

# Tutoriais de Estatística

Laboratório de Estatística Aplicada

2021-02-09



# Contents

<b>Apresentação</b>	<b>5</b>
<b>Estatística Descritiva</b>	<b>7</b>
<b>Regressao Linear Múltipla</b>	<b>9</b>
<b>Regressão Logística</b>	<b>11</b>
<b>Estatística Não-Paramétrica</b>	<b>13</b>



# Apresentação

Essa página é uma iniciativa do Laboratório de Estatística Aplicada do Departamento de Estatística da UFRN. Nessa página é possível encontrar tutoriais em R para diversos métodos estatísticos no R.

Coordenação:

- Carla A. Vivacqua
- Marcus Nunes
- Talita Barros

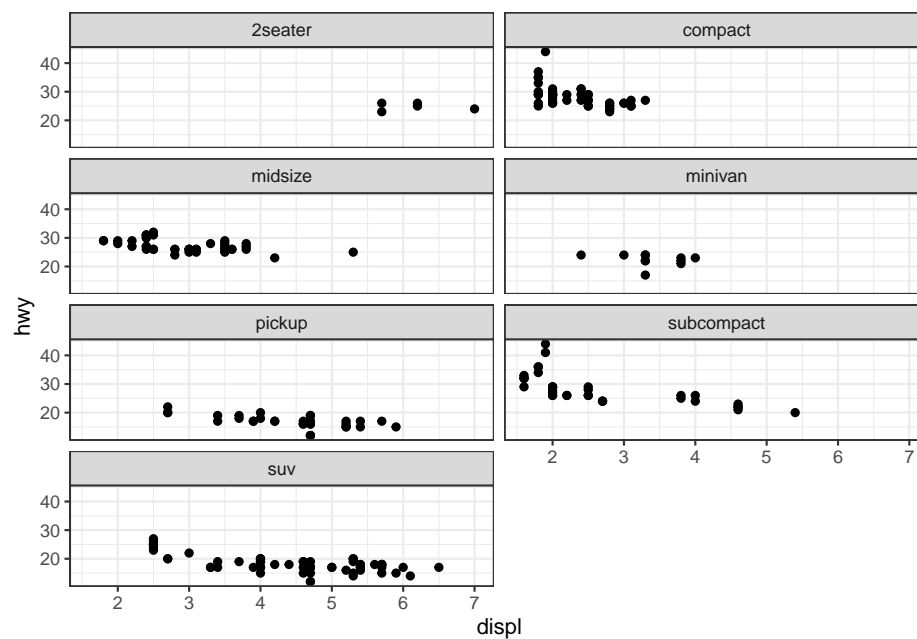


# Estatística Descritiva

Técnicas de visualização de dados.

```
library(ggplot2)
theme_set(theme_bw())

ggplot(mpg, aes(x = displ, y = hwy)) +
  geom_point() +
  facet_wrap(~ class, ncol = 2)
```







# Regressao Linear Múltipla

Alguma coisa sobre esse assunto.

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i$$

```
ajuste <- lm(Petal.Width ~ Petal.Length + Sepal.Width + Sepal.Length,  
             data = iris)  
summary(ajuste)
```

```
##  
## Call:  
## lm(formula = Petal.Width ~ Petal.Length + Sepal.Width + Sepal.Length,  
##     data = iris)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.60959 -0.10134 -0.01089  0.09825  0.60685   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  -0.24031    0.17837  -1.347    0.18      
## Petal.Length  0.52408    0.02449  21.399 < 2e-16 ***  
## Sepal.Width   0.22283    0.04894   4.553 1.10e-05 ***  
## Sepal.Length -0.20727    0.04751  -4.363 2.41e-05 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.192 on 146 degrees of freedom  
## Multiple R-squared:  0.9379, Adjusted R-squared:  0.9366   
## F-statistic: 734.4 on 3 and 146 DF,  p-value: < 2.2e-16
```



# Regressão Logística

Texto sobre regressão logística.

```
mydata <- read.csv("https://stats.idre.ucla.edu/stat/data/binary.csv")
mydata$rank <- factor(mydata$rank)

mylogit <- glm(admit ~ gre,
               data = mydata,
               family = "binomial")

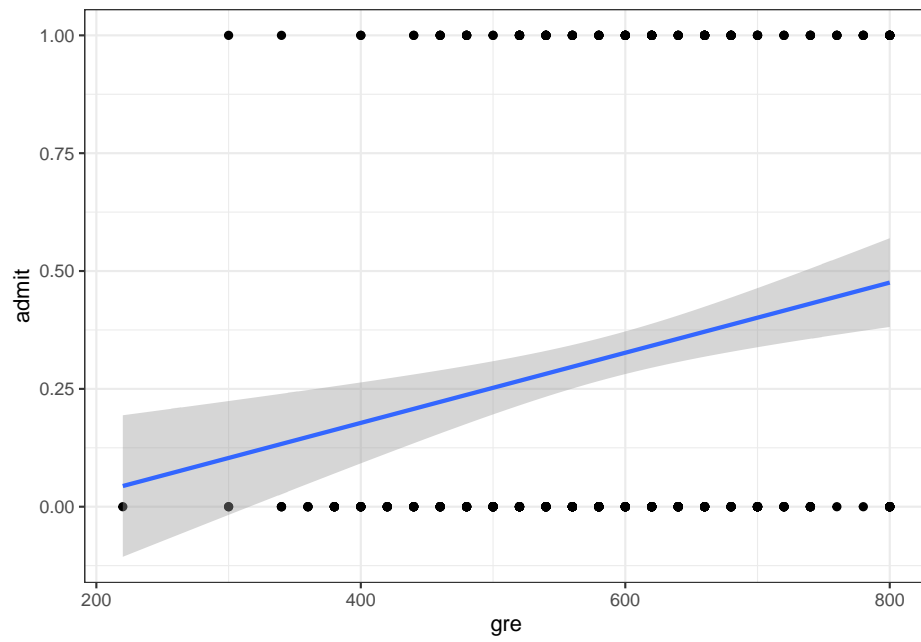
summary(mylogit)
```

```
##
## Call:
## glm(formula = admit ~ gre, family = "binomial", data = mydata)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1623  -0.9052  -0.7547   1.3486   1.9879
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.901344   0.606038  -4.787 1.69e-06 ***
## gre          0.003582   0.000986   3.633 0.00028 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 486.06  on 398  degrees of freedom
## AIC: 490.06
##
## Number of Fisher Scoring iterations: 4
```

```
library(ggplot2)
theme_set(theme_bw())

ggplot(mydata, aes(x = gre, y = admit)) +
  geom_point() +
  geom_smooth(method = "glm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



# Estatística Não-Paramétrica

Métodos usados quando não podemos assumir normalidade ou alguma outra distribuição de probabilidade em nossos dados.

```
x <- c(2.9, 3.0, 2.5, 2.6, 3.2) # normal subjects
y <- c(3.8, 2.7, 4.0, 2.4)      # with obstructive airway disease
z <- c(2.8, 3.4, 3.7, 2.2, 2.0) # with asbestosis

dados <- data.frame(eficiencia = c(x, y, z),
                    grupo = c(rep("normal", length(x)),
                               rep("obstrusivo", length(y)),
                               rep("abestose", length(z)))
                    )

kruskal.test(eficiencia ~ grupo,
             data = dados)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  eficiencia by grupo
## Kruskal-Wallis chi-squared = 0.77143, df = 2, p-value = 0.68

library(ggplot2)
theme_set(theme_bw())
library(reshape2)

ggplot(dados, aes(x = grupo, y = eficiencia)) +
  geom_boxplot()
```

