

Sumário

	Página
1 Introdução	2
1.1 Análise Descritiva Univariada	2
1.1.1 Frequência Relativa	2
1.1.2 Média	3
1.1.3 Mediana	3
1.1.4 Quartis	3
1.1.5 Variância	4
1.1.6 Desvio Padrão	4
1.1.7 Boxplot	5
1.1.8 Gráfico de Dispersão	6
1.1.9 Tipos de Variáveis	6
1.1.10 Coeficiente de Correlação de Pearson	7
2 Análises	8
2.1 Relação entre Taxa Selic, Inflação Acumulada e Juros ao longo do tempo	8
2.2 Estudo da correlação entre Inflação Acumulada e a Taxa Selic	8
2.3 Variação de salário mínimo por mandato presidencial	9
2.4 Conclusão	10

1 Introdução

O seguinte projeto tem por objetivo apresentar as análises estatísticas requisitadas pelo Diretor Financeiro da Phineas & Ferb Capital, Felipe Bretas, buscando compreender os principais fatores que influenciam a inflação para a oferta de subsídios para a formulação de políticas econômicas mais assertivas, por meio de análises detalhadas que evidenciam tendências, variações e correlações ao longo dos anos de 2002 a 2022.

Para cumprir esse objetivo, foi solicitado estudos sobre o impacto da taxa Selic na inflação e nos juros reais, a relação sobre a taxa Selic e a inflação, a distribuição do salário mínimo ao longo dos mandatos presidenciais, a relação entre a taxa Selic e os juros reais, a comparação de índices de inflação, a distribuição da variação mensal da inflação, o impacto da inflação no salário mínimo, a relação entre a taxa de juros e inflação anual e a distribuição do INCC (Índice Nacional de Custo da Construção). Para a realização dessas análises, foi utilizado gráficos de dispersão, linhas e boxplot, além de coeficientes de correlação. Para a realização do projeto foi utilizado o banco de dados inflacao em formato csv, disponibilizado pelo próprio cliente. Para a produção do relatório foram abordadas apenas algumas variáveis presentes no banco de dados, essas sendo a Taxa Selic, a taxa básica de juros da economia brasileira, o IPCA acumulado, que indica a inflação no país, o salário mínimo, que indica o menor salário que um indivíduo pode receber, o Índice Nacional de Custo de Construção (INCC), que indica a variação dos custos da construção civil no Brasil e o ano de cada informação.

O software utilizado para análise estatística dos dados foi o R versão 4.4.2.. O R é um software de programação gratuito largamente usado na área de estatística e visualização de dados que permite não só o manuseio e análise de bancos de dados, como também a confecção de gráficos. # Referencial Teórico

1.1 Análise Descritiva Univariada

1.1.1 Frequência Relativa

A frequência relativa é utilizada para a comparação entre classes de uma variável categórica com c categorias, ou para comparar uma mesma categoria em diferentes estudos.

A frequência relativa da categoria j é dada por:

$$f_j = \frac{n_j}{n}$$

Com:

$$\bullet j = 1, \dots, c$$

- n_j = número de observações da categoria j
- n = número total de observações

Geralmente, a frequência relativa é utilizada em porcentagem, dada por:

$$100 \times f_j$$

1.1.2 Média

A média é a soma das observações dividida pelo número total delas, dada pela fórmula:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Com:

- $i = 1, 2, \dots, n$
- n = número total de observações

1.1.3 Mediana

Sejam as n observações de um conjunto de dados $X = X_{(1)}, X_{(2)}, \dots, X_{(n)}$ de determinada variável ordenadas de forma crescente. A mediana do conjunto de dados X é o valor que deixa metade das observações abaixo dela e metade dos dados acima.

Com isso, pode-se calcular a mediana da seguinte forma:

$$\text{med}(X) = \begin{cases} X_{\frac{n+1}{2}}, & \text{para } n \text{ ímpar} \\ \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}, & \text{para } n \text{ par} \end{cases}$$

1.1.4 Quartis

Os quartis são separatrizes que dividem o conjunto de dados em quatro partes iguais. O primeiro quartil (ou inferior) delimita os 25% menores valores, o segundo representa a mediana, e o terceiro delimita os 25% maiores valores. Inicialmente deve-se calcular a posição do quartil:

- Posição do primeiro quartil P_1 :

$$P_1 = \frac{n + 1}{4}$$

- Posição da mediana (segundo quartil) P_2 :

$$P_2 = \frac{n + 1}{2}$$

- Posição do terceiro quartil P_3 :

$$P_3 = \frac{3 \times (n + 1)}{4}$$

Com n sendo o tamanho da amostra. Dessa forma, $X_{(P_i)}$ é o valor do i -ésimo quartil, onde $X_{(j)}$ representa a j -ésima observação dos dados ordenados.

Se o cálculo da posição resultar em uma fração, deve-se fazer a média entre o valor que está na posição do inteiro anterior e do seguinte ao da posição.

1.1.5 Variância

A variância é uma medida que avalia o quanto os dados estão dispersos em relação à média, em uma escala ao quadrado da escala dos dados.

1.1.5.1 Variância Populacional

Para uma população, a variância é dada por:

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

Com:

- X_i = i -ésima observação da população
- μ = média populacional
- N = tamanho da população

1.1.6 Desvio Padrão

O desvio padrão é a raiz quadrada da variância. Ele avalia o quanto os dados estão dispersos em relação à média.

1.1.6.1 Desvio Padrão Populacional

Para uma população, o desvio padrão é dado por:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$$

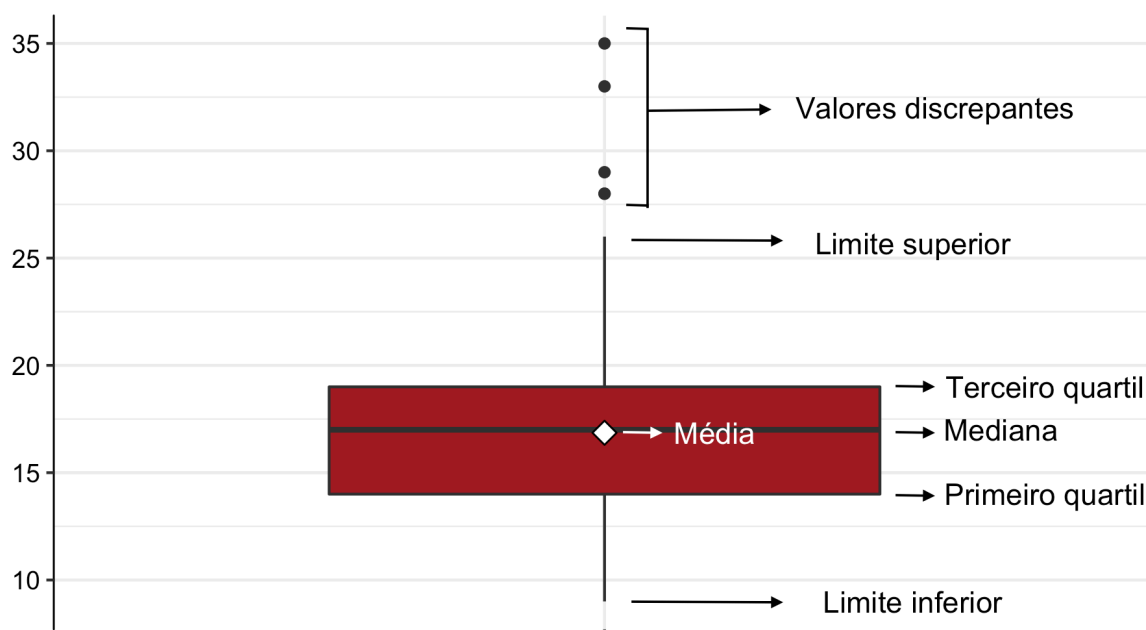
Com:

- X_i = i-ésima observação da população
- μ = média populacional
- N = tamanho da população

1.1.7 Boxplot

O boxplot é uma representação gráfica na qual se pode perceber de forma mais clara como os dados estão distribuídos. A figura abaixo ilustra um exemplo de boxplot.

Figura 1: Exemplo de boxplot

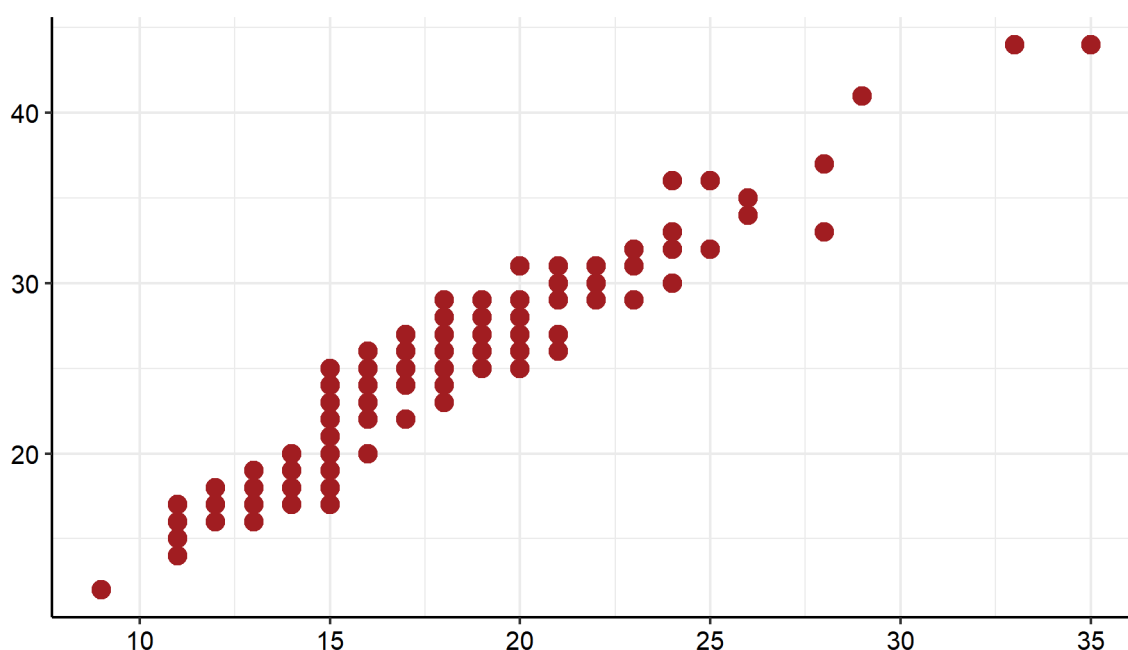


A porção inferior do retângulo diz respeito ao primeiro quartil, enquanto a superior indica o terceiro quartil. Já o traço no interior do retângulo representa a mediana do conjunto de dados, ou seja, o valor em que o conjunto de dados é dividido em dois subconjuntos de mesmo tamanho. A média é representada pelo losango branco e os pontos são *outliers*. Os *outliers* são valores discrepantes da série de dados, ou seja, valores que não demonstram a realidade de um conjunto de dados.

1.1.8 Gráfico de Dispersão

O gráfico de dispersão é uma representação gráfica utilizada para ilustrar o comportamento conjunto de duas variáveis quantitativas. A figura abaixo ilustra um exemplo de gráfico de dispersão, onde cada ponto representa uma observação do banco de dados.

Figura 2: Exemplo de Gráfico de Dispersão



1.1.9 Tipos de Variáveis

1.1.9.1 Qualitativas

As variáveis qualitativas são as variáveis não numéricas, que representam categorias ou características da população. Estas subdividem-se em:

- **Nominais:** quando não existe uma ordem entre as categorias da variável (exemplos: sexo, cor dos olhos, fumante ou não, etc)
- **Ordinais:** quando existe uma ordem entre as categorias da variável (exemplos: nível de escolaridade, mês, estágio de doença, etc)

1.1.9.2 Quantitativas

As variáveis quantitativas são as variáveis numéricas, que representam características numéricas da população, ou seja, quantidades. Estas subdividem-se em:

- **Discretas:** quando os possíveis valores são enumeráveis (exemplos: número de filhos, número de cigarros fumados, etc)

- **Contínuas:** quando os possíveis valores são resultado de medições (exemplos: massa, altura, tempo, etc)

1.1.10 Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson é uma medida que verifica o grau de relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. O valor zero significa que não há relação linear entre as variáveis. Quando o valor do coeficiente r é negativo, diz-se existir uma relação de grandeza inversamente proporcional entre as variáveis. Analogamente, quando r é positivo, diz-se que as duas variáveis são diretamente proporcionais.

O coeficiente de correlação de Pearson é normalmente representado pela letra r e a sua fórmula de cálculo é:

$$r_{Pearson} = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \times \sqrt{\sum_{i=1}^n y_i^2 - n\bar{y}^2}}$$

Onde:

- x_i = i-ésimo valor da variável X
- y_i = i-ésimo valor da variável Y
- \bar{x} = média dos valores da variável X
- \bar{y} = média dos valores da variável Y

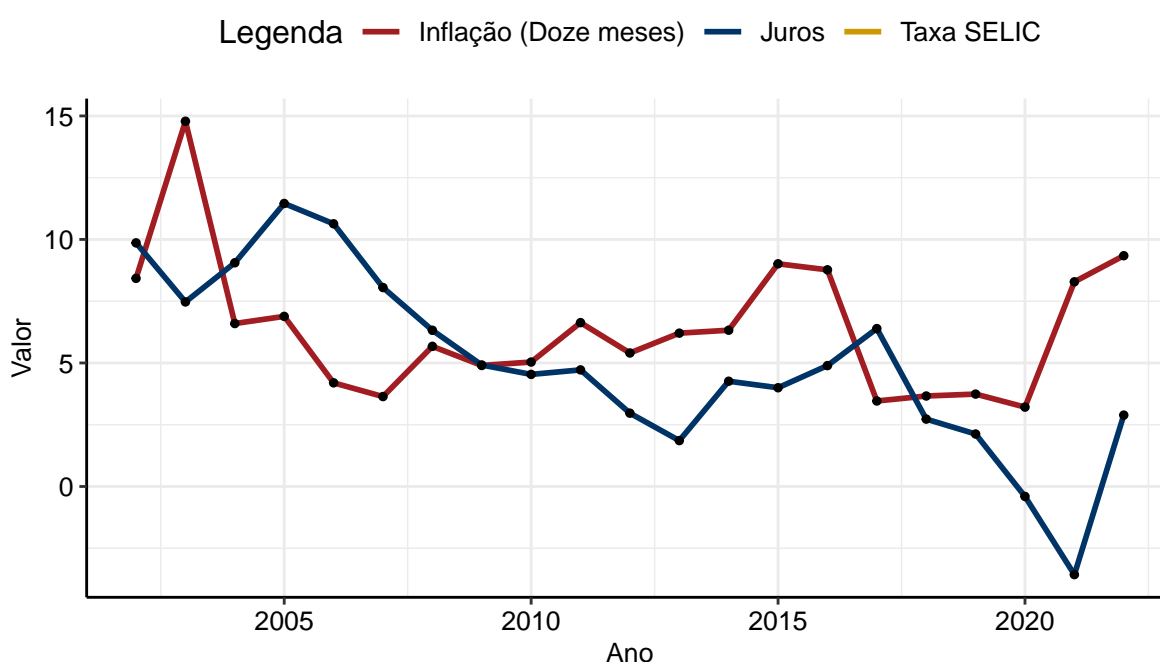
Vale ressaltar que o coeficiente de Pearson é paramétrico e, portanto, sensível quanto à normalidade (simetria) dos dados.

2 Análises

2.1 Relação entre Taxa Selic, Inflação Acumulada e Juros ao longo do tempo

Na primeira análise, foi examinada a relação entre a variável quantitativa contínua Taxa Selic, que representa a taxa básica de juros, e as variáveis quantitativas contínuas Inflação Acumulada e Juros Reais, no período de 2002 a 2022. Para representar esses dados, foi feita uma média dessas variáveis em cada ano para a utilização de um gráfico de linhas multivariado, gráfico esse que é utilizado para análises temporais.

Figura 3: Gráfico de linha do IPCA acumulado em doze meses, juros reais e Taxa Selic ao longo do tempo



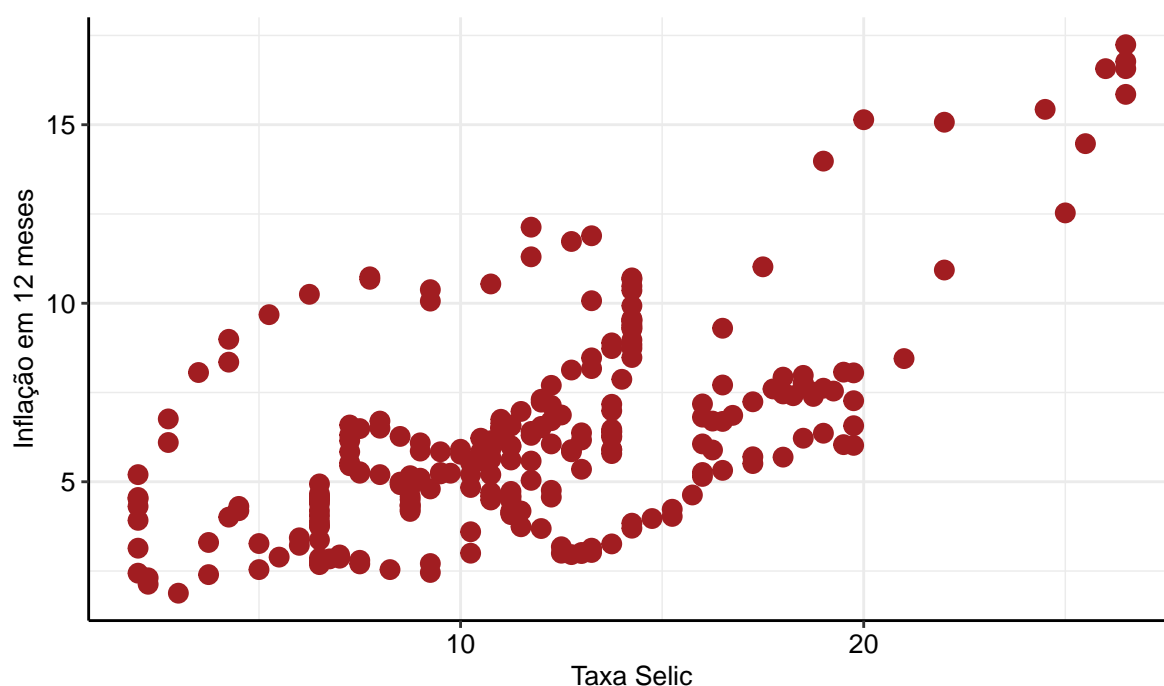
Ao observar a **Figura 3**, é possível notar que, no geral, quando a Taxa Selic sofria uma alteração, a Inflação Acumulada e os Juros sofriam uma alteração similar. Pode-se perceber também que nos anos de 2005, 2006 e 2021, as variáveis apresentaram comportamentos similares, embora o comportamento nos demais anos tenha demonstrado uma relação entre a inflação e a taxa Selic.

2.2 Estudo da correlação entre Inflação Acumulada e a Taxa Selic

O presente estudo tem como objeto de análise a relação entre as variáveis quantitativas contínuas: inflação acumulada, representada pelo IPCA acumulado em doze meses, e a taxa Selic. O objetivo desta investigação é compreender o impacto da in-

flação sobre a taxa Selic, que constitui a taxa básica de juros da economia. Para isso, foram utilizados o gráfico de dispersão bivariado e o Coeficiente de Correlação de Pearson, proporcionando uma melhor compreensão das interações entre as variáveis analisadas.

Figura 4: Gráfico de dispersão de Inflação Acumulada pela Taxa Selic



Pode ser observado na **Figura 4** que existe uma correlação entre as variáveis, o que pode ser comprovado pelo Coeficiente de Correlação de Pearson, utilizado para indentificar a relação linear de variáveis quantitativas, que assume o valor aproximado de 0.635, significando uma relação média a forte entre as váriaveis Inflação Acumulada e a Taxa Selic. Como a taxa Selic representa a taxa básica de juros, pode ser esperado que a inflação seja afetada de forma direta, resultando no seu aumento.

2.3 Variação de salário mínimo por mandato presidencial

Nesta análise, será observada a evolução da variável quantitativa contínua 'salário mínimo' durante o mandato de cada presidente. A variável qualitativa nominal corresponde aos diferentes presidentes no período de 2002 a 2022, incluindo: o final do mandato de Fernando Henrique Cardoso em 2002, os dois mandatos de Luiz Inácio Lula da Silva, abreviado como 'Lula', seguido pelo número do mandato, Dilma Rousseff, Michel Temer e Jair Bolsonaro. As eleições ocorrem a cada quatro anos, sendo o ano eleitoral considerado como o último ano de cada mandato. Isso fez com que as informações referentes a 2002 fossem menos precisas. Para realizar essa análise, foi utilizado um gráfico de boxplot.

Figura 5: Gráfico boxplot do Salário mínimo por mandato

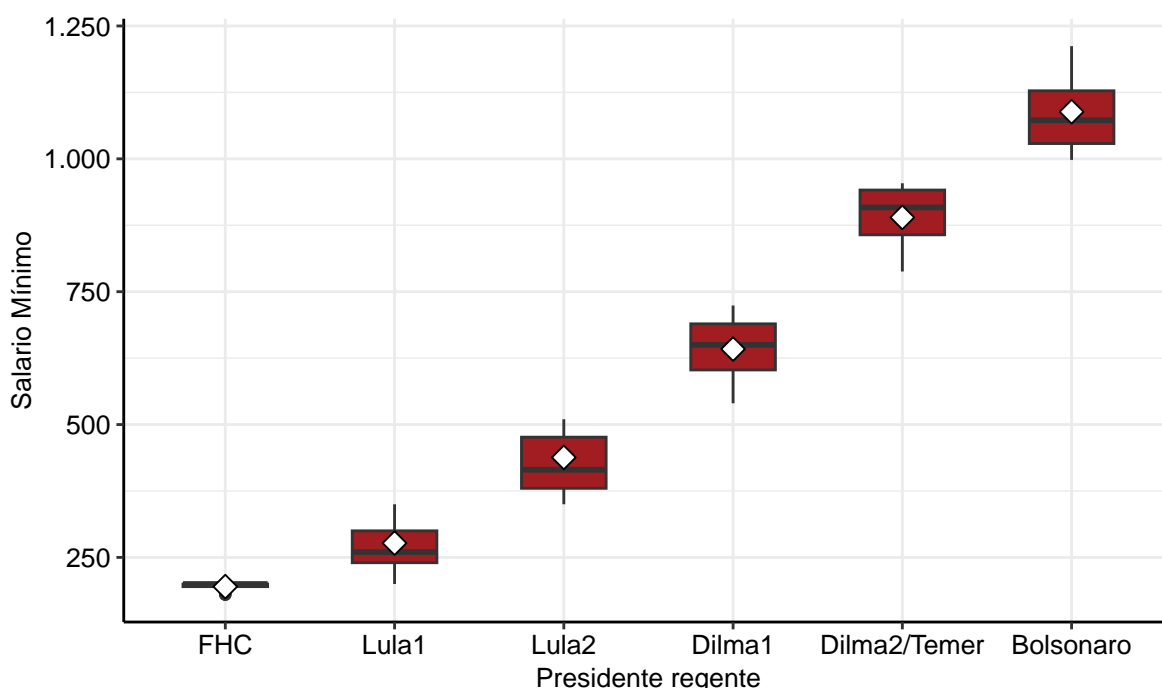


Tabela 1: Quadro de medidas resumo do salário mínimo por mandato presidencial

Medidas Resumo	FHC	Lula1	Lula2	Dilma1	Dilma2/Temer	Bolsonaro
Média	195	276.88	438.12	642.041	889.75	1088.625
Desvio Padrão	9.045	44.44	53.85	67.782	65.510	80.709
Variância	81,902	1974.91	2899,82	4594.399	4291.560	6513.942
1º Quartil	195	240	380	602.75	857	1028.75
Mínimo	180	200	350	540	788	998
Mediana	200	260	415	650	908.5	1072.5
Máximo	200	350	510	724	954	1212
3º Quartil	200	300	476.25	689.5	941.25	1128

Como pode ser percebido pela **Figura 5**, assim como o **Tabela 1** o salário mínimo aumentou com o tempo. Pode ser notado isso olhando como as medianas na **Figura 5**, representadas pelo ponto branco, sobem a cada mandato, o que pode ser visto também na **Tabela 1**. Pode ser notado também o aumento do desvio padrão, medida essa utilizada para entender a dispersão dos valores.

2.4 Conclusão

Após a realização de todas as análises, o objetivo da pesquisa, a compreensão dos fatores que influenciam a inflação, pôde ser bem explorado pelas análises.

A análise da relação entre TaxaSelic, Inflação Acumulada e Jurosa o longo do tempo revela que, de maneira geral, mudanças na Taxa Selic impactaram a Inflação Acumu-

lada e os Juros de forma semelhante. Além disso, nos anos de 2005, 2006 e 2021, essas variáveis apresentaram padrões de comportamento alinhados, enquanto nos demais anos, a relação mais evidente ocorreu entre a inflação e a Taxa Selic.

O estudo da correlação entre Inflação Acumulada e a Taxa Selic indica uma correlação significativa entre a Inflação Acumulada e a Taxa Selic, evidenciada pelo Coeficiente de Correlação de Pearson, que tem um valor aproximado de 0,635. Esse coeficiente aponta para uma relação de média a forte entre as variáveis, sugerindo que alterações na taxa Selic influenciam diretamente a inflação, levando ao seu aumento.

A análise da variação de salário mínimo por mandato presidencial indica que o salário mínimo aumentou ao longo do tempo, com as medianas subindo a cada mandato, representando esse crescimento, assim como as médias.

A análise da relação entre SELIC e juros reais foi visto que embora a Taxa SELIC influencie bastante os juros reais, ela não o afeta completamente.

A comparação de índices de inflação apresentou inflações extremamente similares, algo esperado entre as duas variáveis.

A análise de distribuição da variação mensal da inflação revelou padrões sazonais e anuais, com os boxplots destacando tendências mensais através de medidas estatísticas e outliers. Além disso, o gráfico de linhas mostrou a evolução da média anual do índice, evidenciando suas oscilações ao longo dos anos.

Na análise do impacto da inflação no salário mínimo foi visto pouca relação entre as duas variáveis, já que uma apresenta alta variação enquanto a outra apresenta variação mínima.

A análise da relação entre taxa de juros e inflação anual pode-se perceber pouca relação entre as variáveis, o que demonstra que os juros não indica completamente indícios de inflação.

Por último, na distribuição do INCC por ano foi possível notar uma alta variação em grande parte dos anos, exceto durante o governo de Michel Temer, onde houve uma diminuição da variação.