# UTILIZAÇÃO DE MÉTODOS ESTATÍSTICOS E DA LINGUAGEM PYTHON NO MERCADO FINANCEIRO

Igor Veridiano Andrade<sup>1</sup>

Ciências da Computação



ISSN IMPRESSO 1980-1777
ISSN ELETRÔNICO 2316-3135

#### **RESUMO**

O mercado financeiro é uma área que necessita de muitos cálculos matemáticos e este vem se utilizando da estatística para seus cálculos, os bons investidores sempre estão buscando maneiras, de poderem determinar se uma ação ou carteira de ação, que possua o maior rendimento possível e menor risco possível, já que eles visam ao lucro, estes buscam ativos que representem esses fatores. Para ajudar os investidores a escolherem os melhores ativos, estão sendo utilizadas ferramentas de programação como python, para fazer os cálculos estatísticos, que representam o quão bom é um determinado ativo, tal junção dos conceitos estatísticos e da linguagem de programação, podem ser fortes aliados aos investidores que possuam um mínimo de conhecimento sobre o mercado financeiro.

#### **PALAVRAS-CHAVE**

Mercado Financeiro. Ativos. Carteira de Ações. Python. Estatística.

#### **ABSTRACT**

The financial market is an area that needs a lot of mathematical calculations, and this has been using statistics for its calculations, good investors are always looking for ways to determine if a particular stock or portfolio that has the highest yield possible., and the lowest possible risk, as they aim for profit, they seek assets that represent these factors. To help investors choose the best assets, programming tools such as python are being used to do statistical calculations, which represent how good a particular asset is, such a combination of statistical concepts and programming language can be strong allies. investors who have a minimum knowledge of the financial market.

#### **KEYWORDS**

Financial market. Active. Stock Portfolio. Python. Statistic

# 1 INTRODUÇÃO

No presente artigo abordar-se-á o uso de python para a utilização por pessoas que desejam atuar no mercado financeiro, utilizando conceitos matemáticos e estatísticos, junto da linguagem python, para que uma pessoa que pense em investir, possua melhor conhecimento e ferramentas para ter um melhor desempenho no mercado financeiro.

O artigo é dividido em uma breve explicação de python, ação, portfólio de ações, explicações sobre métodos estatísticos que são a variância, desvio padrão, covariância e correlação; serão abordadadas questões de finanças como Risco x Retorno, retorno de um portfólio, diversificação de um portfólio, risco múltiplo (portfólio) e o Risco sistemático vs Risco Idiossincrático, tudo isso com exemplos em python de como calcular e entender o seu significado nas finanças.

Portanto, o artigo tem como objetivo final fazer com que as pessoas entendam como o python pode ser uma ótima alternativa, para aqueles que desejam trabalhar ou entrar no mundo do mercado financeiro.

#### 2 PYTHON

O python foi criado para ser uma linguagem de fácil leitura, que é utilizada para machine learning, mas que também pode ser utilizada para cálculos matemáticos.

Variáveis em python não precisam ter explícito qual seu tipo na hora de declarar, já que é uma linguagem dinâmica, para se declarar uma variável é feita igual mostrado a seguir:

```
# Setando 'a' como 7

a = 7

# Setando 'b' como 8

b = 8

# Setando 'resultado' como soma de a+ b

resultado = a + b

# Mostrando o resultado na tela

print(resultado)

Fonte: Autor
```

Neste exemplo temos uma variável 'a' que recebe o valor de 7, uma variável 'b' que recebe o valor de 8, e uma variável resultado que recebe a soma de 'a' e 'b', e por fim é mostrado o resultado da soma se utilizando do comando 'print'.

Além da soma como foi mostrada no exemplo anterior pode ser feito os outros tipos de operações que serão mostradas a seguir:

```
# Setando 'a' como 7
a = 7
# Setando 'b' como 8
b = 8
# Setando 'resultado' como multiplicação de a * b
resultado = a * b
# Mostrando o resultado na tela
print(resultado)
# Setando 'resultado' como divisão de a / b
resultado = a / b
# Mostrando o resultado na tela
print(resultado)
# Setando 'resultado' como 'a' elevado a 'b'
resultado = a ** b
# Mostrando o resultado na tela
print(resultado)
# Setando 'resultado' como 'a' elevado a 0.5 para se fazer a raiz guadrada
resultado = a ** 0.5
# Mostrando o resultado na tela
Print (resultado)
Fonte: Autor.
```

O python também é uma ótima alternativa, pois possui muitas bibliotecas que podem auxiliar em diversos tipos de situações, neste artigo será utilizado as bibliotecas pandas e numpy.

## 2.1 AÇÃO

Ação são a menor fração do capital de uma empresa, são papéis que representam o direito que um investidor tem nos recebíveis da distribuição de lucro de uma empresa. Tais ações podem ser divididos em dois tipos que são; ordinárias e preferenciais.

Ações ordinárias são ações que permitem que o seu detentor, tenha participação na decisão da empresa com direito a voto, mas não tem prioridade sobre os dividendos (divisão dos lucros para os acionistas).

Ações preferenciais são ações que oferecem a seu detentor preferência no recebimento dos dividendos da empresa, mas não permite participação de voto nas decisões da empresa.

## 2.2 PORTFÓLIO DE AÇÕES

Um portfólio de ações é uma carteira de um investidor onde, este possui diversos tipos de ações diferentes para que os riscos das ações sejam diminuídos, enquanto tenta-se maximizar os lucros que estas geram.

#### 2.3 VARIÂNCIA

A variância é uma medida estatística, que mede a dispersão de um conjunto de dados, ela mede o quanto esse conjunto de dados se desvia da média dele. Sua fórmula matemática será de:

$$S^2 = \frac{\Sigma(X - X)^2}{N - 1}$$

#### 2.4 DESVIO PADRÃO

O desvio padrão é uma medida estatística, que expressa quanto é o grau de dispersão de um conjunto de dados, tais dados serão homogêneos, quanto mais próximo o desvio for de 0.

#### 2.5 RISCO x RETORNO

Retorno é o quanto um ativo financeiro pode retornar de lucro em um determinado tempo, para o investidor. Esse retorno muitas vezes é mostrado por meio de uma taxa, que é a taxa de retorno de determinado ativo durante um determinado tempo. Essa taxa, pode ser de retorno simples (taxa de retorno simples), que se utiliza guando se tem que se lidar com diversos ativos, em um mesmo período, que é dado pela equação:

taxa de retorno 
$$=\frac{\text{preço final } - \text{preço inicial}}{\text{preço inicial}}$$

Essa equação pode ser simplificada para:

taxa de retorno 
$$=\frac{\text{preço final}}{\text{preço inicial}}$$
 - 1

Essa taxa de retorno é para um único dia, para poder fazer para um determinado período, deve ser feita a conversão, um exemplo seria a anual, que é dado pela equação:

taxa de retorno atual = taxa de retorno \* 250

A taxa anual será a taxa diária multiplicada por 250, pois um ano para o mundo financeiro tem 250 dias.

Para o cálculo da taxa de retorno simples foi se utilizado a api do *yahoo finance,* para se pegar uma ação de exemplo para o cálculo de sua taxa de retorno simples anual, foi também utilizada as bibliotecas do pandas\_datareader como apoio, o código pode ser encontrado a seguir:

```
from pandas_datareader import data as wb
# Recuperando os dados de uma ação na api do yahoo finance
PG = wb.DataReader('PG', data_source='yahoo', start='1995-1-1')
# Aqui se está calculando o retorno simples para cada día dos dados da api
PG['retorno_simples'] = (PG['Adj Close'] / PG['Adj Close'].shift(1)) - 1
# Média dos retornos diários
retorno_diario_medio = PG['retorno_simples'].mean()
print(retorno_diario_medio)
# Como um período de negociação anual é de 250 dias, deve se fazer o
# retorno diário médio vezes 250 dias, para se ter o anual
retorno_anual_medio = retorno_diario_medio * 250
print(retorno_anual_medio)
# Arredondando o retorno anual para 5 casas, e multiplicando por 100, para
# se ter a porcentagem correta anual, da taxa de retorno simples
print(str(round(retorno_anual_medio, 5) * 100) + '%')
Fonte: Autor.
```

O outro tipo de retorno é o logarítmico (taxa de retorno logarítmico), que se utiliza quando se trabalha com um único ativo ao longo de um tempo, que é dado pela equação:

taxa de retorno 
$$= log \left( \frac{preço final}{preço inicial} \right)$$

Assim como a taxa de retorno simples essa taxa é para apenas um dia, e para outro determinado período tem de ser feitas as devidas conversões.

Para o cálculo da taxa de retorno logarítmica foi se utilizado a api do *yahoo finance,* para se pegar uma ação de exemplo para o cálculo de sua taxa de retorno logarítmica anual, foi também utilizada as bibliotecas do pandas\_datareader e do numpy como apoio, o código pode ser encontrado a seguir:

```
import numpy as np
from pandas_datareader import data as wb
# Recuperando os dados de uma ação na api do vahoo finance
PG = wb.DataReader('PG', data_source='vahoo', start='1995-1-1')
# Calculando o retorno logarítmico para cada día dos dados da api
PG['retorno_logaritmico'] = np.log(PG['Adj Close'] / PG['Adj Close'].shift(1))
# Média dos retornos diários
retorno_diario_medio = PG['retorno_logaritmico'].mean()
print(retorno_diario_medio)
# Como um período de negociação anual é de 250 dias, deve se fazer o
# retorno diário médio vezes 250 dias, para se ter o anual
retorno_anual_medio = retorno_diario_medio * 250
print(retorno_anual_medio)
# Arredondando o retorno anual para 5 casas, e multiplicando por 100, para
# se ter a porcentagem correta anual, da taxa de retorno simples
print(str(round(retorno_anual_medio, 5) * 100) + '%')
Fonte: Autor.
```

Risco é a probabilidade de uma ação ou determinado ativo financeiro tem de gerar prejuízo para um investidor, que pode ser causado por valores externos ao mercado financeiro, fatores esses que podem impactar nos preços e nos dividendos dos ativos financeiros.

Para saber se um ativo é mais ou menos arriscado, deve ser calculada a volatilidade do retorno do ativo, para se fazer isso deve se calcular o desvio padrão, pois quanto maior o desvio padrão, maior será a volatilidade do ativo.

Para o cálculo do risco foi se utilizado a api do *yahoo finance*, para se pegar as ações de exemplo para o cálculo do risco de cada uma, foi também utilizada as bibliotecas do pandas, pandas\_datareader e do numpy como apoio, o código pode ser encontrado a seguir:

```
import numpy as np
import pandas as pd
from pandas_datareader import data as wb
# Papéis das ações a serem verificadas
papeis = PG['PG', 'BEI.DE']
# Recuperando os dados financeiros dos papéis, na api do yahoo
```

Fonte: Autor.

# 2.6 RETORNO DE UM PORTFÓI IO

O retorno em um portfólio de ações, é feito por meio de um cálculo que se utiliza as taxas de retorno dos ativos que compõem o portfólio e do peso que cada um desses ativos representa, a expressão para o cálculo seria como mostrado a seguir:

taxa de retorno =  $\Sigma$  taxa de retorno [x] x peso [x]

Neste caso essa taxa de retorno será a anual caso tenha feito o cálculo com as taxas anuais dos ativos, ou seja, depende do tipo de taxa que for usada, ou se pode fazer a conversão se necessário.

Para o cálculo da taxa de retorno simples do portfólio foi se utilizado a api do *yahoo finance*, para se pegar as ações de exemplo para o cálculo taxa de retorno simples do portfólio, foi também utilizada as bibliotecas do pandas, pandas\_datareader e do numpy como apoio, o código pode ser encontrado a seguir:

```
import numpy as np
import pandas as pd
from pandas_datareader import data as wb
# Papéis das ações a serem verificadas
papeis = PG['PG', 'MSFT', 'F', 'GE']
# Recuperando os dados financeiros dos papéis, na api do yahoo
dados = pd.DataFrame()
for t in papeis:
dados[t] = wb.DataReader(t, 'PG', data_source='yahoo', start='1995-1-1')['Adj Close']
# Calculando o retorno logarítmico diário dos papéis
retorno = (dados/ dados.shift(1)) - 1
# Setando os pesos de cada papel
pesos = np.array([0.25, 0.25, 0.25, 0.25])
```

# Retorno médio anual de cada papel
retorno\_anual = retorno.mean() \* 250
# Calculando o retorno anual para o portfólio
np.dot(retorno\_anual, pesos)
# Colocando o retorno anual para porcentagem
portfolio=str(round(np.dot(retorno\_anual, pesos), 5) \* 100) + ' %'
print(portfolio)
Fonte: Autor.

Observação: os pesos dos ativos do portfólio não podem ultrapassar 1, pois a soma dos pesos devem dar 100%, que no caso da notação utilizada neste artigo é 1.

#### 2.7 COVARIÂNCIA

A covariância é uma medida estatística que mede a relação entre duas variáveis, uma covariância de 0 entre duas variáveis significa que as mesmas não têm relação alguma entre si, ou seja, são variáveis independentes, tal relação pode ser calculada por meio da fórmula mostrada a seguir:

$$P_{xy} = \frac{(x \cdot x') \times (y \cdot y)}{\sigma_{\chi} \quad \sigma_{y}}$$

## 2.8 CORRELAÇÃO

A correlação na estatística é uma relação de dependência que exista entre duas variáveis, ou seja, são variáveis que possuem algum grau de relação entre elas, onde que uma explica a outra variável, a correlação pode ser definida em perfeita, imperfeita, positiva, negativa ou pode não existir, no caso da perfeita e imperfeita ela também pode ser negativa ou positiva.

A correlação perfeita será quando as duas variáveis crescerem na mesma proporção, no caso de ser negativa irão crescer inversamente, enquanto uma aumenta a outra diminui, já se forem positivas irão crescer na mesma direção.

Na correlação imperfeita tanto a positiva como a negativa agem igual a perfeita, a diferença é que não será na mesma proporção, pois por ser uma correlação imperfeita, as variáveis irão crescer em proporção diferentes.

## 2.9 DIVERSIFICAÇÃO DE UM PORTFÓLIO

Uma maneira de o investidor diminuir o risco que corre e aumentar o seu ganho dos investimentos, ou seja, aumentar seus retornos, é diversificar os ativos em que investe, isso é investir em tipos diferentes de ações para que possa maximizar seu lucro, enquanto diminui o risco que corre.

Tal diversificação pode ser realizada, ao se criar um portfólio de ações que seja diversificado, ou seja, um portfólio que possui diferentes tipos de ativos de diversos segmentos.

Para saber quais ativos não possuem correlação entre si, é preciso se usar da covariância e correlação, para que então possa se detectar os melhores ativos para se criar um portfólio bem diversificado.

Para o cálculo da taxa de retorno simples do portfólio foi se utilizado a api do vahoo finance, para se pegar as ações de exemplo para o cálculo taxa de retorno simples do portfólio, foi também utilizada as bibliotecas do pandas, pandas\_datareader e do numpy como apoio, o código pode ser encontrado a seguir:

```
import numpy as np
import pandas as pd
from pandas_datareader import data as wb
# Papéis das ações a serem verificadas
papeis = PG['PG', 'BEI.DE']
# Recuperando os dados financeiros dos papéis, na api do vahoo
dados = pd.DataFrame()
for t in papeis:
dados[t] = wb.DataReader(t, 'PG', data_source='yahoo', start='1995-1-1')['Adj Close']
# Calculando o retorno logarítmico diário dos papéis
retorno = np.log(dados/ dados.shift(1))
# Calculando a variância diária
varianciaPG = retorno('PG').var()
print(varianciaPG)
varianciaBEI = retorno['BEI.DE'].var()
print(varianciaBEI)
# Calculando a variância anual
varianciaAnualPG = retorno['PG'].var() * 250
print(varianciaAnualPG)
varianciaAnualBEI = retorno['BEI.DE'].var() * 250
print(varianciaAnualBEI)
# Calculando a covariância diária
covariancia = retorno.cov()
print(covariancia)
# Calculando a covariância diária
covarianciaAnual = retorno.cov() * 250
print(covarianciaAnual)
# Calculando a correlação
correlacao= retorno.corr()
print(correlacao)
Fonte: Autor.
```

Uma observação é que não se deve converter a correlação para uma correlação anual, pois essa não é uma medida diária, ela é na verdade a relação que os preços das ações possuem, no caso desse exemplo.

#### 2.10 RISCO MÚLTIPLO (PORTFÓLIO)

Quando se trabalha com um portfólio de ações, deve se atentar não só o quanto este rende, mais também ao seu risco, por poder possuir vários tipos de ativos diferentes, que por sua vez irão possuir cada um seu risco, deve ser feito o cálculo do risco do portfólio, que assim como a taxa de retorno, dependerá do peso que cada ativo possui no portfólio e, no caso do risco, depende também da variância e da correlação dos ativos, o cálculo para o variância do portfólio pode ser feito por meio da expressão a seguir:

$$(w \square_1 \sigma \square_1 + w \square_2 \sigma \square_2) \square^2$$

$$= (w \square_1 \sigma \square_1) \square^2$$

$$+ 2w \square_1 \sigma \square_1 w \square_2 \sigma \square_2 \rho \square_{12}$$

$$+ (w \square_2 \sigma \square_2) \square^2$$

Para o cálculo do risco do portfólio foi se utilizado a api do *yahoo finance*, para se pegar as ações de exemplo para o cálculo do risco do portfólio foi, também, utilizada as bibliotecas do pandas, pandas\_datareader e do numpy como apoio, o código pode ser encontrado a seguir:

```
import numpy as np
import pandas as pd
from pandas_datareader import data as wb
# Papéis das ações a serem verificadas
papeis = PG['PG', 'BEI.DE']
# Recuperando os dados financeiros dos papéis, na api do yahoo
dados = pd.DataFrame()
for t in papeis:
dados[t] = wb.DataReader(t, 'PG', data_source='yahoo', start='1995-1-1')['Adj Close']
# Calculando o retorno logarítmico diário dos papéis
retorno = np.log(dados/ dados.shift(1))
# Setando os pesos de cada papel
pesos = np.array([0.5, 0.5])
# Calculando a variância do portfólio
portfolio_variancia = np.dot(pesos.T, np.dot(retorno.cov() * 250, pesos))
print(portfolio_variancia)
# Calculando a volatilidade do portfólio
```

#### 2.11 RISCO SISTEMÁTICO VS RISCO IDIOSSINCRÁTICO

O risco sistemático será o aquele risco que independe dos ativos que existem em um portfólio, pois este é um risco de mercado, um risco que independe do ativo, um exemplo seria as incertezas de um mercado local ou do mercado global, este risco é um risco que não pode ser eliminado.

Para o cálculo do risco não diversificável (sistemático) do portfólio foi se utilizado a api do *yahoo finance*, para se pegar as ações de exemplo para o cálculo do risco não diversificável (sistemático) do portfólio foi, também, utilizada as bibliotecas do pandas, pandas\_datareader e do numpy como apoio, o código pode ser encontrado a seguir:

```
import numpy as np
import pandas as pd
from pandas_datareader import data as wb
# Papéis das ações a serem verificadas
papeis = PG['PG', 'BEI.DE']
# Recuperando os dados financeiros dos papéis, na api do yahoo
dados = pd.DataFrame()
for t in papeis:
dados[t] = wb.DataReader(t, 'PG', data_source='yahoo', start='1995-1-1')['Adj Close']
# Calculando o retorno logarítmico diário dos papéis
retorno = np.log(dados/ dados.shift(1))
# Setando os pesos de cada papel
pesos = np.array([0.5, 0.5])
# Calculando a variância do portfólio
portfolio_variancia = np.dot(pesos.T, np.dot(retorno.cov() * 250, pesos))
print(portfolio_variancia)
# Calculando o risco de cada ação
varianciaAnualPG = retorno['PG'].var() * 250
print(varianciaAnualPG)
varianciaAnualBEI = retorno['BEI.DE'].var() * 250
print(varianciaAnualBEI )
# Calculando o risco não diversificável do portfólio
riscoNaoDiversificavel = (pesos[0] ** 2 * varianciaAnualPG) - (pesos[1] ** 2 * varian-
ciaAnualPG)
```

Fonte: Autor.

Enquanto que o risco não sistemático (risco idiossincrático) é um risco que é específico de um ativo ou de um conjunto de ativos, por exemplo seria uma empresa que inventa uma nova tecnologia, está então seria beneficiada por isso, outro exemplo seria uma empresa que ocorre um escândalo de corrupção, está já seria prejudicada, ou seja, o risco não sistemático é um risco que está atrelado ao ativo, diferente do risco sistemático, o não sistemático pode ser diminuído se utilizando a técnica de diversificação de ativos.

Para o cálculo do risco diversificável (idiossincrático) do portfólio foi se utilizado a api do *yahoo finance*, para se pegar as ações de exemplo para o cálculo do risco diversificável (idiossincrático) do portfólio, foi também utilizada as bibliotecas do pandas, pandas\_datareader e do numpy como apoio, o código pode ser encontrado a seguir:

```
import numpy as np
import pandas as pd
from pandas_datareader import data as wb
# Papéis das ações a serem verificadas
papeis = PG['PG', 'BEI.DE']
# Recuperando os dados financeiros dos papéis, na api do yahoo
dados = pd.DataFrame()
for t in papeis:
dados[t] = wb.DataReader(t, 'PG', data_source='yahoo', start='1995-1-1')['Adj Close']
# Calculando o retorno logarítmico diário dos papéis
retorno = np.log(dados/ dados.shift(1))
# Setando os pesos de cada papel
pesos = np.array([0.5, 0.5])
# Calculando a variância do portfólio
portfolio_variancia = np.dot(pesos.T, np.dot(retorno.cov() * 250, pesos))
print(portfolio_variancia)
# Calculando o risco de cada ação
varianciaAnualPG = retorno['PG'].var() * 250
print(varianciaAnualPG)
varianciaAnualBEI = retorno['BEI.DE'].var() * 250
print(varianciaAnualBEI )
# Calculando o risco diversificável do portfólio
riscoDiversificavel = portfolio_variancia - (pesos[0] ** 2 * varianciaAnualPG) - (pe-
sos[1] ** 2 * varianciaAnualPG)
print(riscoDiversificavel)
# Mostrando o risco diversificável em porcentagem
print(str(round(riscoDiversificavel * 100, 3)) + '%)
Fonte: Autor.
```

#### 3 CONCLUSÃO

Por fim como mostrado pelo artigo, o python utilizado junto com conceitos estatísticos e matemáticos, se torna uma ótima alternativa para aqueles que desejam adentrar no mundo das finanças, mais especificamente no mercado financeiro, como foi mostrado pode se utilizar da linguagem python para cálculos que mostram tanto as taxas de rendimento, como conseguir verificar e diminuir o risco que se terá ao investir em determinados ativos e com isso ter uma maior noção de em que ativos investir, para se ter o melhor rendimento com o menor risco possível.

#### **REFERÊNCIAS**

ACÃO (financas). Wikipédia, a enciclopédia livre, 2019. Disponível em: https:// pt.wikipedia.org/wiki/A%C3%A7%C3%A3o\_(finan%C3%A7as). Acesso em: 20 nov. 2019.

O que é uma ação financeira? **Portal Educação**. Disponível em: https://www. portaleducacao.com.br/conteudo/artigos/contabilidade/o-que-e-uma-acaofinanceira/63807. Acesso em: 20 nov. 2019.

GOUVEIA, Rosimar. **Desvio padrão**. Disponível em: https://www.todamateria.com.br/desvio-padrao/. Acesso em: 20 nov. 2019.

COVARIÂNCIA. Wikipédia, a enciclopédia livre, 2019. Disponível em: https:// pt.wikipedia.org/wiki/Covariância. Acesso em: 20 nov. 2019.

CORRELAÇÃO. Wikipédia, a enciclopédia livre, 2019. Disponível em: https:// pt.wikipedia.org/wiki/Correlação. Acesso em: 20 nov. 2019.

RISCO sistemático e não sistemático. Pro educacional. Disponível em: https:// proeducacional.com/ead/curso-cga-modulo-ii/capitulos/teoria-moderna-dascarteiras/aulas/aula-156/. Acesso em: 20 nov. 2019.

RATZAT, Erich. **Python para finanças**: Investimentos & Análise de Dados. 2019. Disponível em: https://www.udemy.com/course/python-para-financasinvestimentos-analise-de-dados/. Acesso em: 20 nov. 2019.

Data do recebimento: 19 de julho de 2019 Data da avaliação: 2 de novembro de 2019 Data de aceite: 3 de dezembro de 2019