

Teste de Wilcoxon-Mann-Whitney

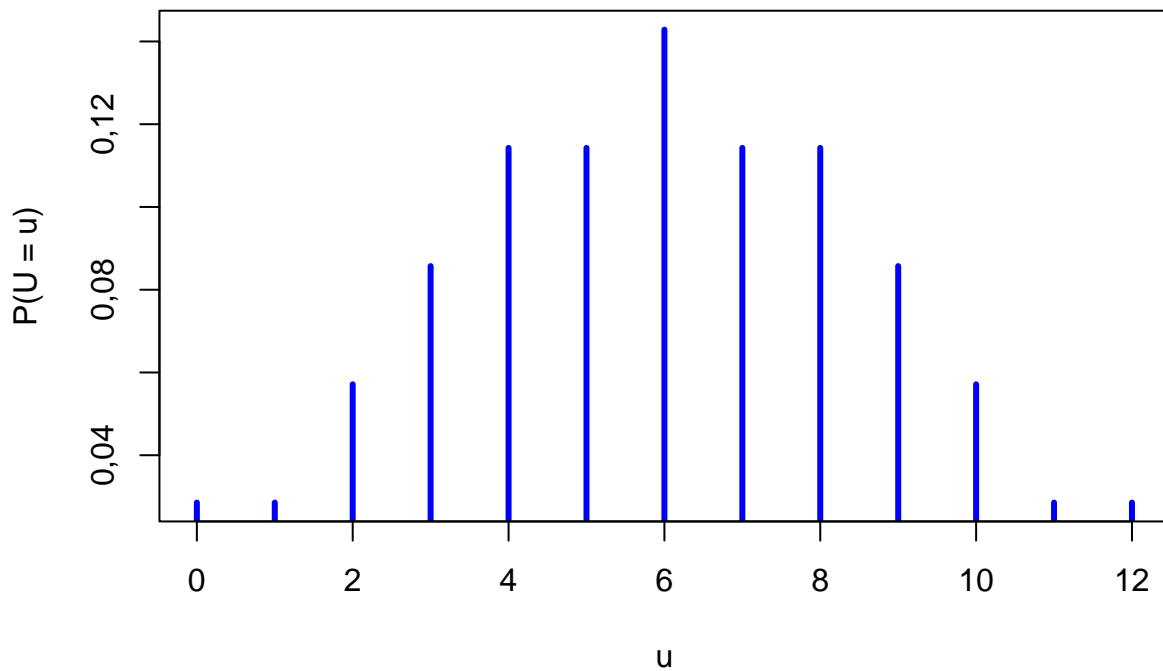
Linguagem R

2023

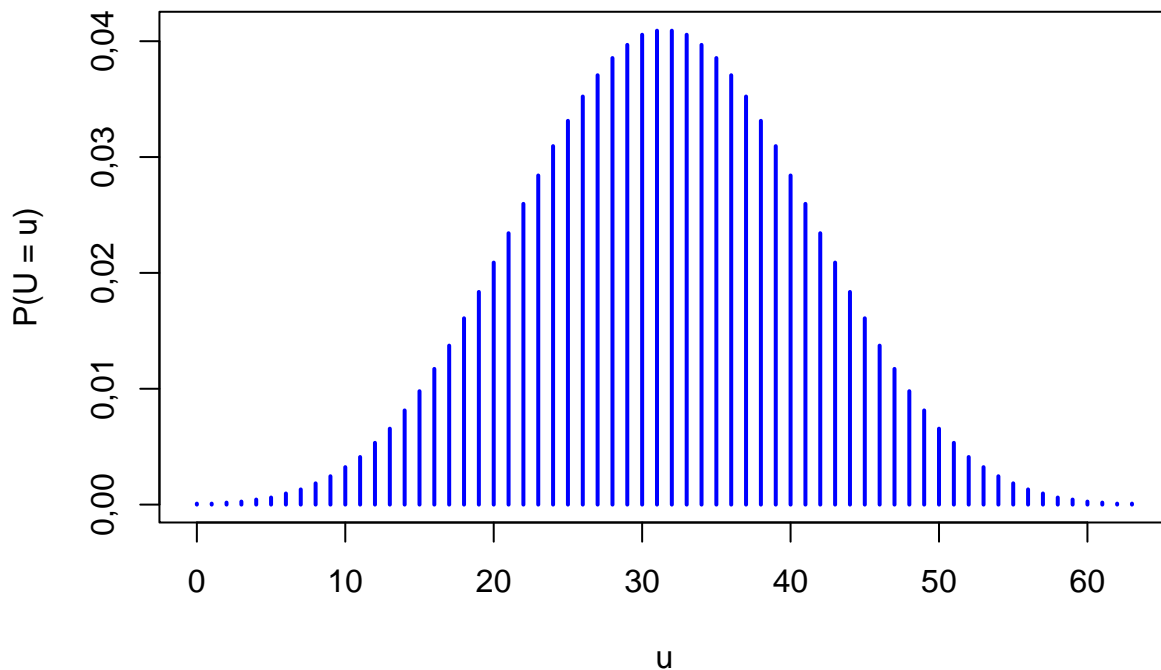
```
# Separador decimal: ","  
options(OutDec = ",")
```

Distribuição exata de U

```
m <- 3  
n <- 4  
faixa <- 0:(m * n)  
d0 <- dwilcox(faixa, m, n)  
plot(faixa, d0, type = "h", xlab = "u", ylab = "P(U = u)", lwd = 3,  
      col = "blue")
```



```
m <- 7  
n <- 9  
faixa <- 0:(m * n)  
d0 <- dwilcox(faixa, m, n)  
plot(faixa, d0, type = "h", xlab = "u", ylab = "P(U = u)", lwd = 2,  
      col = "blue")
```



Exemplo

```
## Dados
## Exercício 5.18, p. 194 em Sprent & Smeeton (3rd ed, 2001)
## Variação percentual no teor de açúcar no sangue em um intervalo de 1h
## Dois níveis de dose de uma droga são comparados
## Deve ser testado se há diferença na resposta entre os dois níveis.

dosei <- c(0.21, -16.2, -10.1, -8.67, -11.13, 1.96, -10.19, -15.87, -12.81)
doseii <- c(1.59, 2.66, -6.27, -2.32, -10.87, 7.23, -3.76, 3.02, 15.01)

m <- length(dosei)
n <- length(doseii)
cat("\ Tamanhos amostrais: m =", m, ", n =", n)

## Tamanhos amostrais: m = 9 , n = 9

summary(dosei)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -16,20 -12,81  -10,19   -9,20   -8,67    1,96

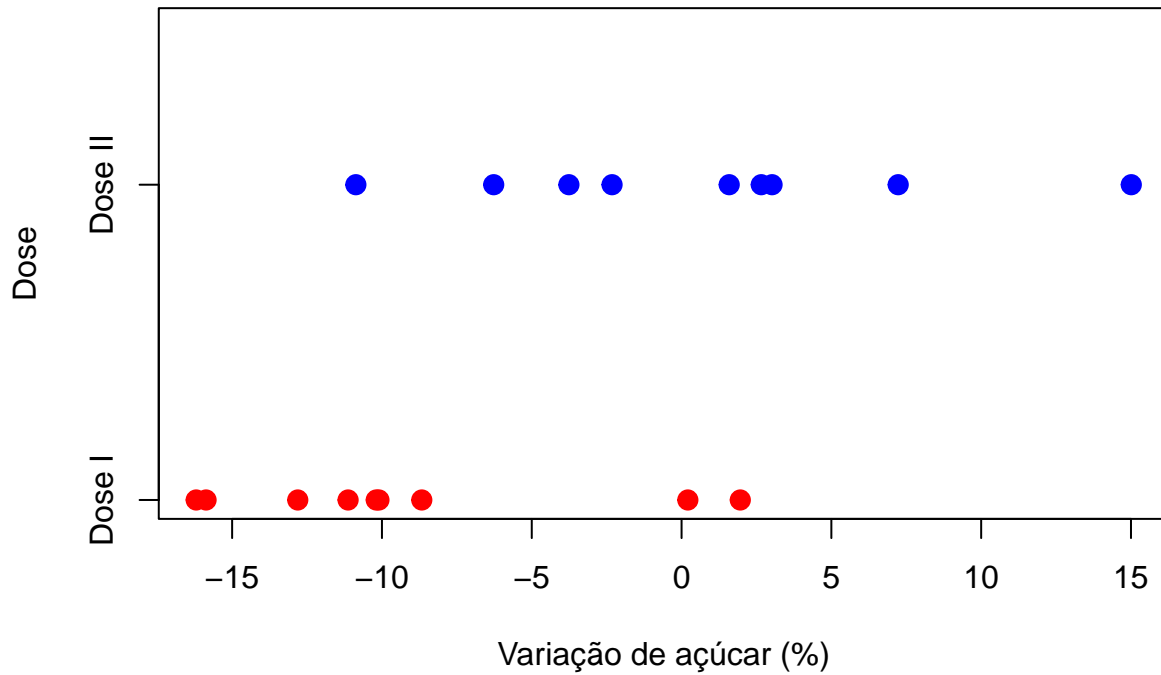
summary(doseii)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -10,8700 -3,7600  1,5900   0,6989  3,0200  15,0100

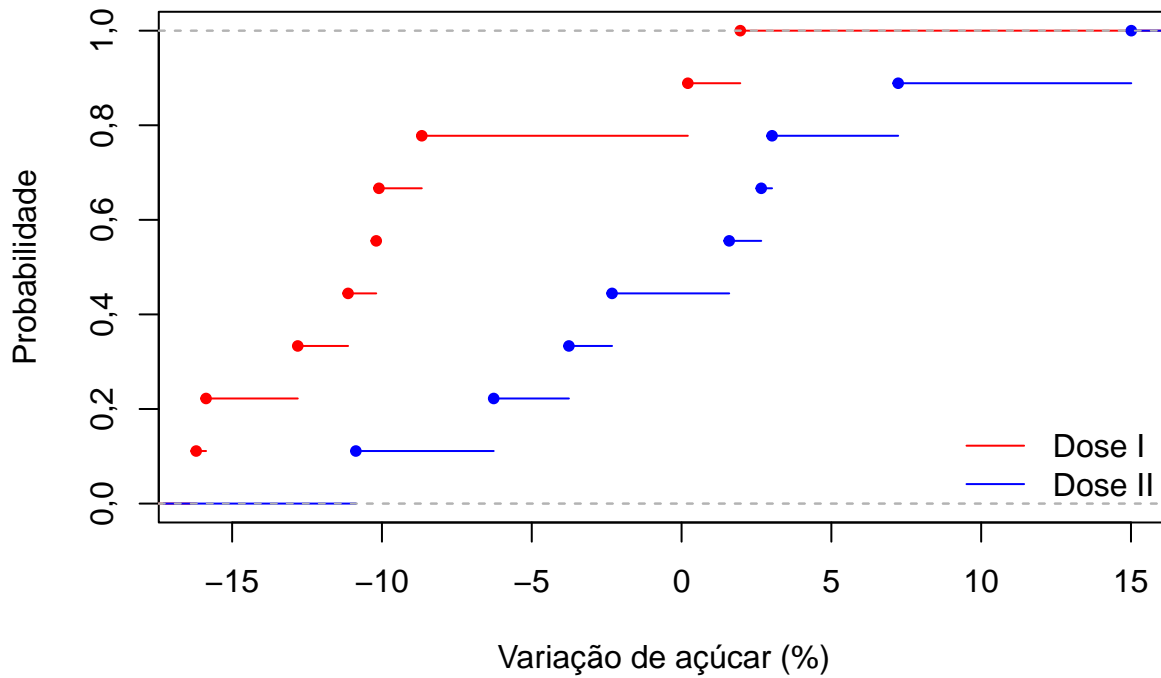
# Amostra combinada
comb <- c(dosei, doseii)

# Gráfico de pontos
dose <- factor(rep(c("Dose I", "Dose II"), times = c(m, n)))
stripchart(comb ~ dose, method = "stack", pch = 20, cex = 2,
```

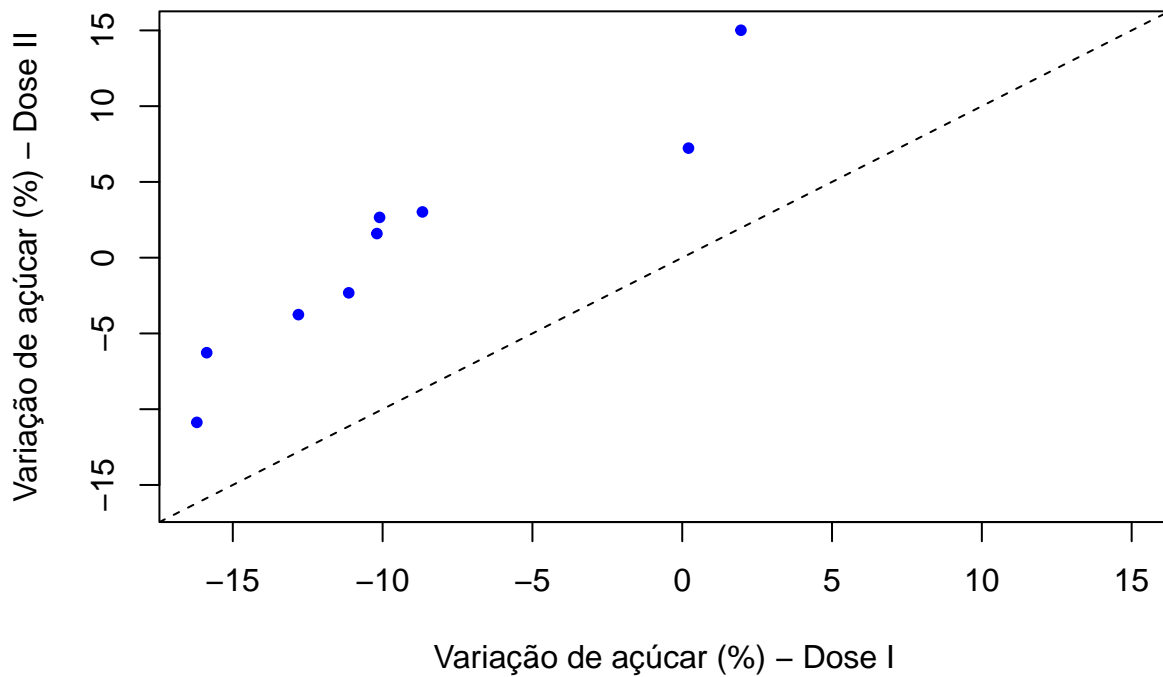
```
xlab = "Variação de açúcar (%)", col = c("red", "blue"),  
ylab = "Dose")
```



```
# Função distribuição empírica  
Fm <- ecdf(dosei)  
Fn <- ecdf(doseii)  
plot(Fm, main = "", pch = 20, xlim = range(comb),  
      xlab = "Variação de açúcar (%)",  
      ylab = "Probabilidade", col = "red")  
lines(Fn, col = "blue", pch = 20)  
legend("bottomright", c("Dose I", "Dose II"), lty = 1,  
      col = c("red", "blue"), bty = "n")
```



```
# Gráfico de quantis (QQ)
faixa <- range(comb)
qqplot(dosei, doseii, pch = 20, xlim = faixa, ylim = faixa,
       xlab = "Variação de açúcar (%) - Dose I",
       ylab = "Variação de açúcar (%) - Dose II", col = "blue")
abline(0, 1, lty = 2)
```



```
# Teste de Wilcoxon-Mann-Whitney
wilcox.test(dosei, doseii, alternative = "two.sided")
```

```
##
```

```
## Wilcoxon rank sum exact test
##
## data: dosei and doseii
## W = 12, p-value = 0,01061
## alternative hypothesis: true location shift is not equal to 0
```

Adotando um nível de significância de 5%, com base nos dados coletados e utilizando o teste de Mann-Whitney, há diferença significativa na variação percentual de açúcar no sangue entre os dois níveis de dose da droga ($p = 0,0106$).

Uma estimativa intervalar do deslocamento entre as duas funções distribuição pode ser obtida com a opção `conf.int = TRUE` (por *default*, o coeficiente de confiança é 0,95).

```
result <- wilcox.test(dosei, doseii, alternative = "two.sided",
                      conf.int = TRUE)
```

Uma estimativa da diferença está no componente `result$estimate`.

```
result$estimate
```

```
## difference in location
##                -10,49
```

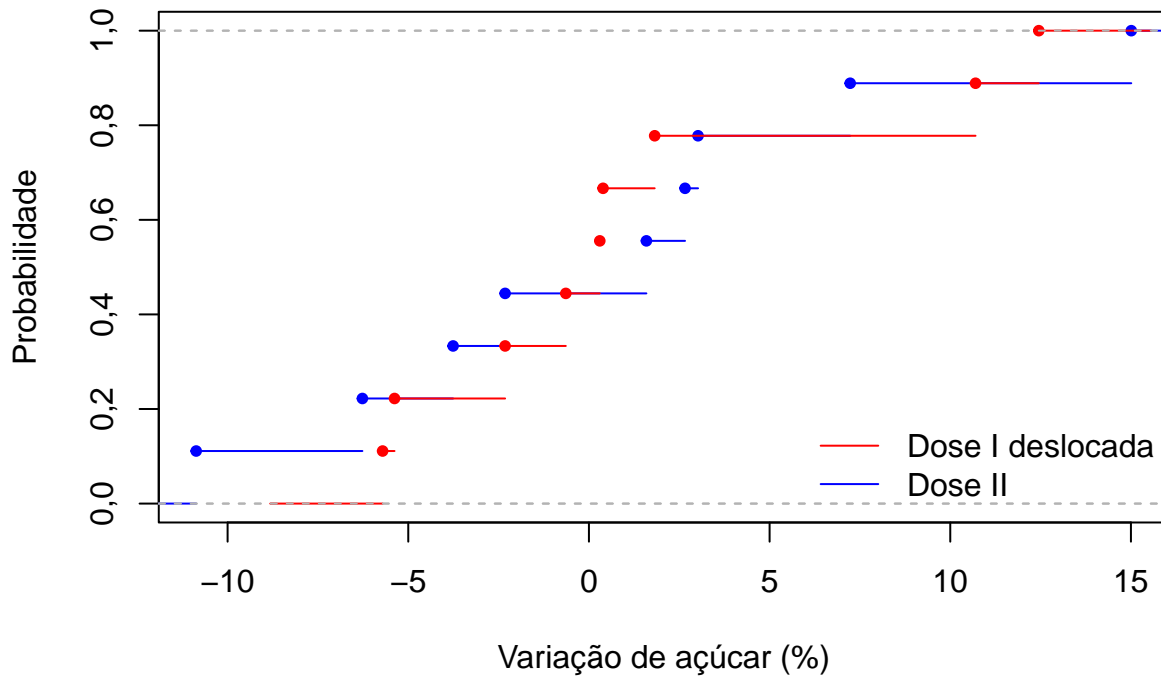
Um intervalo de confiança (IC) para a diferença de localização (diferença mediana) está no componente `result$conf.int`.

```
result$conf.int
```

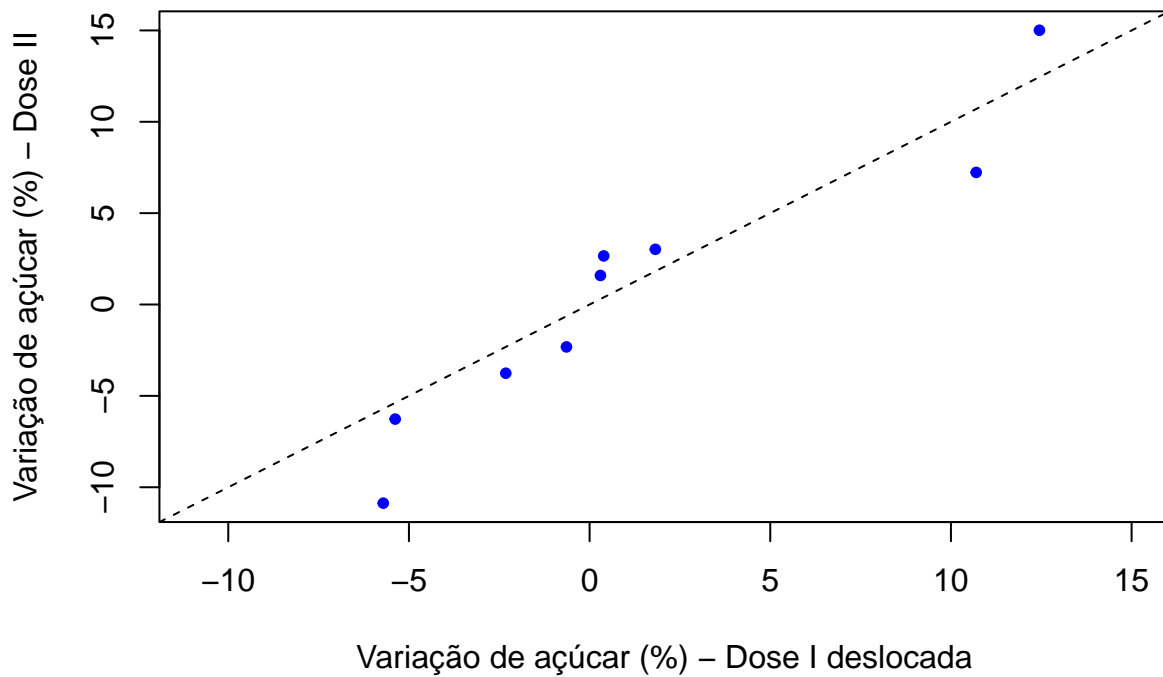
```
## [1] -17,42 -2,40
## attr(,"conf.level")
## [1] 0,95
```

Com a estimativa pontual da diferença podemos obter a função distribuição empírica da variação percentual com a dose I após deslocamento e comparar com a função distribuição empírica da variação percentual com a dose II.

```
# Função distribuição empírica com deslocamento
Fmloc <- ecdf(dosei - result$estimate)
plot(Fn, main = "", pch = 20, xlim = range(doseii,
      dosei - result$estimate),
      xlab = "Variação de açúcar (%)", ylab = "Probabilidade",
      col = "blue")
lines(Fmloc, col = "red", pch = 20)
legend("bottomright", c("Dose I deslocada", "Dose II"), lty = 1,
      col = c("red", "blue"), bty = "n")
```



```
# Gráfico de quantis (QQ) com deslocamento
faixa <- range(dosei - result$estimate, doseii)
qqplot(dosei - result$estimate, doseii, pch = 20, xlim = faixa,
        ylim = faixa, xlab = "Variação de açúcar (%) - Dose I deslocada",
        ylab = "Variação de açúcar (%) - Dose II", col = "blue")
abline(0, 1, lty = 2)
```



Nota 1. Como você justificaria a aplicação do teste de Mann-Whitney a estes dados?

Nota 2. Verifique se é apropriado utilizar o teste t de Student para amostras independentes.

A função `wilcox_test` do pacote `coin` permite obter inferências exatas para a diferença de localização,

mesmo quando há empates. Havendo empates, a distribuição da estatística de teste é exata condicional. Esta função requer uma fórmula com a variável resposta do lado esquerdo e a variável de grupo (da classe `factor`) do lado direito.

```
library(coin)
wilcox_test(comb ~ dose, distribution = "exact", conf.int = TRUE,
            alternative = "two.sided")
```

```
##
## Exact Wilcoxon-Mann-Whitney Test
##
## data:  comb by dose (Dose I, Dose II)
## Z = -2,5166, p-value = 0,01061
## alternative hypothesis: true mu is not equal to 0
## 95 percent confidence interval:
##  -17,42  -2,40
## sample estimates:
## difference in location
##                -10,49
```

A função `wilcox_test` efetua a comparação utilizando a ordem alfabética dos níveis do fator dose. (“Dose I” e “Dose II”) Se for necessário, a ordem dos níveis de um fator pode ser modificada com o argumento `levels` na chamada da função `factor`.