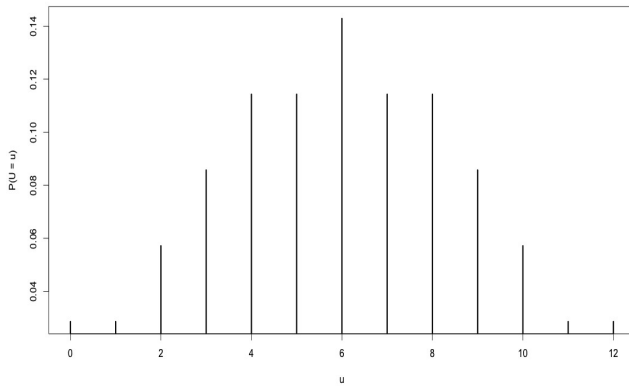


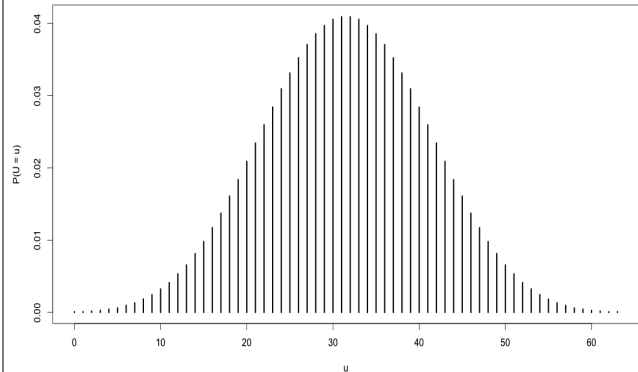
## Teste de Wilcoxon-Mann-Whitney

```
## 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)
```



```
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 = 3)
```



```
## Dados
```

```
# Ex. 21, p. 119 em Noether (1991)
```

```
# Ganho de peso em ratos (em g)
```

```
# Dieta nova é mais econômica.
```

```
# Verifique se o ganho de peso justifica adotar a dieta nova.
```

```
padrao <- c(22.5, 19.1, 11.4, 16.8, 23.5, 22.3, 19.7, 21.2, 16.0, 20.4,
           25.3, 17.5)
```

```
nova <- c(11.4, 13.7, 20.3, 14.7, 11.8, 17.7, 14.5, 12.3, 14.5, 13.0,
         22.0, 13.6)
```

```
cat("\ Tamanhos amostrais: m =", m <- length(padrao), ", n =",
    n <- length(nova))
```

```
Tamanhos amostrais: m = 12 , n = 12
```

```
summary(padrao)
```

```
Min.   1st Qu.  Median    Mean   3rd Qu.    Max.
11.40   17.32   20.05   19.64   22.35   25.30
```

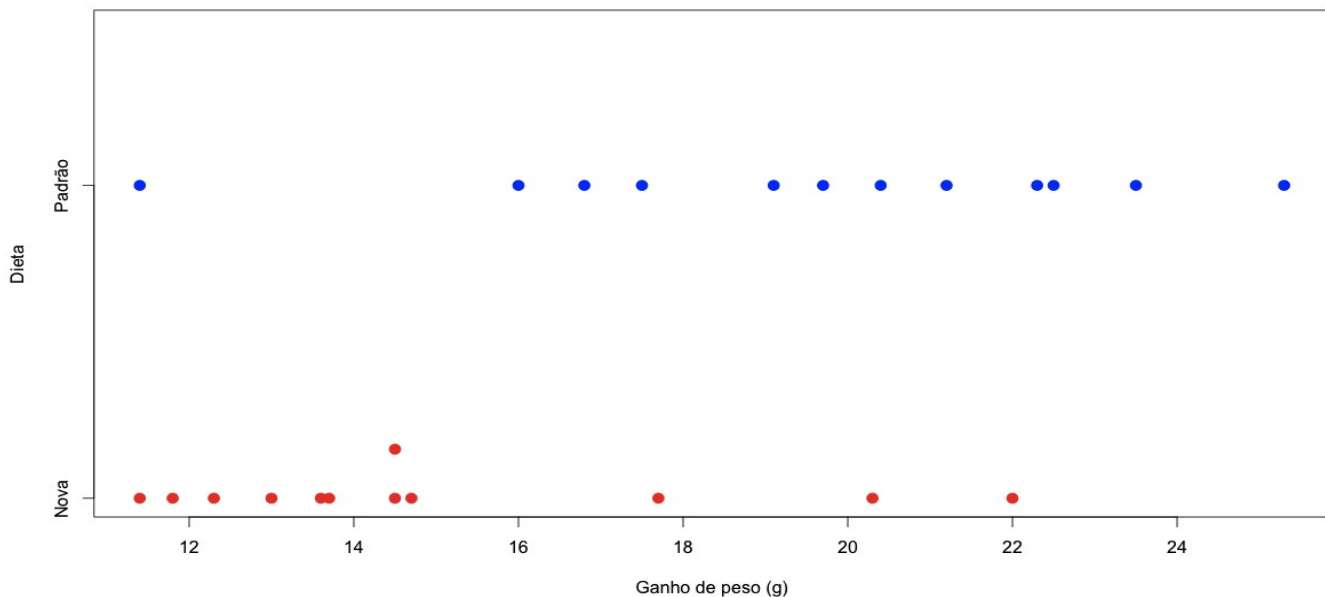
```
summary(nova)
```

```
Min.   1st Qu.  Median    Mean   3rd Qu.    Max.
11.40   12.82   14.10   14.96   15.45   22.00
```

```
# Amostra combinada
```

```
comb <- c(padrao, nova)
```

```
# Gráfico de pontos
dieta <- factor(rep(c("Padrão", "Nova"), times = c(m, n)))
stripchart(comb ~ dieta, method = "stack", pch = 20, cex = 2,
  xlab = "Ganho de peso (g)", col = c("red", "blue"), ylab = "Dieta")
```



Na amostra combinada há empates nos valores 11,4 e 14,5.

```
tabfreq <- table(comb)
tabfreq[which(tabfreq > 1)]
```

```
comb
11.4 14.5
  2   2
```

```
# Teste de Wilcoxon-Mann-Whitney
wilcox.test(nova, padrao, alternative = "greater")
```

```
Wilcoxon rank sum test with continuity correction
```

```
data: nova and padrao
W = 26.5, p-value = 0.9961
alternative hypothesis: true location shift is greater than 0
```

```
Warning message:
```

```
In wilcox.test.default(nova, padrao, alternative = "greater") :
cannot compute exact p-value with ties
```

Com base nos dados coletados e utilizando o teste de Mann-Whitney, não há evidência de aumento no ganho de peso com a dieta nova ( $p = 0,9961$ ).

Neste exemplo, em que há empates, o valor- $p$  é calculado com correção de continuidade (`correct = TRUE` é a opção por *default*).

Nota 1. Verifique se é apropriado utilizar o teste  $t$  de Student para amostras independentes.

Um intervalo de confiança (IC) para a diferença de localização (diferença mediana) pode ser obtido especificando `conf.int = TRUE` na chamada da função `wilcox.test`.

```
wilcox.test(nova, padrao, alternative = "greater", conf.int = TRUE)
```

```
95 percent confidence interval: -7.799986 Inf
```

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 classificação (da classe `factor`) do lado direito.

```
library(coin)
wilcox_test(comb ~ dieta, distribution = "exact", conf.int = TRUE,
            alternative = "greater")
```

```
Exact Wilcoxon Mann-Whitney Rank Sum Test
```

```
data:  comb by dieta (Nova, Padrão)
Z = -2.6281, p-value = 0.9968
alternative hypothesis: true mu is greater than 0
95 percent confidence interval:
 -7.8  Inf
sample estimates:
difference in location
          -5.2
```

Neste exemplo a função `wilcox_test` efetua a comparação utilizando Nova - Padrão respeitando a ordem alfabética dos níveis do fator `dieta`. Se for necessário, a ordem dos níveis de um fator pode ser modificada com o argumento `levels` na chamada da função `factor`.

Nota 2. Procure refazer o exemplo utilizando outros pacotes estatísticos (SAS, SPSS, Minitab e Statística, por exemplo).

A obtenção “manual” das estimativas para a diferença de localização ( $\theta$ ) é apresentada abaixo.

```
## Estimação de teta
alfa <- 0.05 # Nível de sig.
D <- sort(outer(nova, padrao, FUN = "-")) # Diferenças ordenadas
qalfa <- qwilcox(alfa, m, n) # Quantil alfa
```

As diferenças ordenadas podem ser calculadas sem a função `outer`.

```
D <- sort(nova[rep(1:m, each = n)] - padrao[rep(1:n, times = m)])
```

```
# Posição do limite inferior do IC
if (pwilcox(qalfa, m, n) == alfa) k <- qalfa else k <- qalfa - 1
```

```
cat("\n Estimativa pontual de teta =", median(D))
```

```
Estimativa pontual de teta = -5.2
```

```
cat("\n Limite inferior do IC de", 100 * (1 - alfa), "% para teta =", D[k + 1])
```

```
Limite inferior do IC de 95 % para teta = -7.8
```

Nota 3. Apresente um IC bilateral para  $\theta$  com coeficiente de confiança igual a 90%.

Nota 4. Realize um estudo de simulação da probabilidade de cobertura do IC para  $\theta$  baseado na estatística  $U$  de Mann-Whitney.