

Atividade SME0620 - Estatística I (2024)

Leonardo Gueno Risetto NUSP: 13676482

Thiago Kashivagui Gonçalves NUSP: 13676579

Tema da pesquisa

O tema desta pesquisa é "Análise de dados referentes a estudantes do Ensino Superior do Instituto Politécnico de Portalegre".

Problema a ser estudado

A pesquisa visa identificar os padrões e as características dos alunos, bem como identificar a relação entre esses padrões o sucesso acadêmico dos estudantes ao longo do curso. Isso permitirá a implementação de estratégias de apoio direcionadas para os alunos e promover o sucesso acadêmico.

Importancia do tema e objetivos da pesquisa

Esse tipo de pesquisa é crucial para informar políticas educacionais, identificar disparidades entre alunos e otimizar recursos e suporte estudantil. Ao analisar dados sobre o perfil dos alunos, desempenho acadêmico e situação financeira, as instituições podem desenvolver intervenções direcionadas para melhorar a retenção e o sucesso dos alunos, promovendo assim a equidade no acesso e no desempenho acadêmico. Além disso, essas pesquisas contribuem para o avanço do conhecimento científico sobre educação e desenvolvimento humano.

Os objetivos desta pesquisa são:

Identificar Padrões e Tendências: Analisar dados para identificar padrões e tendências relacionados ao desempenho acadêmico, perfil dos alunos e outras variáveis relevantes.

Entender Fatores de Sucesso e Desafios: Investigar os fatores que influenciam o sucesso dos alunos, incluindo fatores socioeconômicos, demográficos e acadêmicos, bem como os desafios que os alunos enfrentam durante seus estudos.

Avaliar Políticas e Programas: Avaliar a eficácia de políticas e programas educacionais, como programas de suporte estudantil, políticas de admissão e iniciativas para promover a equidade educacional.

Contribuir para a Tomada de Decisões: Fornecer informações que possam orientar a tomada de decisões por parte das instituições educacionais e formuladores de políticas, ajudando a desenvolver estratégias para melhorar a qualidade e o sucesso do ensino superior.

Promover a Equidade e a Inclusão: Identificar disparidades e inequidades no acesso e desempenho educacional, visando promover a equidade e a inclusão no sistema educacional.

Gerar Conhecimento Científico: Contribuir para o avanço do conhecimento científico no campo da educação, fornecendo novas evidências e perspectivas sobre questões educacionais relevantes.

Metodologia

Coleta de dados

A coleta de dados será realizada utilizando registros acadêmicos e demográficos dos alunos fornecidos pelo Instituto Politécnico de Portalegre, conforme descrito nas especificações do conjunto de dados. Esses registros incluem informações sobre a trajetória acadêmica dos alunos, dados demográficos, socioeconômicos e familiares, bem como desempenho acadêmico ao longo do curso.

Além dos dados fornecidos pela instituição, também serão utilizadas informações macroeconômicas relevantes, como taxa de desemprego, para contextualizar o ambiente socioeconômico em que os alunos estão inseridos.

Ferramentas Utilizadas

Linguagem de programação: Python será linguagem de programação utilizada no trabalho devido à sua ampla gama de bibliotecas para análise de dados, como Pandas, Numpy e Tabulate.

Local onde foram retirados os dados

Os dados foram retirados do Instituto Politécnico de Portalegre, uma instituição de ensino superior localizada em Portalegre, Portugal. O instituto oferece uma variedade de cursos de graduação em áreas como agronomia, design, educação, enfermagem, jornalismo, gestão, serviços sociais e tecnologia, representados nos dados coletados.

Local de onde foi tirada a database ->

<https://www.kaggle.com/datasets/mikhail1681/student-performance-pip?resource=download>

Descrição dos dados utilizados

Curso: Esta variável indica o curso de graduação em que o aluno está matriculado. Os dados fornecem informações sobre os diferentes cursos oferecidos pelo Instituto Politécnico de Portalegre

33 - Biofuel Production Technologies 171 - Animation and Multimedia Design 8014 - Social Service (evening attendance) 9003 - Agronomy 9070 - Communication Design 9085 - Veterinary Nursing 9119 - Informatics Engineering 9130 - Equiculture 9147 - Management 9238 - Social Service 9254 - Tourism 9500 - Nursing 9556 - Oral Hygiene 9670 - Advertising and Marketing Management 9773 - Journalism and Communication 9853 - Basic Education 9991 - Management (evening attendance).

Mensalidades em Dia: Esta variável indica se o aluno está com as mensalidades em dia ou não.

1 – sim 0 – não.

Gênero: Esta variável indica o gênero do aluno.

1 – masculino 0 – feminino.

Idade na Matrícula: Esta variável indica a idade do aluno no momento da matrícula no curso de graduação. Representa a idade do aluno quando ele ingressou na instituição para iniciar seus estudos.

Nota de Admissão: Esta variável se refere à pontuação acadêmica mínima ou média necessária para aceitação em um programa educacional ou instituição específica.

Taxas de Desemprego: Esta variável representa as taxas de desemprego em uma determinada região ou período de tempo. Reflete a porcentagem de pessoas desempregadas em relação à força de trabalho total. As taxas de desemprego podem influenciar o contexto socioeconômico em que os alunos estão inseridos e podem ter impacto em sua capacidade de pagar as mensalidades e se dedicar aos estudos.

Qualificação anterior (nota): Essa informação é importante para entender o histórico educacional dos alunos e como isso pode influenciar seu desempenho acadêmico no ensino superior.

Estado civil: Esta variável se refere ao estado civil atual.

Classificação das variáveis

Curso: Qualitativa / Nominal

Mensalidades em Dia: Qualitativa / Nominal

Gênero: Qualitativa / Nominal

Idade na Matrícula: Quantitativa / Discreta

Nota de admissão: Quantitativa / Contínua

Taxas de Desemprego: Quantitativa / Contínua

Qualificação anterior (nota): Quantitativa / Contínua

Estado civil: Qualitativa / Nominal

Resultados

Medidas de posição

```
In [ ]: import pandas as pd
import numpy as np
from tabulate import tabulate
import matplotlib.pyplot as plt

In [ ]: # Carrega o arquivo CSV
dados = pd.read_csv('dados.csv')

# Escolhe a variável para calcular as medidas
variaveis = ['Curso', 'Mensalidade em dia', 'Gênero', 'Idade na matrícula', '

# Calcular as medidas para cada variável
for variavel in variaveis:
    # Média
    media = dados[variavel].mean()
    print(f'A média da variável {variavel} é: {media}')

    # Mediana
    mediana = dados[variavel].median()
    print(f'A mediana da variável {variavel} é: {mediana}')

    # Moda
    moda = dados[variavel].mode()[0]
    print(f'A moda da variável {variavel} é: {moda}')

    # Quartis
    quartis = np.percentile(dados[variavel], [25, 50, 75])
    print(f'Os quartis da variável {variavel} são:')
    print(f'Q1: {quartis[0]}, Q2: {quartis[1]}, Q3: {quartis[2]}')

    print("\n")
```

A média da variável Curso é: 8856.642631103074
A mediana da variável Curso é: 9238.0
A moda da variável Curso é: 9500
Os quartis da variável Curso são:
Q1: 9085.0, Q2: 9238.0, Q3: 9556.0

A média da variável Mensalidade em dia é: 0.8806509945750453
A mediana da variável Mensalidade em dia é: 1.0
A moda da variável Mensalidade em dia é: 1
Os quartis da variável Mensalidade em dia são:
Q1: 1.0, Q2: 1.0, Q3: 1.0

A média da variável Gênero é: 0.35171790235081374
A mediana da variável Gênero é: 0.0
A moda da variável Gênero é: 0
Os quartis da variável Gênero são:
Q1: 0.0, Q2: 0.0, Q3: 1.0

A média da variável Idade na matrícula é: 23.265144665461122
A mediana da variável Idade na matrícula é: 20.0
A moda da variável Idade na matrícula é: 18
Os quartis da variável Idade na matrícula são:
Q1: 19.0, Q2: 20.0, Q3: 25.0

A média da variável Nota de admissão é: 126.97811934900544
A mediana da variável Nota de admissão é: 126.1
A moda da variável Nota de admissão é: 130.0
Os quartis da variável Nota de admissão são:
Q1: 117.9, Q2: 126.1, Q3: 134.8

A média da variável Taxa de desemprego é: 11.56613924050633
A mediana da variável Taxa de desemprego é: 11.1
A moda da variável Taxa de desemprego é: 7.6
Os quartis da variável Taxa de desemprego são:
Q1: 9.4, Q2: 11.1, Q3: 13.9

A média da variável Qualificação anterior (nota) é: 132.6133137432188
A mediana da variável Qualificação anterior (nota) é: 133.1
A moda da variável Qualificação anterior (nota) é: 133.1
Os quartis da variável Qualificação anterior (nota) são:
Q1: 125.0, Q2: 133.1, Q3: 140.0

A média da variável Estado civil é: 1.1785714285714286
A mediana da variável Estado civil é: 1.0
A moda da variável Estado civil é: 1
Os quartis da variável Estado civil são:
Q1: 1.0, Q2: 1.0, Q3: 1.0

Medidas de dispersão

```
In [ ]: # Calcular as medidas para cada variável
for variavel in variaveis:
    # Desvio padrão
    desvio_padrao = dados[variavel].std()
    print(f'O desvio padrão da variável {variavel} é: {desvio_padrao}')

    # Amplitude
    amplitude = dados[variavel].max() - dados[variavel].min()
    print(f'A amplitude da variável {variavel} é: {amplitude}')

    # Variância
    variancia = dados[variavel].var()
    print(f'A variância da variável {variavel} é: {variancia}')

    # Intervalo interquartil (IQR)
    quartis = np.percentile(dados[variavel], [25, 50, 75])
    IQR = quartis[2] - quartis[0]
    print(f'O intervalo interquartil (IQR) da variável {variavel} é: {IQR}')

    # Coeficiente de variação
    media = dados[variavel].mean()
    coef_variacao = (desvio_padrao / media) * 100
    print(f'O coeficiente de variação para a variável {variavel} é: {coef_variacao}')

    print("\n")
```

O desvio padrão da variável Curso é: 2063.566416197521
A amplitude da variável Curso é: 9958
A variância da variável Curso é: 4258306.35405828
O intervalo interquartil (IQR) da variável Curso é: 471.0
O coeficiente de variação para a variável Curso é: 4258306.35405828

O desvio padrão da variável Mensalidade em dia é: 0.3242353829723811
A amplitude da variável Mensalidade em dia é: 1
A variância da variável Mensalidade em dia é: 0.10512858357124662
O intervalo interquartil (IQR) da variável Mensalidade em dia é: 0.0
O coeficiente de variação para a variável Mensalidade em dia é: 0.10512858357124662

O desvio padrão da variável Gênero é: 0.47756043706245527
A amplitude da variável Gênero é: 1
A variância da variável Gênero é: 0.2280639710472833
O intervalo interquartil (IQR) da variável Gênero é: 1.0
O coeficiente de variação para a variável Gênero é: 0.2280639710472833

O desvio padrão da variável Idade na matrícula é: 7.587815615029815
A amplitude da variável Idade na matrícula é: 53
A variância da variável Idade na matrícula é: 57.574945807690284
O intervalo interquartil (IQR) da variável Idade na matrícula é: 6.0
O coeficiente de variação para a variável Idade na matrícula é: 57.574945807690284

O desvio padrão da variável Nota de admissão é: 14.482000818849468
A amplitude da variável Nota de admissão é: 95.0
A variância da variável Nota de admissão é: 209.72834771715662
O intervalo interquartil (IQR) da variável Nota de admissão é: 16.900000000000006
O coeficiente de variação para a variável Nota de admissão é: 209.72834771715662

O desvio padrão da variável Taxa de desemprego é: 2.6638504843026354
A amplitude da variável Taxa de desemprego é: 8.6
A variância da variável Taxa de desemprego é: 7.0960994027193856
O intervalo interquartil (IQR) da variável Taxa de desemprego é: 4.5
O coeficiente de variação para a variável Taxa de desemprego é: 7.0960994027193856

O desvio padrão da variável Qualificação anterior (nota) é: 13.188331685876213
A amplitude da variável Qualificação anterior (nota) é: 95.0
A variância da variável Qualificação anterior (nota) é: 173.9320926566865
O intervalo interquartil (IQR) da variável Qualificação anterior (nota) é: 15.0
O coeficiente de variação para a variável Qualificação anterior (nota) é: 173.9320926566865

O desvio padrão da variável Estado civil é: 0.60574694613071
A amplitude da variável Estado civil é: 5
A variância da variável Estado civil é: 0.36692936274668125

0 intervalo interquartil (IQR) da variável Estado civil é: 0.0
0 coeficiente de variação para a variável Estado civil é: 0.36692936274668
125

Tabela de frequência

Variável qualitativa

```
In [ ]: # Escolhe a variável para calcular as medidas
variavel_qualitativa = 'Gênero'

# Criar tabela de frequência para variável qualitativa
tabela_qualitativa = dados[variavel_qualitativa].value_counts().reset_index
tabela_qualitativa.columns = ['Valores', 'Frequência']

# Imprimir tabela de frequência para variável qualitativa
print(f"Tabela de Frequência para a Variável '{variavel_qualitativa}':")
print(tabulate(tabela_qualitativa, headers='keys', tablefmt='pretty'))
```

Tabela de Frequência para a Variável 'Gênero':

	Valores	Frequência
0	0	2868
1	1	1556

Variável quantitativa

```
In [ ]: # Escolhe a variável para calcular as medidas
variavel_quantitativa = 'Idade na matrícula'

# Criar tabela de frequência para variável quantitativa
intervalos = pd.cut(dados[variavel_quantitativa], bins=5)
tabela_quantitativa = dados.groupby(intervalos, observed=True).size().reset_index
tabela_quantitativa.columns = ['Intervalos', 'Frequência']

# Imprimir tabela de frequência para variável quantitativa
print(f"Tabela de Frequência para a Variável '{variavel_quantitativa}':")
print(tabulate(tabela_quantitativa, headers='keys', tablefmt='pretty'))
```

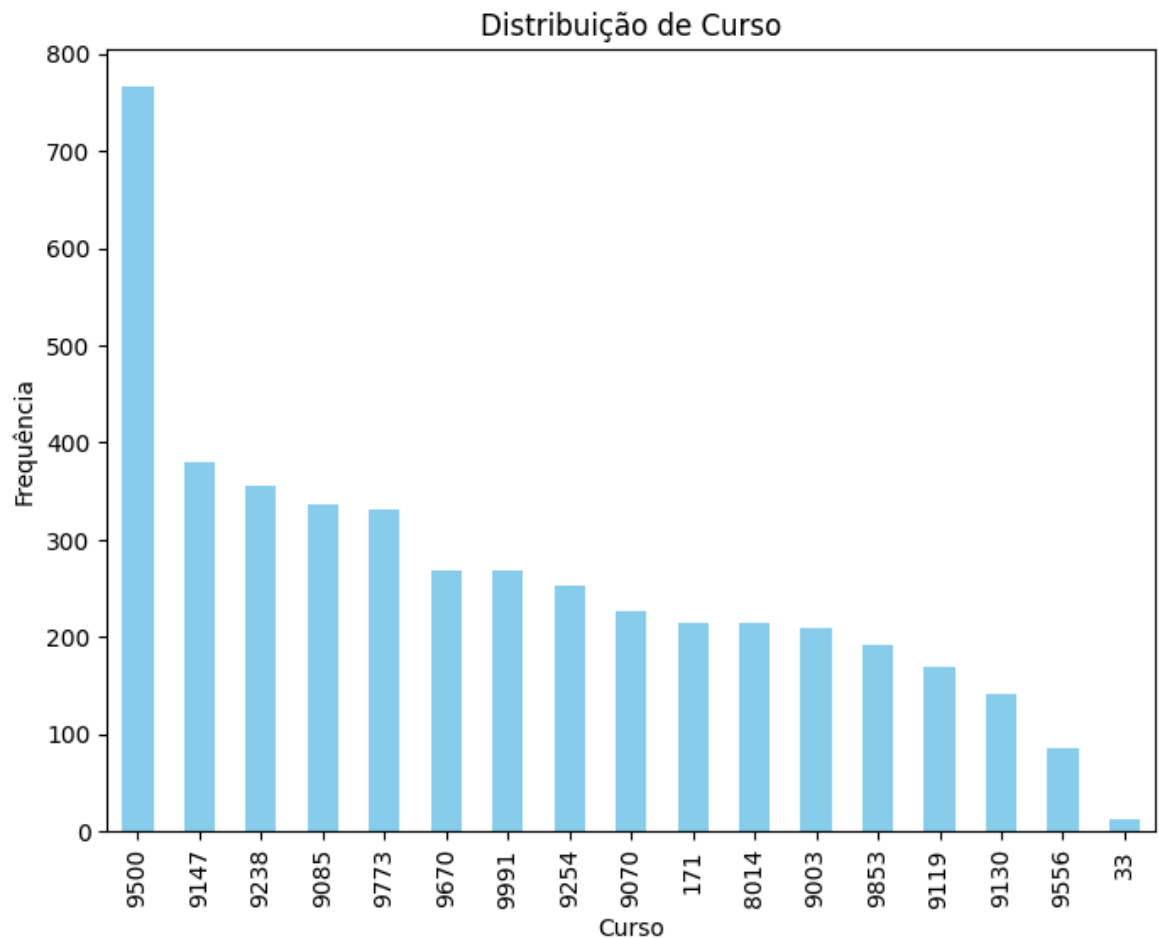
Tabela de Frequência para a Variável 'Idade na matrícula':

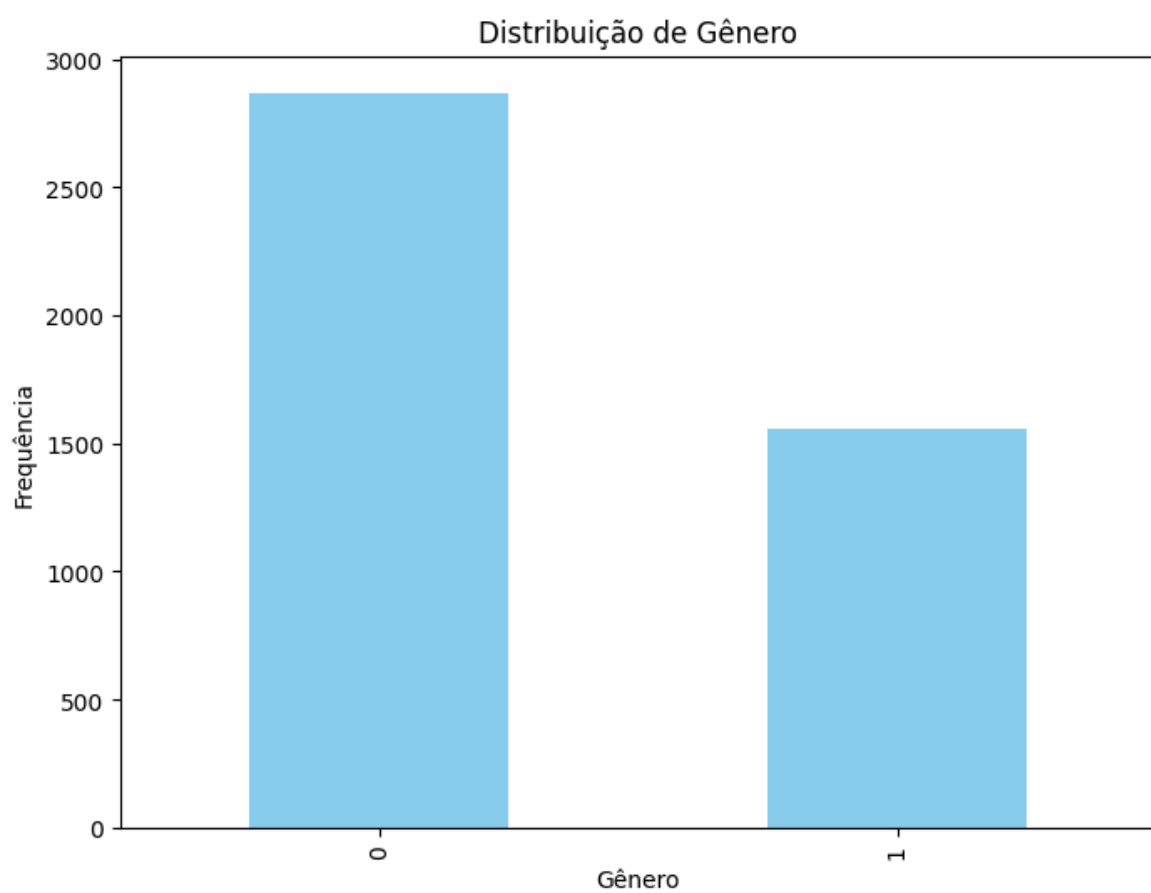
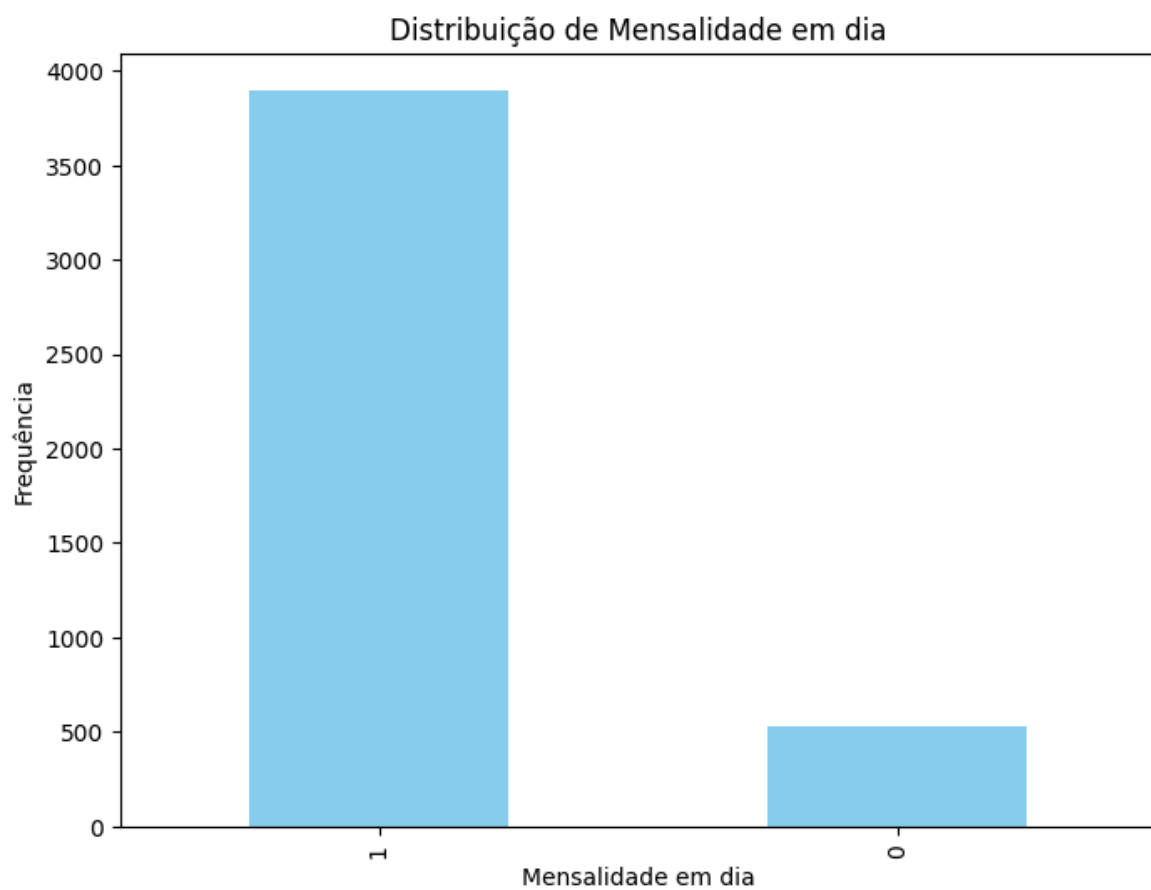
	Intervalos	Frequência
0	(16.947, 27.6]	3564
1	(27.6, 38.2]	574
2	(38.2, 48.8]	214
3	(48.8, 59.4]	67
4	(59.4, 70.0]	5

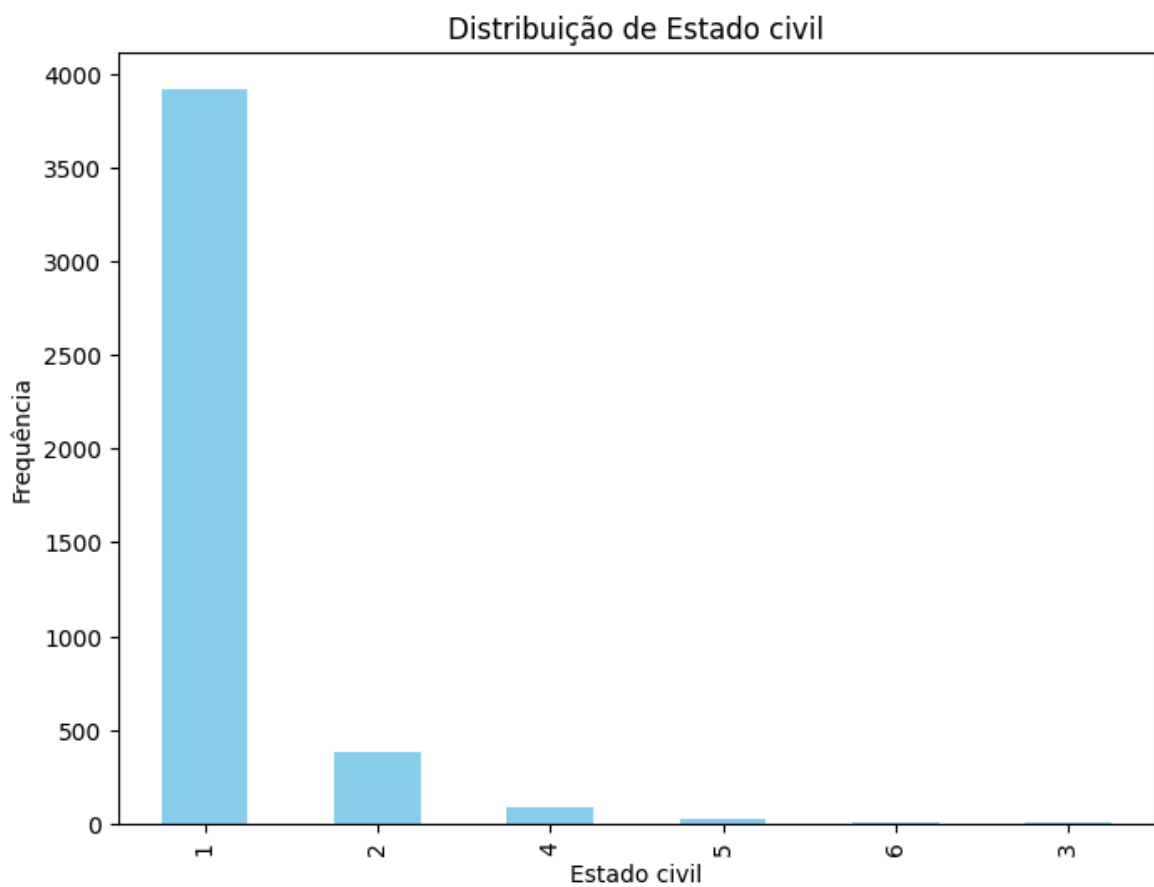
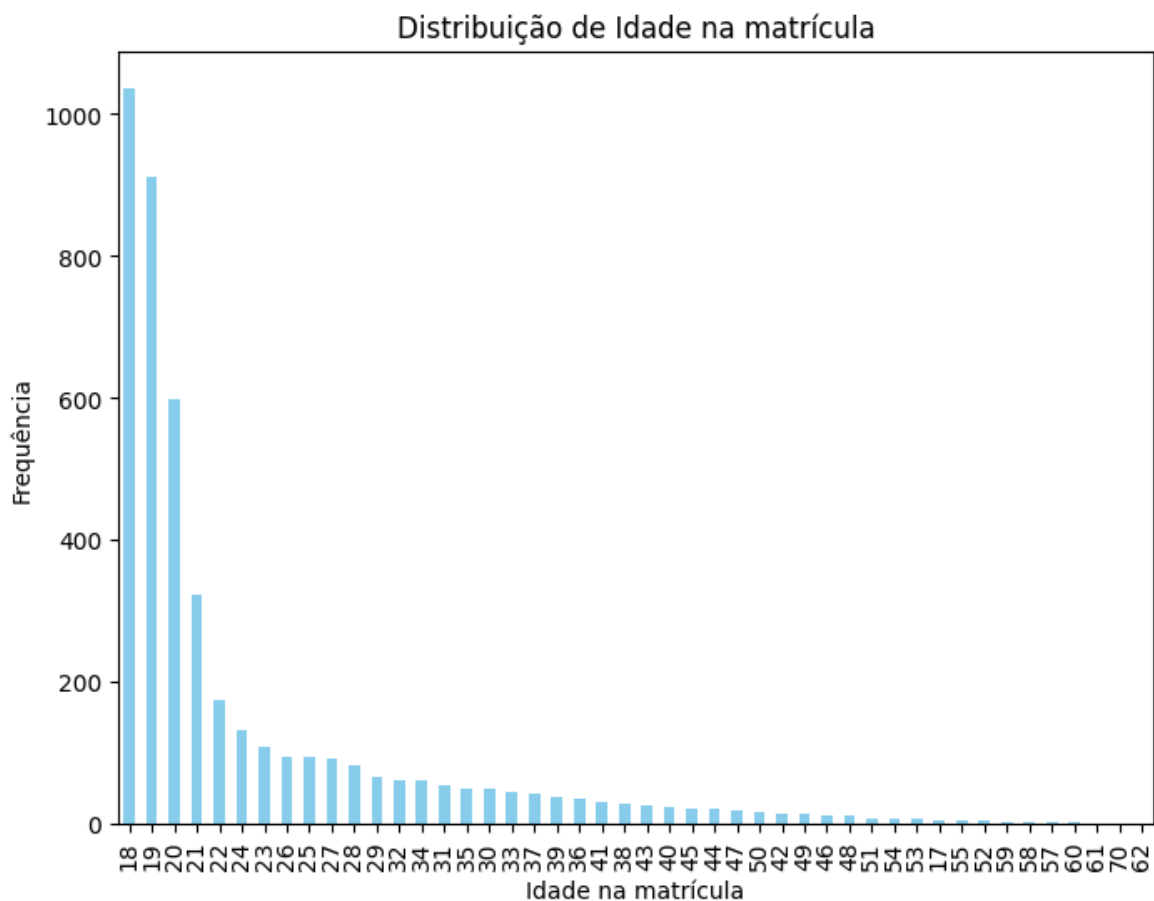
Representação gráfica

Gráfico de barras


```
In [ ]: # Gráfico de barras para variáveis categóricas
for variavel in ['Curso', 'Mensalidade em dia', 'Gênero', 'Idade na matrícula']:
    plt.figure(figsize=(8, 6))
    dados[variavel].value_counts().plot(kind='bar', color='skyblue')
    plt.title(f'Distribuição de {variavel}')
    plt.xlabel(variavel)
    plt.ylabel('Frequência')
    plt.show()
```







Análise dos gráficos de barras:

Distribuição de cursos: Observa-se que os cursos com o maior número de alunos, em ordem decrescente, são os seguintes: 9500, 9147, 9238, 9085, 9773, 9670,

9991, 9254, 9070, 171, 8014, 9003, 9853, 9119, 9130, 9556, 33.

Distribuição de mensalidades em dia: É evidente que a grande maioria dos estudantes está com as mensalidades em dia.

Distribuição de gênero: Conclui-se que a maioria dos estudantes são do sexo feminino.

Distribuição de idade na matrícula: O gráfico demonstra que, à medida que a idade aumenta, a frequência na universidade diminui. Isso sugere que as pessoas têm a tendência de ingressar na universidade o mais cedo possível.

Distribuição de Estado civil: A partir do gráfico é evidente que a grande maioria dos entrevistados são solteiros.

Gráfico circular

```
In [ ]: # Gráfico circular para variáveis categóricas
for variavel in ['Curso', 'Mensalidade em dia', 'Gênero', 'Idade na matrícula']:
    if variavel in {'Idade na matrícula', 'Estado civil'}:
        # Limitar o número de categorias exibidas
        if variavel == 'Idade na matrícula':
            top_categories = 5
        else:
            top_categories = 2
        top_values = dados[variavel].value_counts().nlargest(top_categories)
        other_values_count = dados[variavel].nunique() - top_values.shape[0]

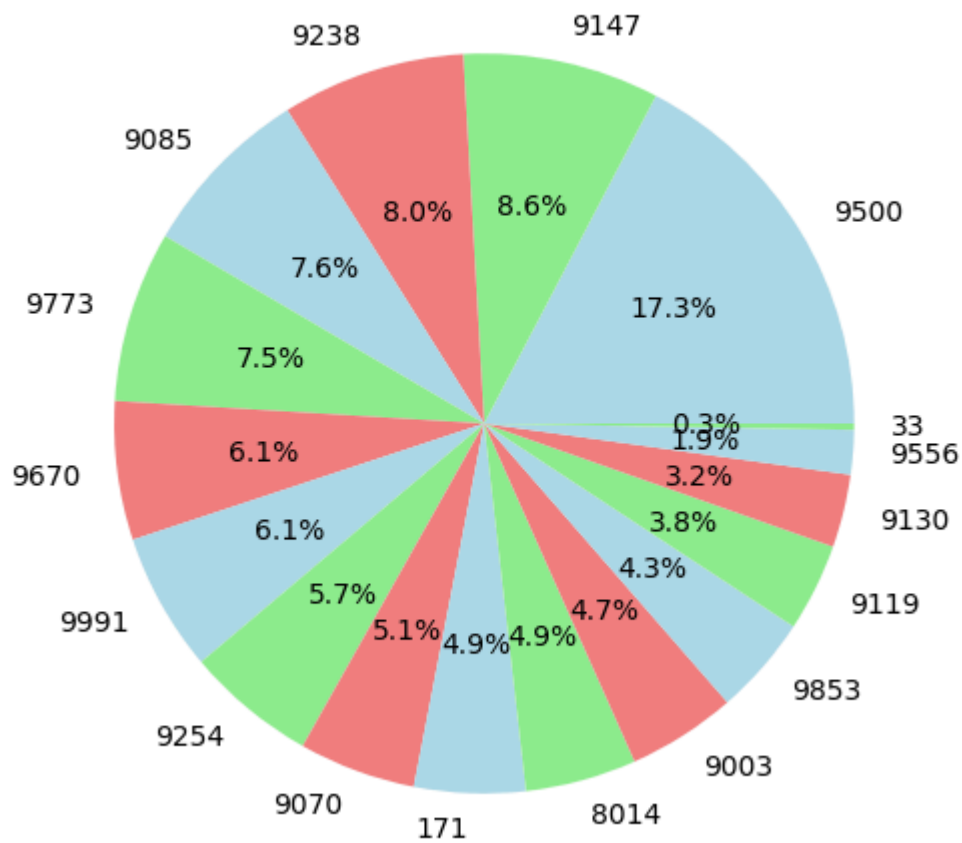
        # Agrupar as outras categorias em uma categoria chamada "Outros"
        other_values = pd.Series([dados[variavel].value_counts().values[t]
                                  for t in range(other_values_count)])

        # Concatenar as categorias mais frequentes com "Outros"
        plot_data = pd.concat([top_values, other_values])

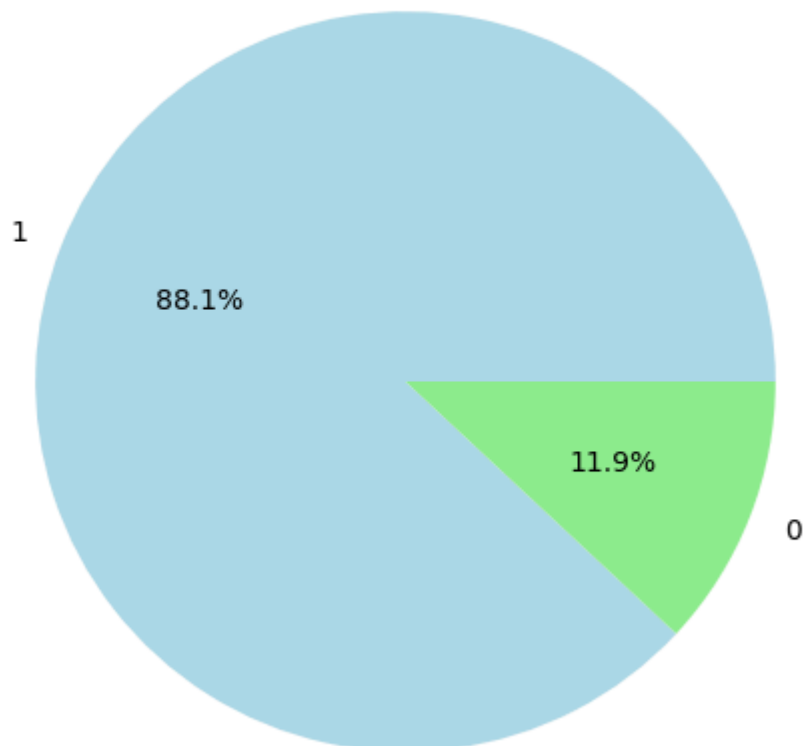
    else:
        plot_data = dados[variavel].value_counts()

plt.figure(figsize=(8, 6))
plot_data.plot(kind='pie', autopct='%1.1f%%', colors=['lightblue', 'lightcoral'])
plt.title(f'Distribuição de {variavel}')
plt.ylabel('')
plt.show()
```

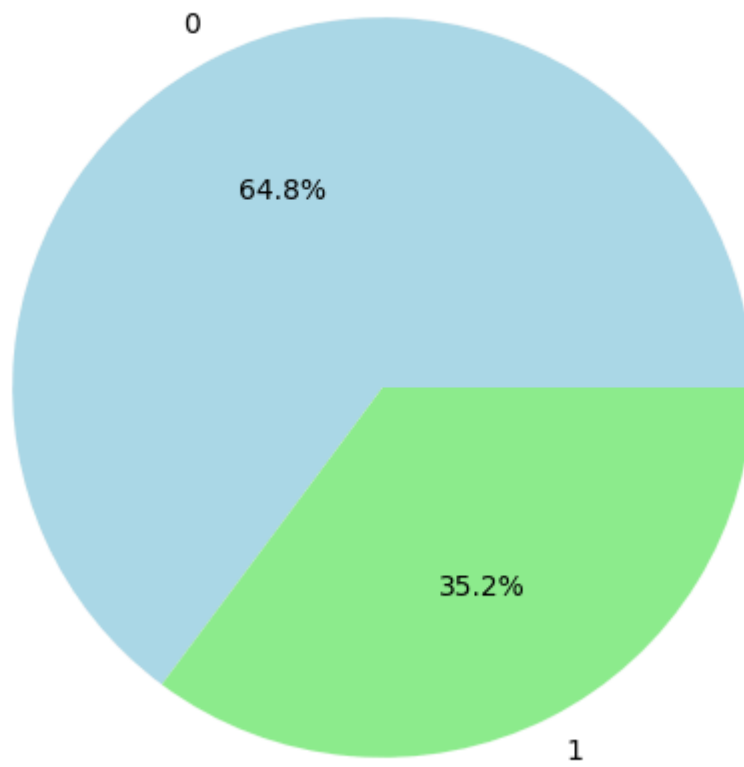
Distribuição de Curso



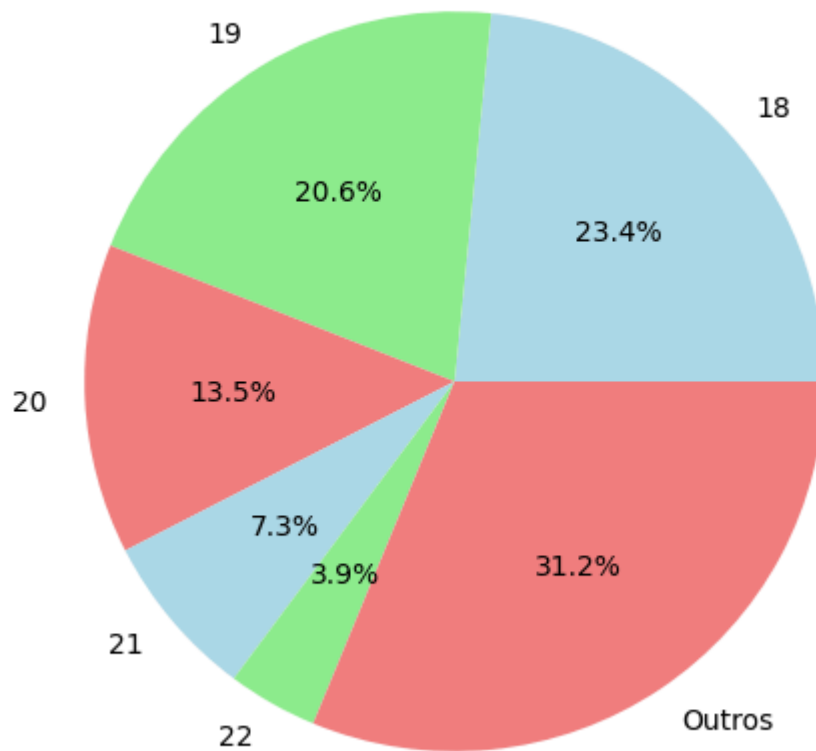
Distribuição de Mensalidade em dia



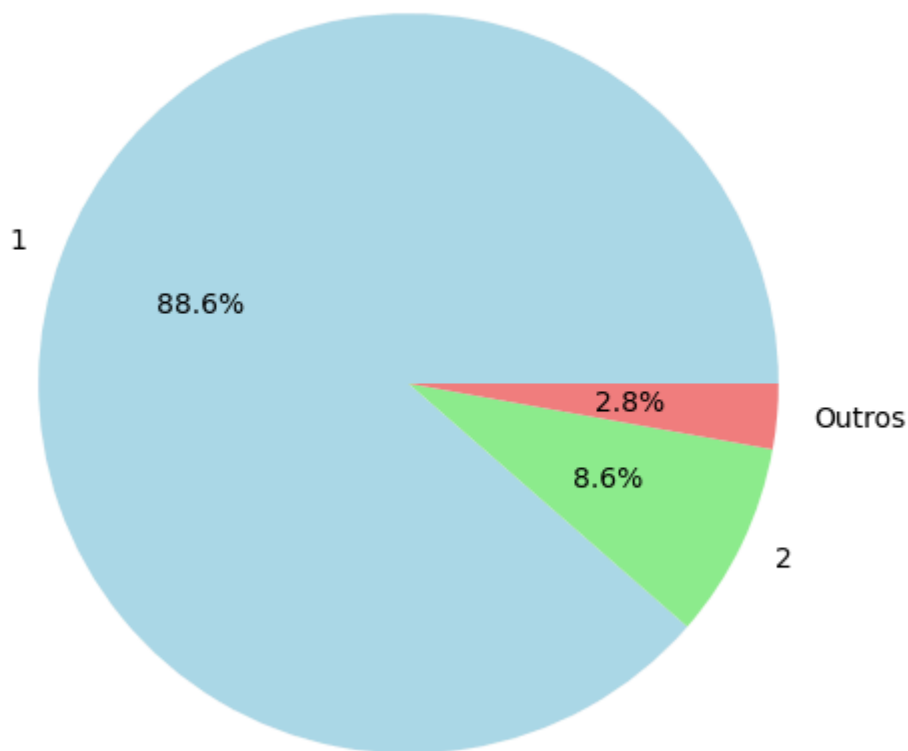
Distribuição de Gênero



Distribuição de Idade na matrícula



Distribuição de Estado civil



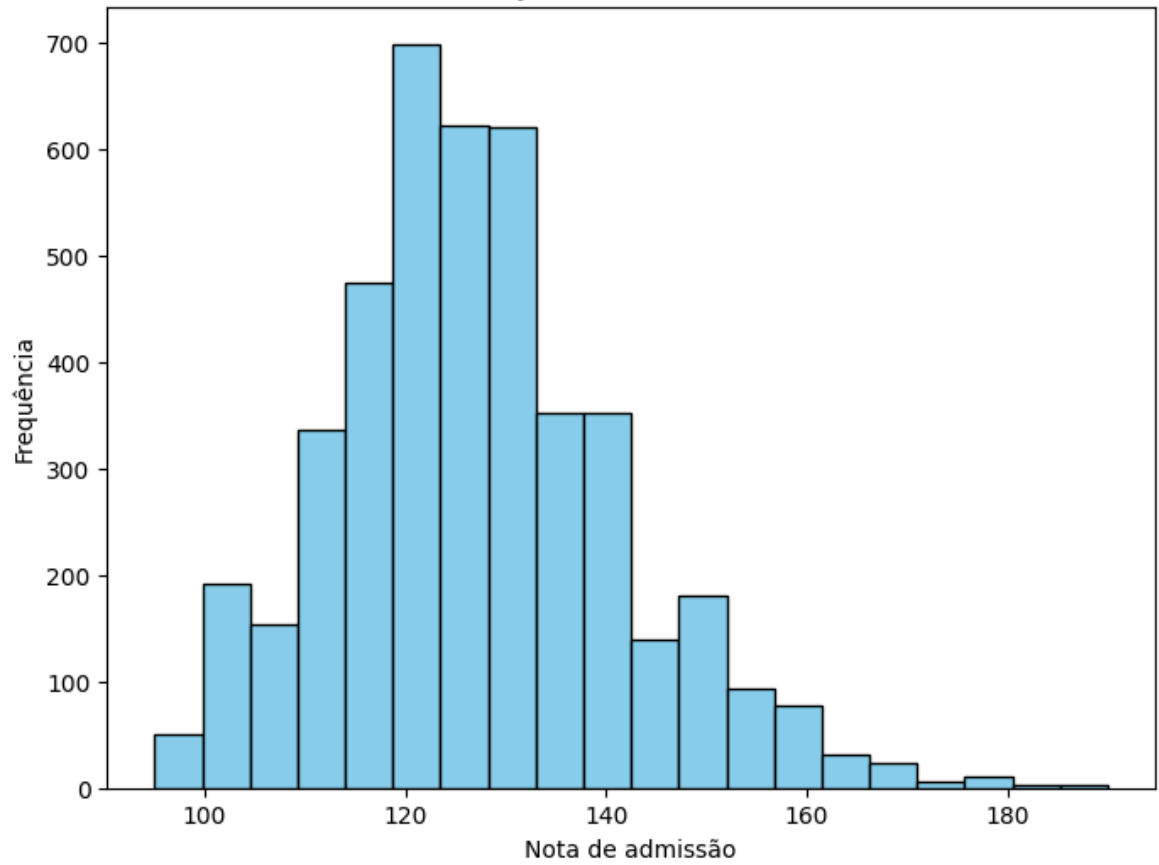
Análise dos gráficos circulares:

Os gráficos circulares a cima demonstram os mesmos resultados do gráfico de barras, mas com um diferente tipo de visualização, agora é possível analisar a porcentagem em relação ao todo das distribuições referentes.

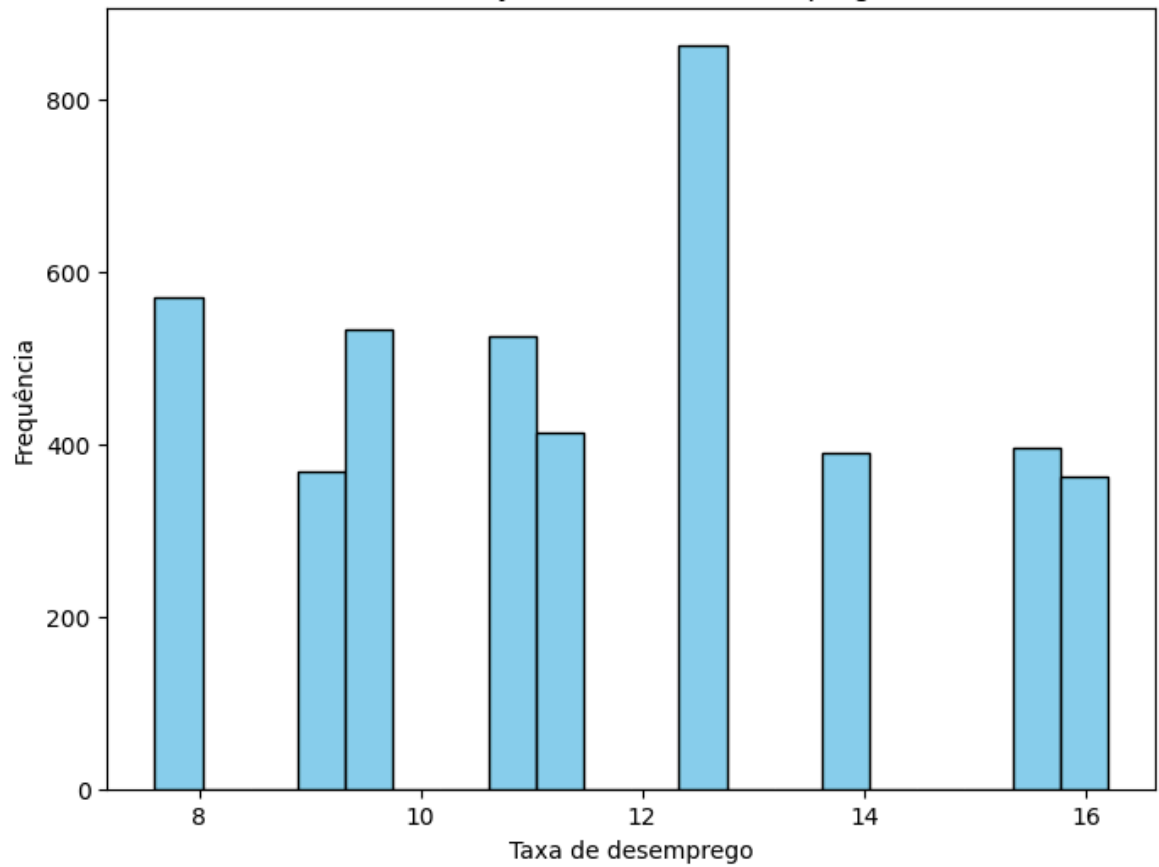
Histograma

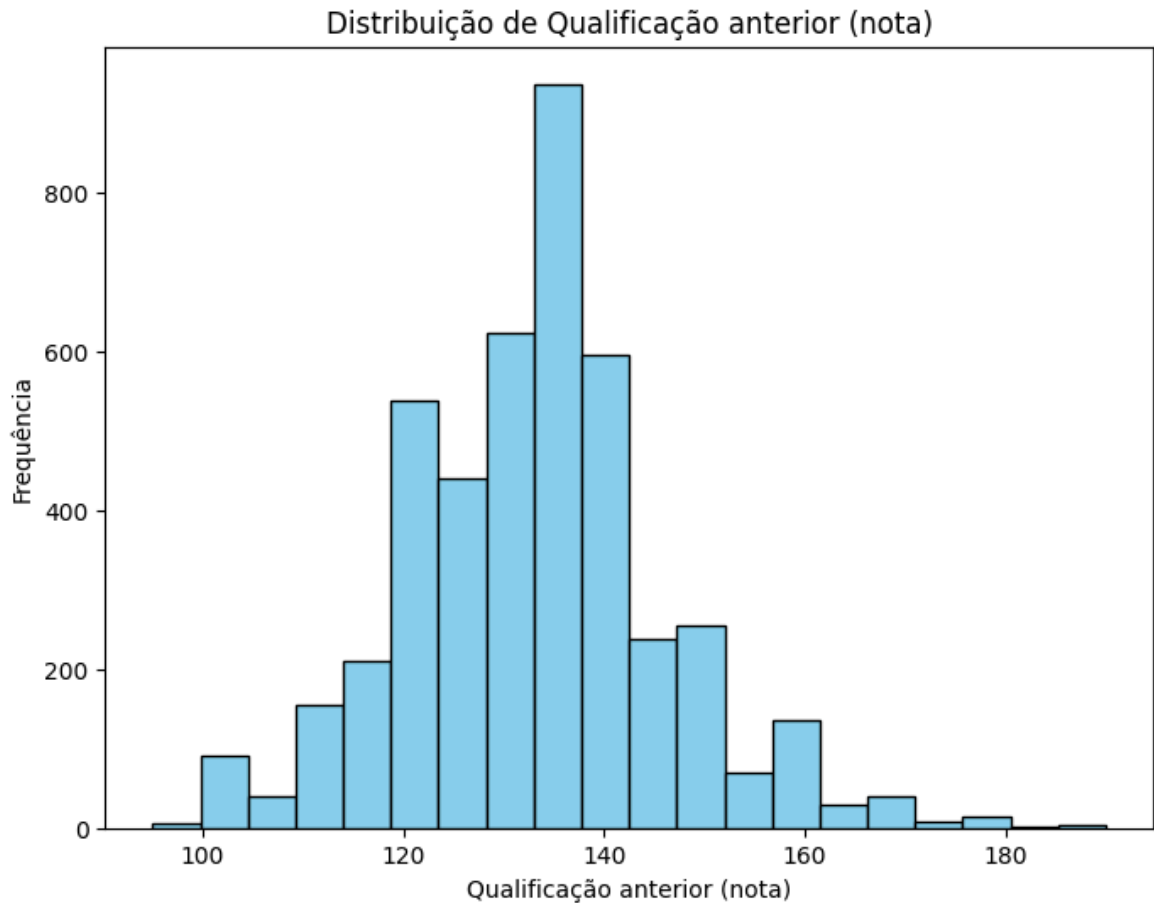
```
In [ ]: # Histograma para variáveis contínuas
for variavel in ['Nota de admissão', 'Taxa de desemprego', 'Qualificação'
plt.figure(figsize=(8, 6))
dados[variavel].plot(kind='hist', bins=20, color='skyblue', edgecolor
plt.title(f'Distribuição de {variavel}')
plt.xlabel(variavel)
plt.ylabel('Frequência')
plt.show()
```

Distribuição de Nota de admissão



Distribuição de Taxa de desemprego





Análise dos histogramas:

Distribuição de nota de admissão: O histograma revela uma distribuição de notas de admissão que se assemelha a uma curva gaussiana, com a maior probabilidade de notas concentradas em torno do valor 120.

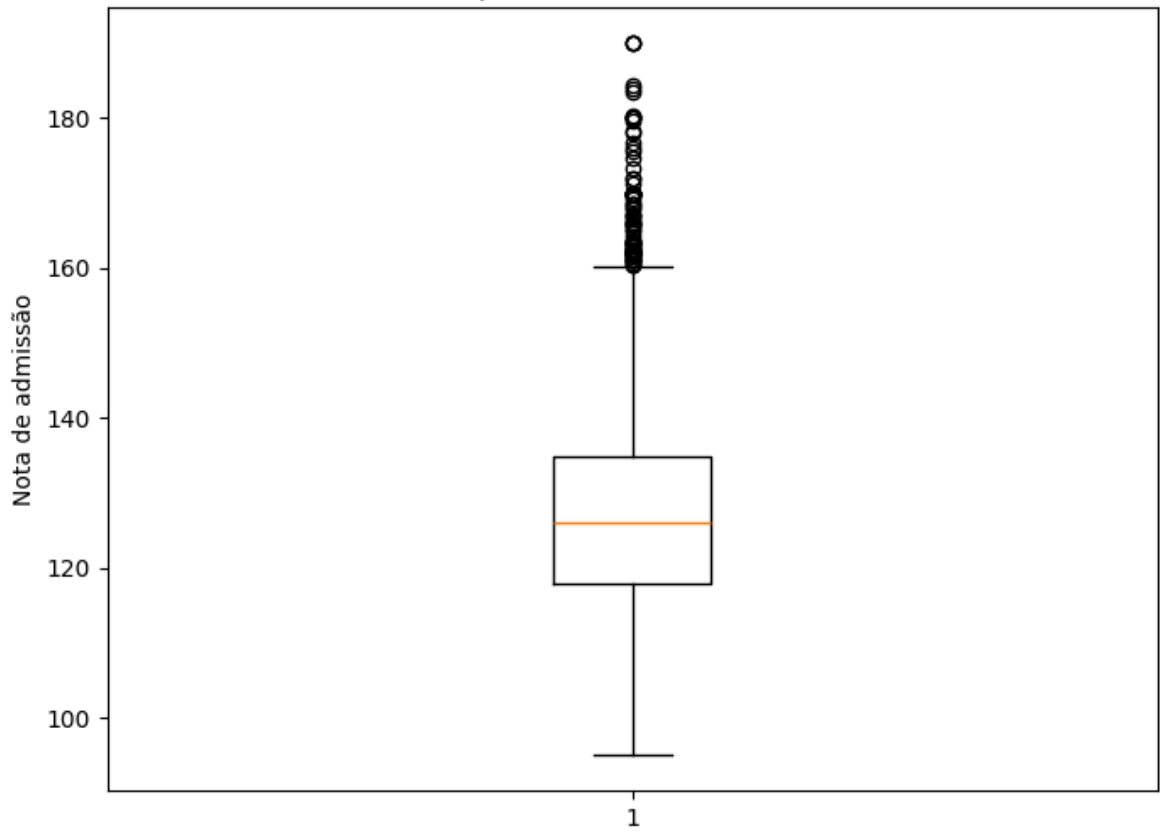
Distribuição de taxa de desemprego: Em relação à taxa de desemprego, fica claro que não há uma relação óbvia entre a frequência e a taxa de desemprego, uma vez que este é um dado altamente volátil e influenciado por uma série de fatores externos relacionados à economia local e global.

Qualificação anterior (nota): O histograma revela uma distribuição de notas que também se assemelha a uma curva gaussiana, com a maior probabilidade de notas concentradas em torno do valor 140.

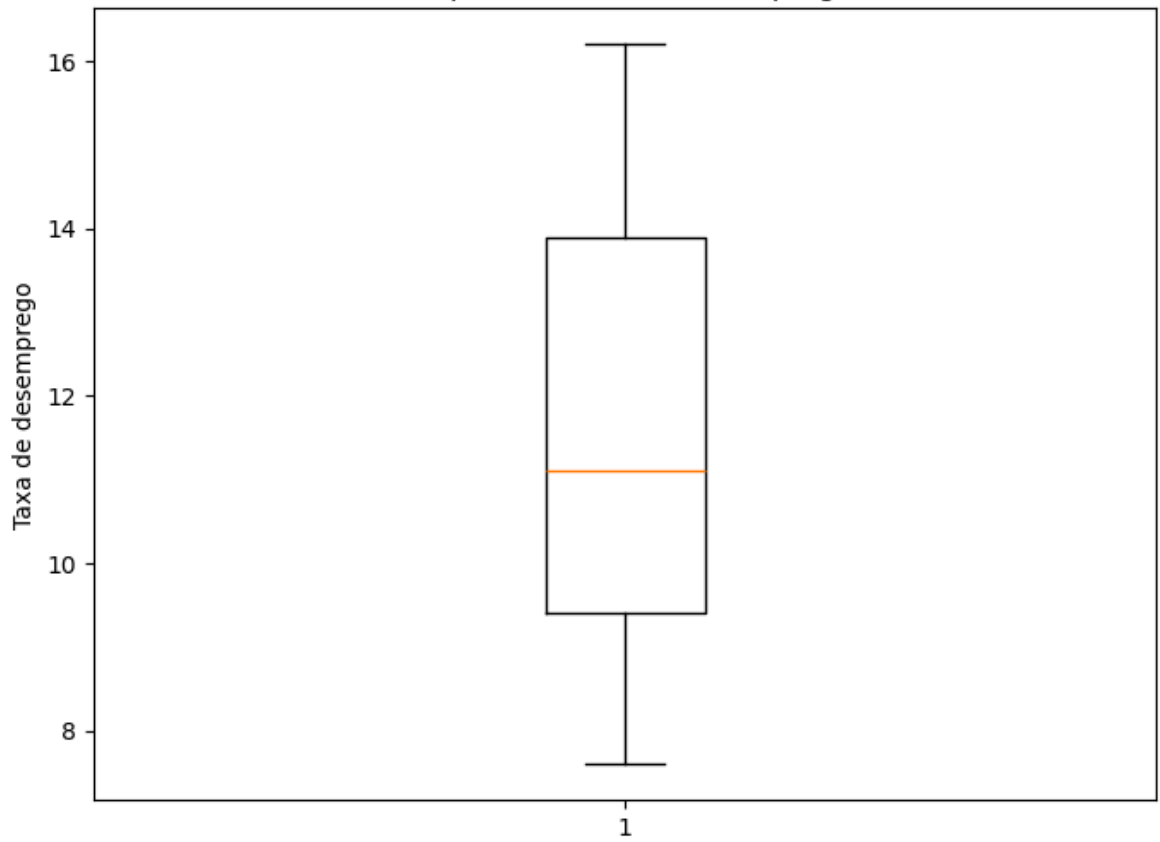
Boxplot

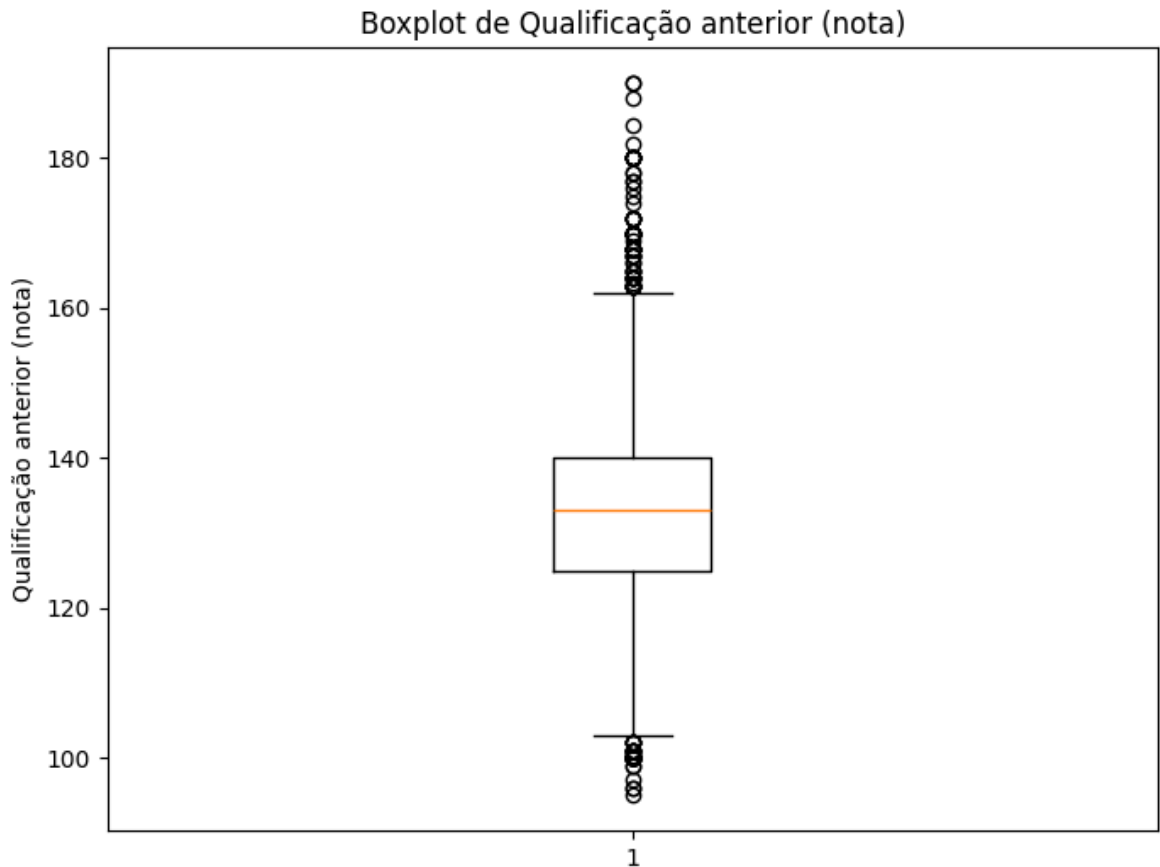
```
In [ ]: # Boxplot para variáveis contínuas
for variavel in ['Nota de admissão', 'Taxa de desemprego', 'Qualificação
plt.figure(figsize=(8, 6))
plt.boxplot(dados[variavel])
plt.title(f'Boxplot de {variavel}')
plt.ylabel(variavel)
plt.show()
```

Boxplot de Nota de admissão



Boxplot de Taxa de desemprego





Análise dos Boxplot:

Boxplot nota de admissão: O boxplot das notas de admissão revela uma distribuição semelhante àquela observada no histograma, com a maioria das notas concentradas em torno do valor 120.

Boxplot taxa de desemprego: Este boxplot indica que as taxas de desemprego locais na época da pesquisa estavam predominantemente entre 10% e 14%.

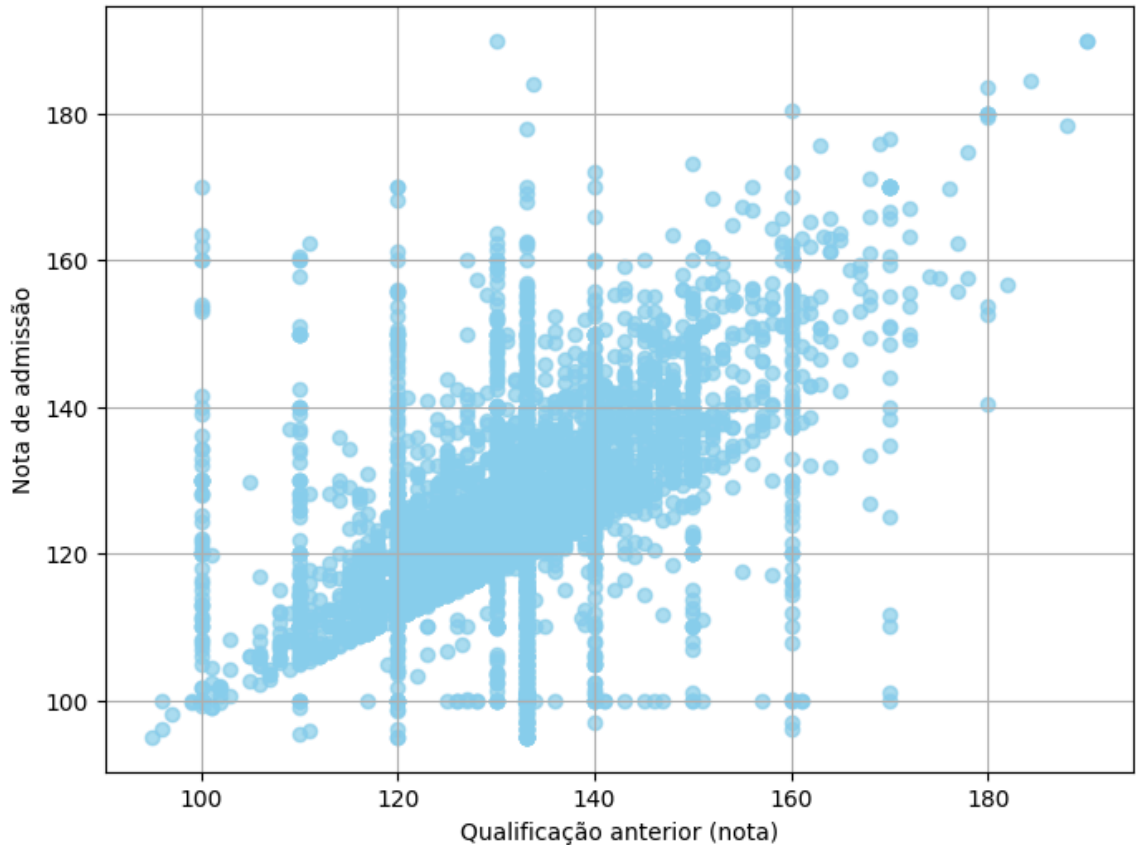
Boxplot Qualificação anterior (nota): O boxplot das notas revela uma distribuição semelhante àquela observada no histograma, com a maioria das notas concentradas em torno de valores entre 120 e 140.

Diagrama de dispersão

```
In [ ]: # Escolhe as duas variáveis quantitativas para o diagrama de dispersão
variavel_x = 'Qualificação anterior (nota)'
variavel_y = 'Nota de admissão'

# Cria o diagrama de dispersão
plt.figure(figsize=(8, 6))
plt.scatter(dados[variavel_x], dados[variavel_y], color='skyblue', alpha=
plt.title('Diagrama de Dispersão entre Qualificação anterior (nota) e Not
plt.xlabel(variavel_x)
plt.ylabel(variavel_y)
plt.grid(True)
plt.show()
```

Diagrama de Dispersão entre Qualificação anterior (nota) e Nota de admissão



Análise do diagrama de dispersão:

Este diagrama de dispersão mostra uma relação linear clara e crescente entre as variáveis representadas. Isso sugere que existe uma conexão direta e proporcional entre o desempenho acadêmico pré-universidade e as notas de admissão. Em outras palavras, à medida que o desempenho acadêmico anterior aumenta, espera-se que as notas de admissão também aumentem de forma consistente. Isso pode indicar que alunos com um histórico acadêmico mais forte têm uma tendência a obter pontuações mais altas nas avaliações de admissão.

Medidas de resumo

Variável quantitativa

```
In [ ]: # Calcular as medidas para a variavel
variavel = 'Nota de admissão'

# Média
media = dados[variavel].mean()
print(f'A média da variável {variavel} é: {media}')

# Mediana
mediana = dados[variavel].median()
print(f'A mediana da variável {variavel} é: {mediana}')

# Moda
moda = dados[variavel].mode()[0]
print(f'A moda da variável {variavel} é: {moda}')
```

```

# Quartis
quartis = np.percentile(dados[variavel], [25, 50, 75])
print(f'Os quartis da variável {variavel} são:')
print(f'Q1: {quartis[0]}, Q2: {quartis[1]}, Q3: {quartis[2]}')

# Desvio padrão
desvio_padrao = dados[variavel].std()
print(f'0 desvio padrão da variável {variavel} é: {desvio_padrao}')

# Amplitude
amplitude = dados[variavel].max() - dados[variavel].min()
print(f'A amplitude da variável {variavel} é: {amplitude}')

# Variância
variancia = dados[variavel].var()
print(f'A variância da variável {variavel} é: {variancia}')

# Intervalo interquartil (IQR)
IQR = quartis[2] - quartis[0]
print(f'0 intervalo interquartil (IQR) da variável {variavel} é: {IQR}')

# Coeficiente de variação
coef_variacao = (desvio_padrao / media) * 100
print(f'0 coeficiente de variação para a variável {variavel} é: {varianci

```

A média da variável Nota de admissão é: 126.97811934900544

A mediana da variável Nota de admissão é: 126.1

A moda da variável Nota de admissão é: 130.0

Os quartis da variável Nota de admissão são:

Q1: 117.9, Q2: 126.1, Q3: 134.8

0 desvio padrão da variável Nota de admissão é: 14.482000818849468

A amplitude da variável Nota de admissão é: 95.0

A variância da variável Nota de admissão é: 209.72834771715662

0 intervalo interquartil (IQR) da variável Nota de admissão é: 16.9000000000000006

0 coeficiente de variação para a variável Nota de admissão é: 209.72834771715662

Variável qualitativa

```

In [ ]: # Calcular as medidas para a variavel
variavel = 'Estado civil'

# Média
media = dados[variavel].mean()
print(f'A média da variável {variavel} é: {media}')

# Mediana
mediana = dados[variavel].median()
print(f'A mediana da variável {variavel} é: {mediana}')

# Moda
moda = dados[variavel].mode()[0]
print(f'A moda da variável {variavel} é: {moda}')

# Quartis
quartis = np.percentile(dados[variavel], [25, 50, 75])
print(f'Os quartis da variável {variavel} são:')
print(f'Q1: {quartis[0]}, Q2: {quartis[1]}, Q3: {quartis[2]}')

```

```

# Desvio padrão
desvio_padrao = dados[variavel].std()
print(f'O desvio padrão da variável {variavel} é: {desvio_padrao}')

# Amplitude
amplitude = dados[variavel].max() - dados[variavel].min()
print(f'A amplitude da variável {variavel} é: {amplitude}')

# Variância
variancia = dados[variavel].var()
print(f'A variância da variável {variavel} é: {variancia}')

# Intervalo interquartil (IQR)
IQR = quartis[2] - quartis[0]
print(f'O intervalo interquartil (IQR) da variável {variavel} é: {IQR}')

# Coeficiente de variação
coef_variacao = (desvio_padrao / media) * 100
print(f'O coeficiente de variação para a variável {variavel} é: {varianci

```

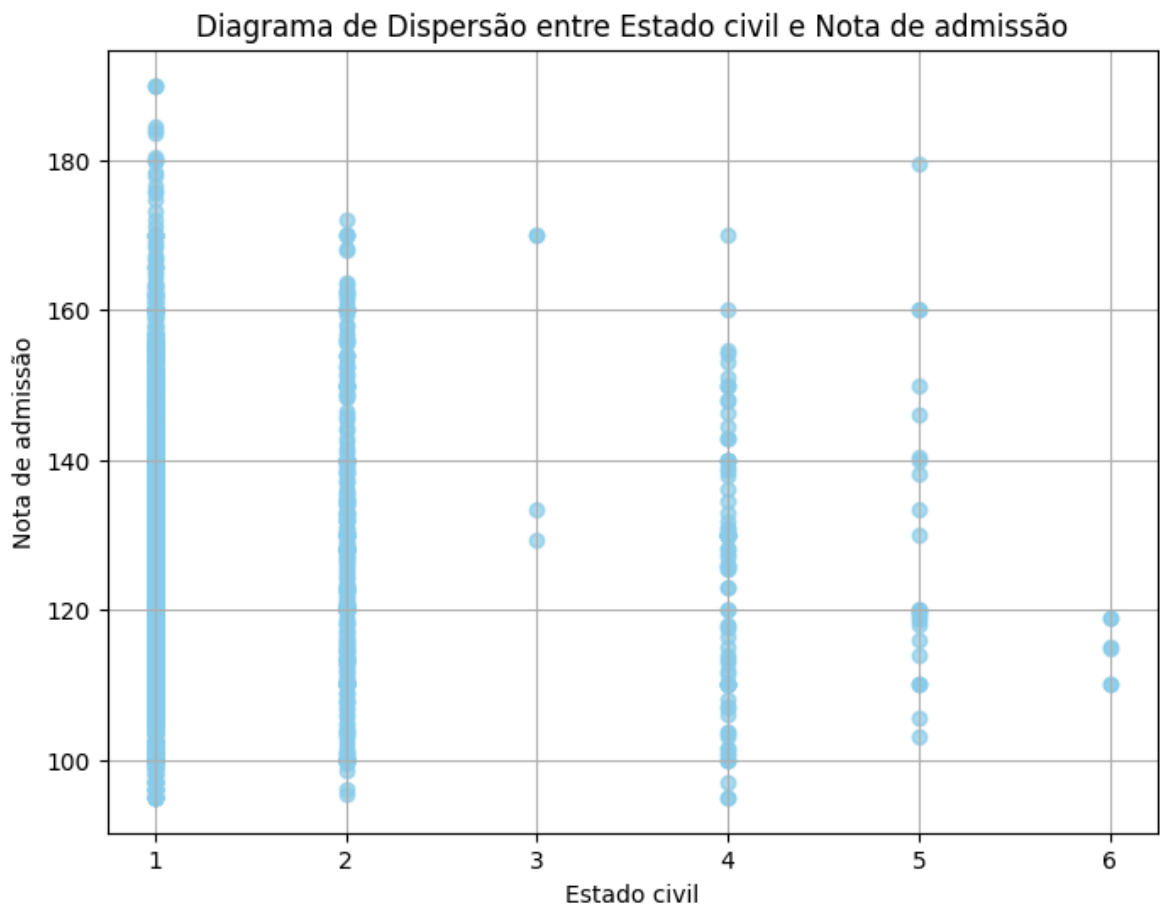
A média da variável Estado civil é: 1.1785714285714286
 A mediana da variável Estado civil é: 1.0
 A moda da variável Estado civil é: 1
 Os quartis da variável Estado civil são:
 Q1: 1.0, Q2: 1.0, Q3: 1.0
 O desvio padrão da variável Estado civil é: 0.60574694613071
 A amplitude da variável Estado civil é: 5
 A variância da variável Estado civil é: 0.36692936274668125
 O intervalo interquartil (IQR) da variável Estado civil é: 0.0
 O coeficiente de variação para a variável Estado civil é: 0.36692936274668125

```

In [ ]: # Escolhe as duas variáveis quantitativas para o diagrama de dispersão
variavel_x = 'Estado civil'
variavel_y = 'Nota de admissão'

# Cria o diagrama de dispersão
plt.figure(figsize=(8, 6))
plt.scatter(dados[variavel_x], dados[variavel_y], color='skyblue', alpha=
plt.title('Diagrama de Dispersão entre Estado civil e Nota de admissão')
plt.xlabel(variavel_x)
plt.ylabel(variavel_y)
plt.grid(True)
plt.show()

```



Análise do diagrama de dispersão:

A partir deste diagrama de dispersão não fica evidente uma relação entre o Estado civil e a nota de admissão, isso pode ter acontecido devido ao fato de possuir uma amostra muito maior de pessoas solteira em relação à outras com estado civil diferente, mas a priori não foi observado nenhuma relação clara entre as variáveis.