



Universidade Federal do Maranhão

Especialização em Análise de Dados e Inteligência Artificial

Atividade III: Análise Comparativa das Taxas Educacionais entre Redes de Ensino no Brasil: Públicas versus Privadas

Disciplina: Inferência Estatística

Professora: Alcione Miranda dos Santos

Autores:

Ernandes Guedes Moura

Hedley Lima Cunha

Poliane Santos Macedo Carvalho

Sirléia Lucena Rodrigues

Fortaleza dos Nogueiras - MA

24 de junho de 2024

Conteúdo

1	Introdução	2
2	Metodologia	2
2.1	Entendimento do Negócio	2
2.2	Entendimento dos Dados	3
2.2.1	2.2 Entendimento dos Dados	3
3	Preparação dos Dados	3
3.1	Verificar dados ausentes	4
4	Modelagem	5
4.1	Análise Descritiva	5
4.1.1	Análise por Rede	7
4.2	Análise gráfica	8
5	Definição das Hipóteses	14
6	Teste de Hipóteses	15
7	Discussão dos Resultados	17
7.1	Taxas de Aprovação	17
7.2	Taxas de Reprovação	17
7.3	Taxas de Abandono	17
8	Conclusão	18
9	Referências	18

1 Introdução

A educação é um dos pilares fundamentais para o desenvolvimento individual e social, sendo essencial para a formação de cidadãos e para o progresso econômico de um país. No Brasil, há um debate contínuo e acalorado sobre as diferenças no desempenho escolar entre alunos de escolas públicas e privadas¹². Estudos anteriores têm mostrado resultados variados e muitas vezes contraditórios³. Este relatório busca contribuir para essa discussão através de uma análise rigorosa de dados educacionais e a aplicação de métodos estatísticos robustos para avaliar as diferenças no desempenho escolar entre essas duas redes de ensino.

2 Metodologia

Para realizar a análise comparativa do desempenho escolar entre escolas públicas e privadas, adotamos a metodologia CRISP-DM (Cross-Industry Standard Process for Data Mining), que é uma abordagem amplamente utilizada para projetos de mineração de dados. Esta metodologia proporciona um framework estruturado para entender os dados, prepará-los para análise, aplicar modelos estatísticos e interpretar os resultados⁴.

2.1 Entendimento do Negócio

Nesta etapa, buscamos compreender o contexto e os objetivos do projeto:

- Objetivo: Comparar o desempenho escolar de alunos de escolas públicas e privadas no Brasil.

¹Franco, C., & Menezes-Filho, N. A. (2013). Desigualdade de desempenho entre escolas públicas e privadas no Brasil: evidências a partir dos microdados da Prova Brasil. *Nova Economia*, 23(3), 507-536. Disponível em: Nova Economia.

²Firpo, S. P., & Pieri, R. (2018). Desempenho escolar e escolha de escolas: uma análise de regressão descontínua para o Brasil. *Economia Aplicada*, 22(4), 145-172. Disponível em: Economia Aplicada.

³Zagury, T., & Damasceno, E. (2018). Análise do desempenho escolar na avaliação SAEB 2015. Disponível em: ResearchGate.

⁴Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526-534. Disponível em: ResearchGate.

- Perguntas-chave:
 - Há diferenças significativas nas taxas de aprovação, reprovação e abandono entre escolas públicas e privadas?

2.2 Entendimento dos Dados

2.2.1 2.2 Entendimento dos Dados

Os dados utilizados neste estudo foram obtidos do Base dos Dados (Base dos Dados), que contém diversas informações incluindo taxa de aprovação, reprovação e abandono dos estudantes, categorizadas por tipo de rede de ensino (estadual, federal, municipal, privada, pública e total)⁵.

```
# Carregar os dados
dados = read.csv("brasil.csv", sep=",")
dim(dados) # Dimensão dos dados
```

```
## [1] 306 214
```

3 Preparação dos Dados

A preparação dos dados é uma etapa crucial que envolve limpeza, transformação e seleção de dados relevantes para a análise. Para este estudo, realizamos a seguinte preparação:

- Seleção de Colunas: Selecionamos e renomeamos as colunas de interesse: rede, taxa_aprovacao_em_3_ano, taxa_reprovacao_em_3_ano e taxa_abandono_em_3_ano.

```
library(dplyr)
library(knitr)

# Selecionar e renomear as colunas de interesse
```

⁵Base dos Dados. Desempenho Escolar por Rede de Ensino no Brasil (3º ano do Ensino Fundamental). Rio de Janeiro: Base dos Dados, 2024. Recuperado de: <https://basedosdados.org/dataset/63f1218f-c446-4835-b746-f109a338e3a1?table=cd65b1d2-45e8-432b-afe8-c3a706addbe8>. Acesso em: 22 jun. 2024.

```
df <- dados %>%
  select(rede, taxa_aprovacao_em_3_ano, taxa_reprovacao_em_3_ano,
         taxa_abandono_em_3_ano)

# Visualizar as primeiras linhas do DataFrame resultante
head(df) %>%
  kable()
```

rede	taxa_aprovacao_em_3_ano	taxa_reprovacao_em_3_ano	taxa_abandono_em_3_ano
estadual	83.8	6.0	10.2
federal	92.4	6.8	0.8
municipal	88.1	3.5	8.4
privada	96.3	2.9	0.8
publica	84.6	5.8	9.6
total	85.0	5.7	9.3

3.1 Verificar dados ausentes

```
# Verificar se há NAs
contagem_na <- colSums(is.na(df))
cat("Contagem de NAs por coluna:\n")
```

```
## Contagem de NAs por coluna:
```

```
print(paste(names(contagem_na), ": ", contagem_na))
```

```
## [1] "rede : 0" "taxa_aprovacao_em_3_ano : 72"
```

```
## [3] "taxa_reprovacao_em_3_ano : 72" "taxa_abandono_em_3_ano : 72"
```

4 Modelagem

Nesta fase, aplicamos técnicas estatísticas para analisar os dados e testar hipóteses.

4.1 Análise Descritiva

Inicialmente, realizamos uma estatística descritiva das taxas de aprovação, reprovação e abandono por rede de ensino:

```
library(kableExtra)

# Selecionar e renomear as colunas de interesse
df2 <- dados %>%
  select(
    Rede = rede,
    `Taxa de Aprovação no 3º Ano` = taxa_aprovacao_em_3_ano,
    `Taxa de Reprovação no 3º Ano` = taxa_reprovacao_em_3_ano,
    `Taxa de Abandono no 3º Ano` = taxa_abandono_em_3_ano
  )

# Estatística descritiva (média) de cada taxa por rede de ensino
estatisticas_descritivas <- df2 %>%
  group_by(Rede) %>%
  summarise(
    `Taxa Média de Aprovação` = mean(`Taxa de Aprovação no 3º Ano`, na.rm =
      TRUE),
    `Taxa Média de Reprovação` = mean(`Taxa de Reprovação no 3º Ano`, na.rm =
      TRUE),
    `Taxa Média de Abandono` = mean(`Taxa de Abandono no 3º Ano`, na.rm =
      TRUE)
  )
```

```

# # Visualizar as estatísticas descritivas
# estatisticas_descritivas %>%
#   kable()

# Gerar a tabela com título usando kableExtra
estatisticas_descritivas %>%
  kable() %>%
  kable_styling(full_width = FALSE) %>%
  add_header_above(c(" " = 1, "Médias das Taxas por Rede de Ensino" = 3))

```

Rede	Médias das Taxas por Rede de Ensino		
	Taxa Média de Aprovação	Taxa Média de Reprovação	Taxa Média de Abandono
estadual	87.90513	5.782051	6.3128205
federal	91.59231	6.941026	1.4666667
municipal	91.46923	3.843590	4.6871795
privada	97.24103	1.941026	0.8179487
publica	88.03590	5.810256	6.1538462
total	89.07436	5.353846	5.5717949

A tabela apresentada revela diferenças marcantes entre as redes de ensino. As escolas privadas se destacam com a maior média de aprovação (97.24%), enquanto as escolas públicas apresentam a menor média (88.03%). Essa disparidade é também evidente nas taxas de reprovação e abandono: as escolas privadas têm as menores taxas (1.94% e 0.82%, respectivamente), enquanto as públicas exibem taxas significativamente mais altas (5.81% e 6.15%, respectivamente).

4.1.1 Análise por Rede

- Rede Estadual:

Aprovação: Média de 87.91%, abaixo da média nacional (89.07%).

Reprovação: Taxa de 5.78%, similar à média nacional (5.35%).

Abandono: Taxa de 6.31%, acima da média nacional (5.57%).

- Rede Federal:

Aprovação: Média de 91.59%, a mais alta entre as redes, superando a média nacional em 3.68%.

Reprovação: Taxa de 6.94%, acima da média nacional (5.35%).

Abandono: Taxa de 1.47%, a menor entre as redes, bem abaixo da média nacional (5.57%).

- Rede Municipal:

Aprovação: Média de 91.47%, segunda mais alta entre as redes, superando a média nacional em 2.56%. Reprovação: Taxa de 3.84%, a menor entre as redes, bem abaixo da média nacional (5.35%). Abandono: Taxa de 4.69%, inferior à média nacional (5.57%).

- Rede Privada:

Aprovação: Média de 97.24%, a mais alta entre todas as redes, superando a média nacional em 8.17%. Reprovação: Taxa de 1.94%, a menor entre as redes, bem abaixo da média nacional (5.35%). Abandono: Taxa de 0.82%, a menor entre todas as redes, consideravelmente abaixo da média nacional (5.57%).

- Rede Pública:

Aprovação: Média de 88.03%, a menor entre as redes, abaixo da média nacional em 1.04%. Reprovação: Taxa de 5.81%, similar à média nacional (5.35%). Abandono: Taxa de 6.15%, acima da média nacional (5.57%).

4.2 Análise gráfica

Os gráficos de barras abaixo ilustram as diferenças nas taxas de aprovação, reprovação e abandono no terceiro ano entre diferentes tipos de rede de ensino.

```
# Carregar pacotes
library(ggplot2)

# Gráfico de barras da taxa de aprovação em 3º ano por tipo de rede de ensino

ggplot(df, aes(x = rede, y = taxa_aprovacao_em_3_ano, fill = rede)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Taxa de Aprovação em 3º Ano por Tipo de Rede de Ensino",
       x = "Rede de Ensino",
       y = "Taxa de Aprovação em 3º Ano") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Removed 72 rows containing missing values or values outside the scale rang
## (`geom_bar()`).
```



```
# Gráfico de barras da taxa de reprovação em 3º ano por tipo de rede de ensino

ggplot(df, aes(x = rede, y = taxa_reprovacao_em_3_ano, fill = rede)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Taxa de Reprovação em 3º Ano por Tipo de Rede de Ensino",
       x = "Rede de Ensino",
       y = "Taxa de Reprovação em 3º Ano") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

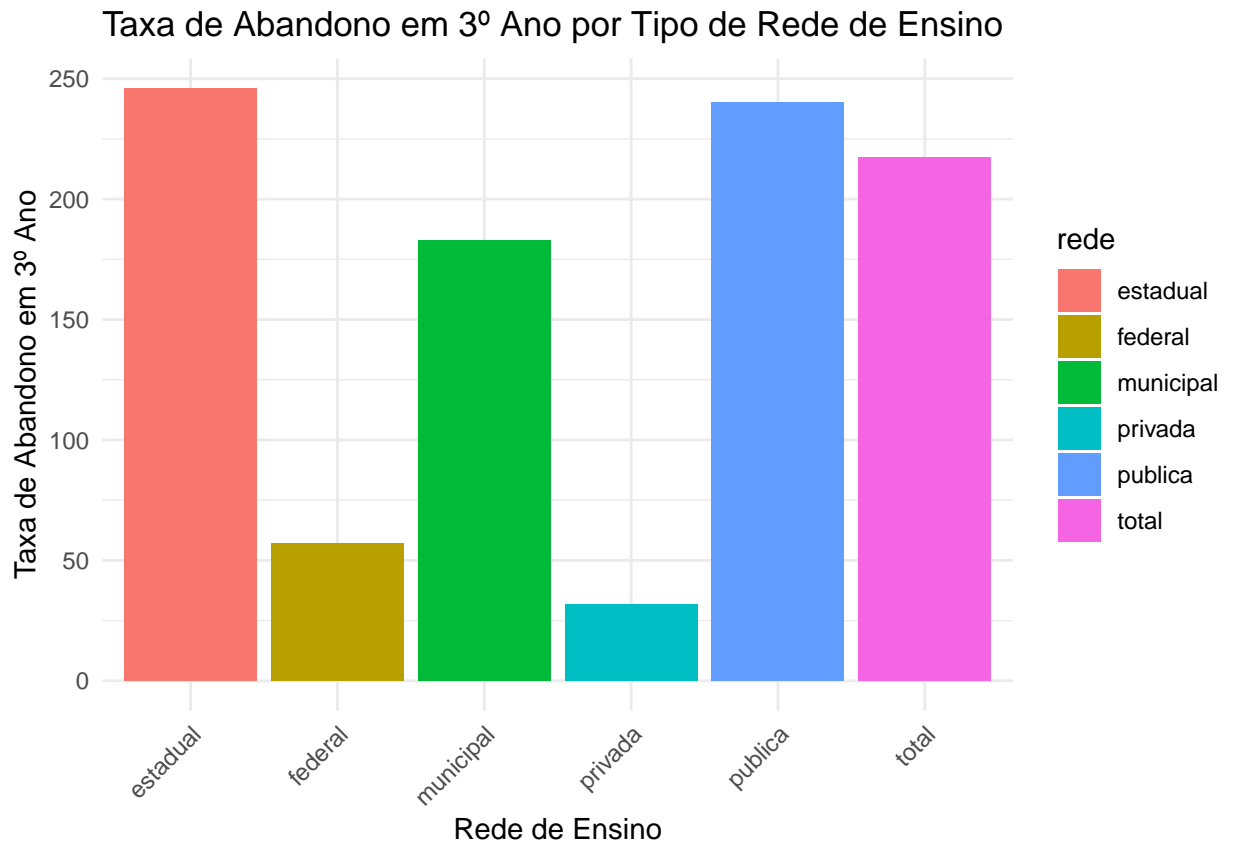
```
## Warning: Removed 72 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```



```
# Gráfico de barras da taxa de abandono em 3º ano por tipo de rede de ensino

ggplot(df, aes(x = rede, y = taxa_abandono_em_3_ano, fill = rede)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Taxa de Abandono em 3º Ano por Tipo de Rede de Ensino",
       x = "Rede de Ensino",
       y = "Taxa de Abandono em 3º Ano") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

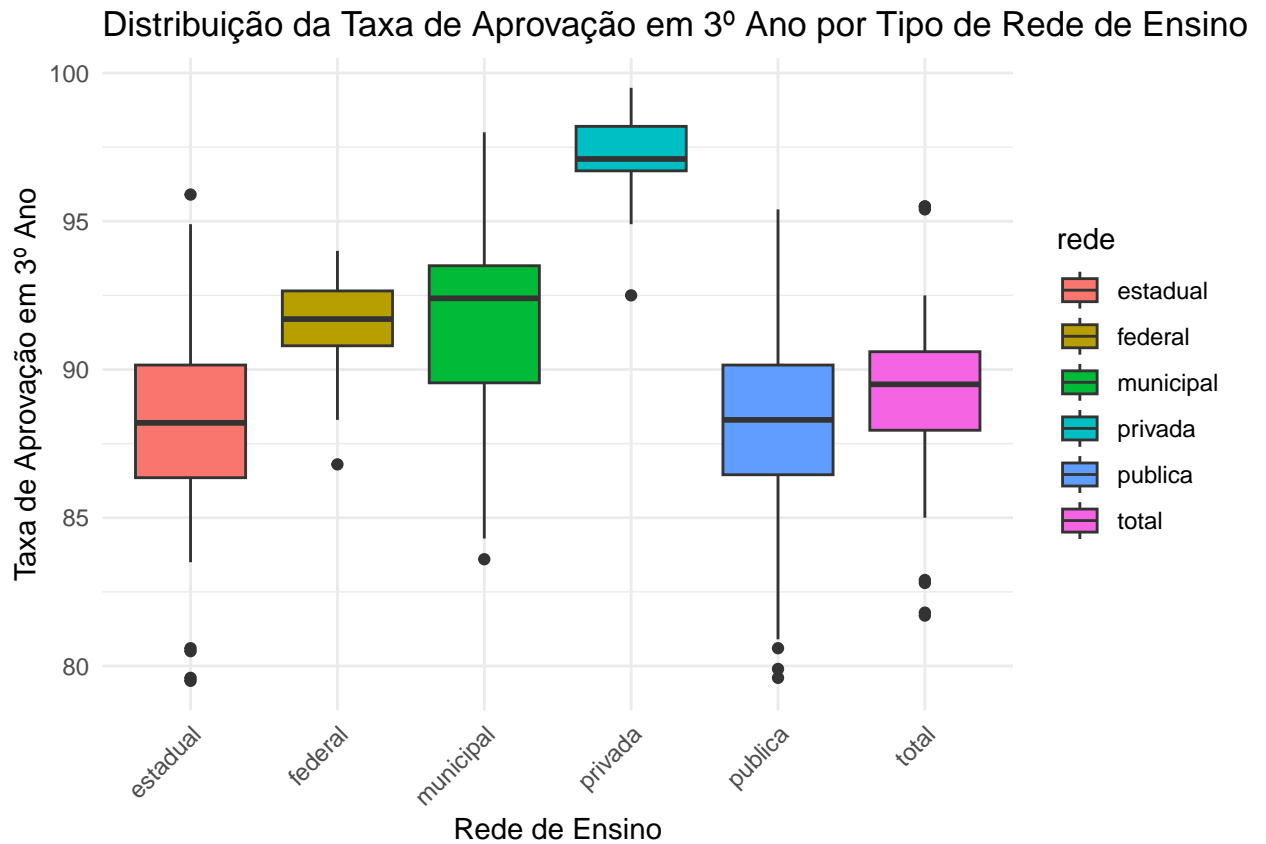
```
## Warning: Removed 72 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```



Além dos gráficos de barras, vamos incluir boxplots para visualizar a distribuição completa dos dados.

```
# Boxplot para análise detalhada
ggplot(df, aes(x = rede, y = taxa_aprovacao_em_3_ano, fill = rede)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Distribuição da Taxa de Aprovação em 3º Ano por Tipo de Rede
    de Ensino",
    x = "Rede de Ensino", y = "Taxa de Aprovação em 3º Ano") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

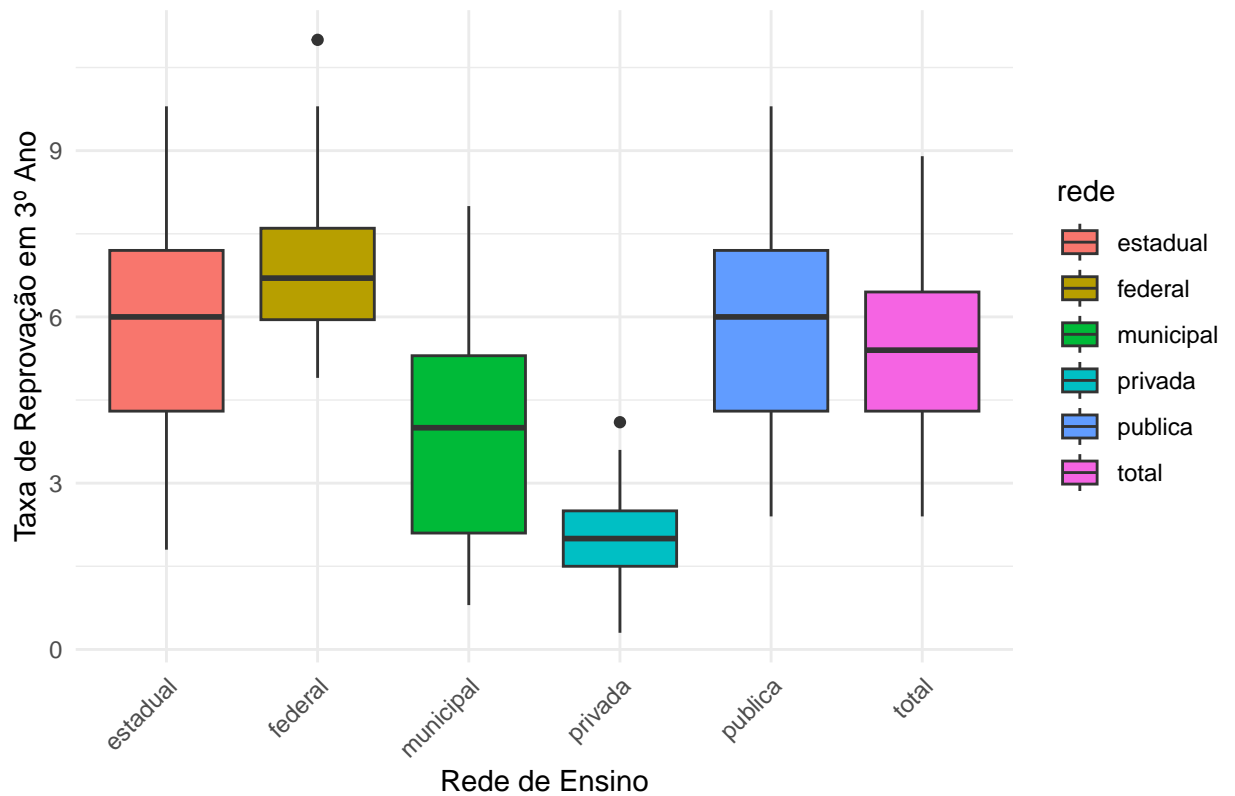
```
## Warning: Removed 72 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



```
ggplot(df, aes(x = rede, y = taxa_reprovacao_em_3_ano, fill = rede)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Distribuição da Taxa de Reprovação em 3º Ano por Tipo de
    Rede de Ensino",
    x = "Rede de Ensino", y = "Taxa de Reprovação em 3º Ano") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

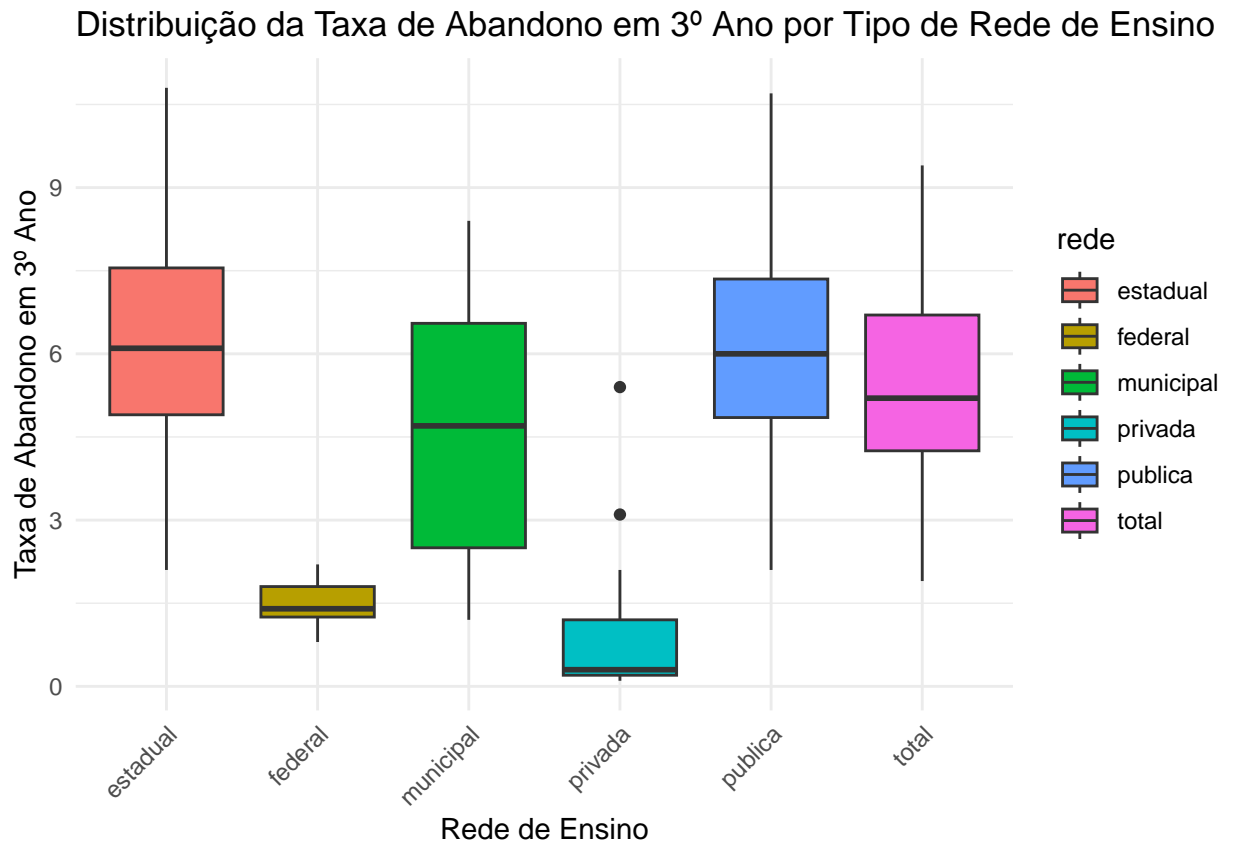
```
## Warning: Removed 72 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```

Distribuição da Taxa de Reprovação em 3º Ano por Tipo de Rede de Ensino



```
ggplot(df, aes(x = rede, y = taxa_abandono_em_3_ano, fill = rede)) +
  geom_boxplot() +
  theme_minimal() +
  labs(title = "Distribuição da Taxa de Abandono em 3º Ano por Tipo de Rede
    de Ensino",
    x = "Rede de Ensino", y = "Taxa de Abandono em 3º Ano") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Warning: Removed 72 rows containing non-finite outside the scale range
## (`stat_boxplot()`).
```



Os gráficos destacam claramente as disparidades nas taxas educacionais entre diferentes tipos de rede de ensino. Escolas privadas e redes federal e municipal mostram-se mais eficazes em manter altas taxas de aprovação e baixas taxas de reprovação e abandono. Por outro lado, as redes públicas estaduais e municipais enfrentam desafios significativos, refletidos em taxas mais baixas de aprovação e mais altas de reprovação e abandono.

5 Definição das Hipóteses

Nesta etapa, definimos as hipóteses a serem testadas:

- Hipótese Nula (H_0): Não há diferença significativa nas taxas de aprovação, reprovação e abandono entre escolas públicas e privadas.
- Hipótese Alternativa (H_1): Existe uma diferença significativa nas taxas de aprovação, reprovação e abandono entre escolas públicas e privadas.

6 Teste de Hipóteses

Realizamos testes estatísticos para verificar a validade das hipóteses definidas. O teste t produziu um valor p (p-value). Para tomada de decisão compara-se o p-value com um nível de significância padrão ($= 0.05$):

- Se $p < 0.05$: Rejeitamos a hipótese nula (H_0), indicando que existe uma diferença significativa entre as médias das taxas de aprovação.
- Se $p \geq 0.05$: Não rejeitamos a hipótese nula, indicando que não há uma diferença significativa entre as médias das taxas de aprovação.

```
# Teste t para comparar a taxa de aprovação entre escolas privadas e  
públicas  
privadas <- df %>% filter(rede == "privada")  
publicas <- df %>% filter(rede == "publica")  
  
test_aprovacao <- t.test(privadas$taxa_aprovacao_em_3_ano,  
                          publicas$taxa_aprovacao_em_3_ano)  
  
# Função para tomada de decisão  
tomada_decisao <- function(teste) {  
  if (teste$p.value < 0.05) {  
    "Rejeitar H0: Diferença significativa entre as médias"  
  } else {  
    "Não rejeitar H0: Não há diferença significativa entre as médias"  
  }  
}  
  
# Aplicar a função
```



```
tomada_decisao(test_aprovacao)
```

```
## [1] "Rejeitar H0: Diferença significativa entre as médias"
```

Com base nos dados fornecidos e no teste t realizado, podemos rejeitar a hipótese nula de que não há diferença na taxa média de aprovação entre escolas privadas e públicas. Há forte evidência de que a diferença verdadeira nas médias não é zero. Isso sugere que a rede de ensino (privada ou pública) influencia significativamente as taxas de aprovação.

```
test_reprovacao<-t.test(privadas$taxa_reprovacao_em_3_ano,
  publicas$taxa_reprovacao_em_3_ano)
# Aplicar a função
tomada_decisao(test_reprovacao)
```

```
## [1] "Rejeitar H0: Diferença significativa entre as médias"
```

O resultado do teste t indica que:

Existe uma diferença significativa entre as médias de reprovação das duas redes. A taxa média de reprovação nas escolas privadas (1.94%) é consideravelmente menor do que nas escolas públicas (5.81%).

Portanto, o valor p extremamente baixo (menor que $2.2e-16$) do teste t indica que essa diferença nas taxas de reprovação provavelmente não ocorreu por acaso.

```
test_abadono<-t.test(privadas$taxa_abandono_em_3_ano,
  publicas$taxa_abandono_em_3_ano)
# Aplicar a função
tomada_decisao(test_abadono)
```

```
## [1] "Rejeitar H0: Diferença significativa entre as médias"
```

Portanto, existe diferença significativa entre as taxas médias de abandono entre as

escolas públicas e privadas.

7 Discussão dos Resultados

A análise comparativa das taxas de aprovação, reprovação e abandono entre escolas públicas e privadas revela diferenças significativas que impactam diretamente o panorama educacional no Brasil.

7.1 Taxas de Aprovação

As escolas privadas se destacam com a maior taxa média de aprovação (97.24%), seguidas pelas redes federal (91.59%) e municipal (91.47%). Em contraste, as redes estadual (87.91%) e pública (88.03%) apresentam taxas mais baixas. Essa disparidade reflete diferentes estratégias pedagógicas, investimentos e condições de ensino entre os setores público e privado.

7.2 Taxas de Reprovação

As escolas privadas também exibem a menor taxa média de reprovação (1.94%), enquanto as redes pública (5.81%) e estadual (5.78%) têm taxas mais altas. A reprovação mais baixa nas escolas privadas sugere uma melhor adaptação aos métodos de ensino ou suporte adicional oferecido aos alunos.

7.3 Taxas de Abandono

O cenário de abandono segue um padrão semelhante, com as escolas privadas apresentando a menor taxa média (0.82%), enquanto as redes pública (6.15%) e estadual (6.31%) enfrentam taxas significativamente mais altas. O abandono escolar pode estar relacionado a diversos fatores, incluindo qualidade de ensino, suporte socioemocional e condições socioeconômicas dos alunos.

8 Conclusão

Os dados analisados corroboram a hipótese de que existem diferenças substanciais no desempenho escolar entre as redes pública e privada no Brasil. As escolas privadas geralmente apresentam melhores resultados em termos de taxas de aprovação, reprovação e abandono, o que pode indicar uma vantagem em recursos, infraestrutura e gestão educacional. Esses resultados sugerem que políticas voltadas para a melhoria das escolas públicas poderiam considerar as práticas bem-sucedidas observadas nas escolas privadas.

9 Referências

Franco, C., & Menezes-Filho, N. A. (2013). Desigualdade de desempenho entre escolas públicas e privadas no Brasil: evidências a partir dos microdados da Prova Brasil. *Nova Economia*, 23(3), 507-536. Disponível em: Nova Economia.

Firpo, S. P., & Pieri, R. (2018). Desempenho escolar e escolha de escolas: uma análise de regressão descontínua para o Brasil. *Economia Aplicada*, 22(4), 145-172. Disponível em: Economia Aplicada.

Zagury, T., & Damasceno, E. (2018). Análise do desempenho escolar na avaliação SAEB 2015. Disponível em: ResearchGate.

Base dos Dados. Desempenho Escolar por Rede de Ensino no Brasil (3º ano do Ensino Fundamental). Rio de Janeiro: Base dos Dados, 2024. Recuperado de: <https://basedosdados.org/dataset/63f1218f-c446-4835-b746-f109a338e3a1?table=cd65b1d2-45e8-432b-afe8-c3a706addbe8>. Acesso em: 22 jun. 2024.

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526-534. Disponível em: ResearchGate.