

Relatório

Handrey h. Izayner z.

Neste relatório, geramos uma amostra com 1.000 observações e a submetemos aos testes de Mann-Whitney e t de Student, com o objetivo de compará-los em diferentes situações, utilizando distribuições simétricas e assimétricas. Será possível observar, ao longo do relatório, como diferentes parâmetros influenciam o desempenho de cada teste. Após essa etapa, os testes são aplicados a um banco de dados real, retirado do Kaggle, permitindo analisar como eles se comportam em um cenário prático.

Mann-Whitney vs. Teste t de Student em Dados Simulados

Dados simétricos

Comparação dos dados simulados $\mu=50$ no grupo A e $\mu=55$ no grupo B e ambos com $sd=10$



OBSERVAÇÕES:

- O teste t é o mais apropriado, pois os dados atendem às suposições paramétricas.
- O Mann-Whitney teve desempenho semelhante, pois embora não use a média diretamente, ele é consistente e também captura a diferença.
- Resultado esperado: os dois testes funcionam bem quando os dados são normais e o efeito é forte ($\mu = 5$).
- O teste de Mann-Whitney é mais robusto nesse cenário, pois não assume normalidade.

O teste como a menor diferença de médias

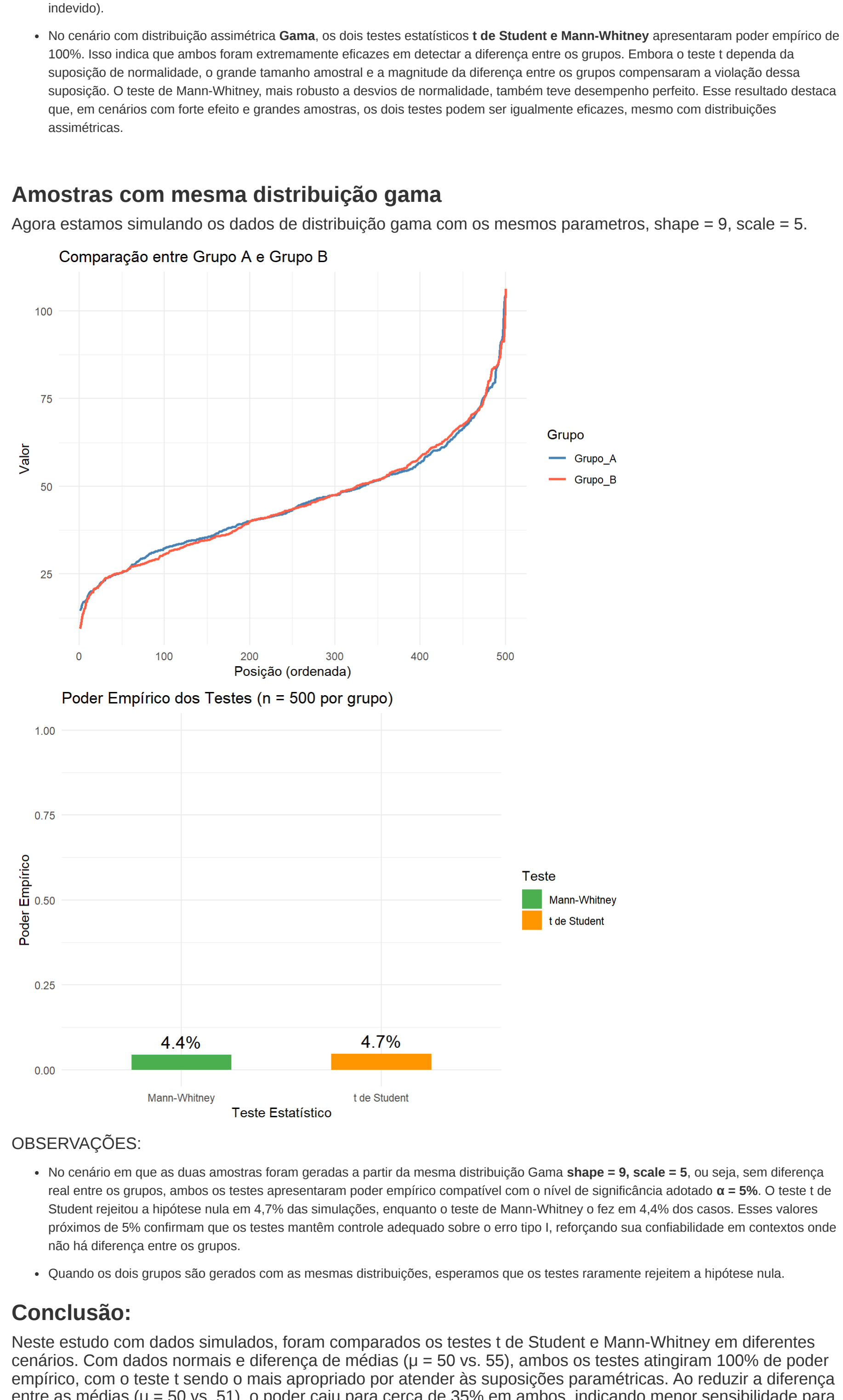
Agora aplicando o teste para $\mu=50$ no grupo A e $\mu=51$ Grupo B. Assim, podemos observar os grupos como a menor diferença nas médias



OBSERVAÇÕES:

- podemos observar que com as médias mais aproximadas obtivemos menor diferença entre os grupos.
- O poder empírico dos testes continuou muito semelhantes, isso se deve ao tamanho dos grupos, pois os testes se ajustam melhor em grupos menores.
- Um poder de 35% significa que, mesmo havendo uma diferença real entre os grupos, o teste só detecta essa diferença em cerca de 1/3 das simulações.
- Isso não significa que o teste está errado, mas que nas condições simuladas, a diferença entre os grupos não foi "grande o suficiente" para ser detectada com confiança em 95% dos casos.

Dados simulados com assimetria



OBSERVAÇÕES:

- O teste t ainda teve alto poder, mas aqui ele viola a suposição de normalidade.
- O teste de Mann-Whitney é mais robusto nesse cenário, pois não assume normalidade.
- Quando os dados são assimétricos, o teste de Mann-Whitney mantém o poder e ainda é mais confiável (menor risco de erro tipo I indevido).
- No cenário com distribuição assimétrica, **Gama**, os dois testes estatísticos t de Student e Mann-Whitney apresentaram poder empírico de 100%. Isso indica que ambos foram extremamente eficazes em detectar a diferença entre os grupos. Embora o teste t dependa da suposição de normalidade, o grande tamanho amostral e a magnitude da diferença entre os grupos compensaram a violação dessa suposição. O teste de Mann-Whitney, mais robusto a desvios de normalidade, também teve desempenho perfeito. Esse resultado destaca que, em cenários com forte efeito e grandes amostras, os dois testes podem ser igualmente eficazes, mesmo com distribuições assimétricas.

Amostras com mesma distribuição gama

Agora estamos simulando os dados de distribuição gama com os mesmos parâmetros, shape = 9, scale = 5.



OBSERVAÇÕES:

- No cenário em que as duas amostras foram geradas a partir da mesma distribuição Gama **shape = 9, scale = 5**, ou seja, sem diferença real entre os grupos, ambos os testes apresentaram poder empírico compatível com o nível de significância adotado $\alpha = 5\%$. O teste t de Student rejeitou a hipótese nula em 4,7% das simulações, enquanto o teste de Mann-Whitney o fez em 4,4% dos casos. Esses valores próximos de 5% confirmam que os testes mantêm controle adequado sobre o erro tipo I, reforçando sua confiabilidade em contextos onde não há diferença entre os grupos.
- Quando os dois grupos são gerados com as mesmas distribuições, esperamos que os testes raramente rejeitem a hipótese nula.

Conclusão:

Neste estudo com dados simulados, foram comparados os testes t de Student e Mann-Whitney em cenários com dados normais e diferença de médias ($\mu = 50$ vs. 55), ambos os testes atingiram 100% de poder empírico, com o teste t sendo o mais apropriado por atender às suposições paramétricas. Ao reduzir a diferença entre as médias ($\mu = 50$ vs. 51), o poder caiu para cerca de 35% em ambos, indicando menor sensibilidade para detectar efeitos pequenos. Já com distribuições assimétricas Gama (μ diferentes), o poder permaneceu alto (100%), evidenciando a robustez do teste de Mann-Whitney, mesmo sem suposição de normalidade. Por fim, com dados Gama idênticos (sem diferença real), os testes rejeitaram a hipótese nula em cerca de 4,4%–4,7%, dentro do esperado pelo nível de significância ($\alpha = 5\%$). Assim, ambos os testes mostraram-se confiáveis, sendo o t de Student mais sensível com normalidade e o Mann-Whitney mais robusto em cenários assimétricos.

Mann-Whitney vs. Teste t de Student em Dados Reais

Esses dados tirados do kaggle são da bolsa de valores da Amazon de 1997-2025.



Conclusão:

Embora tanto o teste t de Student quanto o teste de Mann-Whitney tenham identificado uma diferença significativa entre os grupos antes e depois da crise, é importante considerar a distribuição dos dados. Os testes de Shapiro-Wilk indicaram uma forte violação da normalidade em ambos os grupos ($p < 0.05$), o que compromete a validade do teste t, que assume normalidade. Nesse cenário, o teste de Mann-Whitney é mais apropriado, por ser não paramétrico e robusto a essas violações. Assim, embora os dois testes levem à mesma conclusão, o resultado do Mann-Whitney deve ser considerado mais confiável para este conjunto de dados reais.

Caso se interesse breve resumo da Amazon:

A Amazon começou como uma livraria online e, até 2008, era uma empresa promissora, mas ainda pequena no mercado. Com a crise financeira daquele ano, suas ações caíram junto com o resto da bolsa. No entanto, foi após a crise que a empresa deu um salto: investiu pesado em logística, expandiu serviços como o Prime e criou a AWS, sua divisão de computação em nuvem – que se tornou altamente lucrativa. A valorização foi tão grande que, em 2022, a Amazon fez um desdobramento (split) de ações de 20:1. Assim, o que era uma empresa ariscada antes da crise virou uma das maiores potências do mundo depois dela.

CODIGOS:

Teste com simulação simétrica

```
library(tidyverse)
library(ggplot2)

set.seed(42)

# Parâmetros
n <- 500 # tamanho de cada grupo
n_sim <- 1000 # número de simulações
alpha <- 0.05 # nível de significância

rej_t <- 0
rej_mw <- 0

for (i in 1:n_sim) {
  grupo_1 <- rnorm(n, mean = 50, sd = 10)
  grupo_2 <- rnorm(n, mean = 55, sd = 10)

  df_base <- data.frame(
    grupo = c(rep("Grupo_A", n), rep("Grupo_B", n)),
    valor = c(grupo_1, grupo_2)
  )

  # Teste t de Student
  p_t <- t.test(valor ~ grupo, data = df_base, var.equal = TRUE)$p.value

  # Teste de Mann-Whitney
  p_mw <- wilcox.test(valor ~ grupo, data = df_base)$p.value

  if (p_t < alpha) rej_t <- rej_t + 1
  if (p_mw < alpha) rej_mw <- rej_mw + 1
}

df_plot <- data.frame(
  Posição = 1:n,
  Grupo_A = sort(grupo_1),
  Grupo_B = sort(grupo_2)
)

df_long <- pivot_longer(df_plot, cols = c("Grupo_A", "Grupo_B"),
  names_to = "Grupo", values_to = "Valor")

ggplot(df_long, aes(x = Posição, y = Valor, color = Grupo)) +
  geom_line(linetype = "solid") +
  labs(title = "Comparação entre Grupo A e Grupo B",
    x = "Posição (ordenada)",
    y = "Valor",
    color = "Grupo") +
  theme_minimal() +
  scale_color_manual(values = c("steelblue", "tomato"))

poder_t <- rej_t / n_sim
poder_mw <- rej_mw / n_sim

cat("Teste t de Student: ", round(poder_t * 100, 2), "%\n")
cat("Teste de Mann-Whitney: ", round(poder_mw * 100, 2), "%\n")

library(ggplot2)

df_plot <- data.frame(
  Teste = c("t de Student", "Mann-Whitney"),
  Poder = c(poder_t, poder_mw)
)

ggplot(df_plot, aes(x = Teste, y = Poder, fill = Teste)) +
  geom_bar(stat = "identity", width = 0.5) +
  ylim(0, 1) +
  geom_text(aes(label = paste0(round(Poder * 100, 1), "%")),
    vjust = -0.5, size = 5) +
  labs(title = "Poder Empírico dos Testes (n = 500 por grupo)",
    x = "Teste Estatístico",
    y = "Poder Empírico") +
  scale_fill_manual(values = c("#4682B4", "#FF6347")) +
  theme_minimal()
```

Teste com simulação assimétrica

```
library(tidyverse)
library(ggplot2)

set.seed(43)

# Parâmetros
n <- 500 # tamanho de cada grupo
n_sim <- 1000 # número de simulações
alpha <- 0.05 # nível de significância

rej_t <- 0
rej_mw <- 0

for (i in 1:n_sim) {
  grupo_1 <- rgamma(n, shape = 9, scale = 5)
  grupo_2 <- rgamma(n, shape = 9, scale = 6)

  df_base <- data.frame(
    grupo = c(rep("Grupo_A", n), rep("Grupo_B", n)),
    valor = c(grupo_1, grupo_2)
  )

  # Teste t de Student
  p_t <- t.test(valor ~ grupo, data = df_base, var.equal = TRUE)$p.value

  # Teste de Mann-Whitney
  p_mw <- wilcox.test(valor ~ grupo, data = df_base)$p.value

  if (p_t < alpha) rej_t <- rej_t + 1
  if (p_mw < alpha) rej_mw <- rej_mw + 1
}

df_plot <- data.frame(
  Posição = 1:n,
  Grupo_A = sort(grupo_1),
  Grupo_B = sort(grupo_2)
)

df_long <- pivot_longer(df_plot, cols = c("Grupo_A", "Grupo_B"),
  names_to = "Grupo", values_to = "Valor")

ggplot(df_long, aes(x = Posição, y = Valor, color = Grupo)) +
  geom_line(linetype = "solid") +
  labs(title = "Comparação entre Grupo A e Grupo B",
    x = "Posição (ordenada)",
    y = "Valor",
    color = "Grupo") +
  theme_minimal() +
  scale_color_manual(values = c("steelblue", "tomato"))

poder_t <- rej_t / n_sim
poder_mw <- rej_mw / n_sim

cat("Teste t de Student: ", round(poder_t * 100, 2), "%\n")
cat("Teste de Mann-Whitney: ", round(poder_mw * 100, 2), "%\n")

library(ggplot2)

df_plot <- data.frame(
  Teste = c("t de Student", "Mann-Whitney"),
  Poder = c(poder_t, poder_mw)
)

ggplot(df_plot, aes(x = Teste, y = Poder, fill = Teste)) +
  geom_bar(stat = "identity", width = 0.5) +
  ylim(0, 1) +
  geom_text(aes(label = paste0(round(Poder * 100, 1), "%")),
    vjust = -0.5, size = 5) +
  labs(title = "Poder Empírico dos Testes (n = 500 por grupo)",
    x = "Teste Estatístico",
    y = "Poder Empírico") +
  scale_fill_manual(values = c("#4682B4", "#FF6347")) +
  theme_minimal()
```

Teste com dados reais

```
dados <- read.csv("Amazon_stock_data.csv")

dados$Date <- as.Date(dados$Date)

# Separar os grupos antes e depois da crise de 2008
grupo_antes <- subset(dados, Date < as.Date("2008-09-15"))$Close
grupo_depois <- subset(dados, Date >= as.Date("2008-09-15"))$Close

shapiro_antes <- shapiro.test(grupo_antes)
shapiro_depois <- shapiro.test(grupo_depois)

cat("# Shapiro-Wilk - Antes da crise:\n"); print(shapiro_antes)
cat("# Shapiro-Wilk - Depois da crise:\n"); print(shapiro_depois)

# Teste t de Student (paramétrico)
teste_t <- t.test(grupo_antes, grupo_depois, var.equal = FALSE)
cat("#\nTeste t de Student:\n")
print(teste_t)

# Teste de Mann-Whitney (não paramétrico)
teste_mw <- wilcox.test(grupo_antes, grupo_depois)
cat("#\nTeste de Mann-Whitney:\n")
print(teste_mw)

boxplot(grupo_antes, grupo_depois,
  names = c("Antes da crise", "Depois da crise"),
  col = c("lightblue", "red"),
  main = "Preço de Fechamento (ajustado) - Antes e Depois da Crise",
  ylab = "Preço ajustado por ação (USD)")

mtext("Dados ajustados pelo split de 20:1 (jun/2022)", side = 3, line = 0.5, cex = 0.8)

cat("#\n--- Interpretação ---\n")

if (teste_t$p.value < 0.05) {
  cat("✅ O teste t indica diferença significativa entre os grupos.\n")
} else {
  cat("❌ O teste t não indica diferença significativa entre os grupos.\n")
}

if (teste_mw$p.value < 0.05) {
  cat("✅ O teste de Mann-Whitney indica diferença significativa entre os grupos.\n")
} else {
  cat("❌ O teste de Mann-Whitney não indica diferença significativa entre os grupos.\n")
}
```