

## ✓ Medidas de Estatística Descritiva

### ✓ 1 - Medidas de Tendência Central

```
1 # 1.1 Média Aritmética
2 '''
3 O que é a média aritmética?
4
5 Em sua essência, a média aritmética é a soma de todos os valores em um conjunto
6 de dados dividida pelo número de elementos. Imagine que você tem as notas de uma
7 turma em uma prova: somar todas as notas e dividir pelo número de alunos te dará
8 a média da turma, um indicador de seu desempenho geral.
9
10 Fórmula e Cálculo:
11
12 A fórmula da média aritmética é simples:
13
14 Média =  $\Sigma x / n$ 
15
16 Onde:
17
18  $\Sigma x$  representa a soma de todos os valores (x) no conjunto de dados.
19 n é o número de elementos no conjunto.
20 Aplicações na Análise de Dados:
21
22 A média aritmética possui diversas aplicações na análise de dados:
23
24 Comparação de conjuntos de dados: Permite comparar o desempenho de diferentes
25 grupos, como a média de vendas de duas lojas ou a média de idade de dois times.
26 Identificação de tendências: Ao analisar a média ao longo do tempo, é possível
27 observar se um valor está aumentando, diminuindo ou se mantém estável.
28 Avaliação de desempenho: A média serve como referência para avaliar o desempenho
29 individual em relação ao grupo, como a média salarial de uma profissão.
30 Detecção de outliers: Valores muito acima ou abaixo da média podem indicar erros
31 nos dados ou outliers, que merecem atenção especial.
32 Limitações da Média Aritmética:
33
34 É importante lembrar que a média aritmética, como qualquer ferramenta, tem suas
35 limitações:
36
37 Sensibilidade a outliers: Valores extremos podem distorcer a média, levando a
38 interpretações errôneas. Falta de informação sobre a distribuição dos dados:
39 A média não fornece informações sobre a dispersão dos dados, apenas sobre o
40 valor central. Inadequação para dados qualitativos: A média só pode ser utilizada
41 com dados numéricos, não com dados qualitativos como cores ou gostos.
42 Exemplos práticos:
43
44 Analisando vendas: Calcular a média de vendas por mês para identificar períodos
45 de alta e baixa demanda.
46 Comparando salários: Comparar a média salarial de diferentes cargos para
47 verificar se há disparidades.
48 Avaliando desempenho de alunos: Calcular a média de notas de uma turma para
49 avaliar o desempenho geral e identificar alunos com dificuldades.
50
51 Conclusão:
52
53 A média aritmética é uma ferramenta fundamental para o analista de dados,
54 fornecendo informações valiosas sobre o centro de um conjunto de dados.
55 Ao utilizá-la em conjunto com outras medidas estatísticas, é possível obter uma
56 visão completa e precisa dos dados, embasando decisões estratégicas e otimizando
57 resultados.
58
59 Lembre-se:
```

```
60
61 A média aritmética é uma ferramenta poderosa, mas não é a única. Utilize-a em
62 conjunto com outras medidas para obter uma visão completa dos dados. Esteja
63 atento às suas limitações, como a sensibilidade a outliers e a inadequação para
64 dados qualitativos.
65 '''
66
67 # Exemplo: Calculando a média de notas de uma amostra de 5 alunos
68
69 notas = [5,6,7,8,9]
70 media = sum(notas)/len(notas)
71 print(f'A media calculada manualmente é {media:.2f}')
72
73 # calculando a média utilizando o Numpy
74 import numpy as np
75 medianp = np.mean(notas)
76 print(f'A media calculada pelo numpy é {medianp:.2f}')
77
78 # calculando a média utilizando o statistics
79 import statistics as st
80 mediast = st.mean(notas)
81 print(f'A media calculada pelo statistics é {mediast:.2f}')
```



A media calculada manualmente é 7.00  
A media calculada pelo numpy é 7.00  
A media calculada pelo statistics é 7.00

```
1 # 1.2 Mediana
2
3 '''
4 O que é a mediana?
5
6 Em termos simples, a mediana é o valor que separa a metade superior da metade
7 inferior em um conjunto de dados ordenados. Imagine uma turma de alunos ordenada
8 por altura: a altura do aluno que está no meio da fila é a mediana.
9
10 Cálculo da Mediana:
11
12 O cálculo da mediana é intuitivo:
13
14 Ordene os dados em ordem crescente ou decrescente.
15 Identifique o valor central:
16 Se o número de dados for ímpar, a mediana é o valor central.
17 Se o número de dados for par, a mediana é a média dos dois valores centrais.
18 Aplicações na Análise de Dados:
19
20 A mediana possui diversas aplicações na análise de dados:
21
22 Comparação de conjuntos de dados: Permite comparar o desempenho de diferentes
23 grupos, como a mediana de renda de duas cidades ou a mediana de idade de dois
24 times. Identificação de valores atípicos (outliers): Valores muito acima ou
25 abaixo da mediana podem indicar outliers, que merecem atenção especial.
26 Análise de dados assimétricos: A mediana é mais robusta que a média para lidar
27 com dados que não possuem uma distribuição normal (simétrica).
28 Avaliação de pesquisas de opinião: A mediana é frequentemente utilizada em
29 pesquisas para determinar a opinião majoritária sobre um determinado tema.
30 Exemplos práticos:
31
32 Analisando salários: Calcular a mediana salarial para ter uma ideia do salário
33 "típico" em uma profissão. Comparando tempo de resposta: Comparar a mediana do
34 tempo de resposta de dois sistemas para verificar qual é mais rápido.
35 Avaliando pesquisas de satisfação: Calcular a mediana da satisfação dos clientes
36 com um produto ou serviço.
37
38 Vantagens da Mediana:
39
40 Simplicidade de cálculo: A mediana pode ser facilmente calculada manualmente ou
41 com ferramentas de análise de dados.
42 Robustez: A mediana não é afetada por outliers, tornando-a mais confiável em
43 alguns casos que a média.
44 Interpretabilidade: A mediana é um valor fácil de interpretar, representando o
45 "centro" dos dados.
46
47 Limitações da Mediana:
48
49 Perda de informação: A mediana não fornece informações sobre a dispersão dos
50 dados, apenas sobre o valor central.
51 Menos precisa que a média em alguns casos: Quando os dados são normalmente
52 distribuídos, a média é mais precisa que a mediana.
53
54 Conclusão:
55
56 A mediana é uma ferramenta valiosa para o analista de dados, fornecendo
57 informações sobre o ponto central dos dados e ajudando a identificar outliers.
58 Ao utilizá-la em conjunto com outras medidas estatísticas, é possível obter uma
59 visão completa e precisa dos dados, embasando decisões estratégicas e otimizando
60 resultados.
61
62 Lembre-se:
63
64 A mediana é uma ferramenta complementar à média, fornecendo diferentes
65 perspectivas sobre os dados. Utilize a mediana em conjunto com outras medidas
66 para obter uma visão completa dos dados. Esteja atento às suas limitações, como a
67 perda de informação e a menor precisão em alguns casos.
68 '''
```

```
69
70 # Exemplo: Calculando a mediana de notas de uma amostra de 5 alunos
71
72 notas = [5,6,7,8,9]
73 mediana = notas[int((len(notas)+1)/2)-1]
74 print(f'A mediana calculada manualmente é {mediana:.2f}')
75
76 # calculando a média utilizando o Numpy
77 import numpy as np
78 mediananp = np.median(notas)
79 print(f'A mediana calculada pelo numpy é {mediananp:.2f}')
80
81 # calculando a média utilizando o statistics
82 import statistics as st
83 medianast = st.median(notas)
84 print(f'A mediana calculada pelo statistics é {medianast:.2f}')

A mediana calculada manualmente é 7.00
A mediana calculada pelo numpy é 7.00
A mediana calculada pelo statistics é 7.00
```

1 # 1.3 Moda  
2 '''  
3 O que é a moda?  
4  
5 A moda é o valor que aparece com mais frequência em um conjunto de dados.  
6 Imagine uma pesquisa sobre a cor favorita de um grupo de pessoas: a cor que mais  
7 se repete nas respostas é a moda.  
8  
9 Cálculo da Moda:  
10  
11 O cálculo da moda é simples:  
12  
13 Organize os dados em uma tabela de frequências, contando quantas vezes cada  
14 valor aparece. Identifique o valor com a maior frequência: este é o valor da  
15 moda.  
16  
17 Aplicações na Análise de Dados:  
18  
19 A moda possui diversas aplicações na análise de dados:  
20 Identificação de tendências: A moda pode revelar quais valores são mais  
21 frequentes em um conjunto de dados, indicando tendências de mercado,  
22 comportamentos de consumo ou preferências do público.  
23 Análise de dados categóricos: A moda é especialmente útil para analisar dados  
24 categóricos, como cores, tamanhos ou marcas, que não podem ser facilmente  
25 ordenados ou medidos.  
26 Segmentação de mercado: A moda pode auxiliar na segmentação de mercado,  
27 identificando grupos de consumidores com características em comum.  
28 Desenvolvimento de produtos: A moda pode ser utilizada para identificar quais  
29 produtos ou serviços são mais demandados pelo público, direcionando o  
30 desenvolvimento de novas ofertas.  
31  
32 Exemplos práticos:  
33  
34 Analisando tamanhos de roupas: Calcular a moda dos tamanhos de roupas vendidos  
35 em uma loja para identificar os tamanhos mais procurados.  
36 Comparando marcas de carros: Identificar a moda das marcas de carros em um  
37 determinado país para verificar qual marca é a mais popular.  
38 Avaliando respostas de pesquisa: Calcular a moda das respostas a uma pergunta  
39 de pesquisa para identificar a opinião mais comum sobre um determinado tema.  
40 Vantagens da Moda:  
41  
42 Simplicidade de cálculo: A moda pode ser facilmente calculada manualmente ou  
43 com ferramentas de análise de dados.  
44 Fácil interpretação: A moda é um valor fácil de interpretar, representando o  
45 valor mais popular em um conjunto de dados.  
46 Útil para dados categóricos: A moda é especialmente útil para analisar dados que  
47 não podem ser facilmente ordenados ou medidos.  
48 Limitações da Moda:  
49  
50 Sensibilidade a outliers: Valores muito frequentes podem distorcer a moda,  
51 levando a interpretações errôneas.  
52 Falta de informação sobre a dispersão dos dados: A moda não fornece informações  
53 sobre a dispersão dos dados, apenas sobre o valor mais frequente.  
54 Inadequação para dados numéricos: A moda é mais adequada para dados categóricos  
55 e pode não ser útil para dados numéricos.  
56  
57 Conclusão:  
58  
59 A moda é uma ferramenta valiosa para o analista de dados, fornecendo informações  
60 sobre o valor mais frequente em um conjunto de dados e ajudando a identificar  
61 tendências e padrões. Ao utilizá-la em conjunto com outras medidas estatísticas,  
62 é possível obter uma visão completa e precisa dos dados, embasando decisões  
63 estratégicas e otimizando resultados.  
64  
65 Lembre-se:  
66  
67 A moda é uma ferramenta complementar à média e à mediana, fornecendo diferentes  
68 perspectivas sobre os dados. Utilize a moda em conjunto com outras medidas para

```
69 obter uma visão completa dos dados. Esteja atento às suas limitações, como a
70 sensibilidade a outliers e a inadequação para dados numéricos.
71 '''
72
73 # Exemplo: Calculando a moda de número de filhos de uma amostra de 5 funcionários
74
75 numero_filhos = [0,1,2,2,3]
76 moda = 2 # 2 filhos é a quantidade que aparece mais vezes nessa amostra
77 print(f'A moda calculada manualmente é {moda}')
78
79 # calculando a moda utilizando o statistics
80 import statistics as st
81 modast = st.mode(numero_filhos)
82 print(f'A moda calculada pelo statistics é {modast}')
```

A moda calculada manualmente é 2  
A moda calculada pelo statistics é 2

## ✓ 2 - Medidas de Dispersão

```
1 # 2.1 Variância
2 '''
3 O que é a variância?
4
5 Em termos simples, a variância é a média dos quadrados das distâncias entre cada
6 valor e a média do conjunto de dados. Imagine uma turma de alunos com diferentes
7 notas em uma prova: a variância das notas indica o quão dispersas as notas estão
8 em relação à média da turma.
9
10 Fórmula e Cálculo da Variância:
11
12 A fórmula da variância é:
13
14 Variância =  $\Sigma[(x - \mu)^2] / (n - 1)$ 
15
16 Onde:
17
18  $\Sigma$  representa a soma de todos os valores.
19 x é cada valor no conjunto de dados.
20  $\mu$  é a média do conjunto de dados.
21 n é o número de elementos no conjunto.
22 Interpretação da Variância:
23
24 Uma variância alta indica que os valores estão mais dispersos em torno da média.
25 Uma variância baixa indica que os valores estão mais concentrados em torno da
26 média.
27
28 Aplicações da Variância na Análise de Dados:
29
30 Comparação de conjuntos de dados: Permite comparar a dispersão dos dados em
31 diferentes grupos, como a variância das notas de duas turmas ou a variância dos
32 salários de duas empresas.
33 Avaliação de risco: A variância é utilizada para avaliar o risco em
34 investimentos, pois quanto maior a variância, maior o risco.
35 Análise de regressão: A variância é utilizada na análise de regressão para
36 determinar a qualidade do modelo de ajuste aos dados.
37 Controle de qualidade: A variância é utilizada no controle de qualidade para
38 monitorar a variabilidade de um processo produtivo.
39 Exemplos práticos:
40
41 Analisando a altura de plantas: Calcular a variância da altura de um conjunto de
42 plantas para verificar se elas são homogêneas em tamanho.
43 Comparando o desempenho de ações: Comparar a variância do retorno de duas ações
44 para verificar qual é mais volátil.
45 Avaliando a confiabilidade de um teste: Calcular a variância das notas de um
46 teste para verificar se ele é confiável.
47 Vantagens da Variância:
48
49 Fornece uma medida quantitativa da dispersão: A variância permite comparar a
50 dispersão dos dados em diferentes conjuntos de forma precisa.
51 Útil para análise de risco: A variância é uma ferramenta importante para avaliar
52 o risco em investimentos e outras áreas.
53 Utilizada em diversas áreas: A variância é utilizada em diversas áreas como
54 finanças, estatística, engenharia e controle de qualidade.
55
56 Limitações da Variância:
57
58 Sensível a outliers: Valores muito altos ou baixos podem distorcer a variância,
59 levando a interpretações errôneas.
60 Unidade de medida: A variância depende da unidade de medida dos dados,
61 dificultando a comparação entre conjuntos com diferentes unidades.
62 Dificuldade de interpretação: A variância pode ser difícil de interpretar para o
63 público leigo.
64
65 Conclusão:
66
67 A variância é uma ferramenta valiosa para o analista de dados, fornecendo
68 informações sobre a dispersão dos dados em relação à média. Ao utilizá-la em
```

69 conjunto com outras medidas estatísticas, é possível obter uma visão completa e  
70 precisa dos dados, embasando decisões estratégicas e otimizando resultados.

71

72 Lembre-se:

73

74 A variância é uma ferramenta complementar à média, mediana e moda, fornecendo  
75 diferentes perspectivas sobre os dados. Utilize a variância em conjunto com  
76 outras medidas para obter uma visão completa dos dados. Esteja atento às suas  
77 limitações, como a sensibilidade a outliers e a dificuldade de interpretação.

78 '''

79

80 notas = [5,6,7,8,9]

81

82 # calculando a variância manualmente

83 n = len(notas)

84 mean = sum(notas)/n

85 notasq = []

86 for i in range(n):

87 notasq.append((notas[i]-mean)\*\*2)

88 var = sum(notasq)/(n-1)

89 print(f'A variância calculada manualmente é {var:.2f}')

90

91 # calculando a variância utilizando o Numpy

92 import numpy as np

93 varnp = np.var(notas, ddof=1)

94 print(f'A variância calculada pelo numpy é {varnp:.2f}')

95

96 # calculando a variância utilizando o statistics

97 import statistics as st

98 varst = st.variance(notas)

99 print(f'A variância calculada pelo statistics é {varst:.2f}')

A variância calculada manualmente é 2.50

A variância calculada pelo numpy é 2.50

A variância calculada pelo statistics é 2.50



```
1 # 2.1 Desvio Padrão
2 '''
3 O que é o Desvio Padrão?
4
5 Em termos simples, o desvio padrão é a raiz quadrada da variância. Imagine uma
6 turma de alunos com diferentes notas em uma prova: o desvio padrão das notas
7 indica o quão dispersas as notas estão em relação à média da turma, em termos de
8 unidades de nota.
9
10 Fórmula e Cálculo do Desvio Padrão:
11
12 A fórmula do desvio padrão é:
13
14 Desvio Padrão =  $\sqrt{(\sum[(x - \mu)^2] / (n - 1))}$ 
15
16 Onde:
17
18  $\sqrt{\phantom{x}}$  representa a raiz quadrada.
19  $\sum$  representa a soma de todos os valores.
20  $x$  é cada valor no conjunto de dados.
21  $\mu$  é a média do conjunto de dados.
22  $n$  é o número de elementos no conjunto.
23 Interpretação do Desvio Padrão:
24
25 Um desvio padrão alto indica que os valores estão mais dispersos em torno da
26 média. Um desvio padrão baixo indica que os valores estão mais concentrados em
27 torno da média.
28
29 Aplicações do Desvio Padrão na Análise de Dados:
30
31 Comparação de conjuntos de dados: Permite comparar a dispersão dos dados em
32 diferentes grupos, como o desvio padrão das notas de duas turmas ou o desvio
33 padrão dos salários de duas empresas.
34 Avaliação de risco: O desvio padrão é utilizado para avaliar o risco em
35 investimentos, pois quanto maior o desvio padrão, maior o risco.
36 Análise de regressão: O desvio padrão é utilizado na análise de regressão para
37 determinar a qualidade do modelo de ajuste aos dados.
38 Controle de qualidade: O desvio padrão é utilizado no controle de qualidade para
39 monitorar a variabilidade de um processo produtivo.
40
41 Exemplos práticos:
42
43 Analisando a altura de plantas: Calcular o desvio padrão da altura de um
44 conjunto de plantas para verificar se elas são homogêneas em tamanho.
45 Comparando o desempenho de ações: Comparar o desvio padrão do retorno de duas
46 ações para verificar qual é mais volátil.
47 Avaliando a confiabilidade de um teste: Calcular o desvio padrão das notas de um
48 teste para verificar se ele é confiável.
49
50 Vantagens do Desvio Padrão:
51
52 Fornece uma medida quantitativa da dispersão em unidades da mesma escala dos
53 dados originais: Isso facilita a interpretação e comparação da dispersão entre
54 diferentes conjuntos de dados.
55 Útil para análise de risco: O desvio padrão é uma ferramenta importante para
56 avaliar o risco em investimentos e outras áreas.
57 Utilizada em diversas áreas: O desvio padrão é utilizado em diversas áreas como
58 finanças, estatística, engenharia e controle de qualidade.
59 Limitações do Desvio Padrão:
60
61 Sensível a outliers: Valores muito altos ou baixos podem distorcer o desvio
62 padrão, levando a interpretações errôneas.
63 Dificuldade de interpretação para o público leigo: O desvio padrão pode ser
64 difícil de interpretar para quem não tem conhecimento estatístico.
65
66 Conclusão:
67
68 O desvio padrão é uma ferramenta valiosa para o analista de dados, fornecendo
```

```
69 informações sobre a dispersão dos dados em relação à média em termos práticos.
70 Ao utilizá-lo em conjunto com outras medidas estatísticas, é possível obter uma
71 visão completa e precisa dos dados, embasando decisões estratégicas e otimizando
72 resultados.
73
74 Lembre-se:
75
76 O desvio padrão é uma ferramenta complementar à média, mediana, moda e
77 variância, fornecendo diferentes perspectivas sobre os dados.Utilize o
78 desvio padrão em conjunto com outras medidas para obter uma visão completa dos
79 dados.Esteja atento às suas limitações, como a sensibilidade a outliers e a
80 dificuldade de interpretação para o público leigo.
81 '''
82 notas = [5,6,7,8,9]
83
84 # calculando a variância manualmente
85 n = len(notas)
86 mean = sum(notas)/n
87 notasq = []
88 for i in range(n):
89     notasq.append((notas[i]-mean)**2)
90 desvio = (sum(notasq)/(n-1))**(1/2)
91 print(f'O desvio padrão calculado manualmente é {desvio:.2f}')
92
93 # calculando a variância utilizando o Numpy
94 import numpy as np
95 desvionp = np.std(notas, ddof=1)
96 print(f'A variância calculada pelo numpy é {desvionp:.2f}')
97
98 # calculando a variância utilizando o statistics
99 import statistics as st
100 desviost = st.stdev(notas)
101 print(f'A variância calculada pelo statistics é {desviost:.2f}')

    O desvio padrão calculado manualmente é 1.58
    A variância calculada pelo numpy é 1.58
    A variância calculada pelo statistics é 1.58
```

### 3 - Medidas Separatrizes

```
1 # 3.1 - 1º Quartil - Q1
2 '''
3 Introdução:
4
5 Em sua jornada como analista de dados, você encontrará conjuntos de dados com
6 diversas formas e distribuições. O primeiro quartil (Q1) surge como uma
7 ferramenta poderosa para dividir um conjunto de dados ordenados em quatro partes
8 iguais, revelando o ponto em que 25% dos valores estão abaixo e 75% acima.
9
10 O que é o Primeiro Quartil (Q1)?
11
12 Em termos simples, o Q1 é o valor que separa os 25% menores valores do restante
13 em um conjunto de dados ordenado. Imagine uma fila de alunos ordenados por
14 altura: o Q1 seria a altura do aluno que está na posição 25% da fila, dividindo
15 os alunos mais baixos dos mais altos.
16
17 Cálculo do Primeiro Quartil (Q1):
18
19 O cálculo do Q1 é intuitivo:
20
21 Ordene os dados em ordem crescente.
22 Calcule a posição do Q1:
23 Se o número de dados for ímpar, a posição do Q1 é  $(n + 1) / 4$ , arredondando para
24 o número inteiro mais próximo.
25 Se o número de dados for par, a posição do Q1 é a média das posições
26  $n / 4$  e  $(n + 2) / 4$ , arredondando para o número inteiro mais próximo.
27 Encontre o valor na posição calculada: este é o valor do Q1.
28
29 Aplicações do Primeiro Quartil (Q1) na Análise de Dados:
30
31 Comparação de conjuntos de dados: Permite comparar a distribuição dos dados em
32 diferentes grupos, como o Q1 da renda de duas cidades ou o Q1 da idade de dois
33 times.
34 Identificação de outliers: Valores muito abaixo do Q1 podem indicar outliers,
35 que merecem atenção especial.
36 Análise de dados assimétricos: O Q1 é mais robusto que a média para lidar com
37 dados que não possuem uma distribuição normal (simétrica).
38 Avaliação de pesquisas de opinião: O Q1 pode ser utilizado em pesquisas para
39 determinar a opinião da minoria sobre um determinado tema.
40
41 Exemplos práticos:
42
43 Analisando salários: Calcular o Q1 dos salários para ter uma ideia do salário
44 "mínimo" em uma profissão.
45 Comparando tempo de resposta: Comparar o Q1 do tempo de resposta de dois
46 sistemas para verificar qual é mais rápido para a maioria dos usuários.
47 Avaliando pesquisas de satisfação: Calcular o Q1 da satisfação dos clientes com
48 um produto ou serviço para identificar os pontos de maior insatisfação.
49
50 Vantagens do Primeiro Quartil (Q1):
51
52 Simplicidade de cálculo: O Q1 pode ser facilmente calculado manualmente ou com
53 ferramentas de análise de dados.
54 Robustez: O Q1 não é afetado por outliers, tornando-o mais confiável em alguns
55 casos que a média.
56 Interpretabilidade: O Q1 é um valor fácil de interpretar, representando o ponto
57 de corte dos 25% menores valores.
58
59 Limitações do Primeiro Quartil (Q1):
60
61 Perda de informação: O Q1 não fornece informações sobre a dispersão dos dados,
62 apenas sobre o ponto de corte dos 25% menores valores.
63 Menos preciso que a média em alguns casos: Quando os dados são normalmente
64 distribuídos, a média é mais precisa que o Q1.
65
66 Conclusão:
67
68 O primeiro quartil é uma ferramenta valiosa para o analista de dados, fornecendo
```

```
69 informações sobre a distribuição dos dados e ajudando a identificar outliers.
70 Ao utilizá-lo em conjunto com outras medidas estatísticas, como a mediana e o
71 terceiro quartil, é possível obter uma visão completa e precisa dos dados,
72 embasando decisões estratégicas e otimizando resultados.
73
74 Lembre-se:
75
76 O Q1 é uma ferramenta complementar à média, mediana e moda, fornecendo
77 diferentes perspectivas sobre os dados. Utilize o Q1 em conjunto com outras
78 medidas para obter uma visão completa dos dados. Esteja atento às suas
79 limitações, como a perda de informação e a menor precisão em alguns casos.
80 '''
81
82 # Exemplo: Calculando Q1 das notas de uma amostra de 5 alunos
83
84 notas = [5,6,7,8,9]
85 q1 = notas[int((round(len(notas)+1,0)/4))]
86 print(f'Q1 calculado manualmente é {q1:.2f}')
87
88 # calculando a Q1 utilizando o Numpy
89 import numpy as np
90 q1np = np.quantile(notas,0.25)
91 print(f'Q1 calculado pelo numpy é {q1np:.2f}')
92
93 # calculando Q1 utilizando o statistics
94 import statistics as st
95 q1st = st.quantiles(notas, n=4)[0]
96 print(f'''Q1 calculado pelo statistics é {q1st:.2f} nesse caso o
97 resultado diferente não representa um erro, e sim uma fórmula diferente, mas que
98 conceitualmente também está correta''')
```

Q1 calculado manualmente é 6.00  
Q1 calculado pelo numpy é 6.00  
Q1 calculado pelo statistics é 5.50 nesse caso o  
resultado diferente não representa um erro, e sim uma fórmula diferente, mas que  
conceitualmente também está correta  
''

1 # 3.2 - 2º Quartil - Q2  
2  
3 '''  
4 O que é o Segundo Quartil (Q2)?  
5  
6 Em termos simples, o Q2 é o valor que separa os 50% menores valores dos 50%  
7 maiores em um conjunto de dados ordenado. Imagine uma fila de alunos ordenados  
8 por altura: o Q2 seria a altura do aluno que está na posição 50% da fila,  
9 dividindo os alunos mais baixos dos mais altos.  
10  
11 Cálculo do Segundo Quartil (Q2):  
12  
13 O cálculo do Q2 é intuitivo:  
14  
15 Ordene os dados em ordem crescente.  
16 Calcule a posição do Q2:  
17 O Q2 é sempre a posição  $(n + 1) / 2$ , arredondando para o número inteiro mais  
18 próximo.  
19 Encontre o valor na posição calculada: este é o valor do Q2.  
20 Aplicações do Segundo Quartil (Q2) na Análise de Dados:  
21  
22 Comparação de conjuntos de dados: Permite comparar a distribuição dos dados em  
23 diferentes grupos, como o Q2 da renda de duas cidades ou o Q2 da idade de dois  
24 times.  
25 Identificação de outliers: Valores muito abaixo ou acima do Q2 podem indicar  
26 outliers, que merecem atenção especial.  
27 Análise de dados assimétricos: O Q2 é mais robusto que a média para lidar com  
28 dados que não possuem uma distribuição normal (simétrica).  
29 Avaliação de pesquisas de opinião: O Q2 pode ser utilizado em pesquisas para  
30 determinar a opinião da maioria sobre um determinado tema.  
31 Exemplos práticos:  
32  
33 Analisando salários: Calcular o Q2 dos salários para ter uma ideia do salário  
34 "mediano" em uma profissão.  
35 Comparando tempo de resposta: Comparar o Q2 do tempo de resposta de dois  
36 sistemas para verificar qual é mais rápido para a maioria dos usuários.  
37 Avaliando pesquisas de satisfação: Calcular o Q2 da satisfação dos clientes com  
38 um produto ou serviço para identificar os pontos de maior e menor satisfação.  
39 Vantagens do Segundo Quartil (Q2):  
40  
41 Simplicidade de cálculo: O Q2 pode ser facilmente calculado manualmente ou com  
42 ferramentas de análise de dados.  
43 Robustez: O Q2 não é afetado por outliers, tornando-o mais confiável em alguns  
44 casos que a média.  
45 Interpretabilidade: O Q2 é um valor fácil de interpretar, representando o ponto  
46 de corte dos 50% menores valores.  
47 Limitações do Segundo Quartil (Q2):  
48  
49 Perda de informação: O Q2 não fornece informações sobre a dispersão dos dados,  
50 apenas sobre o ponto de corte dos 50% menores valores.  
51 Menos preciso que a média em alguns casos: Quando os dados são normalmente  
52 distribuídos, a média é mais precisa que o Q2.  
53  
54 Conclusão:  
55  
56 O segundo quartil é uma ferramenta valiosa para o analista de dados, fornecendo  
57 informações sobre a distribuição dos dados e ajudando a identificar outliers.  
58 Ao utilizá-lo em conjunto com outras medidas estatísticas, como a mediana e o  
59 terceiro quartil, é possível obter uma visão completa e precisa dos dados,  
60 embasando decisões estratégicas e otimizando resultados.  
61  
62 Lembre-se:  
63  
64 O Q2 é uma ferramenta complementar à média, mediana, moda e primeiro quartil,  
65 fornecendo diferentes perspectivas sobre os dados.  
66 Utilize o Q2 em conjunto com outras medidas para obter uma visão completa dos  
67 dados. Esteja atento às suas limitações, como a perda de informação e a menor  
68 precisão em alguns casos.

```
69 '''
70 # Exemplo: Calculando Q2 das notas de uma amostra de 5 alunos
71
72 notas = [5,6,7,8,9]
73 q2 = notas[int((round(len(notas)+1,0)/2))-1]
74 print(f'Q2 calculado manualmente é {q2:.2f}')
75
76 # calculando a Q2 utilizando o Numpy
77 import numpy as np
78 q2np = np.quantile(notas,0.5)
79 print(f'Q2 calculado pelo numpy é {q2np:.2f}')
80
81 # calculando Q2 utilizando o statistics
82 import statistics as st
83 q2st = st.quantiles(notas, n=4)[1]
84 print(f'''Q2 calculado pelo statistics é {q2st:.2f}''')

    Q2 calculado manualmente é 7.00
    Q2 calculado pelo numpy é 7.00
    Q2 calculado pelo statistics é 7.00
```

1 # 3.3 - 3º Quartil - Q3  
2  
3 '''  
4 O que é o Terceiro Quartil (Q3)?  
5  
6 Em termos simples, o Q3 é o valor que separa os 75% menores valores dos 25%  
7 maiores em um conjunto de dados ordenado. Imagine uma fila de alunos ordenados  
8 por altura: o Q3 seria a altura do aluno que está na posição 75% da fila,  
9 dividindo os alunos mais baixos dos mais altos.  
10  
11 Cálculo do Terceiro Quartil (Q3):  
12  
13 O cálculo do Q3 é intuitivo:  
14  
15 Ordene os dados em ordem crescente.  
16 Calcule a posição do Q3:  
17 Se o número de dados for ímpar, a posição do Q3 é  $(3n + 1) / 4$ , arredondando  
18 para o número inteiro mais próximo.  
19 Se o número de dados for par, a posição do Q3 é a média das posições  
20  $(3n + 2) / 4$  e  $(3n + 4) / 4$ , arredondando para o número inteiro mais próximo.  
21 Encontre o valor na posição calculada: este é o valor do Q3.  
22  
23 Aplicações do Terceiro Quartil (Q3) na Análise de Dados:  
24  
25 Comparação de conjuntos de dados: Permite comparar a distribuição dos dados em  
26 diferentes grupos, como o Q3 da renda de duas cidades ou o Q3 da idade de dois  
27 times.  
28 Identificação de outliers: Valores muito abaixo do Q1 ou muito acima do Q3  
29 podem indicar outliers, que merecem atenção especial.  
30 Análise de dados assimétricos: O Q3 é mais robusto que a média para lidar com  
31 dados que não possuem uma distribuição normal (simétrica).  
32 Avaliação de pesquisas de opinião: O Q3 pode ser utilizado em pesquisas para  
33 determinar a opinião da maioria sobre um determinado tema.  
34  
35 Exemplos práticos:  
36  
37 Analisando salários: Calcular o Q3 dos salários para ter uma ideia do salário  
38 "máximo" em uma profissão.  
39 Comparando tempo de resposta: Comparar o Q3 do tempo de resposta de dois  
40 sistemas para verificar qual é mais rápido para a maioria dos usuários.  
41 Avaliando pesquisas de satisfação: Calcular o Q3 da satisfação dos clientes com  
42 um produto ou serviço para identificar os pontos de maior e menor satisfação.  
43 Vantagens do Terceiro Quartil (Q3):  
44  
45 Simplicidade de cálculo: O Q3 pode ser facilmente calculado manualmente ou com  
46 ferramentas de análise de dados.  
47 Robustez: O Q3 não é afetado por outliers, tornando-o mais confiável em alguns  
48 casos que a média.  
49 Interpretabilidade: O Q3 é um valor fácil de interpretar, representando o ponto  
50 de corte dos 75% menores valores.  
51  
52 Limitações do Terceiro Quartil (Q3):  
53  
54 Perda de informação: O Q3 não fornece informações sobre a dispersão dos dados,  
55 apenas sobre o ponto de corte dos 75% menores valores.  
56 Menos preciso que a média em alguns casos: Quando os dados são normalmente  
57 distribuídos, a média é mais precisa que o Q3.  
58  
59 Conclusão:  
60  
61 O terceiro quartil é uma ferramenta valiosa para o analista de dados, fornecendo  
62 informações sobre a distribuição dos dados e ajudando a identificar outliers.  
63 Ao utilizá-lo em conjunto com outras medidas estatísticas, como a mediana e o  
64 primeiro quartil, é possível obter uma visão completa e precisa dos dados,  
65 embasando decisões estratégicas e otimizando resultados.  
66 Lembre-se:  
67  
68 O Q3 é uma ferramenta complementar à média, mediana, moda e primeiro quartil,

69 fornecendo diferentes perspectivas sobre os dados. Utilize o Q3 em conjunto com

1 # 3.4 Valor Mínimo

2 '''

3 O que é o Valor Mínimo?

4

5 Em termos simples, o valor mínimo é o menor valor presente em um conjunto de

6 dados. Imagine um conjunto de notas de uma prova: o valor mínimo seria a menor

7 nota obtida por um aluno.

8

9 Cálculo do Valor Mínimo:

10

11 O cálculo do valor mínimo é intuitivo:

12

13 Ordene os dados em ordem crescente.

14 O primeiro valor da lista ordenada é o valor mínimo.

15 Aplicações do Valor Mínimo na Análise de Dados:

16

17 Comparação de conjuntos de dados: Permite comparar a amplitude dos dados em

18 diferentes grupos, como o valor mínimo da temperatura em duas cidades ou o valor

19 mínimo da idade em dois times.

20 Identificação de outliers: Valores muito abaixo do valor mínimo podem indicar

21 outliers, que merecem atenção especial.

22 Análise de dados assimétricos: O valor mínimo é útil para lidar com dados que

23 não possuem uma distribuição normal (simétrica).

24 Avaliação de desempenho: O valor mínimo pode ser utilizado para avaliar o

25 desempenho mínimo aceitável em diferentes áreas, como o valor mínimo de vendas

26 para um vendedor ou o valor mínimo de pontos para ser aprovado em um exame.

27

28 Exemplos práticos:

29

30 Analisando custos: Calcular o valor mínimo dos custos de produção para

31 identificar áreas de otimização.

32 Comparando tempo de resposta: Comparar o valor mínimo do tempo de resposta de

33 dois sistemas para verificar qual é o mais rápido.

34 Avaliando pesquisas de satisfação: Calcular o valor mínimo da satisfação dos

35 clientes com um produto ou serviço para identificar os pontos de maior

36 insatisfação.

37 Vantagens do Valor Mínimo:

38

39 Simplicidade de cálculo: O valor mínimo pode ser facilmente calculado

40 manualmente ou com ferramentas de análise de dados.

41 Interpretabilidade: O valor mínimo é um valor fácil de interpretar,

42 representando o menor valor presente no conjunto de dados.

43 Robustez: O valor mínimo não é afetado por outliers.

44

45 Limitações do Valor Mínimo:

46

47 Perda de informação: O valor mínimo não fornece informações sobre a dispersão

48 dos dados, apenas sobre o menor valor presente.

49 Menos informativo que outras medidas em alguns casos: Em alguns casos, outras

50 medidas como a média ou a mediana podem ser mais informativas que o valor mínimo