

Universidade Federal do Rio Grande do Norte
Instituto Metr pole Digital
IMD0601 - Bioestat stica

Estat stica Descritiva em R

Prof. Dr. Tetsu Sakamoto
Instituto Metr pole Digital - UFRN
Sala A224, ramal 182
Email: tetsu@imd.ufrn.br



Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601
```

```
> git pull
```

Estatística Descritiva

Medidas de Tendência Central

- Média
- Moda
- Mediana

Medidas de Dispersão

- Amplitude
- Desvio entre quartis
- Variância
- Desvio Padrão

Análise de dados normais

- Z-score
- Histogramas
- QQ-plot
- Teste Shapiro-Wilk
- Teste Kolmogorov-Smirnov

Medidas de associação

- Covariância
- Correlação de Pearson

Estatística descritiva em R

- `data(iris)`
- `summary(iris$Sepal.Length)`

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.300	5.100	5.800	5.843	6.400	7.900

Medidas de Tendência Central

- Média
 - `mean(iris$Sepal.Length)`
- Mediana
 - `median(iris$Sepal.Length)`
- Moda
 - `mfv(iris$Sepal.Length)` # do pacote **modeest**

Medidas de Dispersão

- Amplitude

- `max(iris$Sepal.Length)` # retorna valor máximo
- `min(iris$Sepal.Length)` # retorna valor mínimo
- `range(iris$Sepal.Length)` # retorna vetor com valor máximo e mínimo

- IQR (amplitude entre quartis)

- `quantile(iris$Sepal.Length)`
 - # retorna vetor com mínimo, máximo e os três quartis (0.25, 0.5, 0.75)
- `IQR(iris$Sepal.Length)`

Medidas de Dispersão

- Variância
 - `var(iris$Sepal.Length)` # retorna a variância amostral
- Desvio padrão
 - `sd(iris$Sepal.Length)` # retorna o desvio padrão amostral

stat.desc (da biblioteca pastecs)

<code>install.packages("pastecs")</code>		Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
	<code>nbr.val</code>	150.00	150.00	150.00	150.00
<code>library(pastecs)</code>	<code>nbr.null</code>	0.00	0.00	0.00	0.00
	<code>nbr.na</code>	0.00	0.00	0.00	0.00
<code>res <- stat.desc(iris[, -5])</code>	<code>min</code>	4.30	2.00	1.00	0.10
	<code>max</code>	7.90	4.40	6.90	2.50
<code>res</code>	<code>range</code>	3.60	2.40	5.90	2.40
	<code>sum</code>	876.50	458.60	563.70	179.90
<code>res["mean",]</code>	<code>median</code>	5.80	3.00	4.35	1.30
	<code>mean</code>	5.84	3.06	3.76	1.20
	<code>SE.mean</code>	0.07	0.04	0.14	0.06
	<code>CI.mean.0.95</code>	0.13	0.07	0.28	0.12
	<code>var</code>	0.69	0.19	3.12	0.58
	<code>std.dev</code>	0.83	0.44	1.77	0.76
	<code>coef.var</code>	0.14	0.14	0.47	0.64

Casos em que há dados faltantes

A função **mean()**, por exemplo, retornará **NA** se houver um dado faltante. Para calcular a média ignorando estes valores basta colocar **TRUE** para o parâmetro **na.rm**:

```
mean(iris$Sepal.Length, na.rm = TRUE)
```

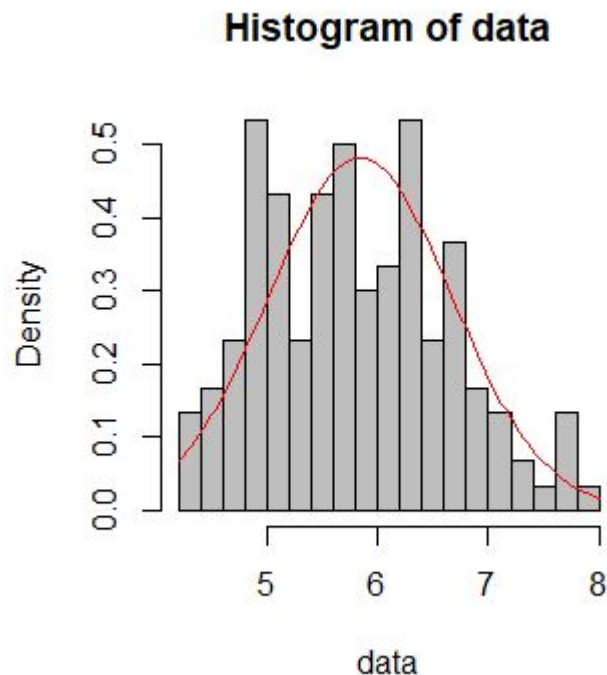
Verificar se os dados seguem uma normal

```
# histograma com curva da normal
```

```
data <- iris$Sepal.Length
```

```
hist(data, breaks=20,freq = FALSE, col  
= "grey")
```

```
curve(dnorm(x,mean=mean(data),sd=sd(da  
ta)), col = 2, add = TRUE)
```



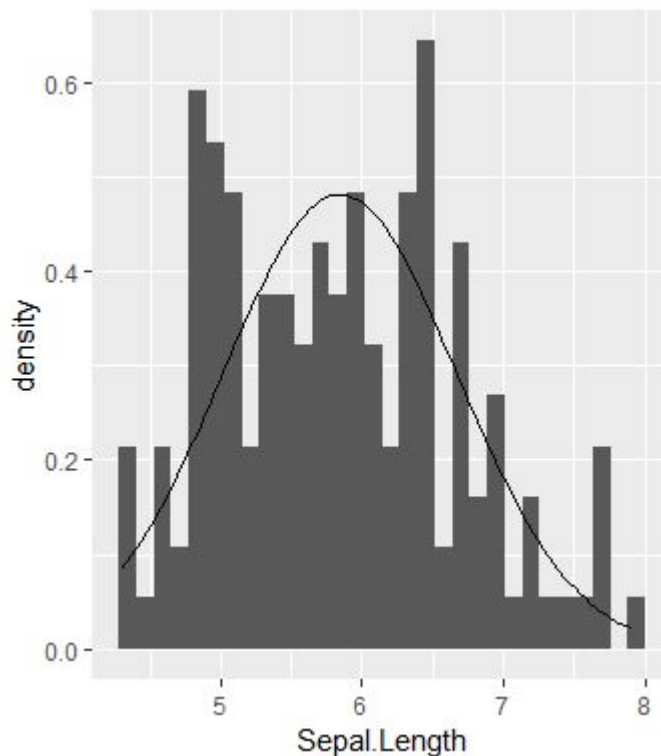
Verificar se os dados seguem uma normal

```
# histograma com curva da normal

gg <- ggplot(iris,
aes(x=Sepal.Length))

gg <- gg +
geom_histogram(aes(y=..density..))

gg <- gg + stat_function(fun=dnorm,
args =
list(mean=mean(iris$Sepal.Length), sd
= sd(iris$Sepal.Length)))
```

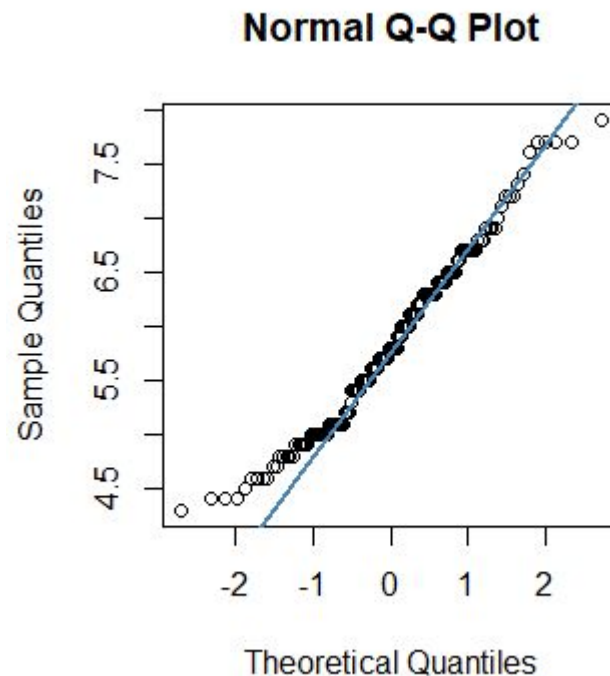


Verificar se os dados seguem uma normal

```
# qq-plot with R basic plot
```

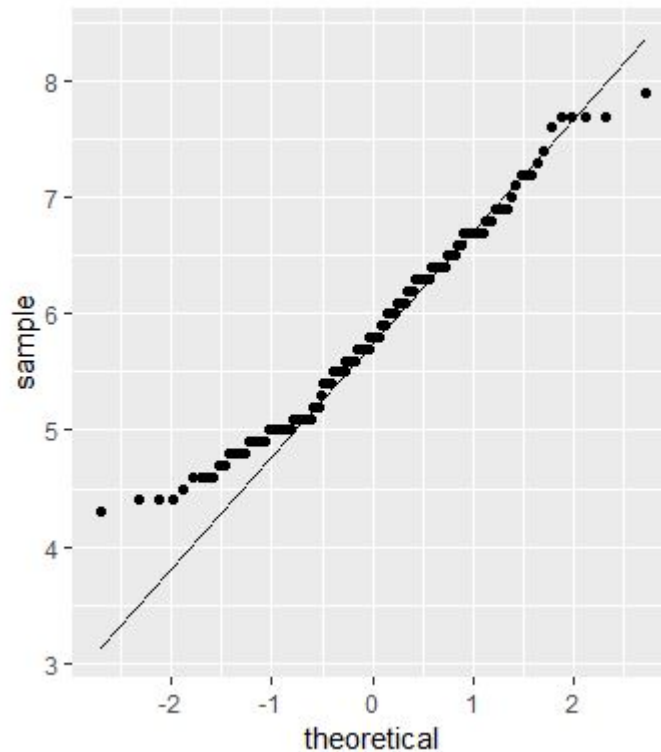
```
qqnorm(iris$Sepal.Length)
```

```
qqline(iris$Sepal.Length, col =  
"steelblue", lwd = 2)
```



Verificar se os dados seguem uma normal

```
# qq-plot with ggplot  
  
g <- ggplot(iris, aes(sample =  
  Sepal.Length))  
  
g + geom_qq + geom_qq_line()
```

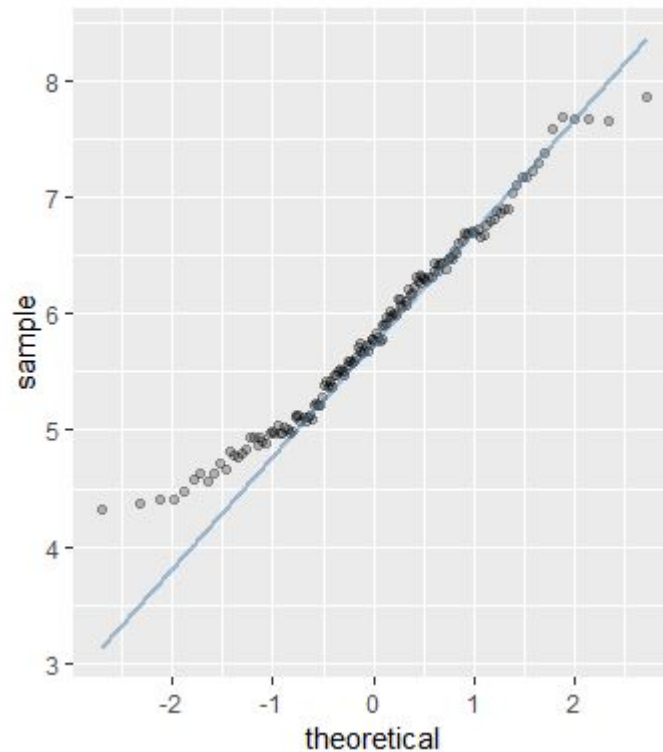


Verificar se os dados seguem uma normal

```
# qq-plot with ggplot
```

```
g <- ggplot(iris, aes(sample =  
Sepal.Length))
```

```
g+ geom_qq(position="jitter",  
alpha=0.25) + geom_qq_line(size=1,  
color="steelblue",alpha=0.5)
```



Verificar se os dados seguem uma normal

```
# Teste de Shapiro-Wilk
```

```
shapiro.test(iris$Sepal.Length)
```

```
# Teste Kolmogorov-Smirnov
```

```
data <- iris$Sepal.Length
```

```
ks.test(data, "pnorm", mean(data),  
sd(data))
```

Shapiro-Wilk normality test

data: iris\$Sepal.Length
W = 0.97609, p-value = 0.01018

One-sample Kolmogorov-Smirnov test

data: data
D = 0.088654, p-value = 0.1891
alternative hypothesis: two-sided

Covariância e Correlação

- Covariância
 - `cov(iris$Sepal.Length, iris$Sepal.Width)`
- Correlação de Pearson
 - `cor(iris$Sepal.Length, iris$Sepal.Width)`