

Projeto I

Bárbara

Nathalia Gabriella Ferreira dos Santos

2023-09-22

Pacotes

```
library(tidyverse)
library(readxl)
library(patchwork) # para parcionar os graficos
```

Análises

Saneamento de dados

Extraíndo dados

Colunas usadas: Dados Atuais, *Dados Gestacionais*, *ELISAs*, *Imunoturbidimetria*, *Doenças Crônicas*

```
rm(list = ls()) # limpa enviroment

df1 <- read_xlsx("dados.xlsx", sheet = "Dados Atuais")
df2 <- read_xlsx("dados.xlsx", sheet = "Dados Gestacionais")
df3 <- read_xlsx("dados.xlsx", sheet = "ELISAs")
df4 <- read_xlsx("dados.xlsx", sheet = "Imunoturbidimetria")
df5 <- read_xlsx("dados.xlsx", sheet = "Doenças Crônicas")
```

- DF1: Dados Atuais

```
#df1 %>% colnames()
colunas_interesse <- c(
  "GRUPO      MN:0      MP:1",
  "IDADE      anos",
  "ESCOLARIDADE COMPLETA  0: 4ª série completa  1:fundamental completo 2:médio completo 3:ensino técnico",
  "NÚMERO GESTAÇÕES",
  "Peso (kg)",
  "IMC",
  "GORDURA (%)",
  "PRESSÃO SISTÓLICA (mmHg)",
  "PRESSÃO DIASTÓLICA (mmHg)",
  "PRESSÃO ARTERIAL MÉDIA",
  "CINTURA (cm)",
  "ABDÔMEN (cm)",
  "QUADRIL (cm)",
  "RELAÇÃO CINTURA/QUADRIL",
  "NEU (10³/µL)",
  "LYM (10³/µL)",
  "NLR",
```

```

"Colesterol total (mg/dL)",
"HDL (mg/dL)",
"LDL (mg/dL) DOSADO",
"VLDL (mg/dL)",
"Não-HDL (mg/dL)",
"Triglicérides (mg/dL)",
"HbA1c (%)",
"Creatinina (mg/dL)",
"Ritmo de filtração glomerular calculado",
"TGP/ALT",
"TGO/AST",
"Ácido Úrico",
"CK",
"Gama GT"
)

renomear_colunas <- c(
  "grupo",
  "idade",
  "escolaridade",
  "num_gestacoes",
  "peso",
  "imc",
  "gordura",
  "pressao_sis",
  "pressao_dias",
  "pressao_media",
  "cintura",
  "abdomen",
  "quadril",
  "cintura_quadril",
  "neu",
  "lym",
  "nlr",
  "colesterol_total",
  "hdl",
  "ldl_dosado",
  "vldl",
  "nao_hdl",
  "triglicerides",
  "hba1c",
  "creatinina",
  "filtr_glomerular",
  "tgp_alt",
  "tgo_ast",
  "acido_urico",
  "ck",
  "gama_gt"
)

df1 <- df1 %>% select(all_of(colunas_interesse))
colnames(df1) <- renomear_colunas

```

```

coluna_na_numerica <- function(lista, caracter = ".", substituto = ""){
  lista %>%
    str_replace(. , pattern = ",", replacement = ".") %>%
    str_replace_na(. , replacement = "") %>%
    ifelse(. == caracter, substituto, .) %>% as.numeric()
}

coluna_numerica <- function(lista){
  lista %>%
    ifelse(. == 99999, NA, .) %>% as.numeric()
}

grupos_nulos <- df1 %>% select(grupo) %>% is.na()
df1 <- df1 %>%
  filter(! grupos_nulos) %>%
  mutate(
    grupo = as.factor(grupo),
    escolaridade = as.factor(escolaridade),
    peso = coluna_numerica(idade),
    imc = coluna_na_numerica(imc, "#DIV/0!", NA) %>% coluna_numerica(),
    gordura = coluna_na_numerica(gordura) %>% coluna_numerica(),
    pressao_sis = coluna_numerica(pressao_sis),
    pressao_dias = coluna_numerica(pressao_dias),
    pressao_media = coluna_numerica(pressao_media),
    cintura = coluna_numerica(cintura),
    abdomen = coluna_numerica(abdomen),
    quadril = coluna_numerica(quadril),
    cintura_quadril = coluna_na_numerica(cintura_quadril, "#DIV/0!", NA) %>% coluna_numerica(),
    nlr = coluna_na_numerica(nlr, "#DIV/0!", NA) %>% coluna_numerica(),
    ldl_dosado = coluna_na_numerica(ldl_dosado, "#DIV/0!", NA) %>% coluna_numerica(),
    vldl = coluna_na_numerica(vldl, "#DIV/0!", NA) %>% coluna_numerica(),
    hba1c = coluna_numerica(hba1c),
    creatinina = coluna_numerica(creatinina),
    filtr_glomerular = coluna_na_numerica(filtr_glomerular, "#DIV/0!", NA) %>% coluna_numerica(),
    ck = coluna_numerica(ck)
  )

df1 %>% summary()

```

```

## grupo      idade      escolaridade num_gestacoes      peso
## 0:102      Min.    :24.0      0      : 7      Min.    :0.00      Min.    :24.0
## 1: 86      1st Qu.:37.0      1      : 4      1st Qu.:1.00      1st Qu.:37.0
##          Median :40.0      2      :35      Median :2.00      Median :40.0
##          Mean   :40.2      3      :10      Mean   :2.14      Mean   :40.2
##          3rd Qu.:44.0      4      :66      3rd Qu.:3.00      3rd Qu.:44.0
##          Max.   :57.0      5      :64      Max.   :7.00      Max.   :57.0
##          NA's    :2      NA's: 2      NA's    :2      NA's    :2
##          imc      gordura      pressao_sis      pressao_dias
## Min.    :16.05      Min.    :16.50      Min.    : 90.0      Min.    : 41.00
## 1st Qu.:23.23      1st Qu.:26.40      1st Qu.:115.0      1st Qu.: 76.00
## Median :26.20      Median :32.50      Median :127.0      Median : 84.00
## Mean   :27.06      Mean   :33.45      Mean   :127.9      Mean   : 85.25
## 3rd Qu.:29.85      3rd Qu.:39.02      3rd Qu.:138.0      3rd Qu.: 94.00
## Max.   :46.68      Max.   :58.50      Max.   :185.0      Max.   :130.00

```

```
## NA's :1      NA's :4      NA's :2      NA's :2
## pressao_media      cintura      abdomen      quadril
## Min. : 0.00      Min. : 52.00      Min. : 68.00      Min. : 77
## 1st Qu.: 89.67      1st Qu.: 75.00      1st Qu.: 85.00      1st Qu.:101
## Median : 97.33      Median : 83.00      Median : 92.00      Median :106
## Mean : 98.95      Mean : 83.79      Mean : 93.45      Mean :107
## 3rd Qu.:108.00      3rd Qu.: 90.50      3rd Qu.:100.00      3rd Qu.:113
## Max. :144.67      Max. :117.00      Max. :132.00      Max. :141
## NA's :1      NA's :1      NA's :1      NA's :1
## cintura_quadril      neu      lym      nlr
## Min. :0.4769      Min. :0.650      Min. :0.090      Min. : 0.3869
## 1st Qu.:0.7333      1st Qu.:3.085      1st Qu.:1.657      1st Qu.: 1.4923
## Median :0.7843      Median :3.835      Median :2.045      Median : 1.8632
## Mean :0.7833      Mean :4.031      Mean :2.074      Mean : 2.2919
## 3rd Qu.:0.8271      3rd Qu.:4.770      3rd Qu.:2.440      3rd Qu.: 2.3743
## Max. :1.0253      Max. :9.180      Max. :4.030      Max. :49.7778
## NA's :1
## colesterol_total      hdl      ldl_dosado      vldl
## Min. : 93.0      Min. : 27.33      Min. : 30.65      Min. : 7.143
## 1st Qu.:144.8      1st Qu.: 54.22      1st Qu.: 78.46      1st Qu.: 15.714
## Median :173.0      Median : 67.04      Median : 97.11      Median : 18.964
## Mean :175.3      Mean : 68.96      Mean :100.65      Mean : 21.613
## 3rd Qu.:192.0      3rd Qu.: 80.98      3rd Qu.:122.77      3rd Qu.: 24.515
## Max. :391.0      Max. :126.00      Max. :218.45      Max. :130.448
##
## nao_hdl      triglicerides      hba1c      creatinina
## Min. : 16.13      Min. : 25.0      Min. : 3.050      Min. : 0.03
## 1st Qu.: 75.57      1st Qu.: 74.0      1st Qu.: 4.200      1st Qu.: 0.60
## Median : 97.39      Median : 97.5      Median : 4.700      Median : 0.82
## Mean :105.64      Mean :121.4      Mean : 4.784      Mean : 1.09
## 3rd Qu.:127.49      3rd Qu.:147.2      3rd Qu.: 5.200      3rd Qu.: 1.00
## Max. :340.52      Max. :874.0      Max. :10.720      Max. :52.00
##
## NA's :44      NA's :3
## filtr_glomerular      tgp_alt      tgo_ast      acido_urico
## Min. : 1.77      Min. : 6.00      Min. :12.0      Min. : 1.900
## 1st Qu.: 79.38      1st Qu.:12.00      1st Qu.:18.0      1st Qu.: 3.900
## Median : 98.27      Median :14.95      Median :21.0      Median : 4.750
## Mean : 989.28      Mean :17.93      Mean :22.8      Mean : 4.882
## 3rd Qu.: 143.16      3rd Qu.:20.00      3rd Qu.:25.0      3rd Qu.: 5.325
## Max. :153470.69      Max. :70.60      Max. :63.0      Max. :17.900
## NA's :4
## ck      gama_gt
## Min. : 21.0      Min. : 10.00
## 1st Qu.: 62.0      1st Qu.: 18.00
## Median : 84.0      Median : 24.00
## Mean :105.8      Mean : 29.79
## 3rd Qu.:119.0      3rd Qu.: 33.00
## Max. :322.0      Max. :157.00
## NA's :143
```

- DF2: Dados Gestacionais

```
#df2 %>% colnames()
colunas_interesse <- c(
  "GRUPO      MN:0      MP:1",
```

```

"Há quantos anos",
"PESO BEBÊ (kg)",
"PARTO                                0:normal    1:cesariana    2:fórceps    3: induzido",
"PARTO PREMATURO 0:não              1:sim"
)

renomear_colunas <- c(
  "grupo",
  "quantos_anos_evento",
  "peso_bebe",
  "tipo_parto",
  "prematuro"
)

df2 <- df2 %>% select(all_of(colunas_interesse))
colnames(df2) <- renomear_colunas

grupos_nulos <- df2 %>% select(grupo) %>% is.na()
df2 <- df2 %>%
  filter(! grupos_nulos) %>%
  mutate(
    grupo = as.factor(grupo),
    quantos_anos_evento = quantos_anos_evento %>% as.numeric(),
    peso_bebe = coluna_numerica(peso_bebe),
    tipo_parto = as.factor(coluna_numerica(tipo_parto)),
    prematuro = as.factor(prematuro)
  )

df2 %>% summary()

```

```

##  grupo    quantos_anos_evento    peso_bebe    tipo_parto prematuro
##  0:102    Min.      : 5.000        Min.      :0.500    0      : 57    0      :133
##  1: 86    1st Qu.: 8.000        1st Qu.:2.400    1      :121    1      : 53
##          Median : 9.000        Median :2.965    2       : 4    NA's:  2
##          Mean   : 9.737        Mean   :2.795    3       : 1
##          3rd Qu.:12.000        3rd Qu.:3.375    NA's:  5
##          Max.    :15.000        Max.    :4.230
##          NA's    :2            NA's    :6

```

Descritivas e Testes de Médias

Funções que serão aplicadas:

```

cor_verde = c("#4fb6a7")
cor_roxa = c("#652177")

cores = colorRampPalette(c("#4fb6a7", "#652177"))
tema = theme_minimal() +
  theme(axis.title.y = element_text(angle = 90, margin = unit(c(0.3, 0.5, 0.3, 0), "cm")),
        axis.title.x = element_text(margin = unit(c(0.3, 0, 0, 0), "cm")),
        axis.line.x = element_line(linewidth = 0.6, linetype=1),
        axis.line.y = element_line(linewidth = 0.6, linetype=1),
        axis.ticks.y = element_line(linewidth = 0.6, linetype = 1),

```

```

    axis.text = element_text(size=10),
    legend.position = "right",
    plot.margin = unit(c(0.3, 0.4, 0.3, 0.4), "cm"),
    plot.subtitle = element_text(hjust = 0.5, vjust= 4))
theme_set(tema)

```

- Funções de gráficos numéricos e categóricos

```

graficos <- function(dados, grupo, variavel, texto = "variável"){

  dados <- dados %>%
    mutate(grupo = ifelse(grupo == "0", "Controle", "Tratamento"))

  tabela <- dados %>%
    dplyr::summarise(
      `Média` = mean(!!sym(variavel) , na.rm = TRUE) %>% round(2),
      `Desvio Padrão` = sd(!!sym(variavel) , na.rm = TRUE) %>% round(2),
      `Valores nulos` = is.na(!!sym(variavel) ) %>% sum() %>% round(2),
      `Tamanho da amostra` = n() %>% round()
    ) %>%
    tidyr::pivot_longer(dplyr::everything(), names_to = "Variável", values_to = "Valores")

  dados <- dados %>% filter(! is.na(variavel) )

  g1 <- dados %>%
    ggplot(aes(x = !!sym(grupo), y = !!sym(variavel), fill = !!sym(grupo))) +
    geom_boxplot() +
    labs(title = "Descritivas\n", x = "Grupos", y = str_to_upper(variavel), fill = "Grupos") +
    tema +
    scale_fill_manual(values=c(cor_verde, cor_roxa))

  g2 <- dados %>%
    ggplot(aes(x = !!sym(variavel))) +
    geom_histogram(fill = cor_verde) +
    labs(y = "Frequência", x = str_to_upper(variavel))

  return(
    (g1 + gridExtra::tableGrob(tabela, rows = c("", "", "", "")))/
    (g2)
  )
}

```

- Função para gráficos categóricos

```

graficos_categoricos <- function(lista1, lista2, variavel){

  tabela <- prop.table(
    table(lista1, lista2),
    margin = 2) #freq da escolaridade dentro dos grupos, ex: freq do grupo 2 no grupo 1

  perc <- data.frame(tabela) %>%
    mutate(lista1 = ifelse(lista1 == "0", "Controle", "Tratamento")) %>%
    rename(`Frequência` = Freq)

```

```

g1 <- perc %>%
  ggplot(aes(x = lista2, y = `Frequência` , fill = lista1)) +
  geom_bar(stat='identity') +
  tema +
  scale_fill_manual(values=c(cor_verde, cor_roxa)) +
  labs(
    x = str_to_title(variavel),
    title = paste("Frequência de", variavel, " por grupo"),
    fill = "Grupo")

return(g1)
}

```

- Teste de Hipóteses

```

realizar_teste <- function(dataframe, variavel) {

  dataframefiltrado <- dataframe %>%
    mutate(grupo = ifelse(grupo == "0", "MN", "MP"))

  grupo_MN<- dataframefiltrado %>%
    filter(grupo == "MN") %>%
    pull(variavel)

  grupo_MP<- dataframefiltrado %>%
    filter(grupo == "MP") %>%
    pull(variavel)

  shapiro1 <- shapiro.test(grupo_MN)
  shapiro2 <- shapiro.test(grupo_MP)

  if (shapiro1$p.value <= 0.05 || shapiro2$p.value <= 0.05) {
    resultado_teste <- wilcox.test(grupo_MP, grupo_MN, alternative = "greater")
    return(resultado_teste)
  } else {
    resultado_teste <- t.test(grupo_MN, grupo_MP, alternative = "greater")
    return(resultado_teste)
  }
}

#realizar_teste(df3,"logpai1")

```

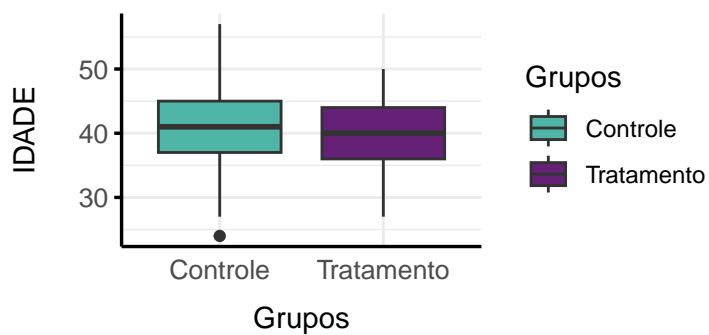
Análise por variável:

Idade:

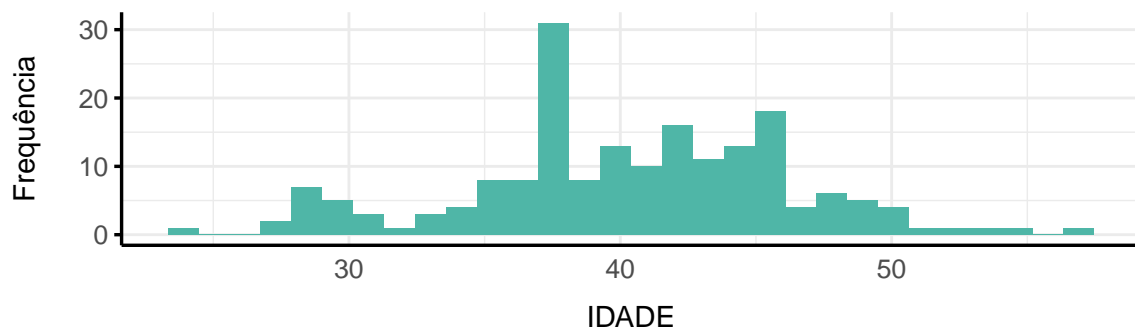
Não rejeita-se a hipótese nula de que as médias sejam iguais.

```
graficos(df1, "grupo", "idade")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 40.20 |
| Desvio Padrão | 5.88 |
| Valores nulos | 2.00 |
| Tamanho da amostra | 188.00 |

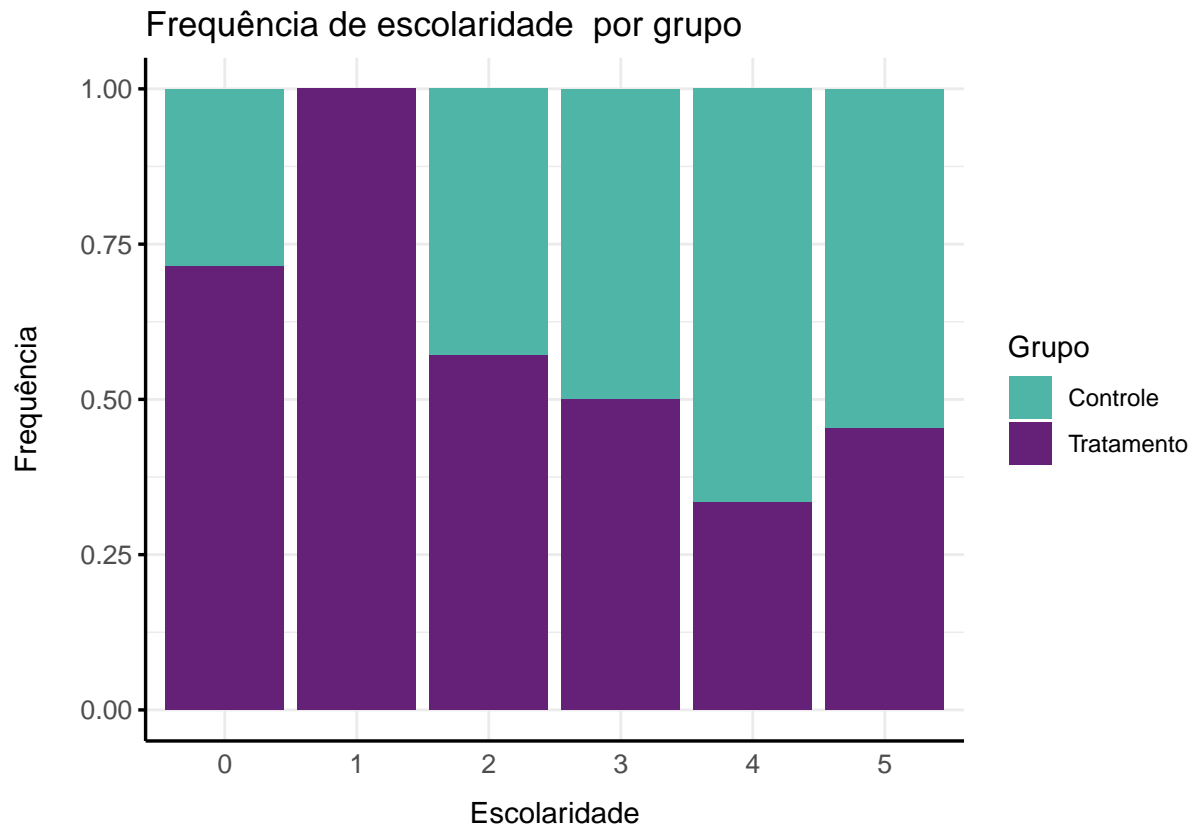


```
realizar_teste(df1, "idade")
```

```
##
## Welch Two Sample t-test
##
## data: grupo_MN and grupo_MP
## t = 1.2435, df = 181.34, p-value = 0.1076
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -0.3524145      Inf
## sample estimates:
## mean of x mean of y
## 40.69307 39.62353
```

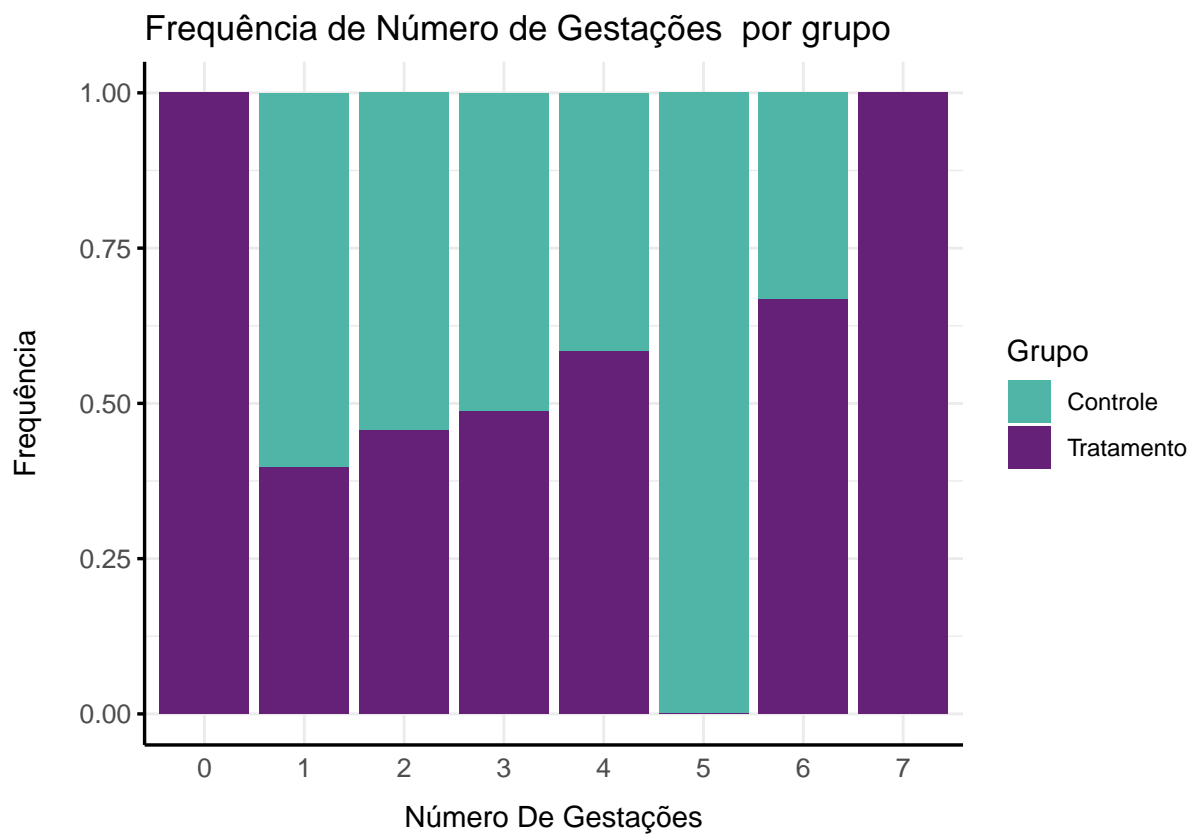
Escolaridade: ** Falta aplicar teste para frequência dos categoricos

```
graficos_categoricos(df1$grupo, df1$escolaridade, "escolaridade")
```

Número de Gestações:

```
graficos_categoricos(df1$grupo, df1$num_gestacoes, "Número de Gestações")
```

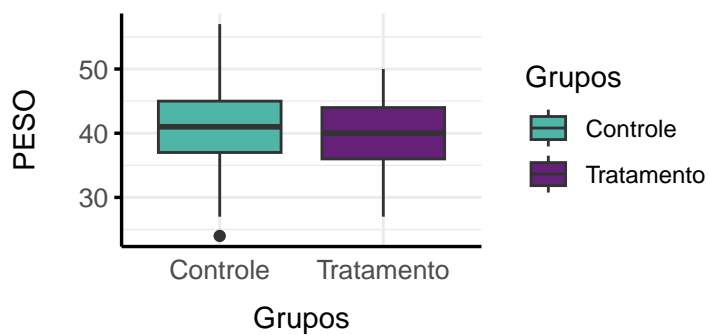


Peso:

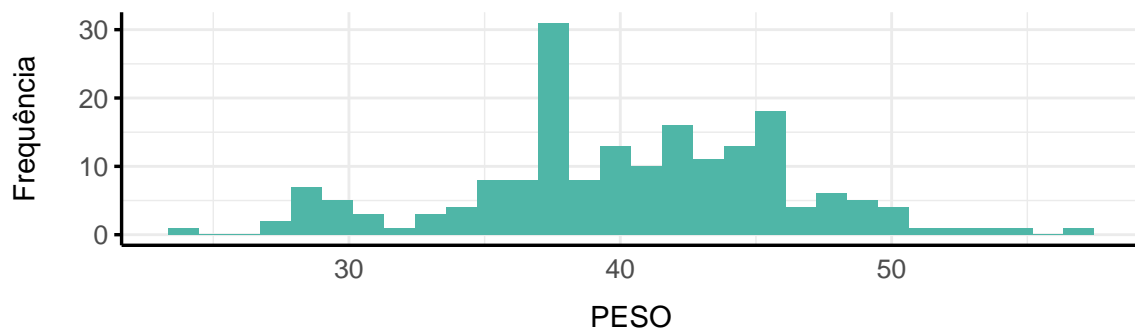
Não rejeita-se a hipótese nula de que as médias sejam iguais.

```
graficos(df1, "grupo", "peso")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 40.20 |
| Desvio Padrão | 5.88 |
| Valores nulos | 2.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "peso")
```

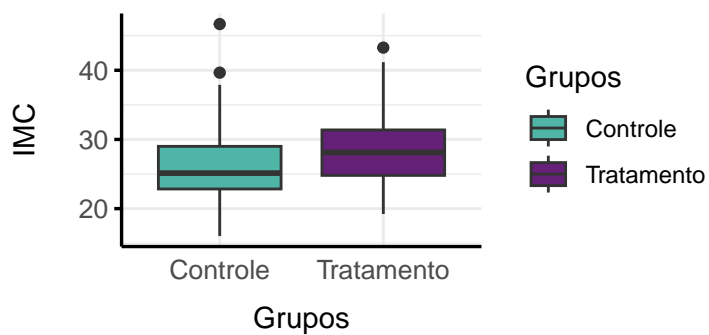
```
##
##  Welch Two Sample t-test
##
## data:  grupo_MN and grupo_MP
## t = 1.2435, df = 181.34, p-value = 0.1076
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.3524145      Inf
## sample estimates:
## mean of x mean of y
##  40.69307  39.62353
```

IMC:

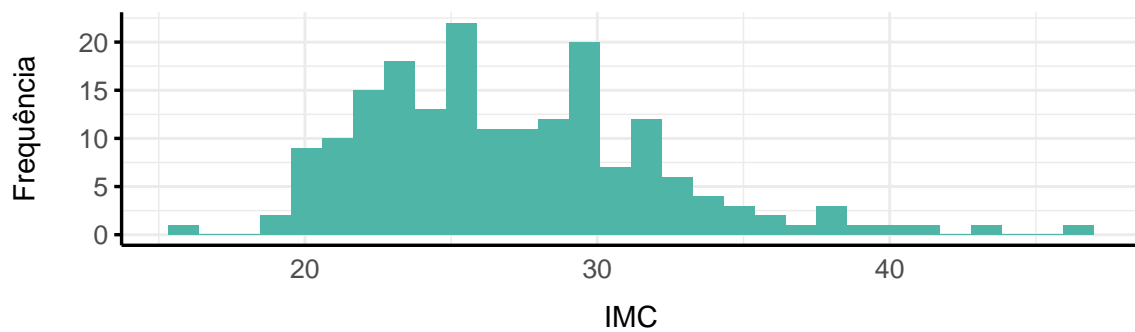
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "imc")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 27.06 |
| Desvio Padrão | 4.99 |
| Valores nulos | 1.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "imc")
```

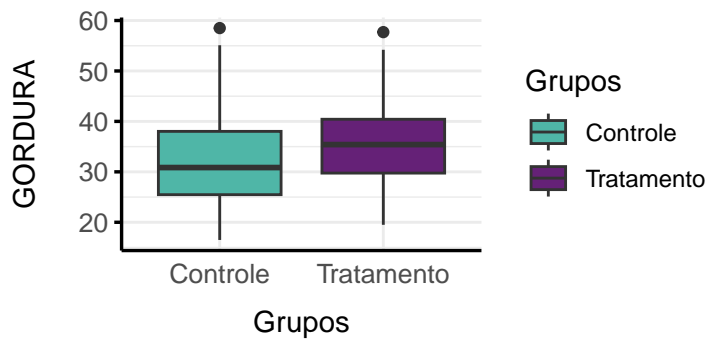
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5413.5, p-value = 0.001862
## alternative hypothesis: true location shift is greater than 0
```

Gordura:

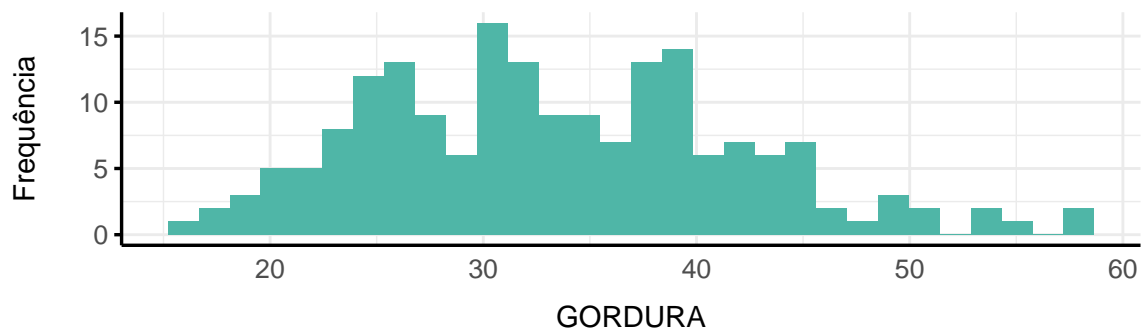
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "gordura")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 33.45 |
| Desvio Padrão | 8.66 |
| Valores nulos | 4.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "gordura")
```

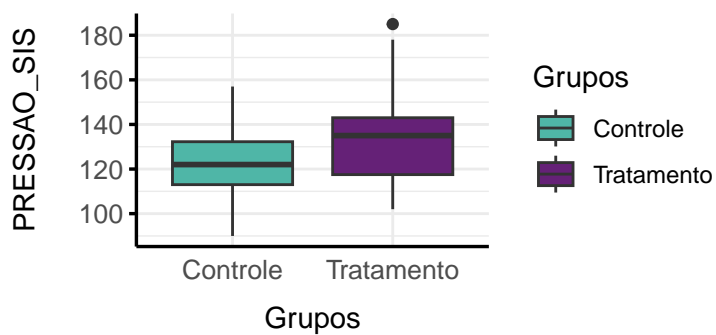
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5162, p-value = 0.00377
## alternative hypothesis: true location shift is greater than 0
```

Pressão Sistólica:

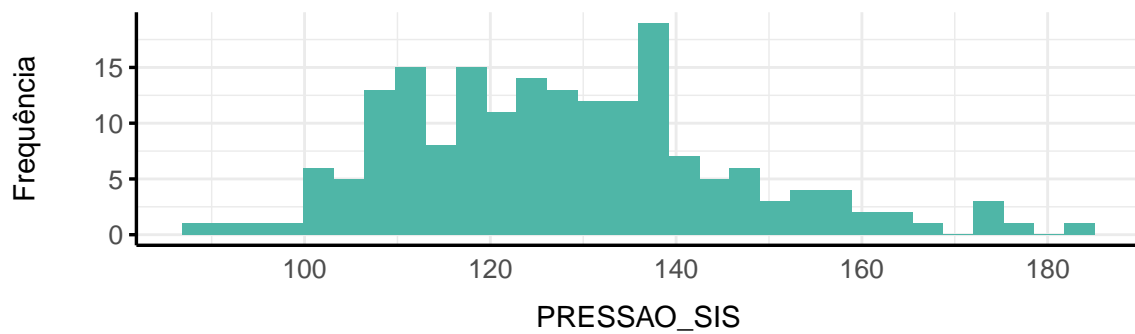
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "pressao_sis")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 127.95 |
| Desvio Padrão | 17.63 |
| Valores nulos | 2.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "pressao_sis")
```

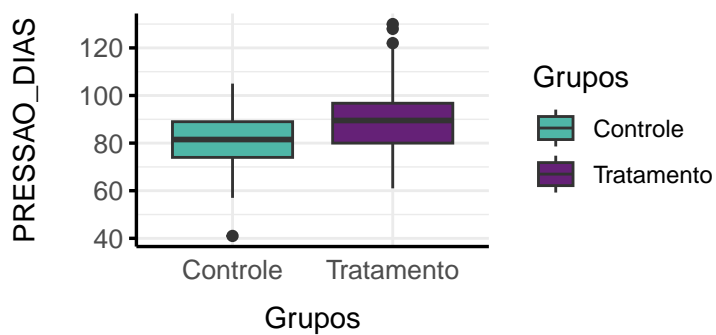
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5704, p-value = 6.283e-05
## alternative hypothesis: true location shift is greater than 0
```

Pressão Diastólica:

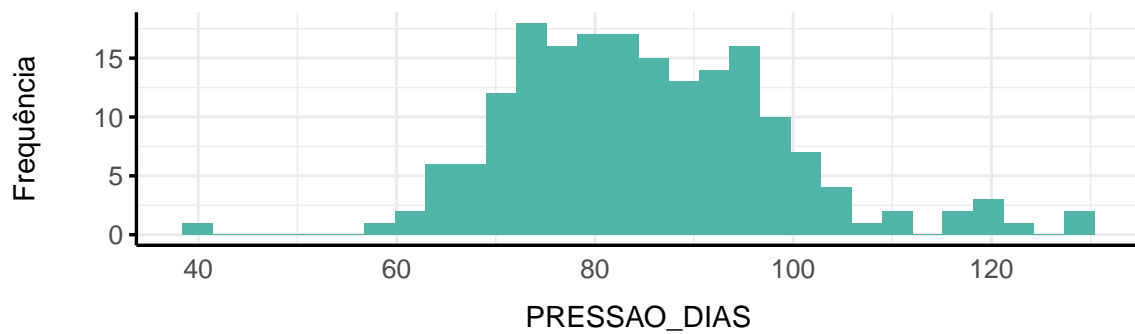
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "pressao_dias")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 85.25 |
| Desvio Padrão | 13.64 |
| Valores nulos | 2.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "pressao_dias")
```

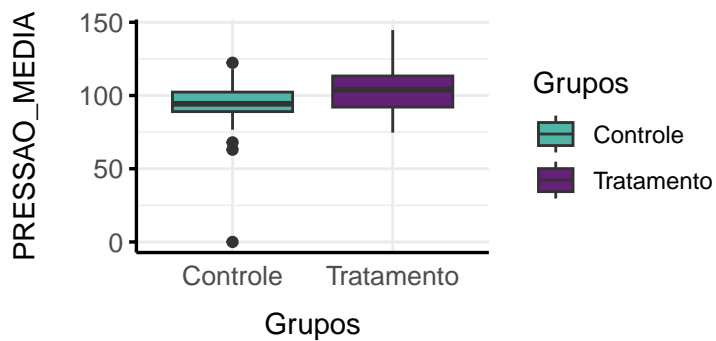
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5739.5, p-value = 4.201e-05
## alternative hypothesis: true location shift is greater than 0
```

Pressão Média:

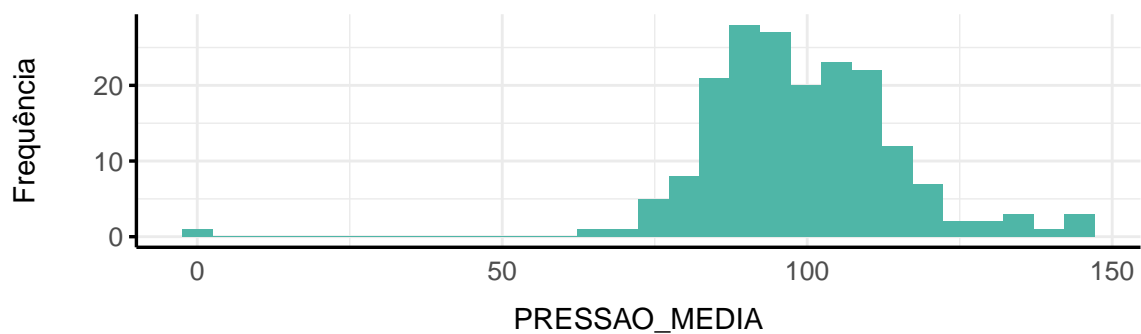
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "pressao_media")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 98.95 |
| Desvio Padrão | 15.98 |
| Valores nulos | 1.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "pressao_media")
```

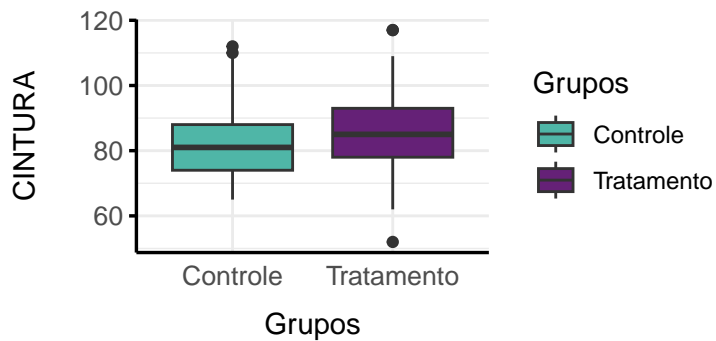
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5884, p-value = 1.481e-05
## alternative hypothesis: true location shift is greater than 0
```

Cintura:

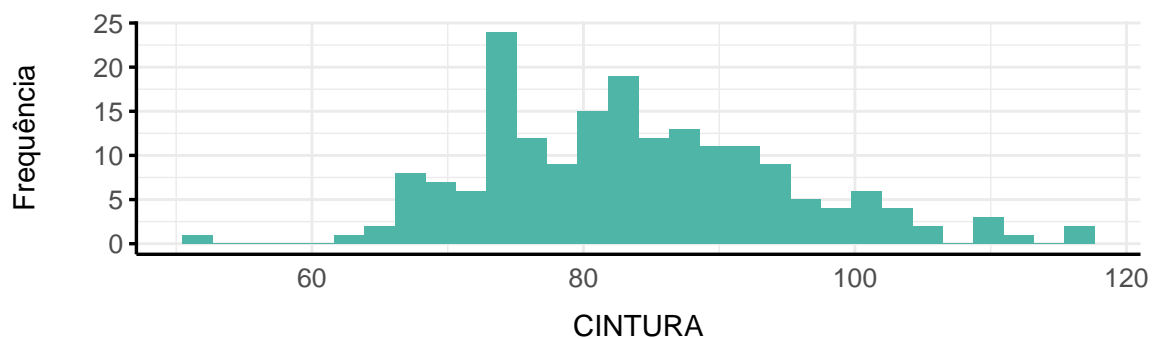
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "cintura")
```


Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 83.79 |
| Desvio Padrão | 11.14 |
| Valores nulos | 1.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "cintura")
```

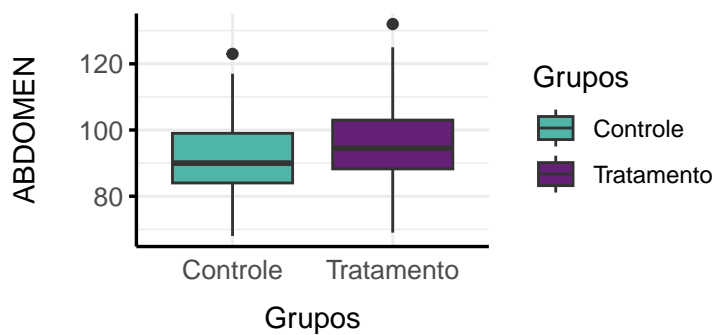
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5113.5, p-value = 0.01837
## alternative hypothesis: true location shift is greater than 0
```

Abdomen:

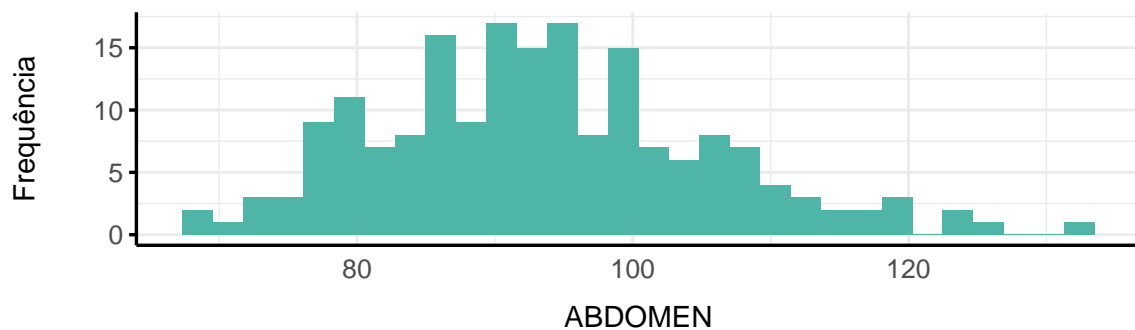
Não rejeita-se a hipótese nula que não haja diferença nas médias dos grupos.

```
graficos(df1, "grupo", "abdomen")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 93.45 |
| Desvio Padrão | 11.96 |
| Valores nulos | 1.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "abdomen")
```

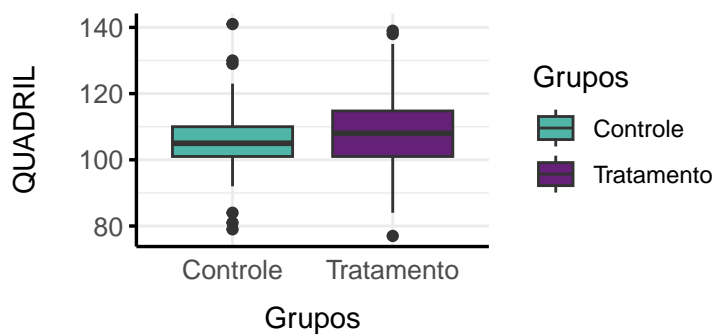
```
##
## Welch Two Sample t-test
##
## data: grupo_MN and grupo_MP
## t = -2.6245, df = 170, p-value = 0.9953
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -7.47221      Inf
## sample estimates:
## mean of x mean of y
## 91.34653 95.93023
```

Quadril:

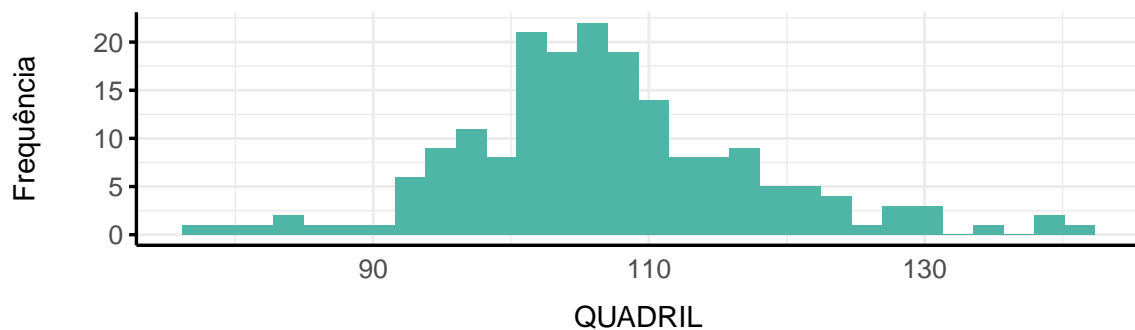
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "quadril")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 107.02 |
| Desvio Padrão | 10.67 |
| Valores nulos | 1.00 |
| Tamanho da amostra | 188.00 |



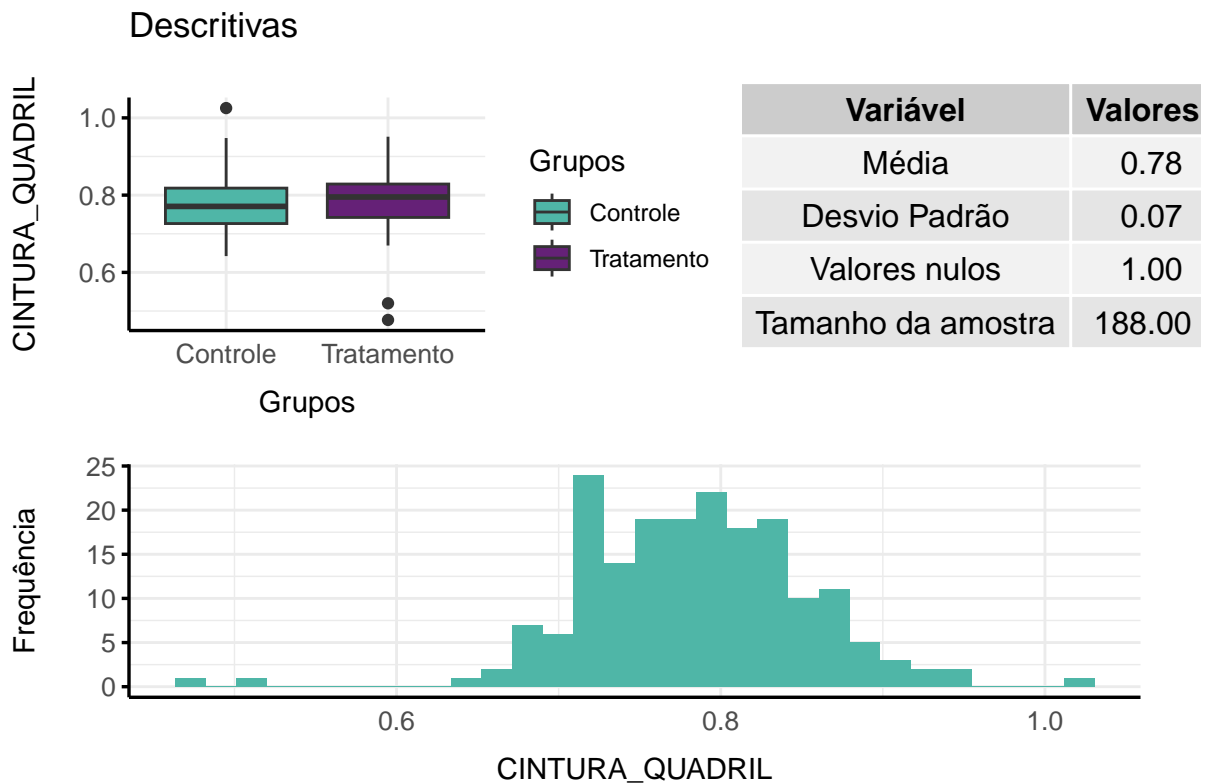
```
realizar_teste(df1, "quadril")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5127, p-value = 0.01677
## alternative hypothesis: true location shift is greater than 0
```

Relação cintura vs quadril:

Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "cintura_quadril")
```



```
realizar_teste(df1, "cintura_quadril")
```

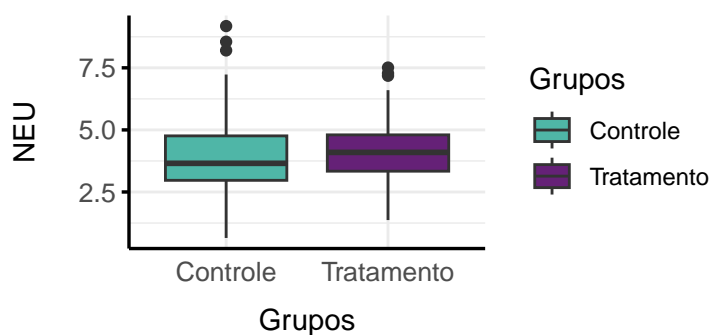
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 4932, p-value = 0.05532
## alternative hypothesis: true location shift is greater than 0
```

neu:

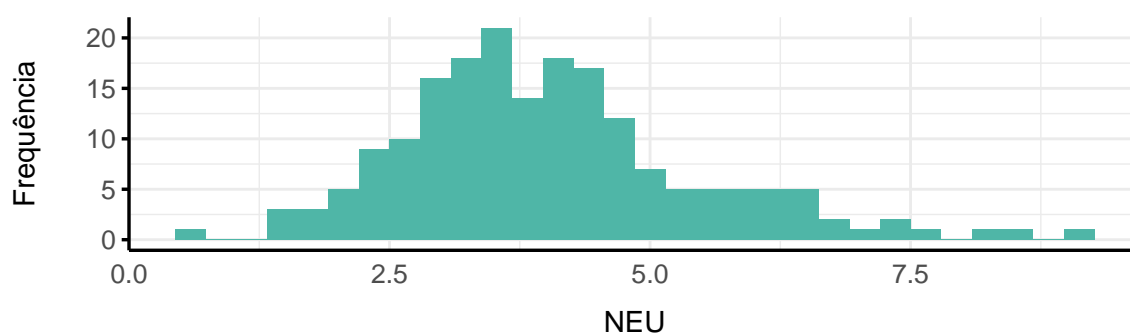
Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "neu")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 4.03 |
| Desvio Padrão | 1.41 |
| Valores nulos | 0.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "neu")
```

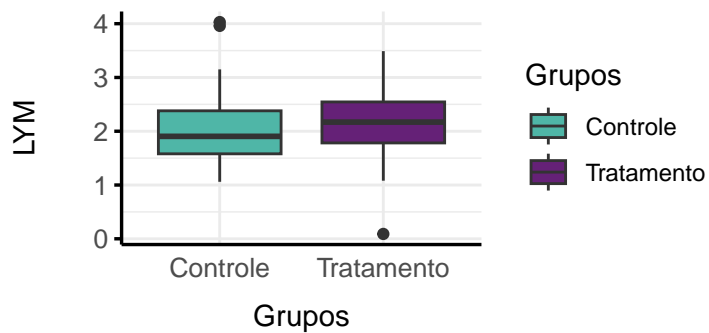
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 4977, p-value = 0.05607
## alternative hypothesis: true location shift is greater than 0
```

lym:

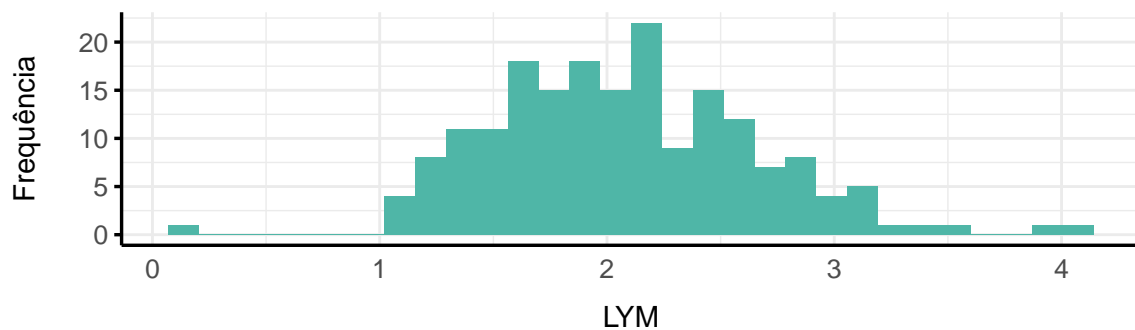
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "lym")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 2.07 |
| Desvio Padrão | 0.58 |
| Valores nulos | 0.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "lym")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5228.5, p-value = 0.01174
## alternative hypothesis: true location shift is greater than 0
```

nlr:

Foi encontrada uma observação outlier, ela será retirada.

- Relação antes da retirada:

```
df1$nlr %>% summary()
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3869  1.4923  1.8632  2.2919  2.3743 49.7778
```

```
df1$nlr[df1$nlr > 6]
```

```
## [1] 49.77778
```

- Após retirada:

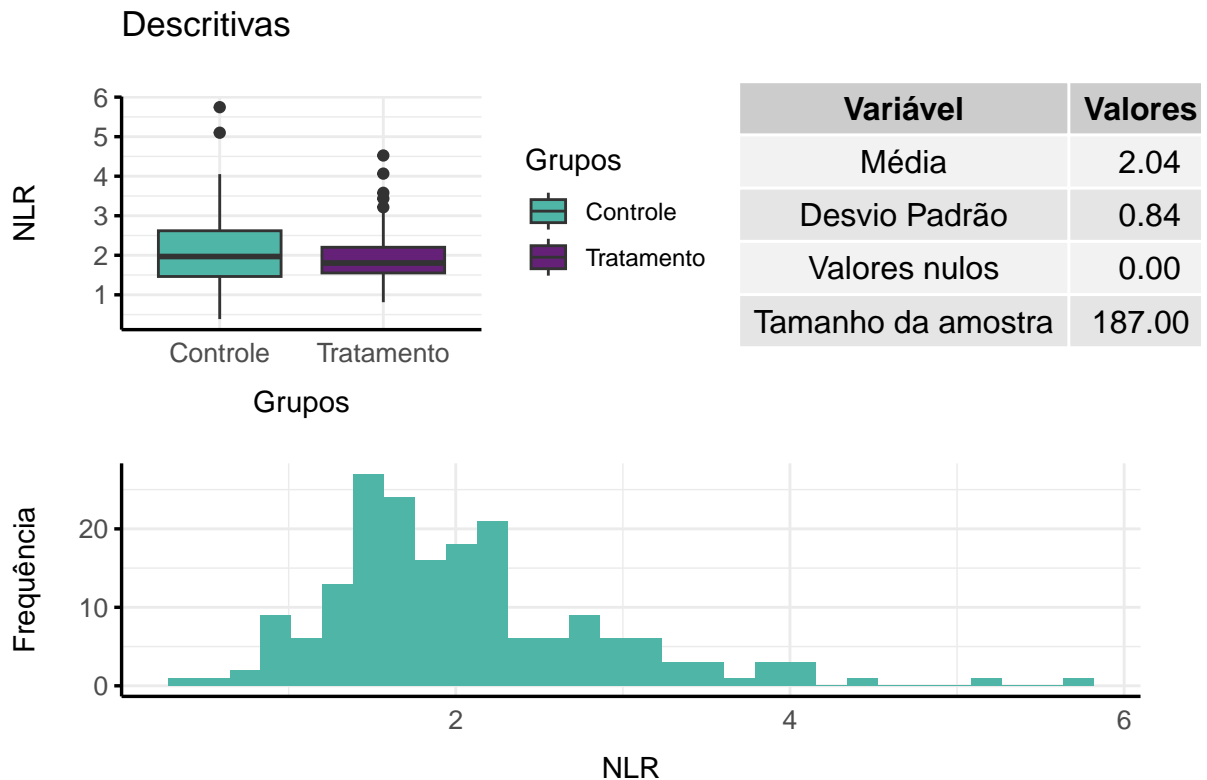
```
df_nlr <- df1 %>%
  filter(nlr < 6)
```

```
df_nlr$nlr %>% summary()
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3869  1.4921  1.8571  2.0379  2.3645  5.7458
```

Não rejeita-se a hipótese nula.

```
graficos(df_nlr, "grupo", "nlr")
```



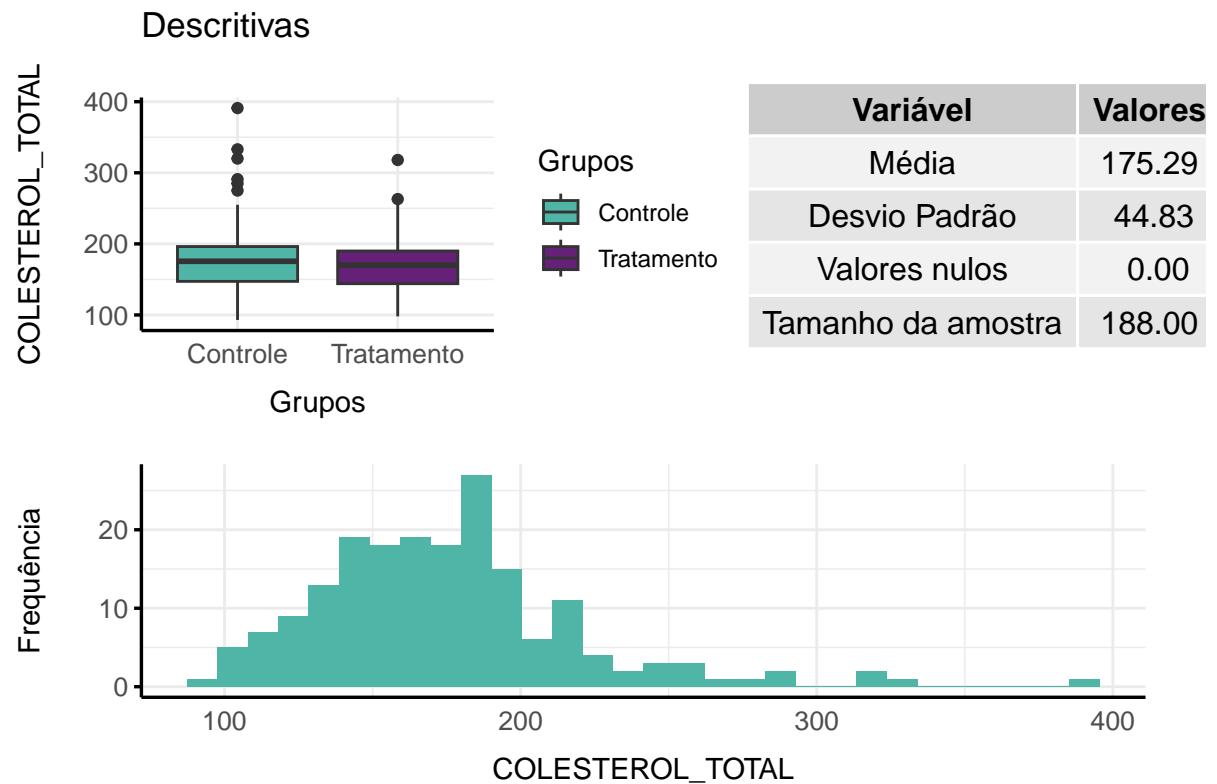
```
realizar_teste(df_nlr, "nlr")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 4062, p-value = 0.771
## alternative hypothesis: true location shift is greater than 0
```

colesterol Total:

Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "colesterol_total")
```



```
realizar_teste(df1, "colesterol_total")
```

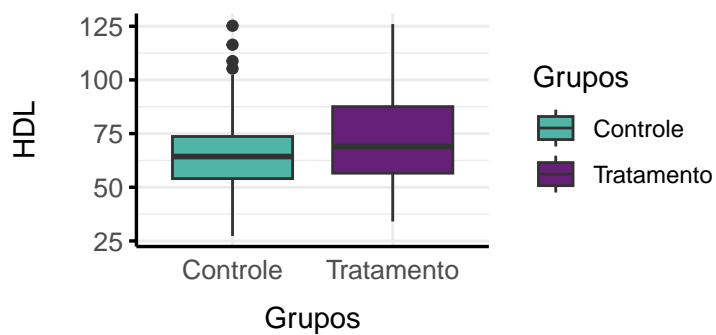
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 4005.5, p-value = 0.8473
## alternative hypothesis: true location shift is greater than 0
```

hdl:

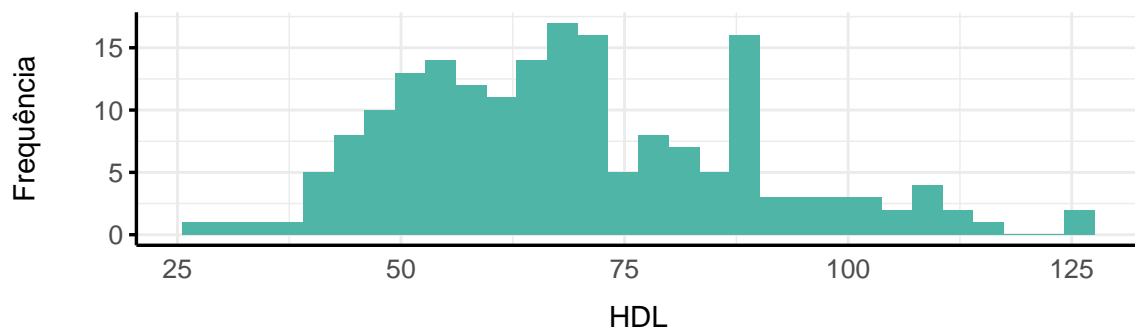
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "hdl")
```


Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 68.96 |
| Desvio Padrão | 19.16 |
| Valores nulos | 0.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "hdl")
```

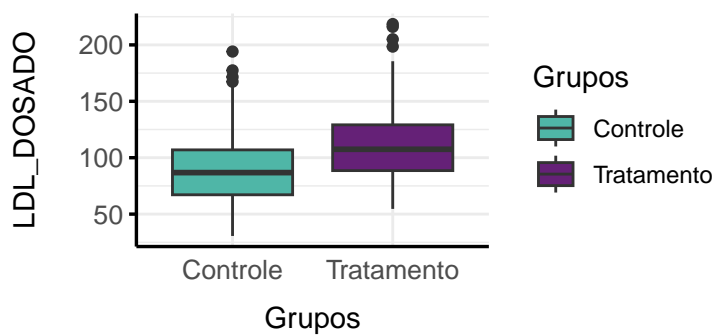
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5024, p-value = 0.04316
## alternative hypothesis: true location shift is greater than 0
```

ldl_dosado:

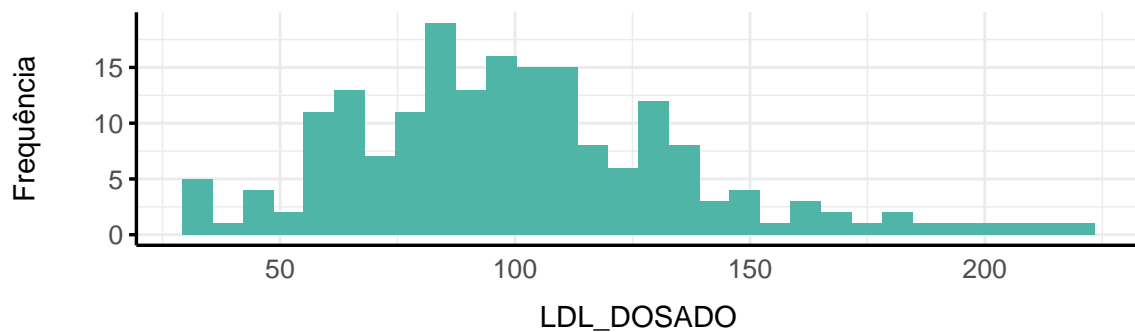
Rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "ldl_dosado")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 100.65 |
| Desvio Padrão | 36.28 |
| Valores nulos | 0.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "ldl_dosado")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5990, p-value = 8.016e-06
## alternative hypothesis: true location shift is greater than 0
```

vldl:

Foi encontrado observação outlier, elas serão retiradas.

- Relação antes da retirada:

```
df1$vldl %>% summary()
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   7.143  15.714  18.964  21.613  24.515  130.448
```

```
df1$vldl[df1$vldl > 45]
```

```
## [1]  58.0000  47.7050 130.4480  62.0339
```

- Após retirada:

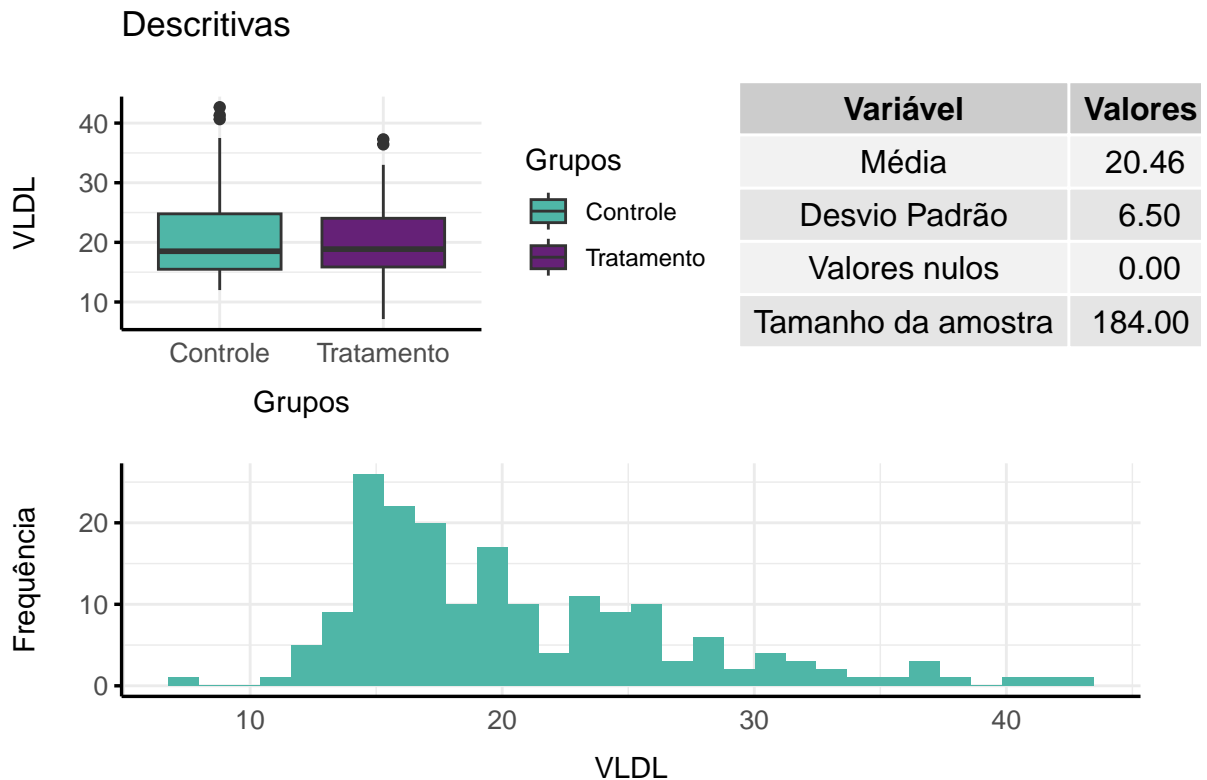
```
df_vldl <- df1 %>%
  filter(vldl < 45)
```

```
df_vldl$vldl %>% summary()
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7.143 15.663 18.855 20.462 24.300 42.667
```

Não rejeita-se a hipótese nula.

```
graficos(df_vldl, "grupo", "vldl")
```



```
realizar_teste(df_vldl, "vldl")
```

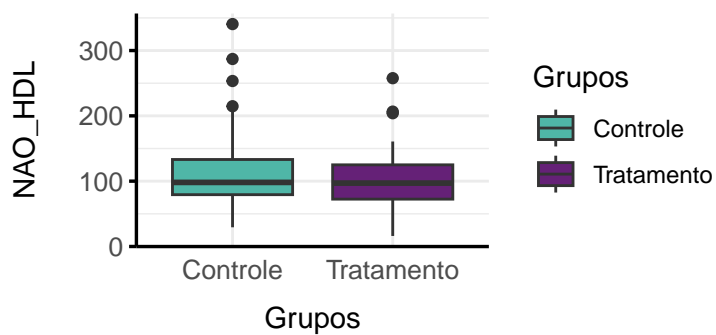
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 4121, p-value = 0.5874
## alternative hypothesis: true location shift is greater than 0
```

nao_hdl:

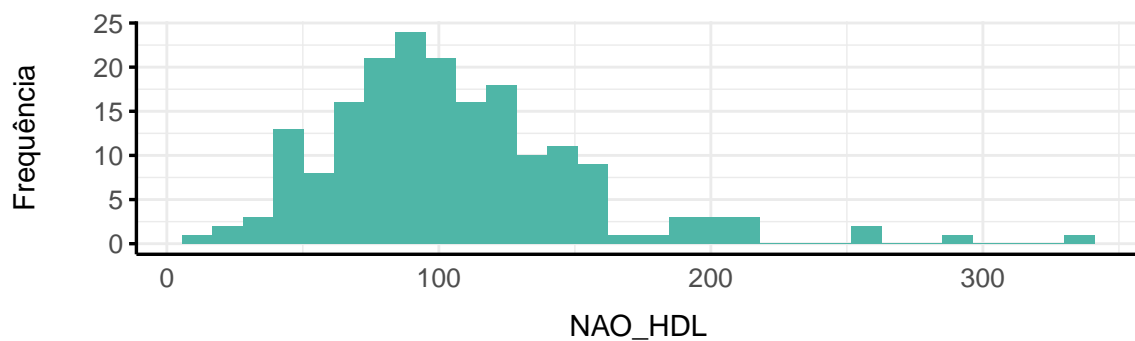
Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "nao_hdl")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 105.64 |
| Desvio Padrão | 48.17 |
| Valores nulos | 0.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "nao_hdl")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 3996.5, p-value = 0.853
## alternative hypothesis: true location shift is greater than 0
```

triglicerides:

Foi encontrada observação outlier, elas serão retiradas.

- Relação antes da retirada:

```
df1$triglicerides %>% summary()
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      25.0   74.0   97.5  121.4  147.2   874.0
```

```
df1$triglicerides[df1$triglicerides > 300]
```

```
## [1] 377 874 366 305 320
```

- Após retirada:

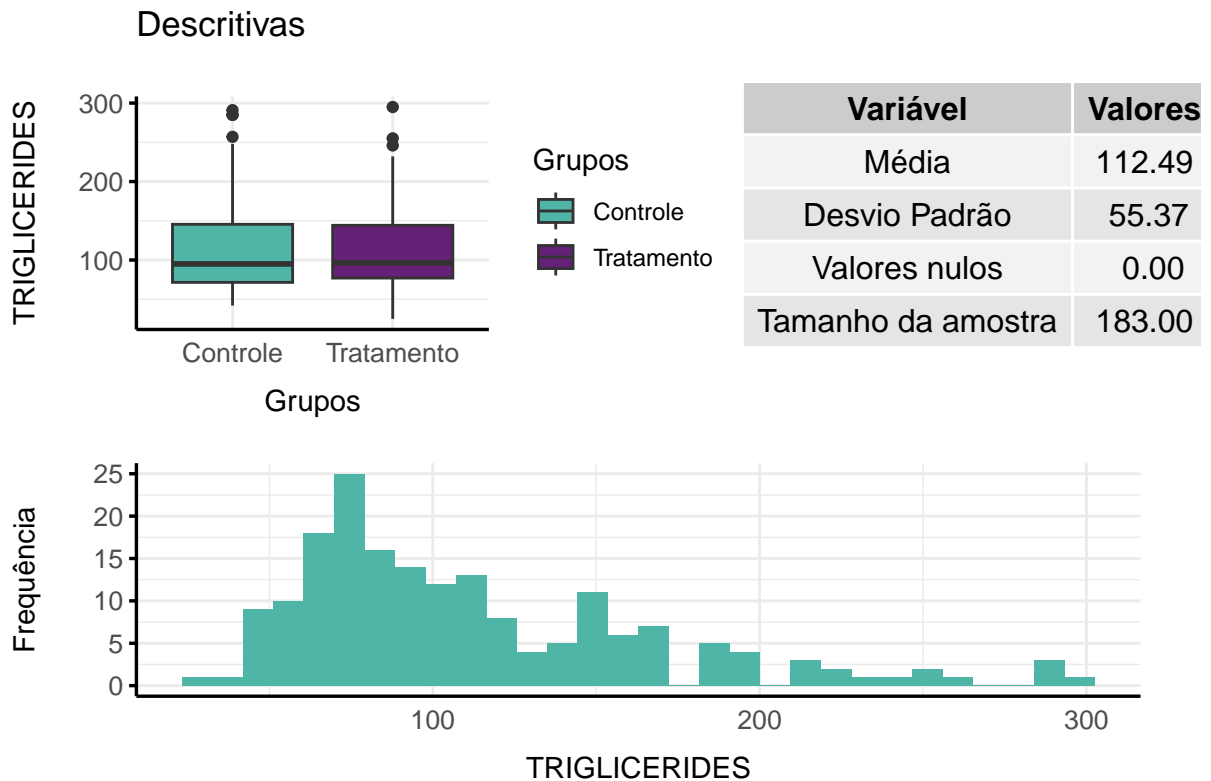
```
df_triglicerides <- df1 %>%
  filter(triglicerides < 300)
```

```
df_triglicerides$triglicerides %>% summary()
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      25.0   73.5   96.0   112.5   145.0   295.0
```

Não rejeita-se a hipótese nula.

```
graficos(df_triglicerides, "grupo", "triglicerides")
```



```
realizar_teste(df_triglicerides, "triglicerides")
```

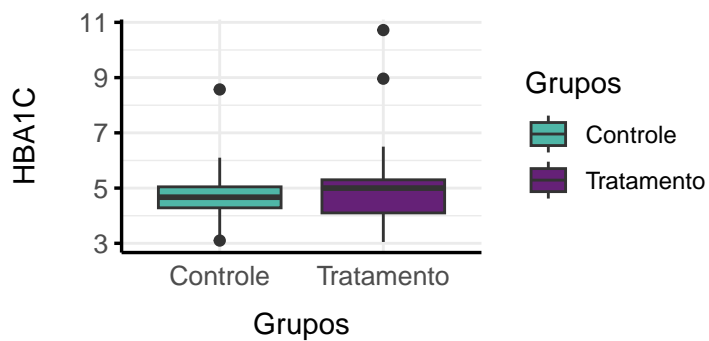
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 4260.5, p-value = 0.3876
## alternative hypothesis: true location shift is greater than 0
```

hba1c: **Aqui também encontramos influentes mas não tiramos

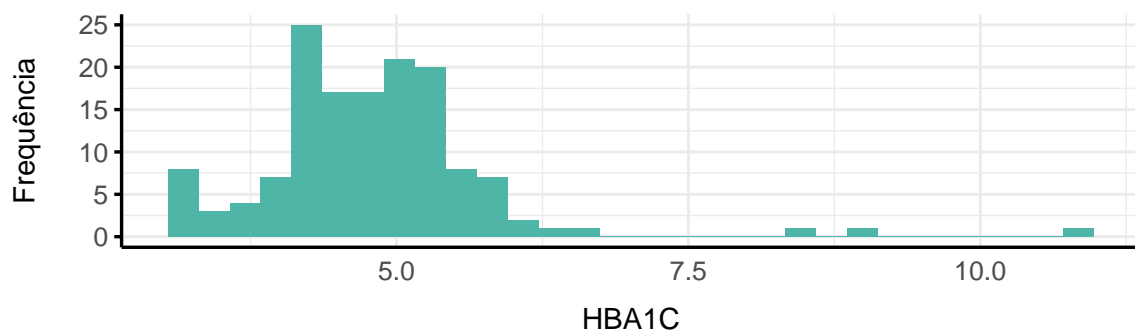
Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "hba1c")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 4.78 |
| Desvio Padrão | 0.98 |
| Valores nulos | 44.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "hba1c")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 2742, p-value = 0.1979
## alternative hypothesis: true location shift is greater than 0
```

creatinina:

Foi encontrado observação outlier, elas serão retiradas.

- Relação antes da retirada:

```
df1$creatinina %>% summary()
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##      0.03   0.60   0.82   1.09   1.00   52.00     3
```

```
df1$creatinina[df1$creatinina > 4]
```

```
## [1] NA NA NA 52
```

- Após retirada:

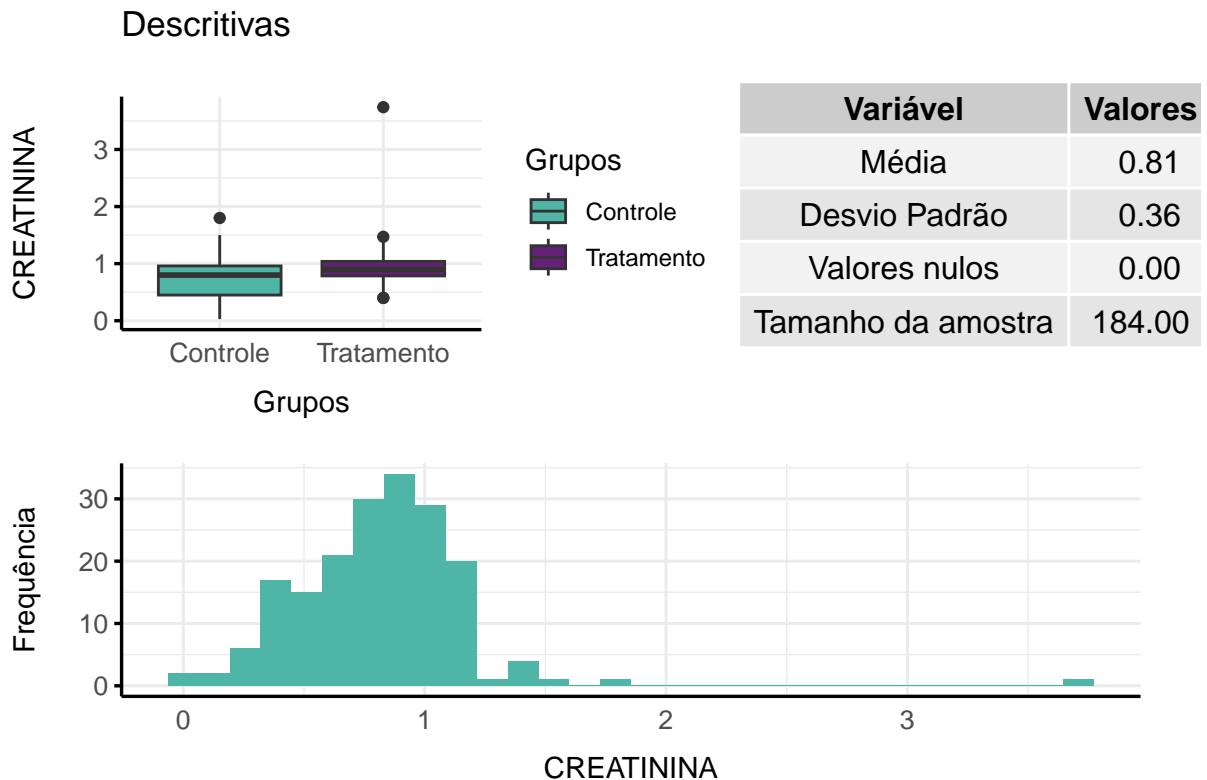
```
df_creatinina <- df1 %>%
  filter(creatinina < 4)
```

```
df_creatinina$creatinina %>% summary()
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0300  0.6000  0.8150  0.8135  1.0000  3.7400
```

Rejeita-se a hipótese nula.

```
graficos(df_creatinina, "grupo", "creatinina")
```



```
realizar_teste(df_creatinina, "creatinina")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 5491, p-value = 0.0001467
## alternative hypothesis: true location shift is greater than 0
```

filtr_glomerular:

Foi encontrada observação outlier, elas serão retiradas.

- Relação antes da retirada:

```
df1$filtr_glomerular %>% summary()
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.     NA's
##    1.77    79.38    98.27   989.28   143.16 153470.69      4
```

```
df1$filtr_glomerular[df1$filtr_glomerular > 3200]
```

```
## [1]      NA      NA      NA      NA 153470.7
```

- Após retirada:

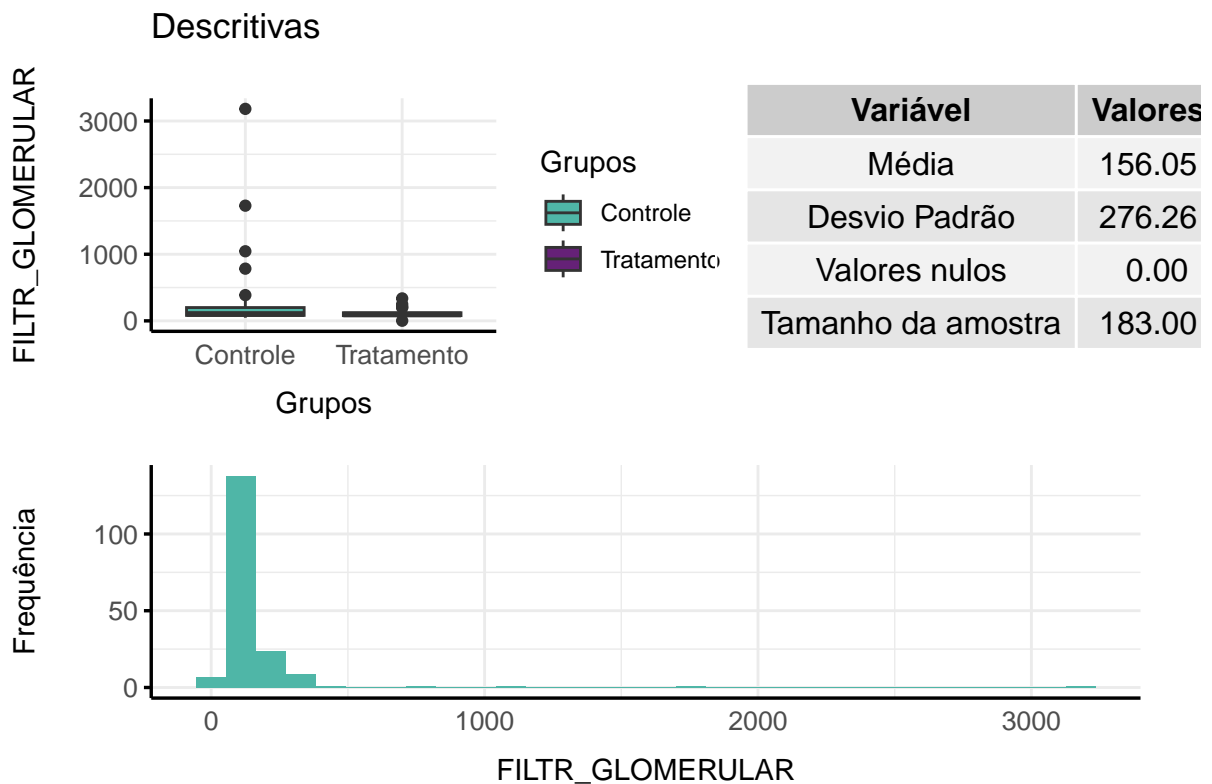
```
df_filtr_glomerular <- df1 %>%
  filter(filtr_glomerular < 3200)
```

```
df_filtr_glomerular$filtr_glomerular %>% summary()
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.771  79.365   98.207  156.051 141.810 3181.597
```

Não rejeita-se a hipótese nula.

```
graficos(df_filtr_glomerular, "grupo", "filtr_glomerular")
```



```
realizar_teste(df_filtr_glomerular, "filtr_glomerular")
```

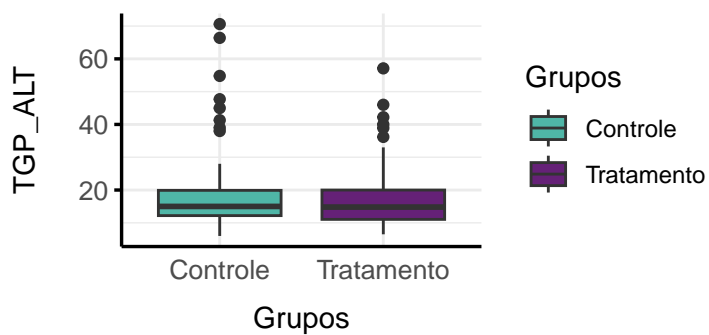
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 3411, p-value = 0.9809
## alternative hypothesis: true location shift is greater than 0
```

tgp_alt:

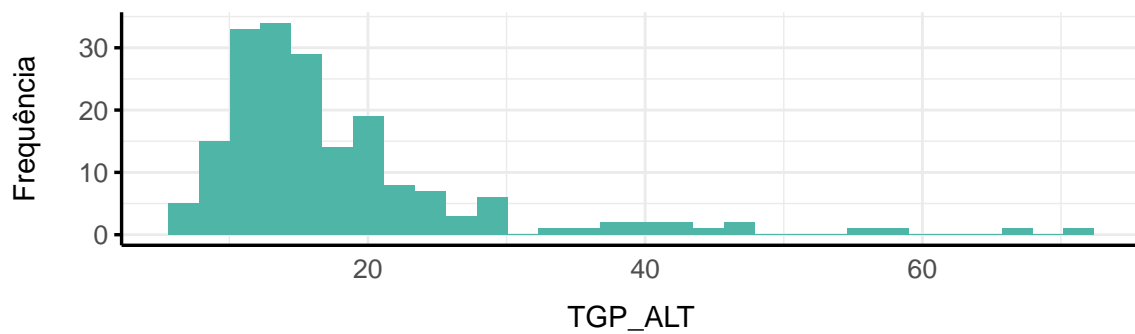
Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "tgp_alt")
```


Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 17.93 |
| Desvio Padrão | 10.20 |
| Valores nulos | 0.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "tgp_alt")
```

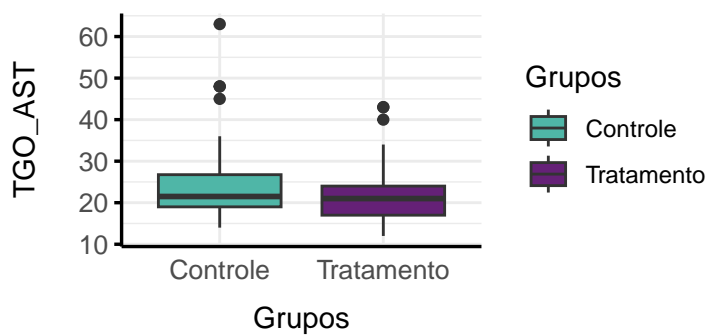
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 4075, p-value = 0.7991
## alternative hypothesis: true location shift is greater than 0
```

tgo_ast:

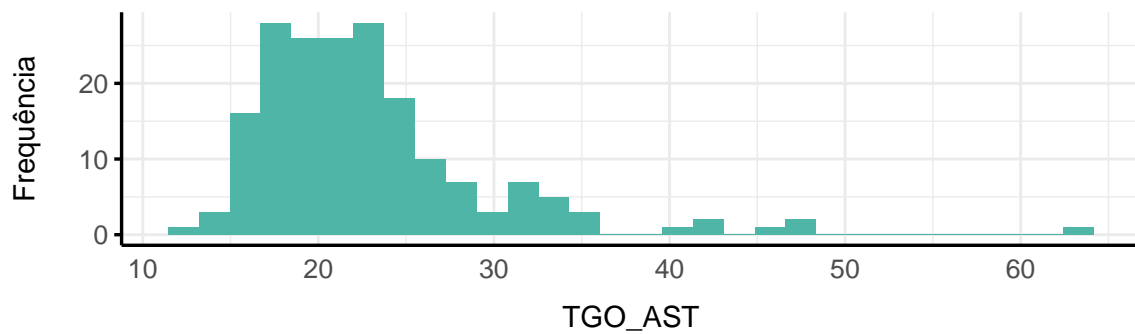
Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "tgo_ast")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 22.80 |
| Desvio Padrão | 6.95 |
| Valores nulos | 0.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "tgo_ast")
```

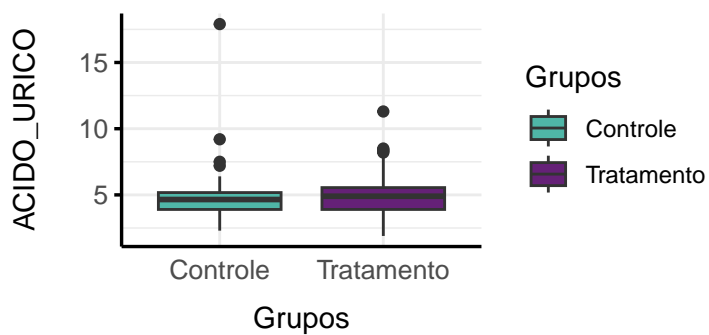
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 3829, p-value = 0.9338
## alternative hypothesis: true location shift is greater than 0
```

acido_urico:

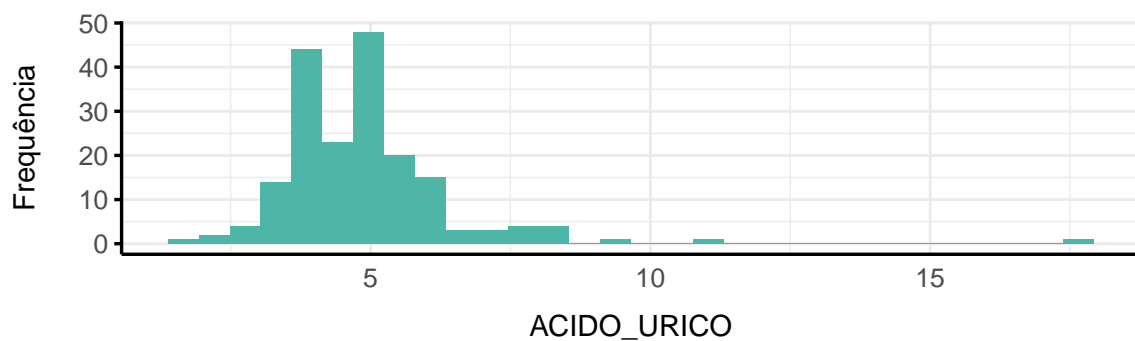
Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "acido_urico")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 4.88 |
| Desvio Padrão | 1.62 |
| Valores nulos | 0.00 |
| Tamanho da amostra | 188.00 |



```
realizar_teste(df1, "acido_urico")
```

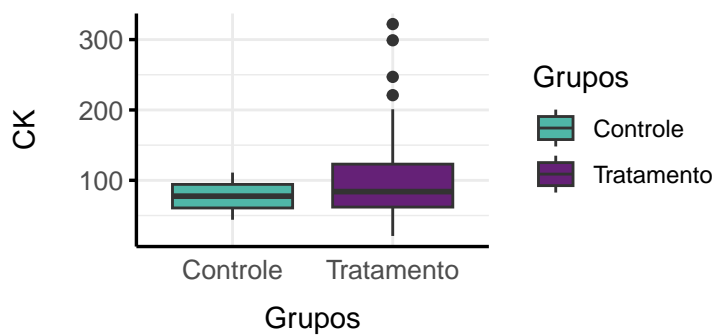
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 4980.5, p-value = 0.0549
## alternative hypothesis: true location shift is greater than 0
```

ck: ** Erro no teste

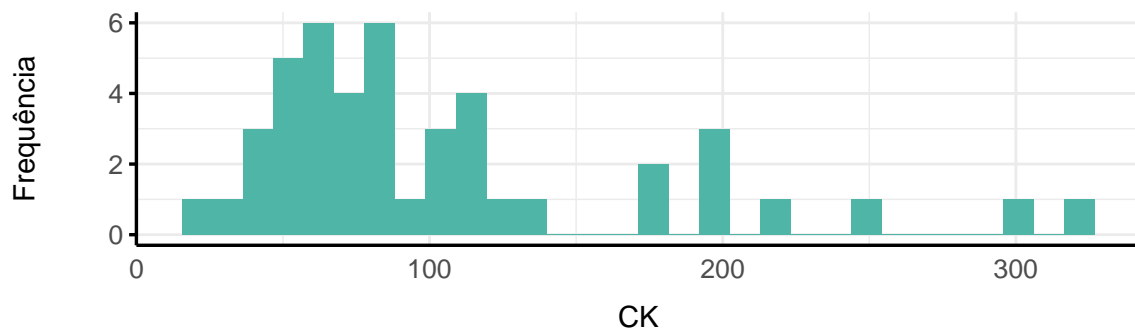
Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "ck")
```

Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 105.82 |
| Desvio Padrão | 69.11 |
| Valores nulos | 143.00 |
| Tamanho da amostra | 188.00 |



```
df_filtro <- df1 %>%
  drop_na(ck)

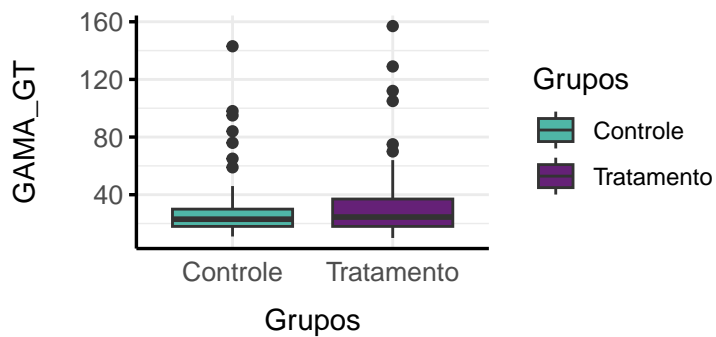
#realizar_teste(df_filtro, "ck")
```

gama_gt: ** Erro no teste

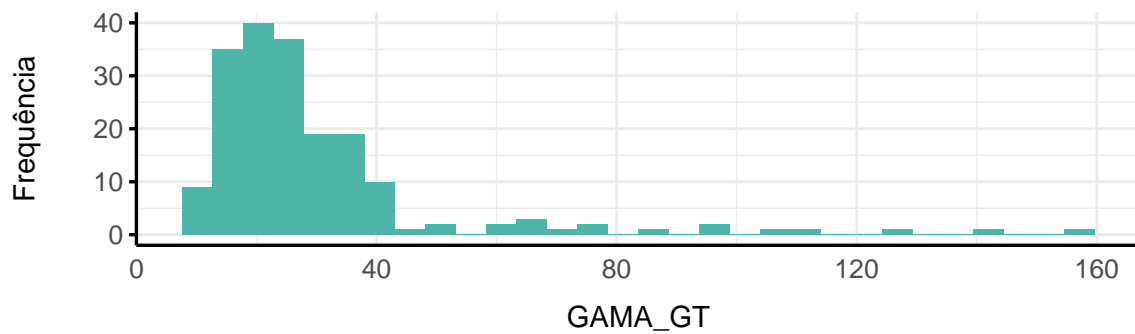
Não rejeita-se a hipótese nula.

```
graficos(df1, "grupo", "gama_gt")
```

Descritivas



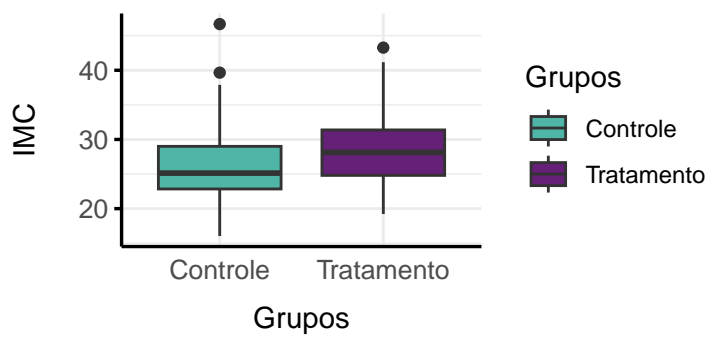
| Variável | Valores |
|--------------------|---------|
| Média | 29.79 |
| Desvio Padrão | 22.27 |
| Valores nulos | 0.00 |
| Tamanho da amostra | 188.00 |



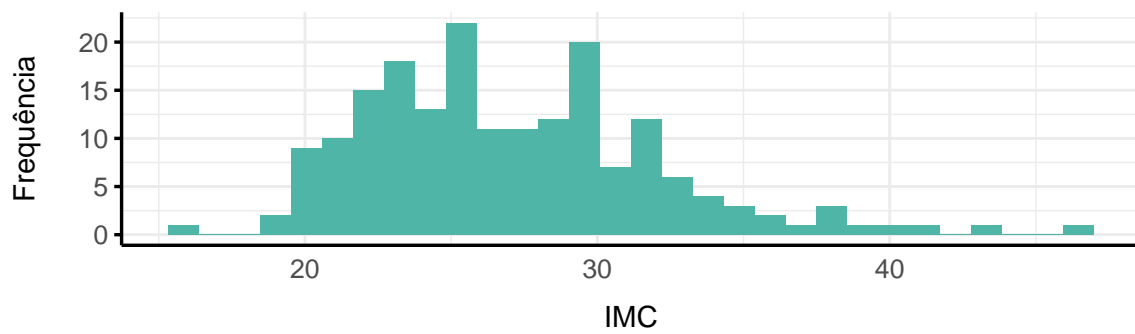
```
realizar_teste(df1, "gama_gt")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: grupo_MP and grupo_MN
## W = 4800.5, p-value = 0.1325
## alternative hypothesis: true location shift is greater than 0
## testes
graficos(df1, "grupo", "imc")
```

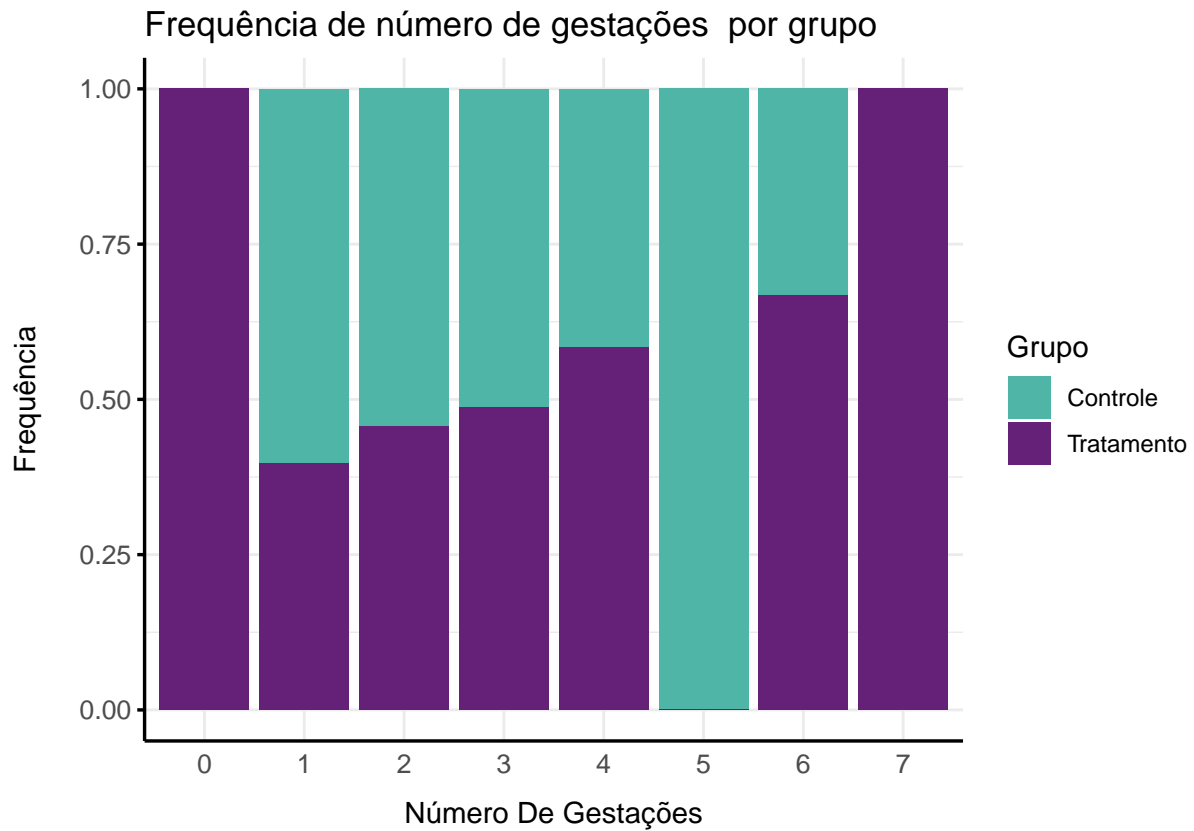
Descritivas



| Variável | Valores |
|--------------------|---------|
| Média | 27.06 |
| Desvio Padrão | 4.99 |
| Valores nulos | 1.00 |
| Tamanho da amostra | 188.00 |



```
graficos_categoricos(df1$grupo, df1$num_gestacoes, "número de gestações")
```



```
realizar_teste(df1, "idade")
```

```
##
## Welch Two Sample t-test
##
## data: grupo_MN and grupo_MP
## t = 1.2435, df = 181.34, p-value = 0.1076
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -0.3524145      Inf
## sample estimates:
## mean of x mean of y
## 40.69307 39.62353
```

Resumo Descritivas

Testes da curva

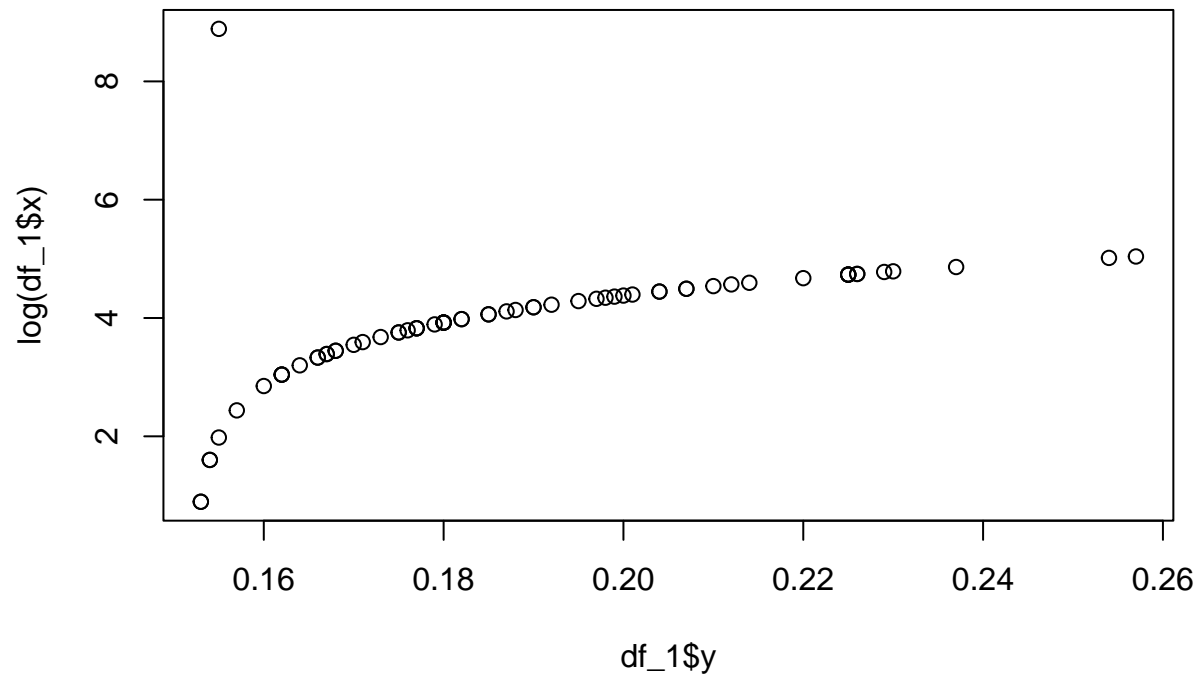
```
#rm(list = ls()) # limpa enviroment
```

- Curva com base nos dados

```
df_1 <- read_xlsx("dados_amostrais.xlsx")
```

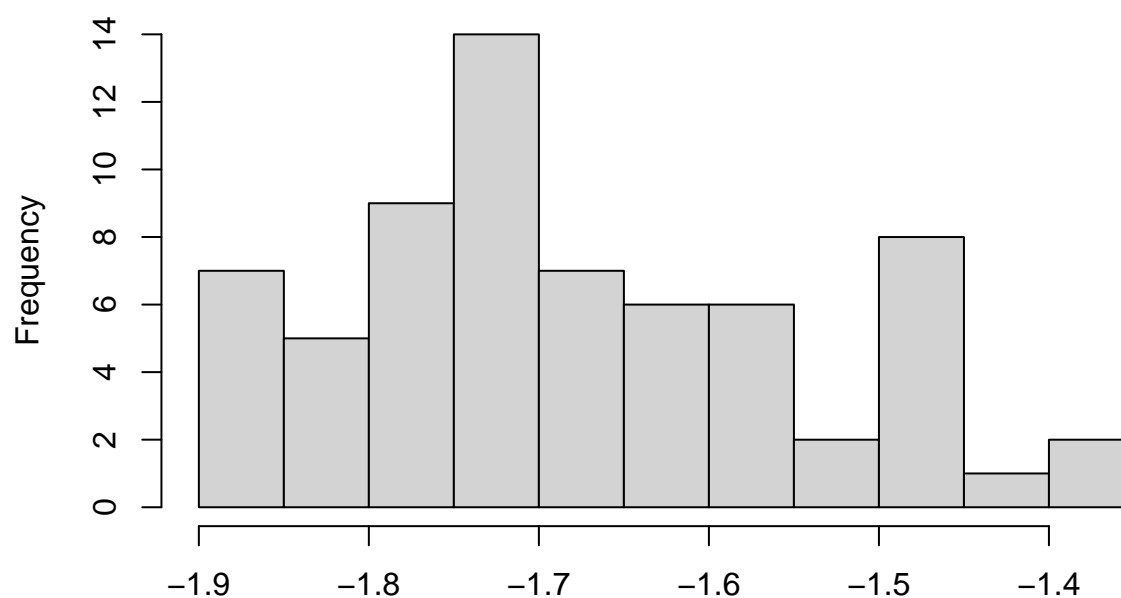
```
df_1 <- df_1 %>% select(x,y)
```

```
plot(df_1$y, log(df_1$x))
```



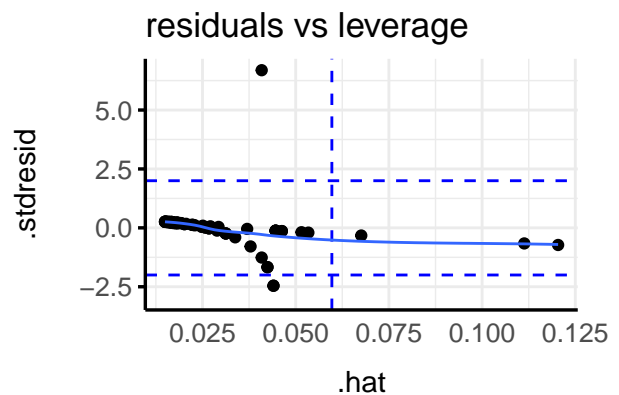
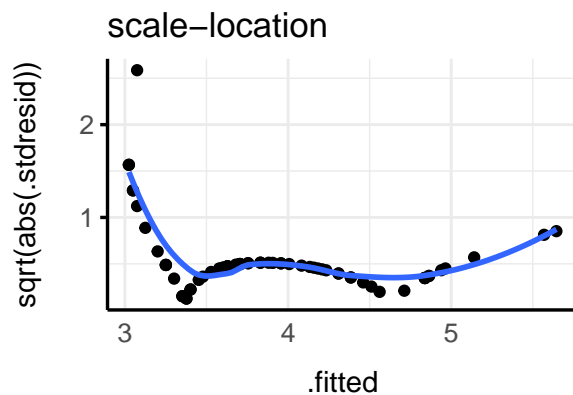
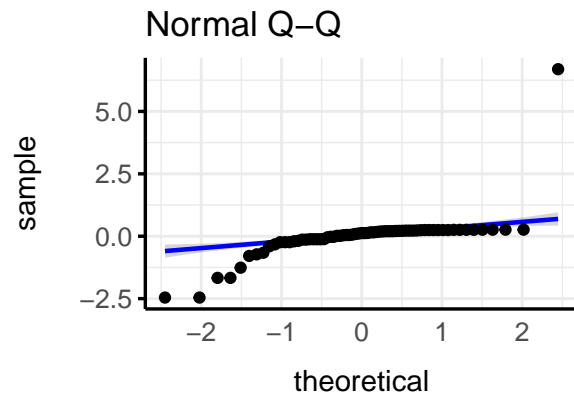
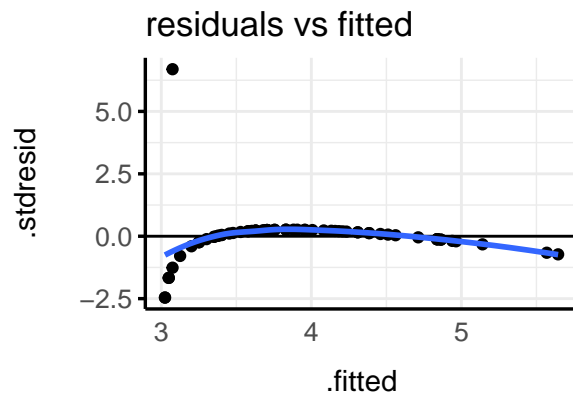
```
df_1$y %>% log() %>% hist()
```


Histogram of .



```
fit1 <- lm(log(x) ~ y, data = df_1)
fit1 %>% summary()
```

```
##
## Call:
## lm(formula = log(x) ~ y, data = df_1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1305 -0.1116  0.1095  0.2005  5.8128
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.8298     0.8052  -1.031   0.307
## y             25.1915     4.2256   5.962 1.13e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8871 on 65 degrees of freedom
## Multiple R-squared:  0.3535, Adjusted R-squared:  0.3436
## F-statistic: 35.54 on 1 and 65 DF,  p-value: 1.127e-07
reglin::ggresiduals(fit1)
```



```
#plot(fit1)

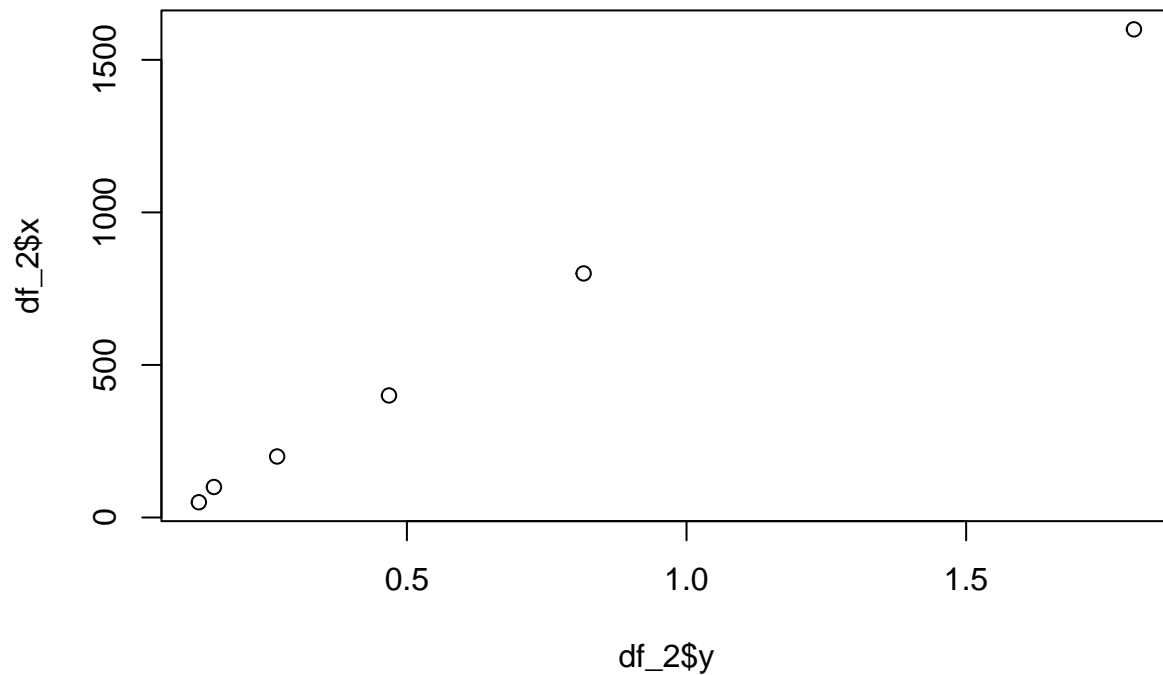
newdata <- data.frame(y = c(0.166, 0.048, .152, 0.210, .207))

newdata$amostra <- predict(fit1, newdata) %>% exp()
```

- Curva com base teórica

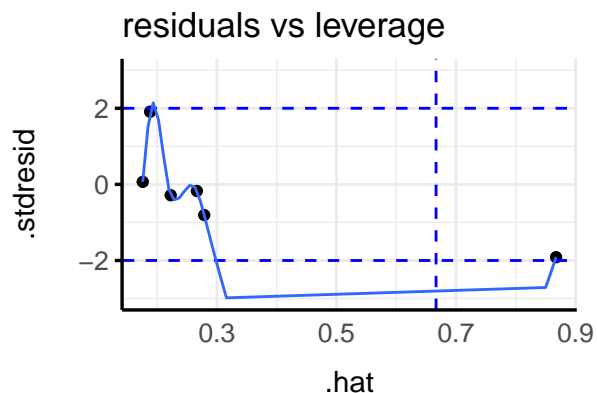
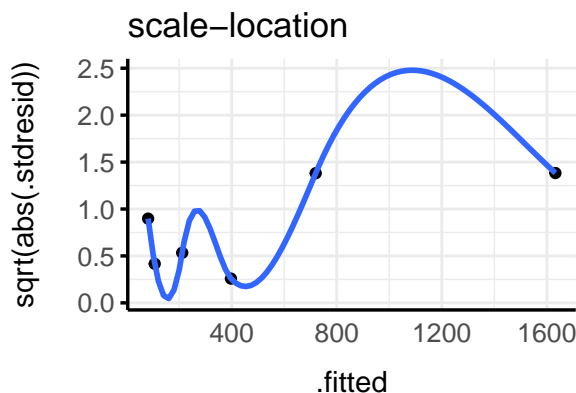
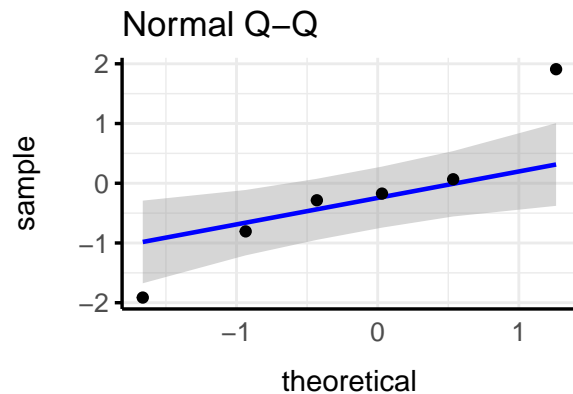
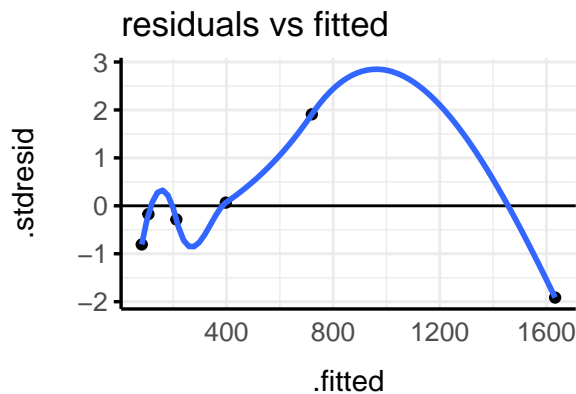
```
df_2 <- read_xlsx("dados_teoricos.xlsx")

plot(df_2$y, df_2$x)
```



```
fit2 <- lm(x ~ y, data = df_2)
fit2 %>% summary()
```

```
##
## Call:
## lm(formula = x ~ y, data = df_2)
##
## Residuals:
##      1      2      3      4      5      6
## -31.869  -6.908 -11.702   2.823  80.097 -32.441
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -36.84      27.44  -1.343   0.251
## y             927.38      32.65  28.405 9.14e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 46.57 on 4 degrees of freedom
## Multiple R-squared:  0.9951, Adjusted R-squared:  0.9938
## F-statistic: 806.8 on 1 and 4 DF, p-value: 9.141e-06
reglin::ggresiduals(fit2)
```



```
#plot(fit2)

newdata$teorico <- predict(fit2, newdata)
```

Resultado curvas

```
newdata
```

```
##      y  amostra  teorico
## 1 0.166 28.559805 117.109255
## 2 0.048  1.461404   7.678923
## 3 0.152 20.071856 104.125996
## 4 0.210 86.524568 157.913786
## 5 0.207 80.226490 155.131659
```

** Curva teorica vs amostral

```
eq1 <- function(x){
  y = 0.1532 + ( (0.8722 - 0.1532) / (1 + (x/606.3139)^(-1.3109)) )
  return(y)
}

eq2 <- function(x){
  y = 0.1173 + ( (0.202658 - 0.1173) / (1 + (x/10746.6371)^(-1.2605)) )
  return(y)
}
```

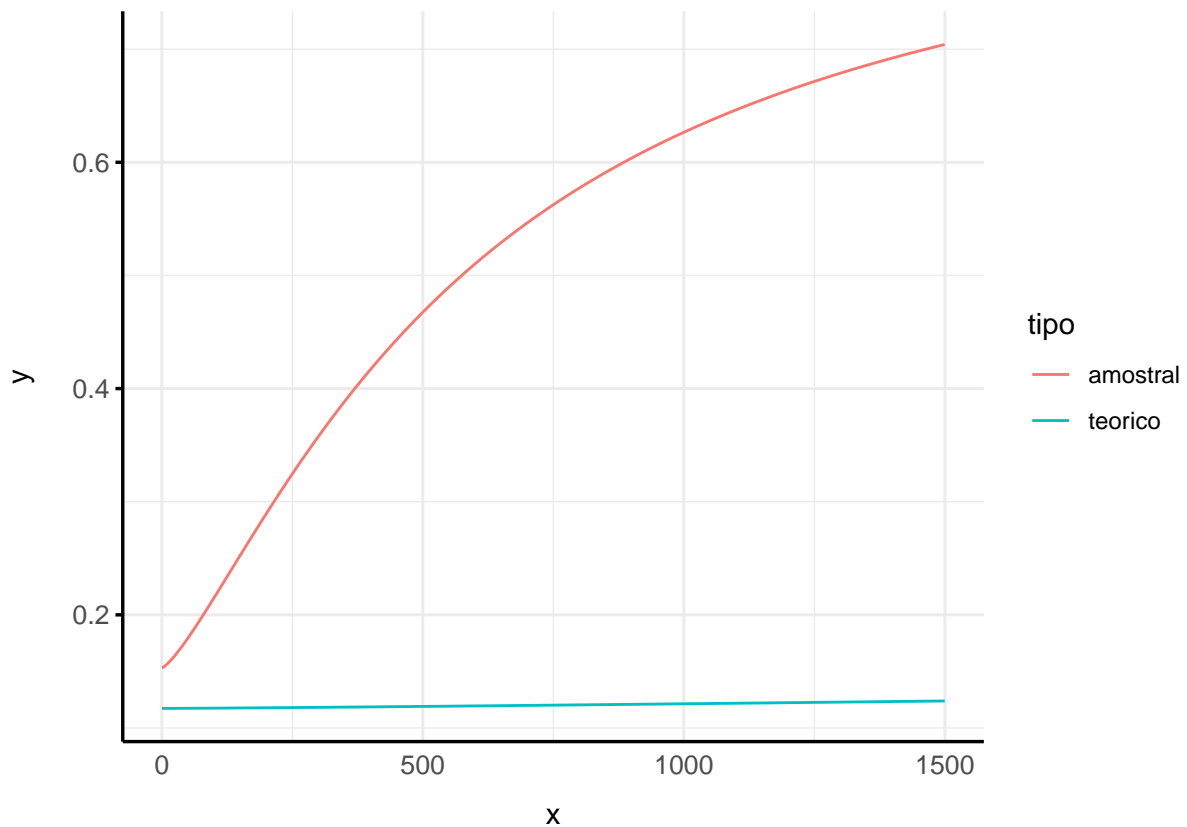
```

teste = tibble(x = seq(0.1,1500,0.5),
               y = eq1(x)
)

teste2 = tibble(x = seq(0.1,1500,0.5),
                y = eq2(x)
)

bind_rows(teste %>% mutate(tipo = 'amostral'),
          teste2 %>% mutate(tipo = 'teorico')) %>%
  ggplot(aes(x = x, y = y, color = tipo)) +
  geom_line(aes(group = tipo))

```



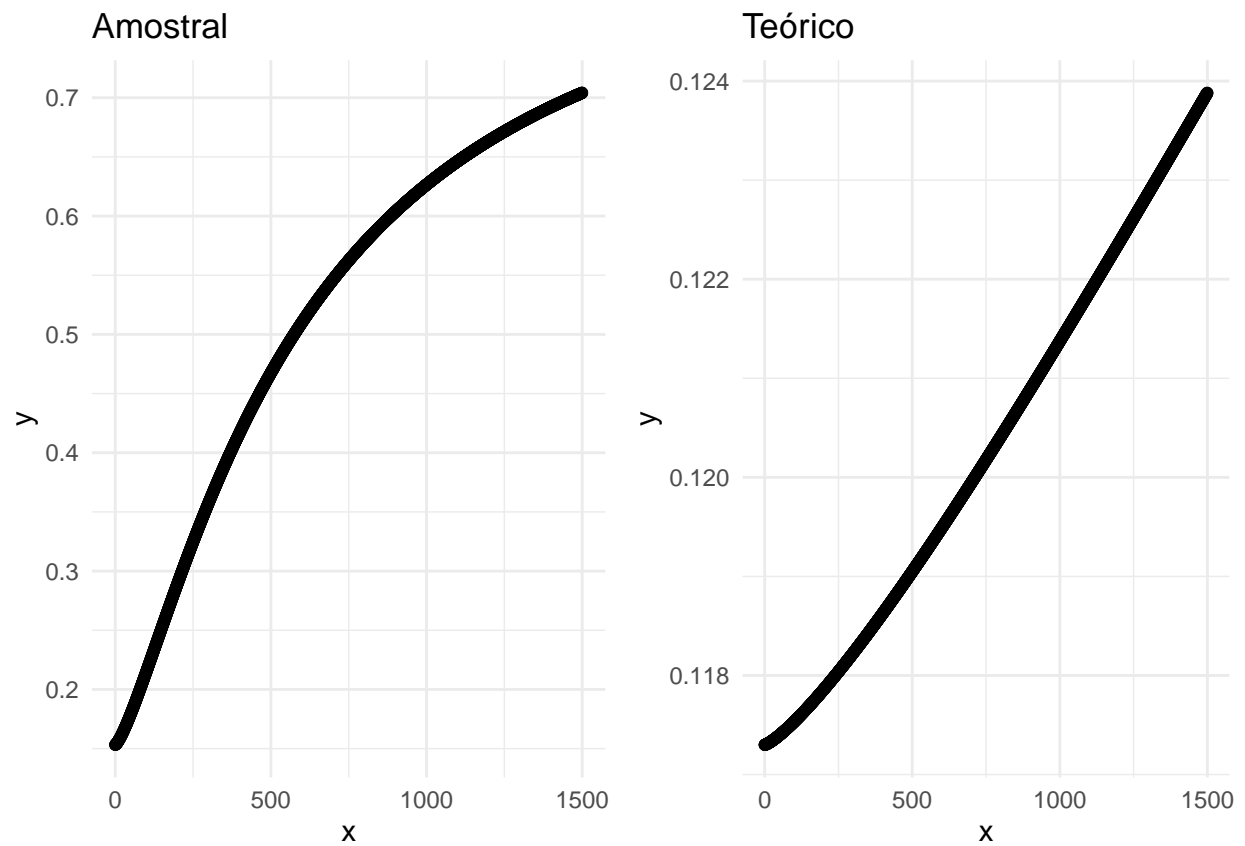
```

g1 <- ggplot(data = teste, aes(x = x, y = y)) +
  geom_line(group = 1) +
  geom_point() +
  theme_minimal() +
  labs(title = "Amostrado")

g2 <- ggplot(data = teste2, aes(x = x, y = y)) +
  geom_line(group = 1) +
  geom_point() +
  theme_minimal() +
  labs(title = "Teórico")

```

```
gridExtra::grid.arrange(g1, g2, ncol = 2)
```



```
# fazer a inversa das eq e dps buscar os x em y  
eq1(0.048)
```

```
## [1] 0.153203
```

```
eq2(0.048)
```

```
## [1] 0.1173
```