

Métodos Não Paramétricos

Profa. Teresa Cristina

01/2025

Teste de Kruskal-Wallis

1. *Teste do psicólogo: ruídos e aprendizado*

Um psicólogo conduz um estudo para determinar se o ruído pode ou não inibir o aprendizado. Cada um das 15 pessoas é, aleatoriamente, designada para um dos três grupos. Cada pessoa tem 20 minutos para memorizar uma lista de 10 sílabas sem sentido, as quais serão usadas no teste no dia seguinte.

- Grupo1 - condição sem ruído: cinco pessoas estudam a lista de sílabas enquanto estão em uma sala silenciosa
- Grupo 2 - condição de ruído moderado: cinco pessoas estudam a lista de sílabas enquanto ouvem música clássica
- Grupo 3 - condição de ruído extremo: cinco pessoas estudam a lista de sílabas enquanto ouvem música rock.

O número de sílabas sem sentido lembradas corretamente pelas 15 pessoas é:

Grupo 1: (8, 10, 9, 10, 9)

Grupo 2: (7, 8, 5, 8, 5)

Grupo 3: (4, 8, 7, 5, 7)

Os dados indicam que o ruído influenciou o desempenho dos sujeitos?

Vamos analisar os dados!

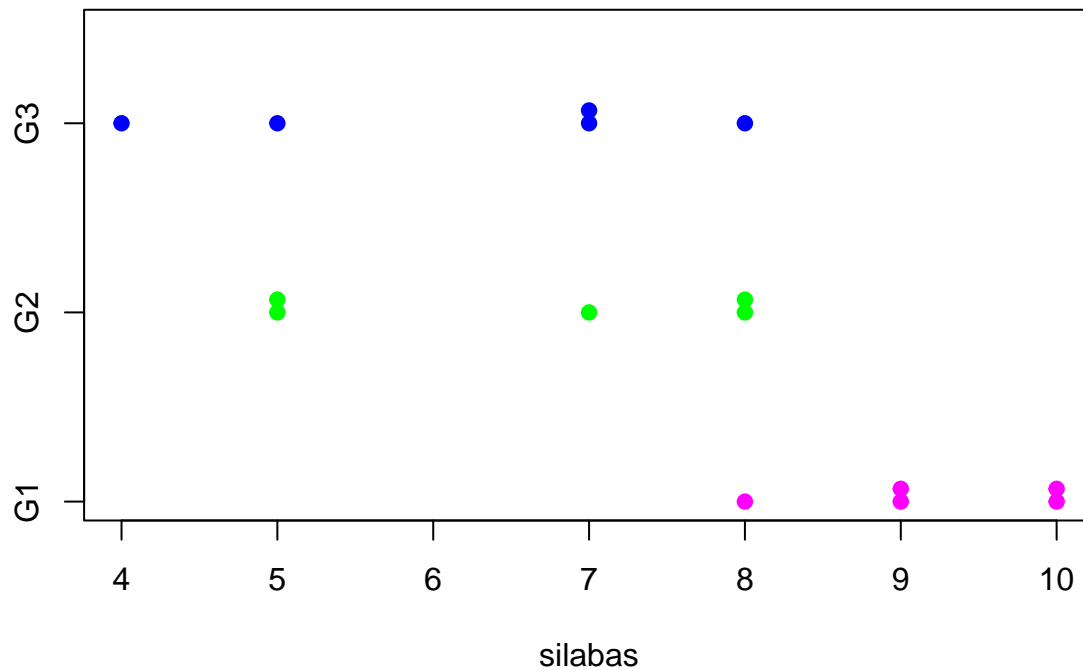
```
silabas <- matrix(c(8, 10, 9, 10, 9, 7, 8, 5, 8, 5, 4, 8, 7, 5, 7),
                  ncol = 3, byrow=FALSE)

grupos <- matrix(c(rep("G1", 5), rep("G2", 5), rep("G3", 5)), ncol = 3,
                  byrow = FALSE)

df <- data.frame(silabas, grupos)
```

Pelo gráfico de pontos, podemos observar o comportamento do número de sílabas lembradas corretamente.

```
stripchart(silabas ~ grupos, col = c("magenta", "green", "blue"),
           pch = c(19, 19, 19), method = "stack", xlab = "silabas")
```



Para aplicação do teste de Kruskal-Wallis, fazemos:

```
G1 <- c(8, 10, 9, 10, 9)
G2 <- c(7, 8, 5, 8, 5)
G3 <- c(4, 8, 7, 5, 7)
kruskal.test(list(G1, G2, G3))
```

```
##
## Kruskal-Wallis rank sum test
##
## data: list(G1, G2, G3)
## Kruskal-Wallis chi-squared = 8.7474, df = 2, p-value = 0.0126
```

Podemos dizer, que não existe evidências na amostra para aceitar H_0 , para p-valor = 0,0126.

Vamos conferir a estatística teste!

- obtendo os postos para a amostra combinada

```
silabas <- matrix(c(8, 10, 9, 10, 9, 7, 8, 5, 8, 5, 4, 8, 7, 5, 7),
                  ncol = 1, byrow=FALSE)

grupos <- matrix(c(rep("G1", 5), rep("G2", 5), rep("G3", 5)), ncol = 1,
                 byrow = FALSE)

df <- data.frame(silabas, grupos, rank(silabas))
```

```
df
```

```
##      silabas grupos rank.silabas.  
## 1         8     G1         9.5  
## 2        10     G1        14.5  
## 3         9     G1        12.5  
## 4        10     G1        14.5  
## 5         9     G1        12.5  
## 6         7     G2         6.0  
## 7         8     G2         9.5  
## 8         5     G2         3.0  
## 9         8     G2         9.5  
## 10        5     G2         3.0  
## 11        4     G3         1.0  
## 12        8     G3         9.5  
## 13        7     G3         6.0  
## 14        5     G3         3.0  
## 15        7     G3         6.0
```

```
attach(df)
```

```
## The following objects are masked _by_ .GlobalEnv:
```

```
##
```

```
##      grupos, silabas
```

- ordenando os postos

```
df%>%arrange(silabas)
```

```
##      silabas grupos rank.silabas.  
## 1         4     G3         1.0  
## 2         5     G2         3.0  
## 3         5     G2         3.0  
## 4         5     G3         3.0  
## 5         7     G2         6.0  
## 6         7     G3         6.0  
## 7         7     G3         6.0  
## 8         8     G1         9.5  
## 9         8     G2         9.5  
## 10        8     G2         9.5  
## 11        8     G3         9.5  
## 12        9     G1        12.5  
## 13        9     G1        12.5  
## 14       10     G1        14.5  
## 15       10     G1        14.5
```

- extraindo postos para os três grupos

```
posto.G1 <-df[df$grupos == "G1",]
posto.G1
```

```
##      silabas grupos rank.silabas.
## 1         8      G1           9.5
## 2        10      G1          14.5
## 3         9      G1          12.5
## 4        10      G1          14.5
## 5         9      G1          12.5
```

```
R.1 <- sum(posto.G1$rank.silabas)
R.1
```

```
## [1] 63.5
```

```
posto.G2 <-df[df$grupos == "G2",]
posto.G2
```

```
##      silabas grupos rank.silabas.
## 6          7      G2           6.0
## 7          8      G2           9.5
## 8          5      G2           3.0
## 9          8      G2           9.5
## 10         5      G2           3.0
```

```
R.2 <- sum(posto.G2$rank.silabas)
R.2
```

```
## [1] 31
```

```
posto.G3 <-df[df$grupos == "G3",]
posto.G3
```

```
##      silabas grupos rank.silabas.
## 11         4      G3           1.0
## 12         8      G3           9.5
## 13         7      G3           6.0
## 14         5      G3           3.0
## 15         7      G3           6.0
```

```
R.3 <- sum(posto.G3$rank.silabas)
R.3
```

```
## [1] 25.5
```

- calculando a estatística teste

```
T <- (12/(15*16)) * ((R.1^2)/5 + (R.2^2)/5 + (R.3^2)/5) - 3*16
T
```

```
## [1] 8.435
```

Observe que T é um valor muito próximo da estatística do teste Kruskal-Wallis obtida acima.

Exercício: *Marcadores de células endoteliais anti-CD34 e anti-fator VIII*

Objetivo: comparar a acurácia do marcador de células endoteliais anti-CD34 e anti-fator VIII em neoplasia cervical uterina, em lesões intra-epiteliais e no colo normal.

Amostra:

- Grupo A: mulheres com diagnóstico anatomopatológico de neoplasia escamosa invasiva
- Grupo B: neoplasia intra-epitelial de alto grau
- Grupo C: neoplasia intra-epitelial de baixo grau
- Grupo D: mulheres sem qualquer processo neoplásico

Os valores correspondentes ao número de vasos avaliados nos carcinomas invasivos de colo uterino (A), lesões intra-epiteliais de alto grau (B) e de baixo grau (C) e colos uterinos normais (D) utilizando a técnica imuno-histoquímica para identificação de estruturas vasculares com anti-CD34, são:

Grupo A: (161,160,128,168,131,107,219,147,175,119,190,203,107,153,128, 209,142,124)

Grupo B: (128,131,125,141,157,132,143,112,131,128,139,135,127,121,168)

Grupo C: (121,98,81,128,91,117,136,95,105,128,90,151,129,95,112)

Grupo D: (109,65,97,96,110,67,106,102,80,96)

O que podemos concluir respeito do número de vasos avaliados nos carcinomas nos diferentes grupos?

Vamos analisar os dados!

- leitura dos valores:

```
GA <- c(161,160,128,168,131,107,219,147,175,119,190,203,107,153,128,209,142,124)
```

```
GB <- c(128,131,125,141,157,132,143,112,131,128,139,135,127,121,168)
```

```
GC <- c(121,98,81,128,91,117,136,95,105,128,90,151,129,95,112)
```

```
GD <- c(109,65,97,96,110,67,106,102,80,96)
```

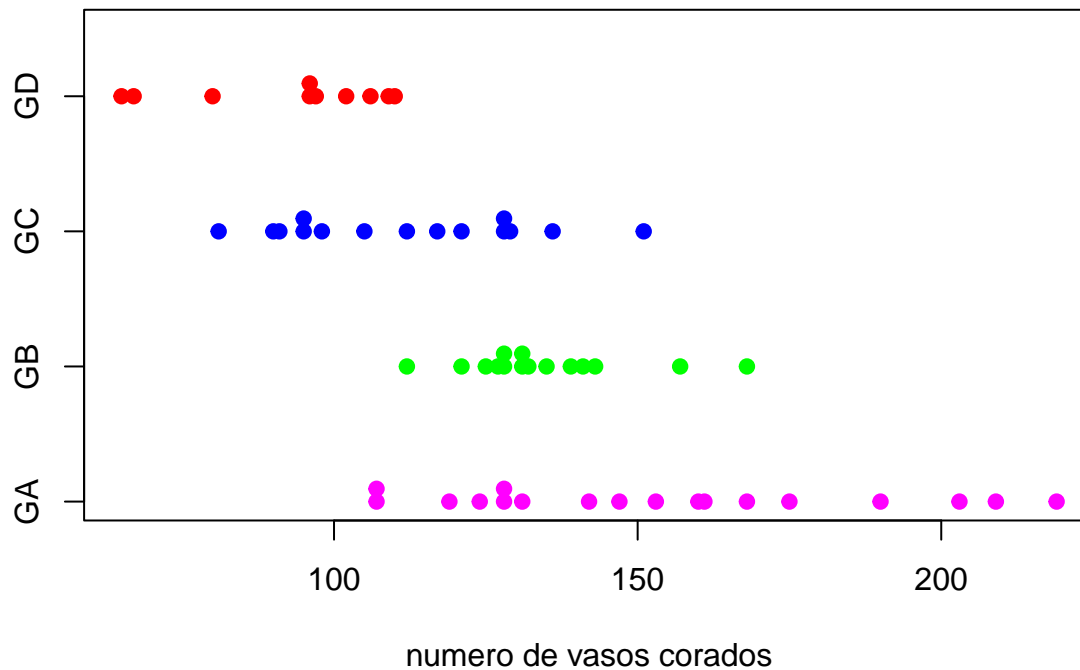
- construção da matriz de dados:

```
vasos.corados <- matrix(c(GA, GB, GC, GD),ncol = 1, byrow = FALSE)
```

```
grupos <- matrix(c(rep("GA", length(GA)), rep("GB", length(GB)),  
                  rep("GC", length(GC)), rep("GD", length(GD))),  
                ncol = 1, byrow = FALSE)
```

- construção do gráfico de pontos:

```
stripchart(vasos.corados ~ grupos, col = c("magenta", "green", "blue",  
      "red"),pch = c(19, 19, 19, 19), method = "stack",  
          xlab = "numero de vasos corados")
```



```
kruskal.test(list(GA, GB, GC, GD))
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  list(GA, GB, GC, GD)
## Kruskal-Wallis chi-squared = 29.628, df = 3, p-value = 1.653e-06
```

Podemos dizer, para estes resultados, que os grupos apresentam comportamento diferentes, para $p\text{-valor} = 1,653e - 06$.

Vamos conferir a estatística teste para esta aplicação

- construção do dataframe e obtenção dos postos:

```
df <- data.frame(vasos.corados, grupos, rank(vasos.corados))
head(df, n=6L)
```

```
##   vasos.corados grupos rank.vasos.corados.
## 1          161     GA              51.0
## 2          160     GA              50.0
## 3          128     GA              31.5
## 4          168     GA              52.5
## 5          131     GA              37.0
## 6          107     GA              16.5
```

```
tail(df, n=6L)
```

```
##   vasos.corados grupos rank.vasos.corados.
## 53          110     GD              19.0
## 54           67     GD               2.0
## 55          106     GD              15.0
## 56          102     GD              13.0
## 57           80     GD               3.0
## 58           96     GD               9.5
```

```
df%>%arrange(vasos.corados)
```

##	vasos.corados	grupos	rank.vasos.corados.
## 1	65	GD	1.0
## 2	67	GD	2.0
## 3	80	GD	3.0
## 4	81	GC	4.0
## 5	90	GC	5.0
## 6	91	GC	6.0
## 7	95	GC	7.5
## 8	95	GC	7.5
## 9	96	GD	9.5
## 10	96	GD	9.5
## 11	97	GD	11.0
## 12	98	GC	12.0
## 13	102	GD	13.0
## 14	105	GC	14.0
## 15	106	GD	15.0
## 16	107	GA	16.5
## 17	107	GA	16.5
## 18	109	GD	18.0
## 19	110	GD	19.0
## 20	112	GB	20.5
## 21	112	GC	20.5
## 22	117	GC	22.0
## 23	119	GA	23.0
## 24	121	GB	24.5
## 25	121	GC	24.5
## 26	124	GA	26.0
## 27	125	GB	27.0
## 28	127	GB	28.0
## 29	128	GA	31.5
## 30	128	GA	31.5
## 31	128	GB	31.5
## 32	128	GB	31.5
## 33	128	GC	31.5
## 34	128	GC	31.5
## 35	129	GC	35.0
## 36	131	GA	37.0
## 37	131	GB	37.0
## 38	131	GB	37.0
## 39	132	GB	39.0
## 40	135	GB	40.0
## 41	136	GC	41.0
## 42	139	GB	42.0
## 43	141	GB	43.0
## 44	142	GA	44.0
## 45	143	GB	45.0
## 46	147	GA	46.0
## 47	151	GC	47.0
## 48	153	GA	48.0
## 49	157	GB	49.0
## 50	160	GA	50.0
## 51	161	GA	51.0

```
## 52      168      GA      52.5
## 53      168      GB      52.5
## 54      175      GA      54.0
## 55      190      GA      55.0
## 56      203      GA      56.0
## 57      209      GA      57.0
## 58      219      GA      58.0
```

```
posto.GA <-df[df$grupos == "GA",]
#posto.GA
R.A <- sum(posto.GA$rank.vasos.corados)
R.A
```

```
## [1] 753.5
```

```
posto.GB <-df[df$grupos == "GB",]
#posto.GB
R.B <- sum(posto.GB$rank.vasos.corados)
R.B
```

```
## [1] 547.5
```

```
posto.GC <-df[df$grupos == "GC",]
#posto.GC
R.C <- sum(posto.GC$rank.vasos.corados)
R.C
```

```
## [1] 309
```

```
posto.GD <-df[df$grupos == "GD",]
#posto.GD
R.D <- sum(posto.GD$rank.vasos.corados)
R.D
```

```
## [1] 101
```

- cálculo da estatística do teste:

```
N = sum(length(GA), (length(GB)), length(GC), length(GD))
N
```

```
## [1] 58
```

```
T <- (12/(N*(N+1))) * ((R.A^2)/length(GA) + (R.B^2)/length(GB) +
                        (R.C^2)/length(GC) + (R.D^2)/length(GD)) - 3*(N + 1)
T
```

```
## [1] 29.58655
```

Observe o valor da estatística obtida no teste.

... Mas, e agora?