**UNIVERSIDADE POSITIVO**

**DATA SCIENCE**

**Atividade avaliativa prática**

**Parte 1: Análise de Posição e Dispersão**

1. Calcule a média das vendas ao longo dos 4 anos e discuta como ela representa o valor central das vendas.

R: A média das vendas ao longo dos 4 anos é de 14,717948717948717 unidades monetárias. Isso significa que, em média, as vendas foram de 14,717948717948717 unidades monetárias por mês. A média representa o valor central das vendas, pois é o valor que divide os dados em duas partes iguais. No caso específico do conjunto de dados fornecido, 50% das vendas estão acima da média e 50% das vendas estão abaixo da média.

1. Calcule a mediana das vendas para cada ano e descreva como ela difere da média.

R: Mediana de 2024: 15.0, Mediana de 2025: 15.3, Mediana de 2026: 15.7, Mediana de 2027: 12.3. A mediana das vendas para cada ano é diferente da média. A mediana é o valor que divide os dados em duas partes iguais, de modo que 50% dos dados estão acima da mediana e 50% dos dados estão abaixo da mediana. No caso específico do conjunto de dados fornecido, a mediana é menor do que a média para o ano de 2027. Isso significa que, para esse ano, a metade das vendas foi menor do que a média e a metade das vendas foi maior do que a média. Para os anos de 2024, 2025 e 2026, a mediana é maior do que a média. Isso significa que, para esses anos, a metade das vendas foi maior do que a média e a metade das vendas foi menor do que a média.

1. Calcule a variância e o desvio padrão das vendas de todos os meses. Interprete o significado dessas medidas em relação à dispersão dos dados.

R: A variância é uma medida da dispersão dos dados em torno da média. Quanto maior a variância, mais dispersos os dados estão. O desvio padrão é a raiz quadrada da variância. Com base nos resultados fornecidos, a variância das vendas de todos os meses é de 5,40249835634451. Isso significa que os dados estão relativamente dispersos. O desvio padrão é de 2,3243275062573496. Isso significa que, em média, os valores estão a cerca de 2,32 unidades de distância da média. Uma interpretação possível desses resultados é que as vendas de todos os meses variam consideravelmente. Em alguns meses, as vendas são muito altas, enquanto em outros meses, as vendas são muito baixas. Essa dispersão pode ser causada por uma variedade de fatores, como condições econômicas, eventos climáticos e mudanças na estratégia de marketing.

1. Determine o mês com a maior venda e o mês com a menor venda. Compare esses valores com a média para identificar os meses atípicos.

R: O mês com a maior venda é dezembro de 2024, com 18 unidades monetárias. O mês com a menor venda é janeiro de 2024, com 10,5 unidades monetárias.

-Dezembro de 2024: é um mês atípico, pois as vendas são muito maiores do que a média.

-Janeiro de 2024: é um mês atípico, pois as vendas são muito menores do que a média.

**Parte 2: Análise de Quartis**

1. Divida os dados de vendas em quatro quartis. Identifique os valores que definem cada quartil.

R: q1 = np.percentile(dados, 25) = Q1: 12.55

q2 = np.percentile(dados, 50) = Q2: 15.2

q3 = np.percentile(dados, 75) = Q3: 16.4

iqr = q3 - q1 = Intervalo interquantil(iqr): 3.849999999999998

1. Calcule os intervalos interquartil (Quadrantes) para as vendas e explique seu significado.

R: IQR = 3.849999999999998. O intervalo interquartil de uma distribuição contínua pode ser calculado integrando a função de densidade de probabilidade, que produz a função de distribuição cumulativa (FDC). O quartil inferior Q 1 é um número tal que a integral da função de distribuição cumulativa de − ∞ a Q 1 é igual a 0,25, enquanto que o quartil superior Q 3 é um número tal que a integral de − ∞ a Q 3 é igual a 0,75.

**Parte 3: Previsão (Forecasting)**

1. Use um método de análise de tendência, para criar uma previsão das vendas para o próximo ano (2028) com base nos dados anteriores.

R: Previsão para 2028 é de 14.76689608636977 vendas.

1. Determine o erro médio e o erro percentual médio da sua análise de tendência.

R: Erro médio: 0.04894736842105114 e Erro percentual médio: 3.0433772593930914

1. Discuta as limitações do método utilizado e sugira como ele pode ser adotado em um cenário real.

R: O método utilizado para calcular a tendência das vendas é uma regressão linear simples. Esse método é simples e fácil de implementar, mas tem algumas limitações. Uma limitação do método é que ele assume que a tendência é linear. Isso pode não ser verdade em todos os casos. Por exemplo, se a tendência for exponencial ou quadrática, o método não será preciso. Outra limitação do método é que ele é sensível a outliers. Um outlier é um valor que está muito distante do resto dos dados. Se um outlier estiver presente no conjunto de dados, ele pode afetar significativamente os resultados do método. Para adotar o método em um cenário real, é importante considerar essas limitações. Para garantir que o método seja preciso, é importante verificar se a tendência é realmente linear. Também é importante remover outliers do conjunto de dados, se possível. Aqui estão algumas sugestões para melhorar a precisão do método:

-Utilizar um método de regressão mais complexo, como uma regressão não linear.

-Remover outliers do conjunto de dados.

-Utilizar um período de treinamento mais longo.

4. Baseado na análise de tendência, faça uma previsão das vendas para os próximos cinco anos (2028 a 2032) e explique as suposições subjacentes.

R: Com base na análise de tendência, a previsão das vendas para os próximos cinco anos é a seguinte:

Ano | Vendas

2028 14.76689608636977

2029 14.81584317274956

2030 14.86478925911935

2031 14.91373434548914

2032 14.96267843185893

Essa previsão é baseada na suposição de que a tendência das vendas continuará a ser linear nos próximos cinco anos. Aqui estão algumas suposições subjacentes à previsão:

-O ambiente econômico não sofrerá mudanças significativas.

-A empresa não lançará novos produtos ou serviços que possam afetar as vendas.

-A empresa não enfrentará competição significativa de novos entrantes.

É importante ressaltar que essa previsão é apenas uma estimativa. As vendas reais podem variar significativamente, dependendo de fatores externos.