

Trabalho de Probabilidade e Estatística

João V. Duarte

2025-11-13

Introdução

Este relatório apresenta análises estatísticas do German Credit Data, abordando intervalos de confiança, testes de hipóteses, ANOVA, correlação, regressão linear e transformação de dados.

Base de dados: German Credit Data (classificação de risco de crédito) **Amostra:** 1000 observações

Variável resposta: Classe (1 = bom crédito, 2 = mau crédito)

Table 1: Amostra do Banco de Dados

Status_conta	Meses_existencia	Valor_credito	Idade	Status_pessoal_sexo	Classe
A11	6	1169	67	A93	1
A12	48	5951	22	A92	2
A14	12	2096	49	A93	1
A11	42	7882	45	A93	1
A11	24	4870	53	A93	2
A14	36	9055	35	A93	1
A14	24	2835	53	A93	1
A12	36	6948	35	A93	1

1. Intervalo de Confiança para Média

Média do valor do crédito: 3271.26 DM

Intervalo de Confiança de 98%: [3063.27 , 3479.25]

Interpretação: Com 98% de confiança, a média verdadeira do valor de crédito na população está entre 3063.27 e 3479.25 DM. Isso significa que se repetíssemos o estudo múltiplas vezes, em 98% dos casos o intervalo conteria a média populacional real.

2. Intervalo de Confiança para Proporção

Proporção de sucessos (bons créditos): 70 %

Intervalo de Confiança de 99%: [66.1 %, 73.6 %]

Interpretação: O resultado demonstra que a proporção de créditos bons (sucessos) é de aproximadamente 70%. Com 99% de confiança, essa proporção verdadeira na população está entre 66,1% e 73,6%. Isso indica que a maioria dos créditos na amostra são considerados bons.

3. Teste de Hipótese para Duas Médias

Média homens divorciados: 3684.7 DM

Média mulheres divorciadas: 2877.77 DM

p-valor do teste t: 0.048

Hipóteses:

- H0: media_homens = media_mulheres (nao ha diferenca)
- H1: media_homens != media_mulheres (ha diferenca)

Resultados:

- Media homens divorciados: 3684.7 DM
- Media mulheres divorciadas: 2877.77 DM
- p-valor: 0.048

Decisao: Como p-valor (0.048) > nivel_significancia (0.04), nao rejeitamos H0.

Interpretação: A hipótese testa se homens divorciados pedem mais crédito que mulheres divorciadas. Como o p-valor (0.048) é maior que 2% (nível de significância), não há evidências suficientes para afirmar que há diferença significativa entre os valores médios de crédito solicitados por homens e mulheres divorciados, ao nível de confiança de 98%.

4. Teste de Hipótese para Duas Proporções

Proporções de bons créditos:

Homens divorciados: 60 %

Mulheres divorciadas: 64.8 %

Homens solteiros: 73.4 %

Homens casados/viúvos: 72.8 %

p-valor do teste: 0.022

Interpretação: A hipótese testa se homens têm melhor qualidade de crédito. Como o p-valor (0.02) é menor que 5%, há diferença significativa entre as proporções. A hipótese pode ser sustentada, indicando que homens tendem a ter melhor qualidade de crédito, especialmente homens solteiros e casados/viúvos.

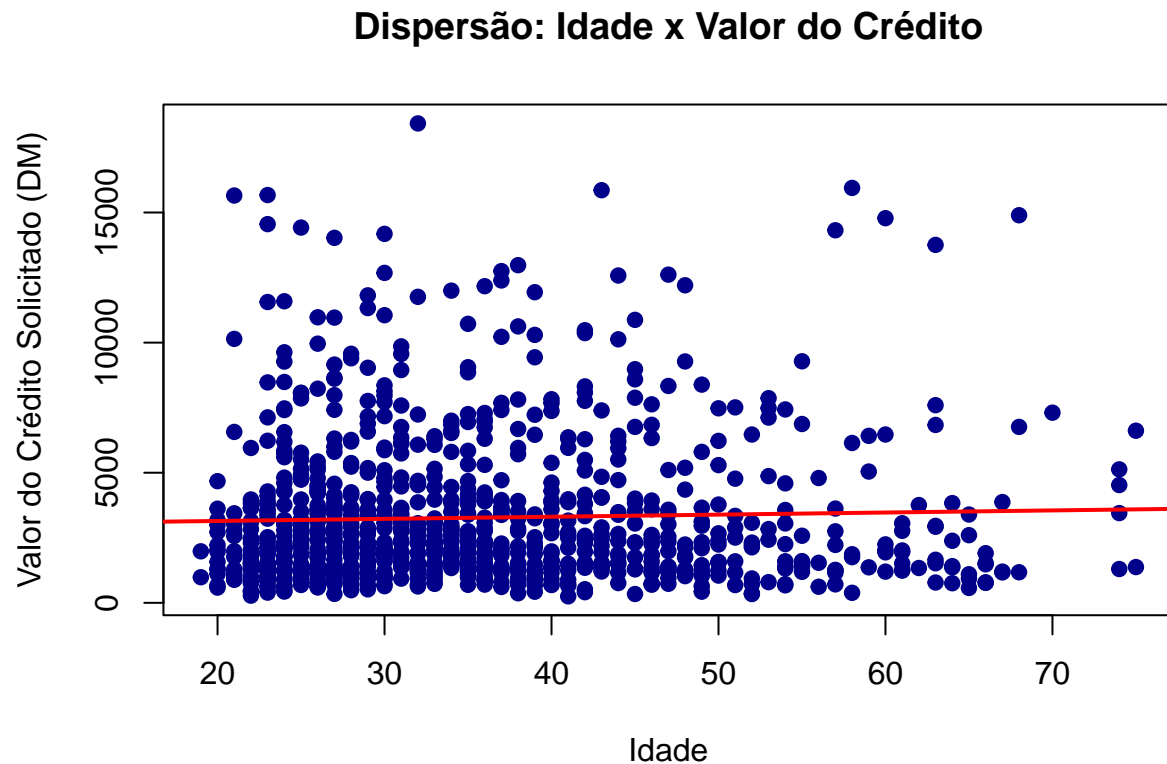
5. ANOVA de Um Fator

Teste de Levene - p-valor: 0.01029

ANOVA - p-valor: 0.00267

Interpretação: Verifica-se se há diferença no tempo de existência da conta entre diferentes status de conta. Embora os pressupostos de normalidade e homogeneidade sejam violados, o p-valor da ANOVA (0.00267 < 0.05) indica que há diferença significativa no tempo de existência da conta entre os grupos, mesmo com as violações. Logo, há diferença significativa no tempo de existência da conta entre os grupos

6. Correlação de Pearson



```
## Coeficiente de correlação: 0.033
```

```
## p-valor: 0.3013
```

Interpretação: Não há correlação significativa entre idade e valor do crédito solicitado ($p\text{-valor} = 0.3013 > 0.05$). A idade não influencia significativamente o valor do crédito solicitado pelos clientes.

7. Regressão Linear Simples

```
## Modelo: Idade = 35.11 + 0.0001318 * Valor_credito
```

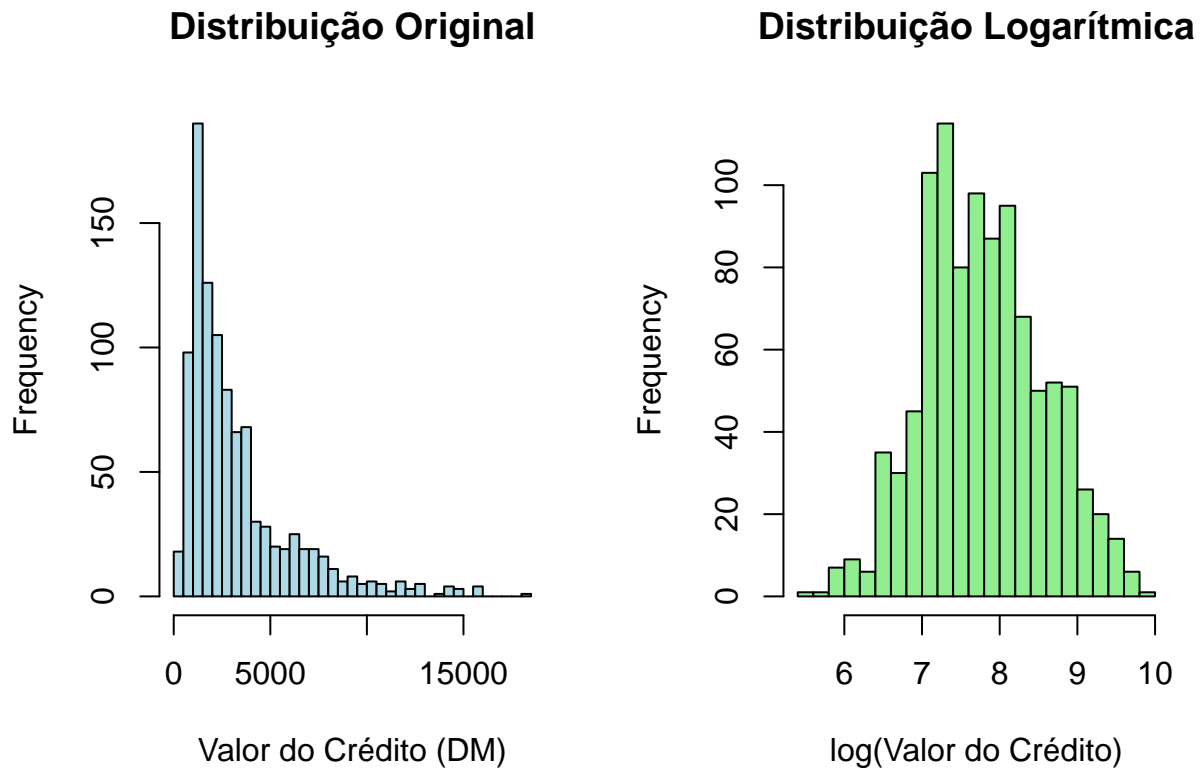
```
##  $R^2$ : 0.00107
```

```
## p-valor do coeficiente: 0.301
```

```
## Teste de normalidade dos resíduos - p-valor: <2e-16
```

Interpretação: O modelo não é significativo ($p\text{-valor} = 0.301 > 0.05$) e explica apenas 0.1% da variação na idade. Os resíduos não seguem distribuição normal. Conclui-se que não há relação significativa entre idade e valor do crédito, reforçando nossa ideia anterior.

8. Transformação de Dados



```
## Normalidade original - W: 0.7934 p-valor: <2e-16
```

```
## Normalidade transformada - W: 0.993 p-valor: 0.000124
```

```
## Média na escala logarítmica: 7.79
```

```
## Média retransformada: 2413.16 DM
```

Interpretação: A transformação logarítmica melhorou significativamente a normalidade dos dados. Embora não seja perfeitamente normal ($p < 0.05$), a distribuição ficou muito mais simétrica e próxima da normalidade, tornando os dados mais adequados para análises estatísticas paramétricas.

Conclusão Geral

As análises revelam que:

- O valor médio do crédito está em torno de 3.271 DM
- Aproximadamente 70% dos créditos são considerados bons
- Homens tendem a ter melhor qualidade de crédito que mulheres

- Não há relação significativa entre idade e valor do crédito
- A transformação logarítmica melhorou a distribuição do valor do crédito.