,008841
3	5	0,003031	-0,007620

Tabela 14 Quadro resumo dos três modelos

Sendo R^2 o coeficiente de determinação, pode-se verificar que $R^2_1 > R^2_3 > R^2_2$, assim, o Modelo 1 ajusta-se melhor à estrutura de dados.

O mesmo se verifica com o $R^2_{ajustado}$, em que, $R^2_{ajustado 1} > R^2_{ajustado 3} > R^2_{ajustado 2}$, no entanto, quer no modelo 2 quer no modelo 3, o $R^2_{ajustado}$ é negativo, o que significa que estes modelos não irão conseguir explicar a estrutura de dados.

3.2. Teste de significância individual (TSI)

Estabelecer as Hipóteses:

$$H_0: \beta_i = 0 \quad \text{VS} \quad H_1: \beta_i \neq 0, i = 1, \dots, 5 \quad (\text{Teste bilateral})$$

Estatística Teste:

$$T = \frac{\hat{\beta}_i}{\sigma\beta_i} \sim t(n - k - 1)$$

Valores observados:

$$\alpha = 0,05$$

$$\text{Valor de } p_{TA} = 0,1297$$

$$\text{Valor de } p_{CP} = 0,2194$$

$$\text{Valor de } p_{VN} = 0,5582$$

$$\text{Valor de } p_{CPF} = 0,4218$$

$$\text{Valor de } p_{RLP} = 0,0606$$

Tomada de decisão:

Como o valor de p de todas as variáveis independentes são $> \alpha$, não se rejeita a H_0 para o nível de significância de 5%, logo podemos afirmar que as 5 variáveis não são estatisticamente importantes para explicar o modelo.

3.3. Teste de significância geral (TSG)

Estabelecer as Hipóteses:

$$H_0: R^2 = 0 \quad \text{VS} \quad H_1: R^2 \neq 0 \quad (\text{Teste unilateral à direita})$$

Estatística-Teste:

$$F = \frac{(n - k - 1)}{k} * \frac{R^2}{1 - R^2} \sim F(k, n - k - 1)$$

Valor observado:

$$f_{obs} = 1,485286$$

Tomada de decisão:

Valor de p = 0,193102

Como o valor de $p > \alpha$, não se rejeita H_0 , podemos concluir que o modelo não é estatisticamente significativo para o nível de significância de 5%, logo, não temos nenhuma variável que explique o modelo.

3.4. Teste de White

Estabelecer as Hipóteses:

$$H_0: \text{Modelo Homocedástico} \quad \text{VS} \quad H_1: \text{Modelo Heterocedástico}$$

$$H_0: (V(E_i|X) = \sigma^2 \quad \text{VS} \quad H_1: V(E_i|X) \neq \sigma^2 \quad (\text{Teste unilateral à direita})$$

Estatística Teste:

$$Q = nR_w^2 \sim X^2(p - 1) \sim X^2(21 - 1)$$

Teste de White para a heterocedasticidade
Mínimos Quadrados (OLS), usando as observações 1-497 (n = 474)
Observações omissas ou incompletas foram ignoradas: 23
Variável dependente: uhat^2

	coeficiente	erro padrão	rácio-t	valor p
const	84356,0	77600,8	1,087	0,2776
TA	-11,0789	17,2644	-0,6417	0,5214
CP	2,12252	21,9287	0,09679	0,9229
VN	1827,13	1084,37	1,685	0,0927 *
CPF	394,036	809,515	0,4868	0,6267
RLP	-8,12011	60,0167	-0,1353	0,8924
sq_TA	1,83501e-05	0,000249205	0,07363	0,9413
X2_X3	7,44440e-05	0,000608070	0,1224	0,9026
X2_X4	0,0449195	0,105360	0,4263	0,6701
X2_X5	0,00514468	0,0165379	0,3111	0,7559
X2_X6	-0,000223792	0,00113352	-0,1974	0,8436
sq_CP	-3,44273e-05	0,000340192	-0,1012	0,9194
X3_X4	-0,0523967	0,116350	-0,4503	0,6527
X3_X5	-0,00906413	0,0208694	-0,4343	0,6643
X3_X6	0,000409012	0,00134239	0,3047	0,7607
sq_VN	-1,36591	1,17628	-1,161	0,2462
X4_X5	-3,19539	5,21801	-0,6124	0,5406
X4_X6	0,0866419	0,541533	0,1600	0,8730
sq_CPF	-0,185014	0,438927	-0,4215	0,6736
X5_X6	0,00998068	0,0510585	0,1955	0,8451
sq_RLP	0,000302187	0,00181541	0,1665	0,8679

R-quadrado não-ajustado = 0,007993

Estatística de teste: $TR^2 = 3,788640$,
com valor p = $P(\text{Qui-quadrado}(20) > 3,788640) = 0,999970$

Figura 4 Teste de White do Modelo 1

Valor observado:

$$q_{obs} = 3,788640$$

Tomada de decisão:

Neste caso temos o valor p = 0,999970, sendo, valor p > α , para o nível de significância de 5%, não se rejeita H_0 e podemos concluir que o modelo é homocedástico.

Se a H_0 tivesse sido rejeitada, a conclusão seria de que o modelo seria heterocedástico e para corrigir a heterocedasticidade utilizaríamos a correção de White para corrigir as variâncias, o que no nosso caso não será necessário, uma vez que o nosso modelo é homocedástico e os erros aleatórios têm variância constante.

3.5. Teste de Chow

Regressão aumentada para o teste de Chow
Mínimos Quadrados (OLS), usando as observações 1-497 (n = 474)
Observações omissas ou incompletas foram ignoradas: 23
Variável dependente: ROE

	coeficiente	erro padrão	rácio-t	valor p
const	39,7488	23,1560	1,717	0,0867 *
TA	-0,00250077	0,00307216	-0,8140	0,4161
CP	0,00216788	0,00359541	0,6030	0,5468
VN	0,0430579	0,0894997	0,4811	0,6307
CPF	0,0699196	0,110113	0,6350	0,5258
RLP	0,00316101	0,00696859	0,4536	0,6503
splitdum	-20,1574	30,5643	-0,6595	0,5099
sd_TA	0,00204219	0,00368493	0,5542	0,5797
sd_CP	-0,00156254	0,00425898	-0,3669	0,7139
sd_VN	0,0809260	0,624451	0,1296	0,8969
sd_CPF	-0,0983321	0,165352	-0,5947	0,5523
sd_RLP	0,0212585	0,0135041	1,574	0,1161
Média var. dependente	20,93916	D.P. var. dependente	299,4342	
Soma resid. quadrados	41370219	E.P. da regressão	299,2423	
R-quadrado	0,024507	R-quadrado ajustado	0,001281	
F(11, 462)	1,055167	valor P(F)	0,396453	
Log. da verosimilhança	-3368,894	Critério de Akaike	6761,788	
Critério de Schwarz	6811,722	Critério Hannan-Quinn	6781,426	

Teste de Chow para a falha estrutural na observação 249
F(6, 462) = 0,701471 com valor p 0,6486

Figura 5 Teste de Chow do Modelo 1

Estabelecer as Hipóteses:

H_0 : Não há alteração de estrutura VS H_1 : Há alteração de estrutura

(Teste unilateral à direita)

Estatística-Teste:

$$F = \frac{\frac{(n_T - (k+1))\delta_T^2 - ((n_{G_1} - (k+1))\delta_{G_1}^2 - (n_{G_2} - (k-1))\delta_{G_2}^2)}{K+1}}{\frac{(n_{G_1} - (k-1))\delta_{G_1}^2 + (n_{G_2} - (k+1))\delta_{G_2}^2}{n_T - 2(k+1)}} \sim F(k+1, n_T - 2(k+1))$$

Valor observado:

$$f_{obs} = 0,701471$$

Tomada de decisão:

Como o valor $p = 0,6486 > \alpha$, não rejeitamos H_0 ao nível de significância de 5%, logo não há alteração da estrutura de dados.

3.6.Multicolinearidade

```
Factores de Inflaccionamento da Variância (VIF)
Valor mínimo possível = 1,0
Valores > 10,0 podem indicar um problema de colinearidade

      TA      8,932
      CP      5,934
      VN      1,048
      CPF     2,688
      RLP     1,030

VIF(j) = 1/(1 - R(j)^2), onde R(j) é o coeficiente de correlação múltipla
entre a variável j e a outra variável independente

Diagnósticos de colinearidade de Belsley-Kuh-Welsch:

proporções de variância

lambda  cond  const    TA      CP      VN      CPF      RLP
2,452   1,000  0,039   0,015   0,018   0,020   0,026   0,000
1,159   1,454  0,064   0,001   0,000   0,241   0,042   0,376
0,909   1,642  0,187   0,003   0,013   0,245   0,001   0,472
0,747   1,811  0,571   0,002   0,020   0,308   0,035   0,075
0,676   1,904  0,116   0,000   0,050   0,170   0,269   0,073
0,057   6,533  0,023   0,978   0,898   0,017   0,627   0,005

lambda = valores próprios da inversa da matriz de covariância (smallest is 0,0574499)
cond    = índice da condição
nota: a soma da coluna das proporções de variância é 1.0

De acordo com BKW, cond >= 30 indica uma "forte" quase dependência linear,
e cond entre 10 e 30 "moderadamente forte". As estimativas de parâmetros
cuja variância está principalmente associada a valores cond problemáticos,
podem elas mesmas serem consideradas problemáticas.

Contagem de índices de condição >= 30: 0
Contagem de índices de condição >= 10: 0

Não há evidência de colinearidade excessiva
```

Figura 6 Colinearidade do modelo 1

O modelo 1 apresenta para todas as variáveis independentes valores <10 , o que significa que os coeficientes de regressão não estão mal estimados, caso estivessem, as soluções possíveis seriam proceder à centralização, retirar as variáveis cujo valor seja >10 ou utilizar outro tipo de regressão.

Relativamente às variáveis TA e CP, os seus valores encontram-se entre 5 e 10 o que indica que poderá existir colinearidade ou multicolinearidade elevada entre as variáveis, devendo ser ignorado.

No caso das variáveis independentes VN, CPF e RLP os valores são <5, sendo estes os valores ideais, apesar de poder indicar a existência de uma correlação moderada, mas insuficiente para ser considerada problemática.

Para não existir multicolinearidade, os valores deveriam ser iguais a 1, neste caso, as variáveis VN e RLP são as que mais se aproximam.

3.7. Valores dos coeficientes estimados

TA: $\hat{C}_1 = -0,00223645$ -> Acréscimo de 1% em TA provoca uma variação de 0,00224%

CP: $\hat{C}_2 = 0,00198465$ -> Acréscimo de 1% em CP provoca uma variação de 0,00198%

VN: $\hat{C}_3 = 0,0499878$ -> Acréscimo de 1% em VN provoca uma variação de 0,04998%

CPF: $\hat{C}_4 = 0,0489666$ -> Acréscimo de 1% em CPF provoca uma variação de 0,04896%

RLP: $\hat{C}_5 = 0,00989891$ -> Acréscimo de 1% em RLP provoca uma variação de 0,00989%

4. CONCLUSÃO

Este trabalho permitiu-nos analisar estatisticamente os retornos de três ativos, e o retorno do portefólio constituído por esses três ativos, que estão cotados na bolsa de valores, através das suas cotações diárias; por outro lado, reproduzir o estudo de um artigo, à escolha, relativamente a uma base de dados de indicadores financeiros de um conjunto de empresas.

Na primeira parte analisámos as cotações diárias, nomeadamente de 364 dias, de três criptomoedas, Ethereum, Binance Coin e Bitcoin, concluímos que o portefólio composto pelos três ativos mencionados, com um peso de 25%, 25% e 50% respetivamente, apresenta um retorno esperado de 21.611,86 euros, concluindo-se assim que é mais vantajoso investir nas ações da Bitcoin, por apresentar um retorno esperado de 41.976,75 euros.

Na segunda parte, tendo-se optado pelo artigo: FINANCIAL PERFORMANCE OF PRIVATE COMMERCIAL BANKS IN INDIA: MULTIPLE REGRESSION ANALYSIS como modelo de análise. Aplicando o Modelo de Regressão Linear Múltiplo e estimando os 3 modelos, concluímos que o Modelo 1 era o que melhor explicava o comportamento da variável dependente ROE, sendo que este modelo explica, aproximadamente, 1,56% do comportamento do ROE.

Verificamos, ainda, que o modelo escolhido é homocedástico pelo que, não foi necessária a aplicação da correção de White.

REFERÊNCIAS BIBLIOGRÁFICAS

Bass, I. (2007). *Six Sigma Statistics with Excel and Minitab*. The McGraw-Hill Companies.

Larguinho, M. (2021). *Infoestudante*. Obtido de Politécnico de Coimbra: infoestudante.ipc.pt

Nataraja, N., Chilale, N. R., & Ganesh, L. (2018). FINANCIAL PERFORMANCE OF PRIVATE COMMERCIAL BANKS IN INDIA: MULTIPLE REGRESSION ANALYSIS.

Navidi, W. (2010). *Statistics for Engineers and Scientists*. The McGraw-Hill Companies.

Royal, J. (2021). *12 most popular types of cryptocurrency*. Obtido de Bankrate:
<https://www.bankrate.com/investing/types-of-cryptocurrency/>

Stephens, L. J., & Spiegel, M. R. (2007). *Theory and problems of statistics*. The McGraw-Hill Companies.