

Análise Exploratória de Dados (revisão)

CE219 - Controle Estatístico de Qualidade

Prof. Cesar Taconeli
taconeli@ufpr.br

Prof. Walmes Zeviani
walmes@ufpr.br

Laboratório de Estatística e Geoinformação
Departamento de Estatística
Universidade Federal do Paraná

Métodos estatísticos para a análise da qualidade

- ▶ O objetivo aqui é fazer uma (breve) revisão de **modelos probabilísticos** e **métodos estatísticos** com aplicação na descrição, modelagem e produção de inferências para processos.
- ▶ Serão abordados:
 1. **Métodos de análise descritiva.**
 2. Probabilidade e principais modelos probabilísticos.
 3. Inferência estatística aplicada à qualidade do processo.
- ▶ As próximas aulas serão intercaladas com ilustrações no R e os scripts disponibilizados na página da disciplina.



Métodos descritivos

Métodos descritivos



- ▶ Há uma grande variedade de métodos descritivos disponíveis, cuja aplicação dependerá dos objetivos da análise, da quantidade de variáveis sob estudo (e de suas respectivas escalas, tipo de valor, etc).
- ▶ As principais ferramentas estatísticas são os:
 - ▶ gráficos,
 - ▶ tabelas e
 - ▶ resumos numéricos.

Métodos descritivos: gráficos

Alguns gráficos úteis na análise de qualidade de processos:

- ▶ Gráfico de ramos e folhas;
- ▶ Histograma;
- ▶ Gráfico de série;
- ▶ Gráfico de dispersão;
- ▶ Gráfico de Pareto;
- ▶ Gráfico de caixa (box-plot);
- ▶ Gráficos de probabilidade (quantil-quantil).

Dados de precipitação de 70 cidades. Veja ?precip.

```
x <- unname(sort(precip))  
x <- append(round(x, digits = 0), rep(NA, 2))  
matrix(x, ncol = 8)
```

1
2
3

```
#      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]  
# [1,]    7   15   30   34   37   41   43   49  
# [2,]    7   16   30   35   38   41   45   54  
# [3,]    8   17   31   35   39   42   46   55  
# [4,]    8   17   31   36   39   42   46   57  
# [5,]   12   21   31   36   39   42   46   59  
# [6,]   13   23   31   36   39   43   48   60  
# [7,]   14   25   32   36   40   43   48   67  
# [8,]   15   26   32   36   40   43   48   NA  
# [9,]   15   29   33   37   40   43   49   NA
```

Diagrama de **ramos e folhas** construído com os dados em precip.

`stem(x)`

1

```
#  
# The decimal point is 1 digit(s) to the right of the |  
#  
# 0 | 7788  
# 1 | 234555677  
# 2 | 13569  
# 3 | 001111223455666667789999  
# 4 | 00011222333333566688899  
# 5 | 4579  
# 6 | 07
```

```
hist(precip, main = NA, col = "orange",  
     xlab = "Precipitação", ylab = "Frequência")  
rug(precip)
```

1
2
3

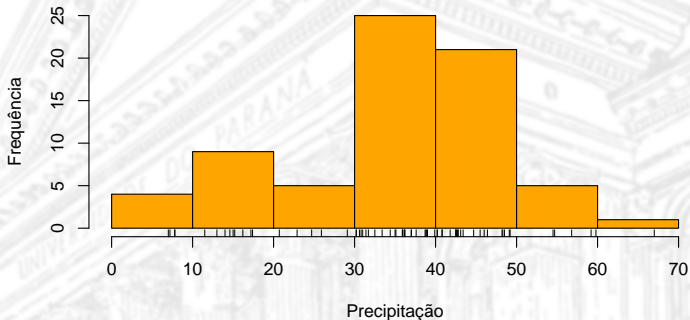


Figura 1. Histograma construído com o precip.

```
library(tidyverse)
tidyverse_packages()
```

1
2

```
# [1] "broom"          "cli"            "crayon"         "dplyr"          "dbplyr"
# [6] "forcats"       "ggplot2"       "haven"          "hms"           "httr"
# [11] "jsonlite"     "lubridate"     "magrittr"      "modelr"        "purrr"
# [16] "readr"        "readxl\n(>=" "reprex"        "rlang"         "rstudioapi"
# [21] "rvest"        "stringr"       "tibble"        "tidyr"         "xml2"
# [26] "tidyverse"
```

```
ggplot(data = data.frame(precip = precip),  
       mapping = aes(x = precip)) +  
  geom_histogram(bins = 7, fill = "orange", col = "black") +  
  geom_rug() + xlab("Precipitação") + ylab("Frequência")
```

1
2
3
4

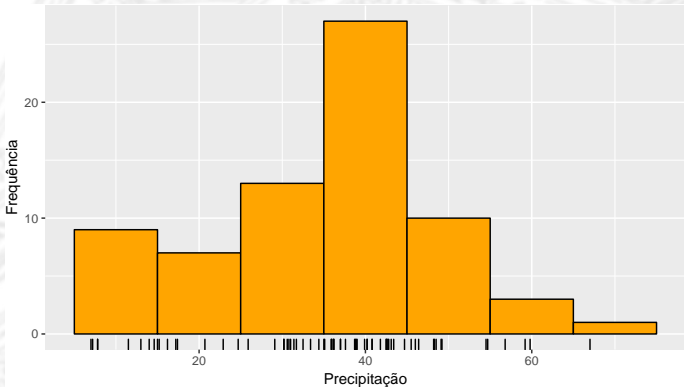


Figura 2. Histograma construído com o precip.

```
plot(Unemployed ~ Year, data = longley,  
     xlab = "Ano", ylab = "Desempregados",  
     type = "o", pch = 19)
```

1
2
3

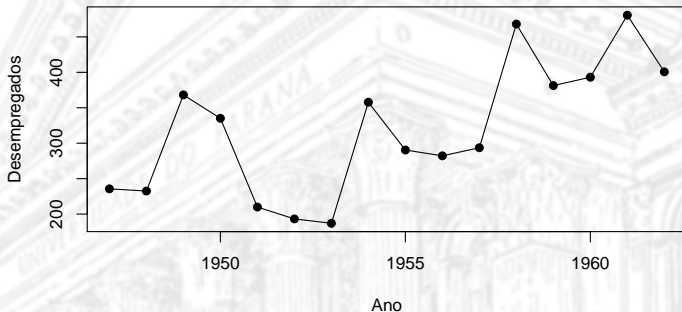


Figura 3. Gráfico de série construído com o longley.

```
ggplot(data = longley,  
       mapping = aes(x = Year, y = Unemployed)) +  
  geom_point() +  
  geom_line() +  
  xlab("Ano") + ylab("Desempregados")
```

1
2
3
4
5



Figura 4. Gráfico de série construído com o longley.

```
plot(dist ~ speed, data = cars,  
      xlab = "Velocidade", ylab = "Distância")
```

1
2

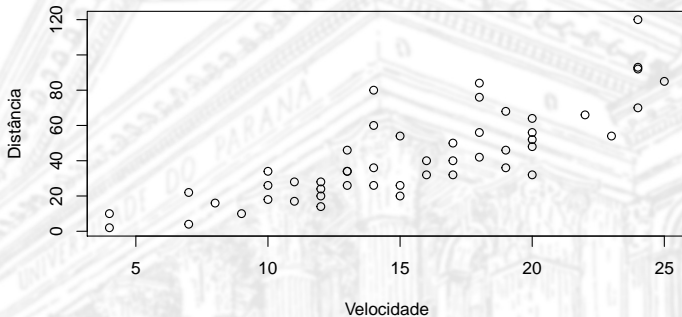


Figura 5. Diagrama de dispersão construído com o cars.

```
ggplot(data = cars,  
       mapping = aes(y = dist, x = speed)) +  
  geom_point() +  
  xlab("Velocidade") + ylab("Distância")
```

1
2
3
4

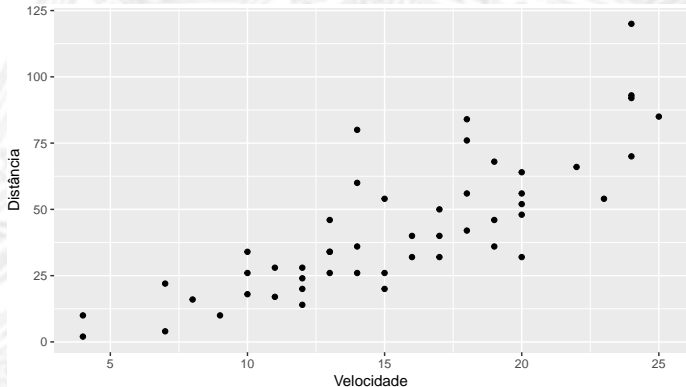


Figura 6. Diagrama de dispersão construído com o cars.

```
barplot(tail(sort(precip), n = 20),  
        horiz = TRUE,  
        las = 1,  
        xlab = "Precipitação", ylab = "Cidade (ordenada)")
```

1
2
3
4

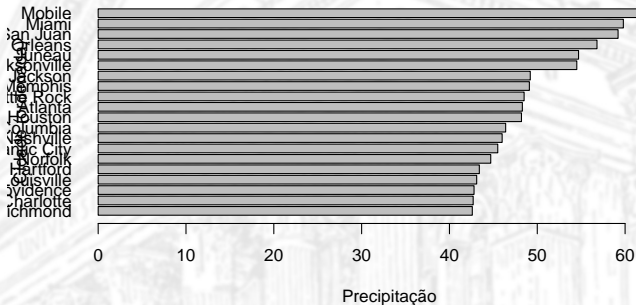


Figura 7. Gráfico de Pareto construído com o precip.

```
da <- cbind(as.data.frame(precip), city = names(precip))
da <- da[order(da$precip, decreasing = TRUE), ]
ggplot(data = head(da, n = 20),
       mapping = aes(x = reorder(city, precip), y = precip)) +
  geom_col() + coord_flip() +
  xlab("Precipitação") + ylab("Cidade (ordenada)")
```

1
2
3
4
5
6

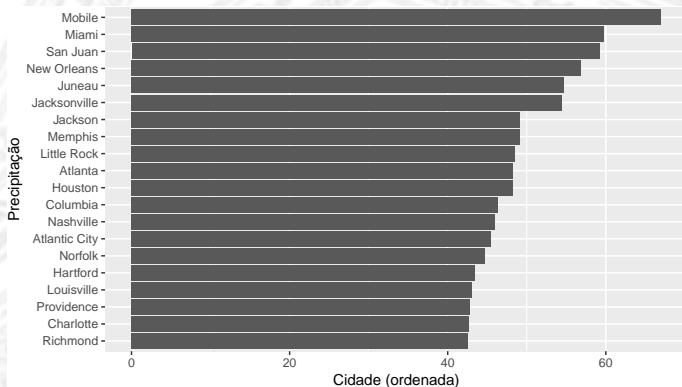


Figura 8. Gráfico de Pareto construído com o precip.


```
boxplot(count ~ spray, data = InsectSprays,  
        xlab = "Tipo de spray", ylab = "Contagem de insetos",  
        col = "orange")
```

1
2
3

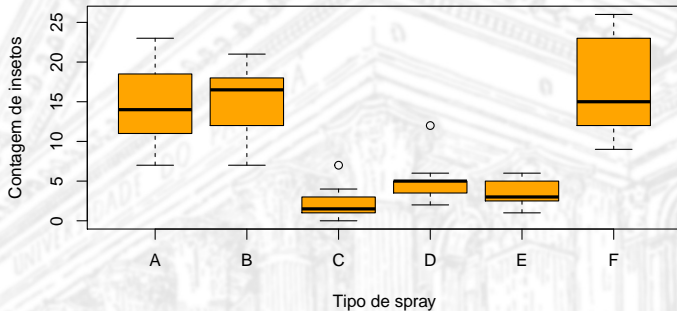


Figura 9. Gráfico de caixas e bigodes construído com o InsectSprays.

```
ggplot(data = InsectSprays,  
       mapping = aes(x = spray, y = count)) +  
  geom_boxplot(fill = "orange") +  
  xlab("Tipo de spray") + ylab("Contagem de insetos")
```

1
2
3
4

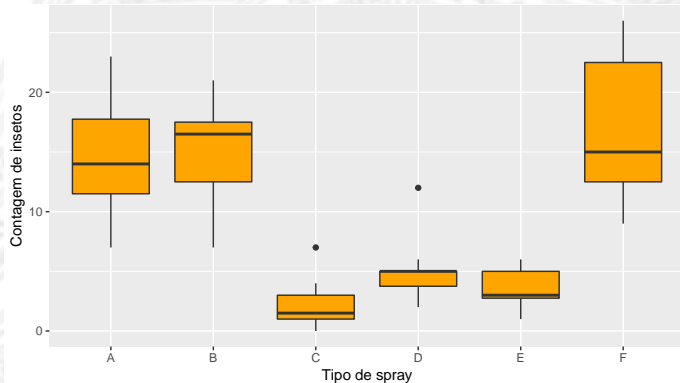


Figura 10. Gráfico de caixas e bigodes construído com o InsectSprays.

```
qqnorm(precip, xlab = "Quantis teóricos", ylab = "Quantis observados")  
qqline(precip, lty = 2)
```

1
2

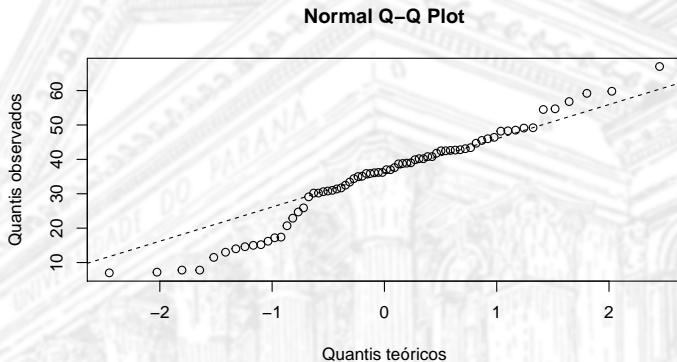


Figura 11. Gráfico quantil-quantil construído com o precip.

```
da <- data.frame(precip = precip)
ggplot(data = da, mapping = aes(sample = precip)) +
  stat_qq() +
  stat_qq_line() +
  xlab("Quantis teóricos") +
  ylab("Quantis observados")
```

1
2
3
4
5
6

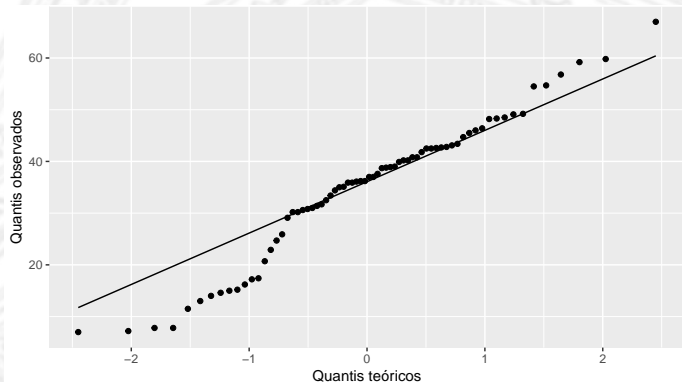


Figura 12. Gráfico quantil-quantil construído com o precip.

Métodos descritivos: medidas resumo

Principais medidas resumo:

- ▶ **Localção:**
 - ▶ Média, média aparada, média ponderada.
 - ▶ Mediana (* também é uma separatriz).
 - ▶ Moda (valor “mais freqüente”).
- ▶ **Dispersão:**
 - ▶ Variância e desvio padrão.
 - ▶ Desvio absoluto médio da mediana.
 - ▶ Coeficiente de variação.
 - ▶ Amplitude.
- ▶ **Separação/delimitação:**
 - ▶ Mínimo e máximo (extremos).
 - ▶ Quartis e demais percentis.