

Análise Inferencial de Dados

Entrega 1

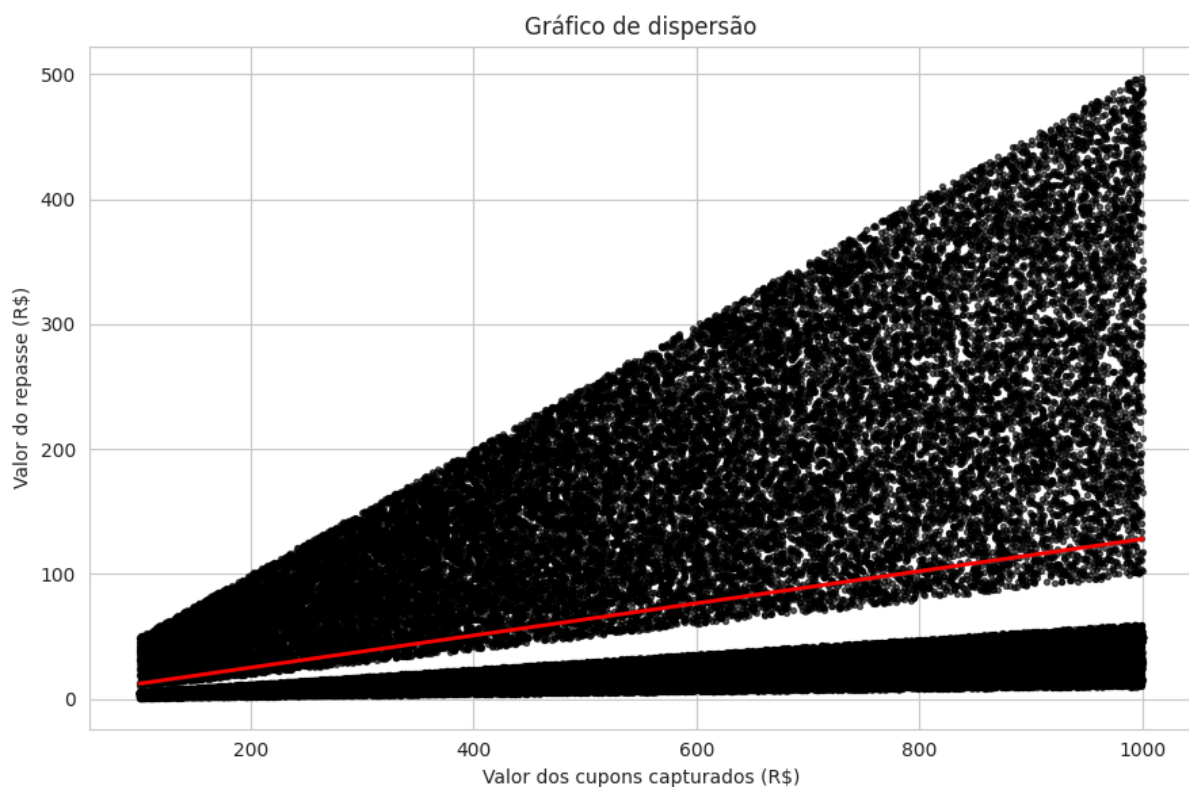
Introdução

Este estudo tem como propósito investigar a relação entre o valor dos cupons capturados (R\$) e o valor do repasse (R\$). A análise foi realizada utilizando a base de dados "PicMoney-Base_de_Transações-CuponsCapturados-10000 linhas.csv". Para isso, foram empregados os seguintes métodos: a construção de um gráfico de dispersão com uma linha de tendência de regressão linear simples, o cálculo do coeficiente de correlação de Pearson (r) e a obtenção do coeficiente de determinação (R^2).

Gráfico de dispersão com a linha de tendência

O gráfico a seguir apresenta a relação entre os valores observados de cupom e repasse, juntamente com a linha de tendência (reta de regressão linear), que foi ajustada pelo método dos mínimos quadrados.

- **Eixo X:** valor dos cupons capturados (R\$).
- **Eixo Y:** valor do repasse (R\$).
- **Linha de tendência:** representa a tendência média da relação entre as duas variáveis.



- **Coeficiente de Correlação de Pearson (r):** 0,3677.
- **Coeficiente de Determinação (R^2):** 0,1352

Interpretação de dados

Gráfico de dispersão:

O gráfico de dispersão ilustra a relação entre o valor do cupom e o valor do repasse. A reta de regressão, que sintetiza a tendência central dos dados, exibe uma inclinação positiva. Isso indica visualmente que, à medida que o valor do cupom aumenta, há uma tendência de crescimento no valor do repasse.

Equação da Regressão Linear

A relação entre as variáveis pode ser descrita pela seguinte equação da reta:

$$y = a + b \cdot x$$

A relação linear é descrita pela equação:

$$y = -0,36 + 0,1286 \cdot x$$

Onde:

- **Intercepto (a = - 0,36):** Representa o valor estimado do repasse quando o valor do cupom (x) é zero .
- **Inclinação (b = 0,1286):** Indica que, para cada R\$ 1,00 adicional no valor do cupom, o valor do repasse aumenta, em média, R\$ 0,13.

Correlação de Pearson

Valor de R

O coeficiente de **correlação(r)** foi calculado entre **valor_cupom** e **repasse_picmoney**:

$$r = 0,3677$$

- **r > 0:** correlação positiva (quanto maior o valor do cupom, maior o repasse).
- **r ≈ 0,3677:** positiva, porém fraca.

Esse valor mostra uma **correlação positiva**, porém **fraca**. Ou seja, existe tendência de aumento do repasse com o aumento do valor do cupom, mas a relação não é forte. O teste de significância: $p < 0,001$ mostra que este resultado é estatisticamente significativo.

Coeficiente de Determinação

Valor de R²

$$r^2 = 0,1352$$

Este valor indica que apenas 13,52% da variação no valor do repasse pode ser explicada pela variação no valor do cupom. Isso demonstra que o modelo possui um baixo poder explicativo, pois a maior parte da variabilidade do repasse (86,48%) é influenciada por outros fatores não incluídos na análise.