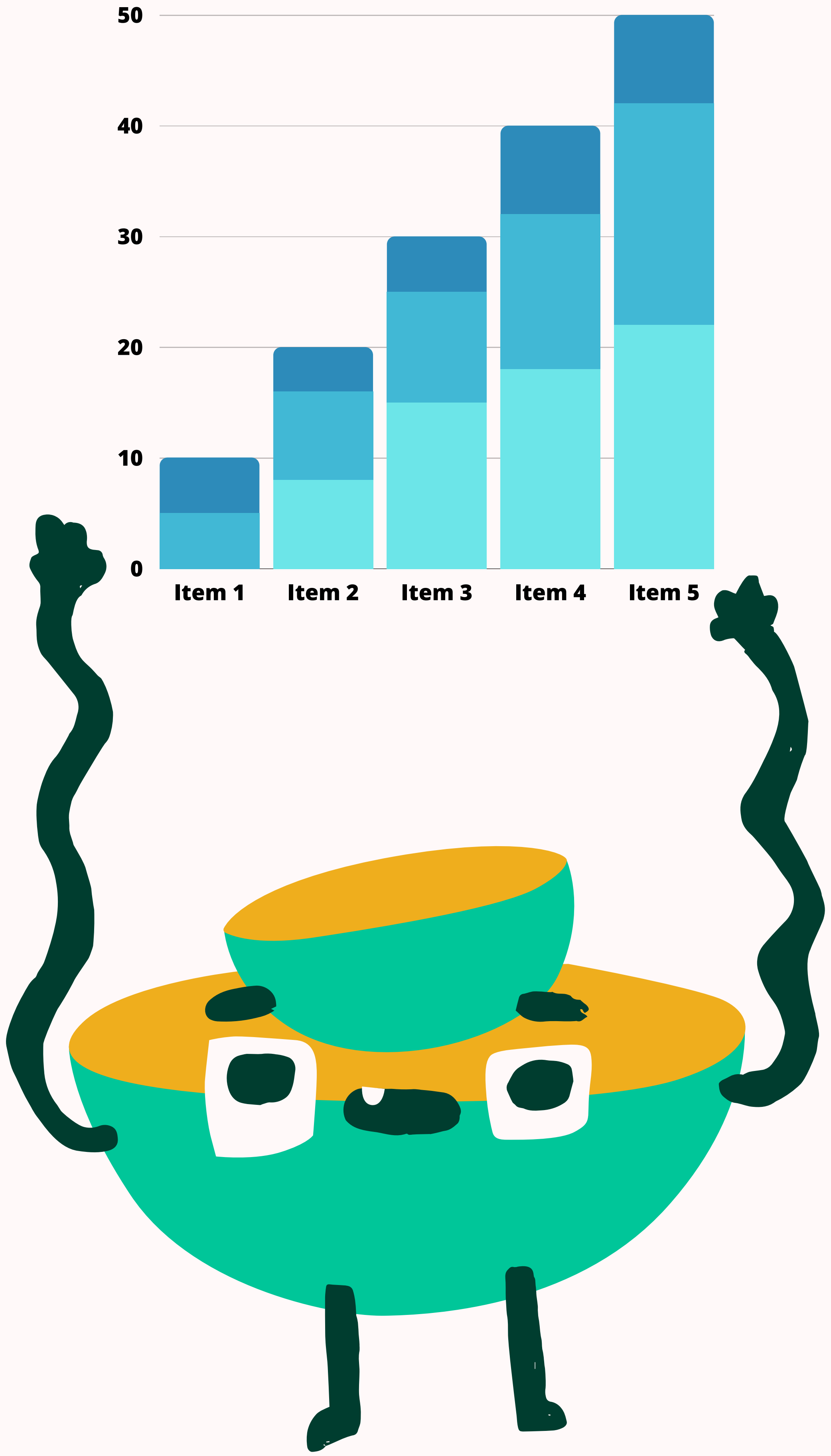
TRABALHO DE PROB ESTATÍSTICA.

Professora Sâmia

Integrantes :

Angelo Bianco
Anna Clara
Giuliana Prado
João Victor Rangel
Sávio Alexandre
Thiago Capuano



Introdução

- Santos, L.d., Corrêa, D.C., Walker, D.M. et al. Characterisation of neonatal cardiac dynamics using ordinal partition network. Med Biol Eng Comput 60, 829-842 (2022).
- Variação da frequência cardíaca -> intervalo RR -> maturação do Sistema Nervoso Autônomo
- Simplificação do modelo probabilístico
- Chance do recém-nascido ter ANS imaturo
- Pode tanto significar quanto acarretar doenças
- Heart rate time series / MIT-BIH Database. Disponível em: <https://www.kaggle.com/datasets/ahmadsaeed1007/heart-rate-time-series-mitbih-database>. Acesso em: 20 jun. 2023.

Teste Levene

É usado para testar a igualdade das variâncias de populações. Então se tivermos:

- Ho= as variâncias são iguais
- H1= as variâncias são diferentes

Calculado o p valor das amostras e considerando um nível de significância de 5% podemos observar que $p > \alpha$ apenas quando comparamos as colunas t3 e t4, nesse caso não rejeitamos Ho, e em todas as outras amostras temos $p < \alpha$, então rejeitamos Ho nesses casos.

```python
# Realizar o teste de Levene para cada coluna em relação às outras colunas
columns = ["T1", "T2", "T3", "T4"]
levene_results = []
for i, col1 in enumerate(columns):
    for j, col2 in enumerate(columns):
        if j > i:
            levene_test = levene(data[col1], data[col2])
            levene_results.append(((col1, col2), levene_test.statistic, levene_test.pvalue))
```

Teste Levene

O p valor das colunas T3 e T4 é de 95% portanto não rejeitamos a hipótese nula, então podemos dizer que as variâncias dessas amostras são iguais estatisticamente

```
Teste de Levene:
Colunas T1 e T2:
Estatística de teste: 25.719714296600515
Valor p: 4.3589493000714685e-07

Colunas T1 e T3:
Estatística de teste: 138.95833193348153
Valor p: 6.154023934442789e-31

Colunas T1 e T4:
Estatística de teste: 128.53541846824396
Valor p: 8.12453659408405e-29

Colunas T2 e T3:
Estatística de teste: 208.6537212494096
Valor p: 8.323845676019624e-45

Colunas T2 e T4:
Estatística de teste: 198.82013512306625
Valor p: 7.009118354307755e-43

Colunas T3 e T4:
Estatística de teste: 0.002724157291917194
Valor p: 0.9583804002693371
```

Teste de Shapiro-Wilk

O teste de Shapiro-Wilk testa a hipótese nula que uma amostra $y_1, y_2, \cdots, y_n$, retirada de uma população, tem distribuição normal. Para calcular o valor da estatística W, dada a amostra aleatória, de tamanho n, deve-se proceder da seguinte maneira:

1. Obter uma amostra ordenada: $y_1 \leq y_2 \leq \cdots \leq y_n$;

2. Calcular a soma de quadrado dos desvios:

$$S^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

3. Uma vez que se tem o número de amostras (n) coletadas, tem-se que:

- **se n é par**, $n = 2k$, calcule:

$$b = \sum_{i=1}^{k} a_{n-i+1}\left(y_{n-i+1} - y_i\right)$$

em que os valores de a_i são obtidos na tabela de coeficientes de Shapiro-Wilk.

- **se n é ímpar**, $n = 2k+1$, o cálculo é exatamente como no item anterior, uma vez que $a_{k+1} = 0$ quando $n = 2k+1$. Assim, determina-se:

$$b = a_n\,(y_n - y_1) + \cdots + a_{k+2}\,(y_{k+2} - y_k)$$

em que o valor de y_{k+1}, que é a mediana, não entra no cálculo de b.

4. Calcule

$$W = \frac{b^2}{S^2}.$$

5. Para tomada de decisão a respeito da normalidade dos dados, compara-se o valor calculado de W com o valor tabelado $W_{n;\alpha}$, obtido da Tabela Shapiro_prob. Se o valor calculado W for menor que o tabelado, rejeita-se a hipótese de normalidade ao nível α de significância.

Aplicação aos resíduos da análise de variância dos dados de colesterol.

As hipóteses testadas são:

$$\begin{cases} H_0 : \text{Os erros têm distribuição normal;} \\ H_1 : \text{Os erros não têm distribuição normal.} \end{cases}$$

```
Teste de Shapiro-Wilk:
Coluna T1:
Estatística de teste: 0.990785539150238
Valor p: 2.2666641598334536e-05

Coluna T2:
Estatística de teste: 0.9154111742973328
Valor p: 6.955894249103389e-22

Coluna T3:
Estatística de teste: 0.8935468196868896
Valor p: 2.5102735323338164e-24

Coluna T4:
Estatística de teste: 0.8576120138168335
Valor p: 1.2929983941431395e-27
```

Tabela ANOVA

A tabela ANOVA, ou Análise de Variância, é uma ferramenta estatística usada para comparar as médias de três ou mais grupos ou populações. Ela é amplamente utilizada em experimentos e estudos científicos para determinar se existe uma diferença significativa entre as médias dos grupos em análise. A tabela ANOVA desempenha um papel fundamental na decomposição da variabilidade total dos dados em componentes associados às diferenças entre os grupos e à variabilidade dentro dos grupos. Ela fornece estatísticas como a soma de quadrados, os graus de liberdade, as médias quadráticas e os valores-p, que ajudam a interpretar os resultados e a tomar decisões estatísticas sobre as diferenças observadas.

Código do teste de variância

Aqui, obtemos o resultado retornado pelo nosso código. Onde os dados dispostos na tabela foram tratados e representados de formas mais clara e precisa quanto ao seu significado.

```
Colunas T2 e T4:
Estatística de teste: 198.82013512306625
Valor p: 7.009118354307755e-43

Colunas T3 e T4:
Estatística de teste: 0.002724157291917194
Valor p: 0.9583804002693371

Teste de Shapiro-Wilk:
Coluna T1:
Estatística de teste: 0.990785539150238
Valor p: 2.2666641598334536e-05

Coluna T2:
Estatística de teste: 0.9154111742973328
Valor p: 6.955894249103389e-22

Coluna T3:
Estatística de teste: 0.8935468196868896
Valor p: 2.5102735323338164e-24

Coluna T4:
Estatística de teste: 0.8576120138168335
Valor p: 1.2929983941431395e-27

Tabela ANOVA:
              df        sum_sq        mean_sq             F  PR(>F)
grupo        3.0  1.148866e+06  382955.245374  19093.925835     0.0
Residual  3560.0  7.140075e+04      20.056391           NaN     NaN
```

```python
# Realizar o teste de Shapiro-Wilk para cada coluna
shapiro_results = []
for col in columns:
    shapiro_test = shapiro(data[col])
    shapiro_results.append((shapiro_test.statistic, shapiro_test.pvalue))

# Realizar a ANOVA
formula = "valor ~ grupo"
long_data = pd.melt(data, var_name='grupo', value_vars=columns, value_name='valor')
model = ols(formula, data=long_data).fit()
anova_table = anova_lm(model)

# Exibir os resultados
print("Teste de Levene:")
for result in levene_results:
    (col1, col2), statistic, pvalue = result
    print(f"Colunas {col1} e {col2}:")
    print("Estatística de teste:", statistic)
    print("Valor p:", pvalue)
    print()

print("Teste de Shapiro-Wilk:")
for i, col in enumerate(columns):
    print(f"Coluna {col}:")
    print("Estatística de teste:", shapiro_results[i][0])
    print("Valor p:", shapiro_results[i][1])
    print()

print("Tabela ANOVA:")
print(anova_table)
```

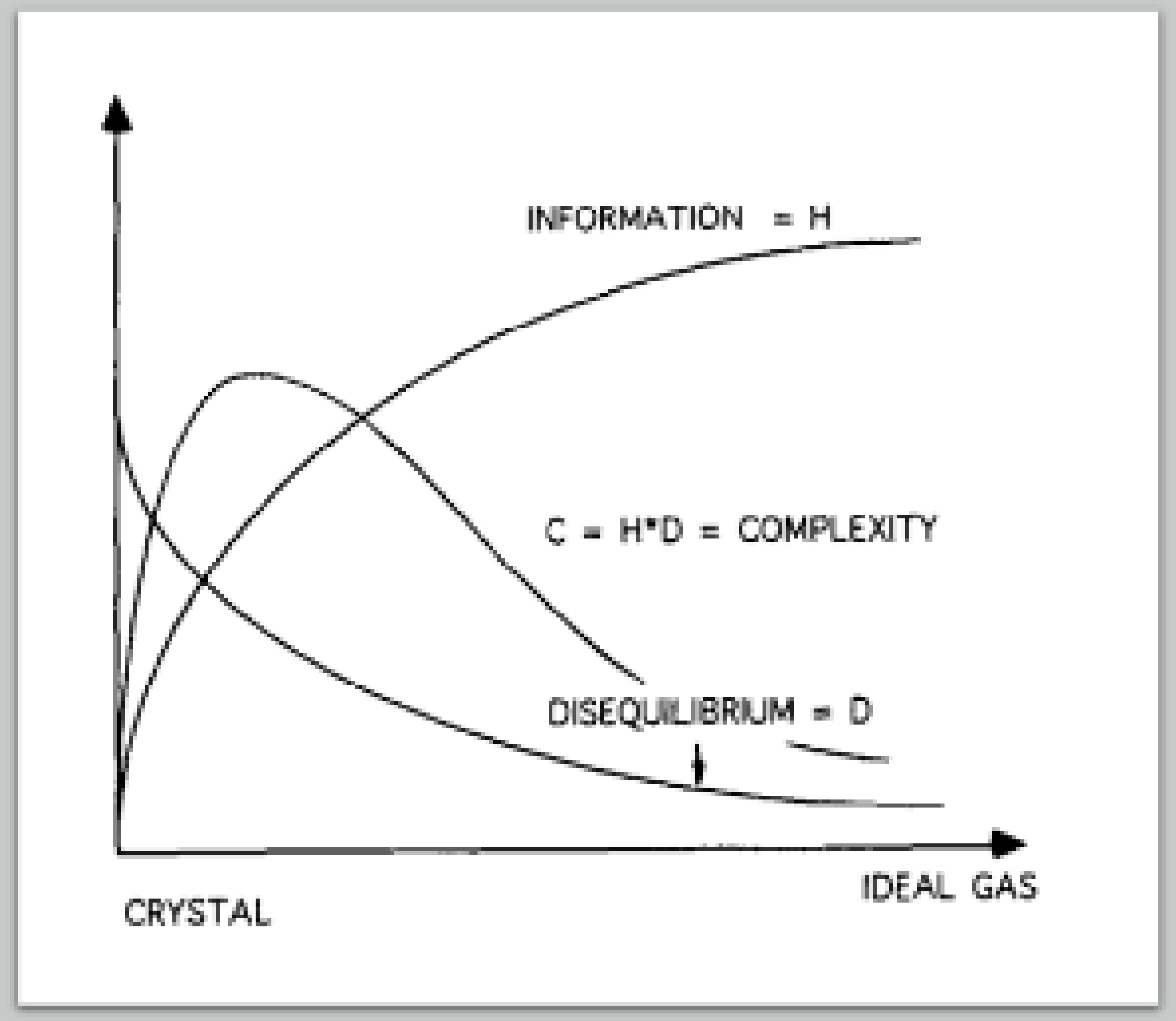
Definição matemática

Definição Matemática

$$C = HD$$

$$H = -K \sum_{i=1}^{N} p_i \log p_i ,$$

$$D = \sum_{i=1}^{N} (p_i - 1/N)^2 .$$



Substituindo na fórmula

$$
\begin{aligned}
C &= HD \\
&= -\left(K\sum_{i=1}^{N} p_i \log p_i\right)\left(\sum_{i=1}^{N}(p_i - 1/N)^2\right). \qquad (3)
\end{aligned}
$$

Para um sistema de duas possibilidades

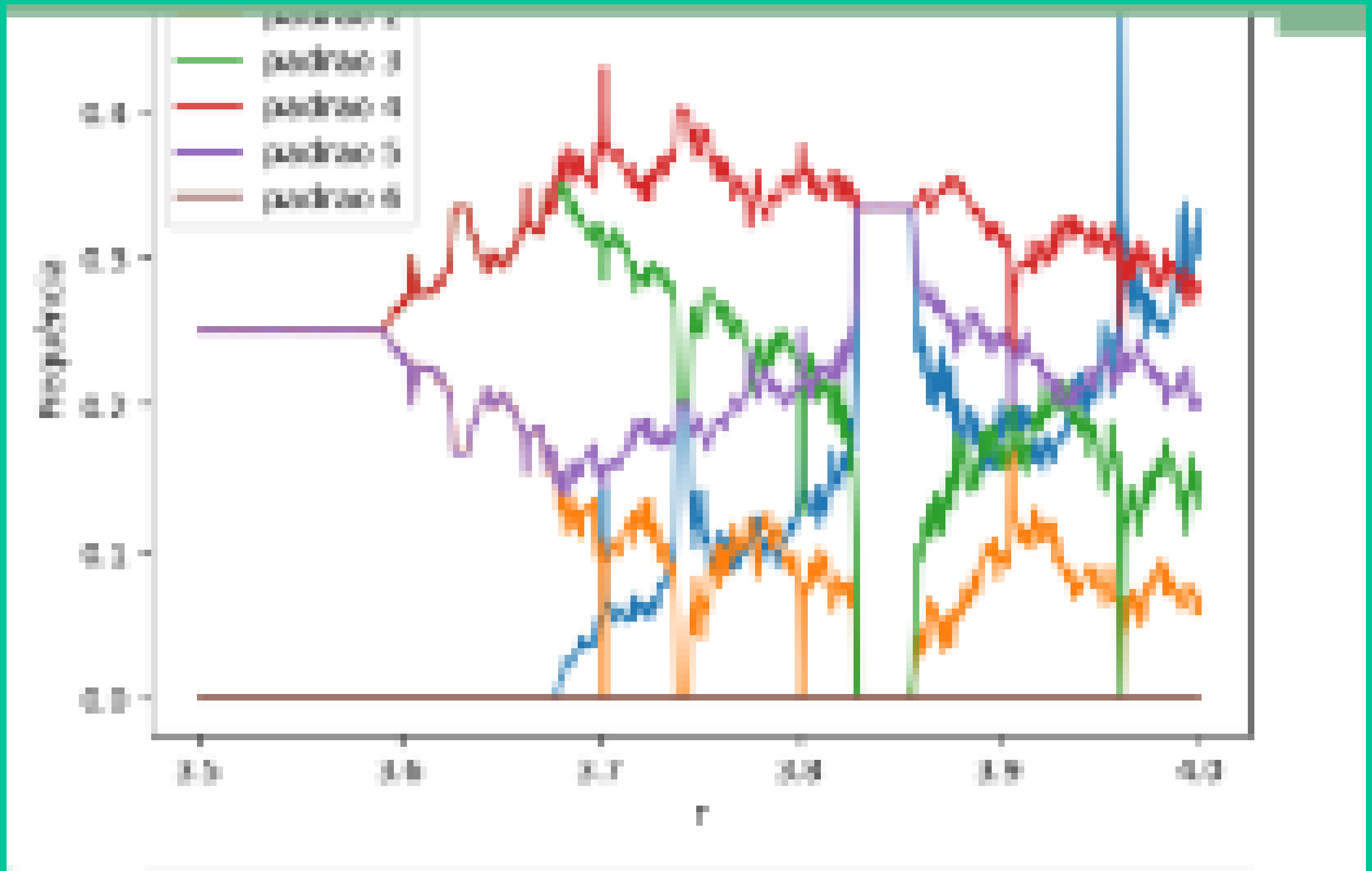
$$\overline{C}(x) = \overline{H}(x)D(x)$$
$$= -\frac{1}{\log 2}\left[x\log\left(\frac{x}{1-x}\right) + \log(1-x)\right]$$
$$\times 2(x-\tfrac{1}{2})^2. \qquad (5)$$

É possivel ser calculado mais para sistemas N-3, N-5, N-7, nos teriamos maior dificuldade.

O mapa de Lorentz

- O mapa imita o comportamento e o desenvolvimento do atrator de Lorentz

- O mapa imita o comportamento e o desenvolvimento do atrator de Lorentz



features of the first return map of this attractor: $\{x_{n+1} = \alpha x_n$ if $x_n < 0.5$ and $x_{n+1} = \alpha(x_n - 1) + 1$ if $x_n > 0.5\}$ where $\alpha \in (0, 2)$. Its dynamic evolution displays three

CONCLUSÃO

Bibliotecas usadas no código:

- Scipy
- Statsmodels

Motivos do código não ter funcionado:

- O teste de variância dá falso, logo a hipótese nula é falsa.
- Apesar das estatísticas terem uma distribuição normal, os dados são complexos e podem variar com vários fatores externos. Por isso, a teoria de Shapiro Wilk não é suficiente para interpretar os dados.

Referências

- Levene, H. (1960). "Robust tests for equality of variances". In Olkin, I., Ghurye, S.G., Hoeffding, W., Madow, W.G., Mann, H.B. (Eds.), Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, Stanford University Press, pp. 278-292
- A statistical measure of complexity. Physics Letters A, v. 209, n. 5-6, p. 321-326, .
- Teste de Shapiro-Wilk. Disponível em: <http://www.uel.br/projetos/experimental/pages/arquivos/Shapiro.html>. Acesso em: 25 Jun. 2023..