ESTATÍSTICA NÃO PARAMÉTRICA

Avaliação Prática da parte 1

Luiz Fernando Coelho Passos

19 de Setembro de 2019

Questão 1

```
set.seed(19092019); x = rnorm(n = 100, mean = 30, sd = 10)
```

Pelo enunciado, nossas hipóteses são

$H_0 : mediana = 30$

$H_1 : mediana > 30$

Teste dos Sinais

Formulação do Teste dos Sinais

$H_0 : p = 0.5$

$H_1 : p > 0.5$

Seja X_i os valores de X na posição i

Temos que, $T = +$, se $X_i > 30$ e -, se $X_i < 30$

```
x1 = x[x>30]
x2 = x[x<30]
x3 = x[x==30]

length(x3)
```

```
## [1] 0
```

Como não temos valores nulos, ou seja, valores onde $X_i - 30 = 0$, então, definindo T = número de sinais (+).

```
length(x1)
```

```
## [1] 46
```

$T_{obs} = 46$

Temos que T ~ Bin(n=100, p=0.5)

Assim, calculando o p-valor para o Teste dos Sinais

```
binom.test(x = 46, n = 100, p = 0.5, alternative = "greater")
```

```
##
##  Exact binomial test
##
## data:  46 and 100
## number of successes = 46, number of trials = 100, p-value = 0.8159
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
```

```
##  0.3747892 1.0000000
## sample estimates:
## probability of success
##                   0.46
```

Como o p-valor obtido foi de 0.8159, não rejeitamos H_0 a um nível de significância de 0.05 (5%). Ou seja, há evidências de que a mediana é igual a 30.

Teste de Wilcoxon

Calculando o p-valor para o Teste de Wilcoxon

```r
wilcox.test(x = x, mu = 30, alternative = "greater")
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  x
## V = 2396, p-value = 0.6719
## alternative hypothesis: true location is greater than 30
```

Como o p-valor obtido foi de 0.6719, não rejeitamos H_0 a um nível de significância de 0.05 (5%). Ou seja, há evidências de que a mediana é igual a 30.

Questão 2

Pelo enunciado, temos que n=10 e nossas hipóteses são

$H_0 : p = 0.5$

$H_1 : p > 0.5$

Região Crítica

```r
tibble::tibble("x" = 0:10, y = sprintf("%.4f", 1-pbinom(-1:9, 10, 0.5), 4))
```

```
## # A tibble: 11 x 2
##        x y
##    <int> <chr>
##  1     0 1.0000
##  2     1 0.9990
##  3     2 0.9893
##  4     3 0.9453
##  5     4 0.8281
##  6     5 0.6230
##  7     6 0.3770
##  8     7 0.1719
##  9     8 0.0547
## 10     9 0.0107
## 11    10 0.0010
```

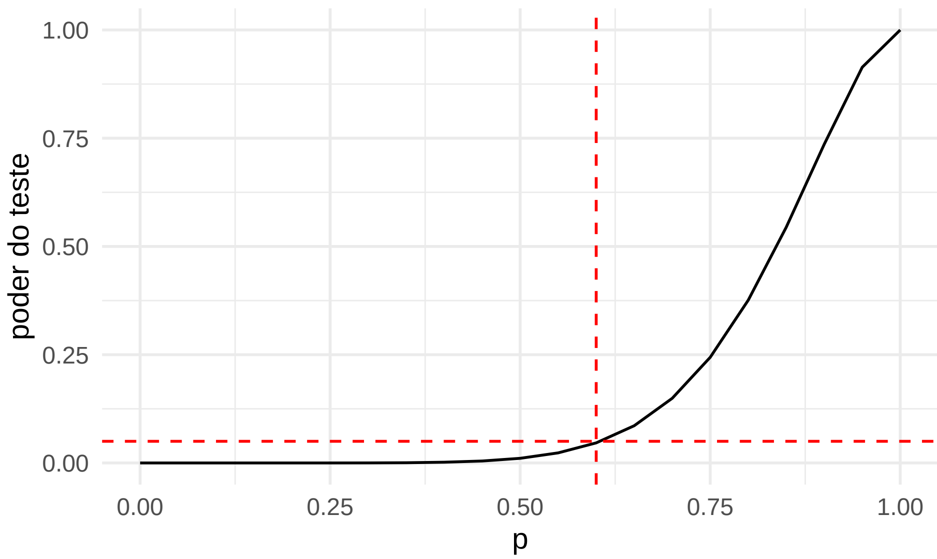
Como podemos observar, com base num nível de significância de 5%, a região crítica para o teste em questão é $P(T > 8)$.

Poder do teste

```r
library(ggplot2)

x = seq(0, 1, 0.05)
y = round(1 - pbinom( 8, 10, x), 4)

ggplot(data.frame(x, y), aes(x, y)) +
  geom_line() +
  geom_hline(yintercept = 0.05, color = "red", linetype = "dashed") +
  geom_vline(xintercept = 0.60, color = "red", linetype = "dashed") +
  labs(x = "p", y = "poder do teste") +
  theme_minimal()
```



Como podemos observar, a partir de p = 0.6, aproximadamente, o poder do teste é superior a 0.05 (nível de significância de 5%). Assim, temos que o teste é não tendencioso para estes valores de p.

Questão 3

Dados

```r
x1 = c( 2, 3, 4,10, 8, 3,10,10, 8, 6, 8, 8)
x2 = c( 2, 2, 4, 6, 6, 2, 4, 8,10, 8, 3, 4)

tibble::tibble("idoso"=1:12, x1, x2)
```

```
## # A tibble: 12 x 3
##    idoso    x1    x2
##    <int> <dbl> <dbl>
##  1     1     2     2
##  2     2     3     2
##  3     3     4     4
##  4     4    10     6
##  5     5     8     6
##  6     6     3     2
##  7     7    10     4
```

```
## 8      8     10      8
## 9      9      8     10
## 10    10      6      8
## 11    11      8      3
## 12    12      8      4
```

Temos que nossas hipóteses são

H_0 : Não houve melhora no nível de estresse

H_1 : Houve melhora no nível de estresse

Ou seja,

H_0 : nível de estresse antes (X_1) = nível de estresse depois (X_2)

H_1 : nível de estresse antes (X_1) > nível de estresse depois (X_2)

Utilizaremos o Teste de Wilcoxon

Foi escolhido o Teste de Wilcoxon, pois além de considerar o sinal, como no Teste do Sinal, também leva em conta a posição das diferenças entre as variáveis X_1 e X_2.

Formulação

Calculando a diferença entre X_1 e X_2

```
x1_x2 = x1-x2
x1_x2
```

```
##  [1]  0  1  0  4  2  1  6  2 -2 -2  5  4
```

Desconsiderando os valores nulos temos que n = 10.

Colocando em ordem crescente os valores absolutos resultante da diferença e desconsiderando os nulos, temos

```
sort(abs(x1_x2[x1_x2!=0]))
```

```
##  [1] 1 1 2 2 2 2 4 4 5 6
```

Como há valores repetidos, devemos fazer a média dos postos entre eles

Postos 1 e 2 (valores iguais a 1)

Média da soma dos postos = $\frac{1+2}{2} = 1.5$

Postos 3 a 6 (valores iguais a 2)

Média da soma dos postos = $\frac{3+4+5+6}{4} = 4.5$

Postos 7 e 8 (valores iguais a 4)

Média da soma dos postos = $\frac{7+8}{2} = 7.5$

Assim, temos que

R_+ = soma dos postos com sinal positivo $= 1.5 + 1.5 + 4.5 + 4.5 + 7.5 + 7.5 + 9 + 10 = 46$

R_- = soma dos postos com sinal negativo $= 4.5 + 4.5 = 9$

Logo, como definido em sala, a estatística de teste é R = $\min\{R_+ , R_-\}$.

P-valor

```r
wilcox.test(x1, x2, paired = TRUE, alternative = "greater")
```

```
## Warning in wilcox.test.default(x1, x2, paired = TRUE, alternative =
## "greater"): cannot compute exact p-value with ties

## Warning in wilcox.test.default(x1, x2, paired = TRUE, alternative =
## "greater"): cannot compute exact p-value with zeroes

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  x1 and x2
## V = 46, p-value = 0.03221
## alternative hypothesis: true location shift is greater than 0
```

OBS: *Pode-se notar que a função* `wilcox.test()` *utilizou o valor de* $R_+ = 46$.

Como o p-valor obtido foi de 0.03221, rejeitamos H_0 a um nível de significância de 0.05 (5%). Ou seja, há evidências de que houve melhora no nível de estresse dos idosos com a adoção do programa.