

Análise Discriminante em Detecção de Risco de Crédito

Data de entrega: 26 de fevereiro de 2026

1 Contexto

Uma instituição financeira deseja desenvolver um sistema de apoio à decisão para identificar clientes com alto risco de inadimplência a partir de informações comportamentais e financeiras recentes.

Cada cliente foi classificado historicamente em duas populações:

- **Adimplente:** clientes que mantêm pagamentos regulares
- **Risco:** clientes que apresentaram atraso ou inadimplência relevante

O objetivo da equipe de análise é construir e avaliar funções discriminantes capazes de separar essas populações e apoiar o processo de análise de crédito.

2 Base de dados

```
'data.frame': 500 obs. of 7 variables:
 $ valor_limite      : num  15896 20105 36154 21086 22349 ...
 $ dias_atraso_medio : num  4.99 6.11 8.07 5.38 10.96 ...
 $ qt_atrasos_12m   : int   0 2 1 0 1 0 0 0 0 0 ...
 $ ratio_divida_renda : num  0.174 0.271 0.321 0.278 0.359 0.339 0.341 0.429 0.394 0.197 ..
```

```

$ dist_agencia_resid : num  7.18 15.35 5.57 19.23 6.56 ...
$ mudancas_endereco_6m: num  0 0 0 1 0 1 0 1 0 0 ...
$ classe              : Factor w/ 2 levels "Adimplente","Risco": 1 2 1 1 1 1 1 1 1 1 ...

```

```

      valor_limite   dias_atraso_medio qt_atrasos_12m  ratio_divida_renda
Min.   : 5123      Min.   : 0.430      Min.   :0.000      Min.   :0.0210
1st Qu.:13590     1st Qu.: 4.810      1st Qu.:0.000     1st Qu.:0.1958
Median :18562     Median : 7.665      Median :0.000     Median :0.2920
Mean   :20660     Mean   : 9.440      Mean   :0.652     Mean   :0.3064
3rd Qu.:25155     3rd Qu.:12.652     3rd Qu.:1.000     3rd Qu.:0.4010
Max.   :70722     Max.   :39.990     Max.   :5.000     Max.   :0.7450
dist_agencia_resid mudancas_endereco_6m      classe
Min.   : 0.540      Min.   :0.000      Adimplente:400
1st Qu.: 7.305     1st Qu.:0.000      Risco      :100
Median :10.730     Median :0.000
Mean   :12.426     Mean   :0.352
3rd Qu.:16.465     3rd Qu.:1.000
Max.   :41.980     Max.   :3.000

```

```

Adimplente      Risco
      0.8        0.2

```

A base de dados contém as seguintes variáveis:

- `valor_limite`: limite total de crédito disponível
- `dias_atraso_medio`: média de dias de atraso nos últimos 12 meses
- `qt_atrasos_12m`: número de atrasos registrados
- `ratio_divida_renda`: razão dívida/renda mensal
- `dist_agencia_resid`: distância entre residência e agência mais próxima
- `mudancas_endereco_6m`: mudanças de endereço recentes
- `classe`: população (Adimplente ou Risco)

3 Etapas obrigatórias da análise

O aluno deverá realizar uma análise discriminante completa para classificar clientes em duas populações (Adimplente *vs* Risco), reproduzindo todas as etapas de um estudo aplicado. Em síntese, espera-se que o aluno:

- Explorar os dados: analisar estrutura, distribuições, boxplots, relações bivariadas e matrizes de correlação por grupo.
- Verificar suposições: aplicar o teste de Box (Box's M) e comparar as matrizes de covariância das populações.
- Estimar modelos discriminantes:
 - Função Discriminante Linear (LDA)
 - Função Discriminante Quadrática (QDA)
- Visualizar e interpretar:
 - projeção na função discriminante
 - separação entre os grupos
 - importância das variáveis.
- Avaliar desempenho preditivo usando três abordagens:
 - método da ressubstituição,
 - divisão treino/teste,
 - validação cruzada.
- Comparar LDA e QDA, discutindo qual modelo é mais adequado ao problema.
- Produzir interpretação final, relacionando resultados estatísticos com o contexto aplicado de risco de crédito.