

# Introdução à Estatística usando o R com Aplicação em Análises Laboratoriais

Profa Carolina & Prof Gilberto

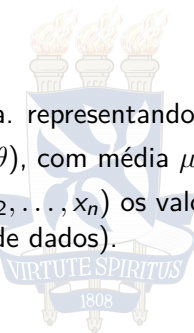
Instituto de Matemática e Estatística  
Universidade Federal da Bahia

26 de outubro de 2019

# Inferência Estatística

Considere:

- $X$  uma v.a. representando uma característica de uma população.
- $X \sim f(x; \theta)$ , com média  $\mu$  e variância  $\sigma^2$ .
- $\mathbf{x} = (x_1, x_2, \dots, x_n)$  os valores observados (amostra observada, conjunto de dados).



Universidade  
Federal da Bahia

# Estimação pontual

**Estimação pontual:** processo que usa os dados amostrais para calcular um único valor que serve como um “melhor palpite” de um parâmetro populacional desconhecido (por exemplo, a média populacional).

Para estimar  $\theta$ , nós selecionamos um estimador e avaliamos o mesmo com os dados observados,  $\mathbf{x}$ .

Se o verdadeiro valor de  $\theta$  é  $\theta_0$  (desconhecido), ao observar  $\mathbf{x}$  e utilizar o estimador, nossa sugestão para  $\theta_0$  (desconhecido) é a estimativa.

# Estimação pontual

## Principais problemas de estimação pontual para uma população:

Estimar a média  $\mu$  de uma única população:

- Estimativa:  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ .

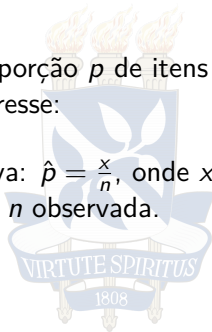
Estimar a variância  $\sigma^2$  (ou desvio-padrão  $\sigma = \sqrt{\sigma^2}$ ) de uma única população:

- Estimativa:  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ .

# Estimação pontual

Estimar a proporção  $p$  de itens em uma população que pertencem à uma classe de interesse:

- Estimativa:  $\hat{p} = \frac{x}{n}$ , onde  $x$  é número de sucessos na amostra de tamanho  $n$  observada.



UFBA  
Universidade  
Federal da Bahia

# Estimação pontual

## Principais problemas de estimação pontual para duas populações independentes:

Estimar a diferença entre as médias de duas populações independentes  $\mu_1 - \mu_2$ :

- Estimativa:  $\bar{x} - \bar{y}$ , onde  $\bar{x}$  é a média da amostra observada da população 1 e  $\bar{y}$  é a média da amostra observada da população 2.

# Estimação pontual

Estimar a diferença entre as proporções de duas populações independentes  $p_1 - p_2$ :

- Estimativa:  $\hat{p}_1 - \hat{p}_2 = \frac{x}{n_1} - \frac{y}{n_2}$ , onde  $x$  é número de sucessos na amostra de tamanho  $n_1$  observada da população 1 e  $y$  é número de sucessos na amostra de tamanho  $n_2$  observada da população 2.

# Estimação intervalar

## Intervalo de confiança (IC):

Para um parâmetro populacional desconhecido  $\theta$ , um IC é um intervalo da forma  $[L(\mathbf{x}), U(\mathbf{x})]$ .

O objetivo ao se construir IC's é ter uma medida de qual é a possível magnitude do erro que está sendo cometido é preciso de uma outra forma para obter estimativas.

Na prática, o intervalo de confiança para um parâmetro populacional  $\theta$  será da forma:

$$\hat{\theta} \pm E,$$

onde  $\hat{\theta}$  é a estimativa pontual e  $E$  é a margem de erro.



# Estimação intervalar

A ideia é construir um intervalo em torno da estimativa pontual, de modo que ele contenha o verdadeiro valor do parâmetro com um grau de confiança pré-fixado.

Seja  $\alpha$  o nível de significância e  $1 - \alpha$  o nível de confiança do IC.

Para o intervalo numérico calculado com os dados observados, dizemos que os dados fornecem evidências de que o verdadeiro valor de  $\theta$  encontra-se no intervalo calculado com confiança de  $1 - \alpha$ . Isto é, se obtivermos 100 amostras com  $n$  observações cada, esperamos que  $(1 - \alpha)100\%$  delas forneçam intervalos numéricos que contenha o verdadeiro valor de  $\theta$ .

# Estimação intervalar

## Interpretação do intervalo de confiança:

```
dados <- read_xlsx("dados.xlsx", sheet = "teor_alcoolico")
dp <- sd(dados$teor_pop)
mu = mean(dados$teor_pop)

dados %>% group_by(amostra) %>%
  summarise(ic_li = MeanCI(teor_amostra, sd = dp) ['lwr.ci'],
            ic_ls = MeanCI(teor_amostra, sd = dp) ['upr.ci'],
            media = mu) %>%
  mutate(ind = ifelse(ic_li < mu & mu < ic_ls, "Sim", "Não"))
```

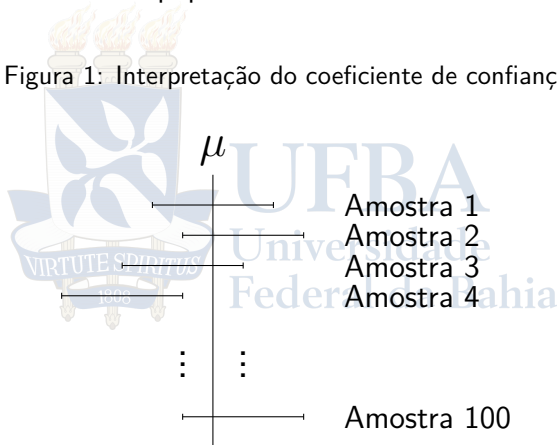
```
## # A tibble: 6 x 5
##   amostra ic_li ic_ls media ind
##   <chr>   <dbl> <dbl> <dbl> <chr>
## 1 amostra1  5.97  7.98  6.37 Sim
## 2 amostra2  5.13  7.14  6.37 Sim
## 3 amostra3  5.79  7.80  6.37 Sim
## 4 amostra4  5.44  7.45  6.37 Sim
## 5 amostra5  4.48  6.48  6.37 Sim
## 6 amostra6  5.35  7.36  6.37 Sim
```

# Estimação intervalar

## Interpretação do intervalo de confiança:

$(1 - \alpha)100\%$  das amostras vão gerar intervalos de confiança que contém o verdadeiro valor da média populacional.

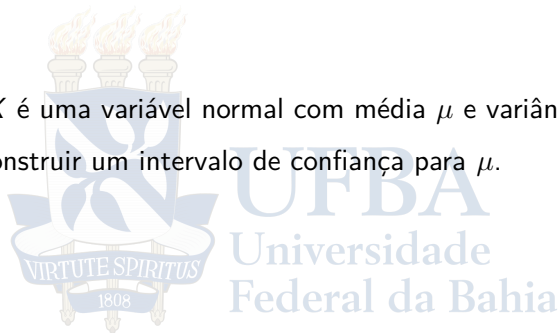
Figura 1: Interpretação do coeficiente de confiança.



# Intervalos de confiança para média

**Problema:**  $X$  é uma variável normal com média  $\mu$  e variância  $\sigma^2$ .

**Interesse:** construir um intervalo de confiança para  $\mu$ .



# Intervalos de confiança para média

**Caso 1: intervalo de confiança para a média de uma população normal com  $\sigma^2$  conhecida.**

Neste caso, o intervalo com  $(1 - \alpha)100\%$  de confiança para  $\mu$  é:

$$\left[ \bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

**Caso 2: intervalo de confiança para a média de uma população normal com  $\sigma^2$  desconhecida.**

Neste caso, o intervalo com  $(1 - \alpha)100\%$  de confiança para  $\mu$  é:

$$\left[ \bar{x} - t_{\frac{\alpha}{2}; (n-1)} \frac{s}{\sqrt{n}}; \bar{x} + t_{\frac{\alpha}{2}; (n-1)} \frac{s}{\sqrt{n}} \right].$$

# Intervalos de confiança para média

## Exemplo:

```
dados <- read_xlsx("dados.xlsx", sheet = "Iris")
```

```
MeanCI(dados$sepala_comp, sd = NULL,  
        conf.level = 0.95)
```

```
##      mean    lwr.ci    upr.ci  
## 5.843333 5.709732 5.976934
```

```
with(dados, MeanCI(sepala_comp, sd = NULL,  
                    conf.level = 0.95))
```

```
##      mean    lwr.ci    upr.ci  
## 5.843333 5.709732 5.976934
```

# Intervalos de confiança para média

## Exemplo:

```
dados %>% group_by(especie) %>%  
  summarise(media = MeanCI(sepala_comp) ['mean'],  
            ic_li = MeanCI(sepala_comp) ['lwr.ci'],  
            ic_ls = MeanCI(sepala_comp) ['upr.ci'])
```

```
## # A tibble: 3 x 4  
##   especie media ic_li ic_ls  
##   <chr>   <dbl> <dbl> <dbl>  
## 1 setosa    5.01  4.91  5.11  
## 2 versicolor  5.94  5.79  6.08  
## 3 virginica  6.59  6.41  6.77
```

# Intervalos de confiança para a proporção populacional ( $p$ )

**Problema:**  $X$  é uma variável Bernoulli com probabilidade de sucesso  $p$ .

**Interesse:** construir um intervalo de confiança para  $p$ .

O intervalo com  $(1 - \alpha)100\%$  de confiança para  $p$  é dado por:

$$\left[ \hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}; \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right].$$



# Intervalos de confiança para a proporção populacional ( $p$ )

## Exemplo:

Uma pesquisadora observou que em uma amostra aleatória de 400 poços em regiões do semi-árido Brasileiro é salobra, 120 deles têm a água salobra.

Obtenha um intervalo de 99% de confiança para a proporção.

```
BinomCI(120, 400, conf.level = 0.99)
```

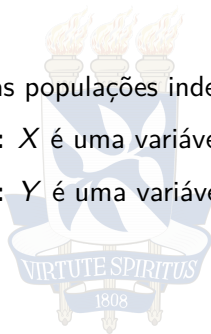
```
##          est      lwr.ci      upr.ci
## [1, ] 0.3 0.2446362 0.3618904
```

# Intervalo para a diferença entre duas médias

Considere duas populações independentes.

**População 1:**  $X$  é uma variável normal com média  $\mu_1$  e variância  $\sigma_1^2$ .

**População 2:**  $Y$  é uma variável normal com média  $\mu_2$  e variância  $\sigma_2^2$ .



Universidade  
Federal da Bahia

# Intervalo para a diferença entre duas médias

**Caso 1:  $X$  e  $Y$  são normais e  $\sigma_1^2$  e  $\sigma_2^2$  são conhecidas.**

Neste caso, temos que o intervalo com  $(1 - \alpha)100\%$  de confiança para  $(\mu_1 - \mu_2)$  é:

$$\left[ (\bar{x} - \bar{y}) - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; (\bar{x} - \bar{y}) + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right].$$

# Intervalo para a diferença entre duas médias

**Caso 2:**  $X$  e  $Y$  são normais e  $\sigma_1^2$  e  $\sigma_2^2$  são desconhecidas e iguais ( $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ).

Neste caso, temos que o intervalo com  $(1 - \alpha)100\%$  de confiança para  $(\mu_1 - \mu_2)$  é obtido de:

$$(\bar{x} - \bar{y}) \pm t_{\frac{\alpha}{2}; (n_1 + n_2 - 2)} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

onde

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

sendo que  $s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^n (x_i - \bar{x})^2$  e  $s_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^n (y_i - \bar{y})^2$ .

# Intervalo para a diferença entre duas médias

**Caso 3:  $X$  e  $Y$  são normais e  $\sigma_1^2$  e  $\sigma_2^2$  desconhecidas e diferentes.**

Neste caso, o intervalo com  $(1 - \alpha)100\%$  de confiança para  $(\mu_1 - \mu_2)$  é obtido de:

$$(\bar{x} - \bar{y}) \pm t_{\frac{\alpha}{2}; \nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

onde  $\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}.$

# Intervalo para a diferença entre duas médias

## Exemplo:

```
dados <- read_xlsx("dados.xlsx", sheet = "Iris")
```

```
with(dados,  
      var.test(sepala_comp[which(especie == "setosa")],  
               sepala_comp[which(especie == "versicolor")],  
               ratio = 1,  
               alternative = "two.sided",  
               conf.level = 0.95))
```

```
with(dados,  
      MeanDiffCI(  
        sepala_comp[which(especie == "setosa")],  
        sepala_comp[which(especie == "versicolor")],  
        conf.level = 0.95))
```

# Intervalo para a diferença entre duas médias

**Exemplo:** (continuação)

```
##  
## F test to compare two variances  
##  
## data:  sepala_comp[which(especie == "setosa")] and  
## F = 0.46634, num df = 49, denom df = 49, p-value =  
## alternative hypothesis: true ratio of variances is  
## 95 percent confidence interval:  
## 0.2646385 0.8217841  
## sample estimates:  
## ratio of variances 0.4663429  
  
## meandiff      lwr.ci      upr.ci  
## -0.9300000 -1.1057074 -0.7542926
```

# Intervalo para a diferença entre duas proporções

Considere duas populações independentes.

**População 1:**  $X$  é uma variável Bernoulli com probabilidade de sucesso  $p_1$ .

**População 2:**  $Y$  é uma variável Bernoulli com probabilidade de sucesso  $p_2$ .

Assim, o intervalo com  $(1 - \alpha)100\%$  de confiança para  $p_1 - p_2$  é obtido de:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}.$$



# Intervalo para a diferença entre duas proporções

## Exemplo:

Vamos utilizar os dados do artigo:

RUTLEDGE, Robert et al. The cost of not wearing seat belts. A comparison of outcome in 3396 patients. **Annals of surgery**, v. 217, n. 2, p. 122, 1993.

População 1: pessoas que usavam cinto.

População 2: pessoas que não usam cinto.

Sucesso: sobreviver.

Fracasso: morrer.

Podemos construir um intervalo de confiança para a diferença entre a proporção de sobreviventes entre os que usavam cinto de segurança e a proporção de mortos entre os que não usavam cinto de segurança.

# Intervalo para a diferença entre duas proporções

## Exemplo: (continuação)

```
sobreviventes <- matrix(c(1443,1781,47,135), ncol=2)
colnames(sobreviventes) <- c('Sobreviveram','Morerram')
rownames(sobreviventes) <- c('Com cinto','Sem cinto')

wald2ci(sobreviventes['Com cinto', 'Sobreviveram'],
        sum(sobreviventes['Com cinto',]),
        sobreviventes['Sem cinto', 'Sobreviveram'],
        sum(sobreviventes['Sem cinto',]),
        conf.level = 0.9, adjust = "Wald")
```

```
##
##
##
## data:
##
## 90 percent confidence interval:
##  0.02675203 0.05107930
## sample estimates:
## [1] 0.03891567
```



UFBA  
Universidade  
Federal da Bahia

# Teste de hipóteses

**Hipótese estatística:** é uma afirmação que se quer testar sobre determinados parâmetros de uma ou mais populações.

$H_0$ : hipótese nula.

$H_1$ : hipótese alternativa.

**Teste de hipóteses:** é um procedimento que utiliza os dados para decidir sobre uma determinada afirmação à respeito de um parâmetro desconhecido da população (decidir entre  $H_0$  e  $H_1$ ).

Decidir se alguma hipótese que tenha sido formulada está correta ou não, consiste apenas na escolha entre duas decisões: **rejeitar** ou **não rejeitar** tal hipótese.

**Ideia:** a partir de uma amostra da população iremos estabelecer uma regra de decisão segundo a qual rejeitaremos ou não a hipótese proposta. Esta regra de decisão é chamada de teste.

# Teste de hipóteses

Ao tomar uma decisão entre  $H_0$  e  $H_1$ , dois tipos de erros podem ser cometidos.

## Tipos de erros:

- ① erro tipo I: consiste em **rejeitar  $H_0$  quando  $H_0$  é verdadeira**;
- ② erro tipo II: consiste em **não rejeitar  $H_0$  quando  $H_0$  é falsa**.

**Nível de significância:** definimos por nível de significância do testes o número pré-fixado  $\alpha \in (0, 1)$ .

# Teste de hipóteses

O procedimento de teste de hipótese consiste das seguintes etapas:

- ① a partir do problema em questão, definimos o parâmetro de interesse;
- ② determinar a hipótese nula ( $H_0$ ) e a hipótese alternativa ( $H_1$ );
- ③ definir o teste adequado;
- ④ defina  $\alpha \in (0, 1)$ , o nível de significância do teste;
- ⑤ usar os dados da amostra para decidir entre  $H_0$  e  $H_1$ .

# Teste de hipóteses

Para os testes discutidos aqui, vamos considerar hipóteses do tipo:

$H_0 : \theta = \theta_0$  contra  $H_1 : \theta \neq \theta_0$  (teste bilateral)

$H_0 : \theta \leq \theta_0$  contra  $H_1 : \theta > \theta_0$  (teste unilateral superior)

$H_0 : \theta \geq \theta_0$  contra  $H_1 : \theta < \theta_0$  (teste unilateral inferior)

**Definição de  $H_0$  e  $H_1$ :**

Usualmente, definimos:

A hipótese alternativa como hipótese de pesquisa.

A hipótese nula como afirmação a ser desafiada.

# Teste de hipóteses

## Como decidir entre $H_0$ e $H_1$ ?

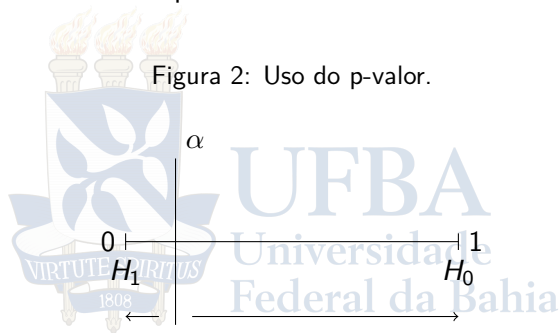
**p-valor:** é o menor nível de significância para o qual rejeita-se  $H_0$  com os dados observados.

Usamos e interpretamos o p-valor do seguinte modo:

- Se o p-valor for grande, a amostra não fornece evidências suficientes para rejeitar  $H_0$ , então podemos decidir por  $H_0$ ;
- Se o p-valor for pequeno, a amostra fornece evidências suficientes para rejeitar  $H_0$ , então podemos decidir por  $H_1$ .

# Teste de hipóteses

A Figura 2 ilustra o uso do p-valor:





# Teste de hipóteses a média de uma população normal

1 -  $H_0 : \mu = \mu_0$  contra  $H_1 : \mu \neq \mu_0$  (teste bilateral)

2 -  $H_0 : \mu \leq \mu_0$  contra  $H_1 : \mu > \mu_0$  (teste unilateral superior)

3 -  $H_0 : \mu \geq \mu_0$  contra  $H_1 : \mu < \mu_0$  (teste unilateral inferior)

**Teste de hipótese para  $\mu$  com  $\sigma^2$  conhecida:** teste Z.

**Teste de hipótese para  $\mu$  com  $\sigma^2$  desconhecida:** teste t.

# Teste de hipóteses a média de uma população normal

## Exemplo:

```
dados <- read_xlsx("dados.xlsx", sheet = "Iris")
```

```
t.test(dados$sepalcomp, sd = NULL,  
       alternative = "two.sided",  
       mu = 6, conf.level = 0.95)
```

```
##  
## One Sample t-test  
##  
## data: dados$sepalcomp  
## t = -2.3172, df = 149, p-value = 0.02186  
## alternative hypothesis: true mean is not equal to 6  
## 95 percent confidence interval:  
## 5.709732 5.976934  
## sample estimates:  
## mean of x  
## 5.843333
```

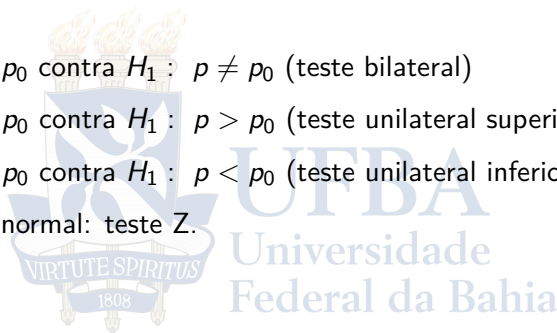
# Teste de hipóteses para a proporção populacional

1 -  $H_0 : p = p_0$  contra  $H_1 : p \neq p_0$  (teste bilateral)

2 -  $H_0 : p \leq p_0$  contra  $H_1 : p > p_0$  (teste unilateral superior)

3 -  $H_0 : p \geq p_0$  contra  $H_1 : p < p_0$  (teste unilateral inferior)

Aproximação normal: teste Z.



# Teste de hipóteses para a proporção populacional

## Exemplo:

Um relatório de uma companhia afirma que 40% de toda água obtida através de poços artesianos em regiões do semi-árido Brasileiro é salobra.

Uma pesquisadora observou que em uma amostra aleatória de 400 poços em regiões do semi-árido Brasileiro é salobra, 120 deles têm a água salobra.

Teste a afirmação da companhia usando um nível de significância 1%?

# Teste de hipótese para a proporção populacional

**Exemplo:** (continuação)

```
prop.test(120, 400, p = 0.4,  
          alternative = 'two.sided',  
          conf.level = 0.99)
```

```
##  
## 1-sample proportions test with continuity correction  
##  
## data: 120 out of 400, null probability 0.4  
## X-squared = 16.253, df = 1, p-value = 5.543e-05  
## alternative hypothesis: true p is not equal to 0.4  
## 99 percent confidence interval:  
## 0.2434753 0.3631882  
## sample estimates:  
## p  
## 0.3
```

# Testes para a diferença de médias de duas populações normais

1 -  $H_0 : \mu_1 - \mu_2 = \Delta_0$  contra  $H_1 : \mu_1 - \mu_2 \neq \Delta_0$  (teste bilateral)

2 -  $H_0 : \mu_1 - \mu_2 \leq \Delta_0$  contra  $H_1 : \mu_1 - \mu_2 > \Delta_0$  (teste unilateral superior)

3 -  $H_0 : \mu_1 - \mu_2 \geq \Delta_0$  contra  $H_1 : \mu_1 - \mu_2 < \Delta_0$  (teste unilateral inferior)

**Quando  $\sigma_1^2$  e  $\sigma_2^2$  são conhecidas: teste Z.**

**Quando  $\sigma_1^2$  e  $\sigma_2^2$  são desconhecidas e iguais: teste t.**

**Quando  $\sigma_1^2$  e  $\sigma_2^2$  são desconhecidas e diferentes: teste t.**

**Teste de igualdade de variâncias de duas populações normais:**

$H_0 : \sigma_1^2 = \sigma_2^2$  contra  $H_1 : \sigma_1^2 \neq \sigma_2^2$  (teste bilateral): teste F.

# Testes para a diferença de médias de duas populações normais

## Exemplo: (continuação)

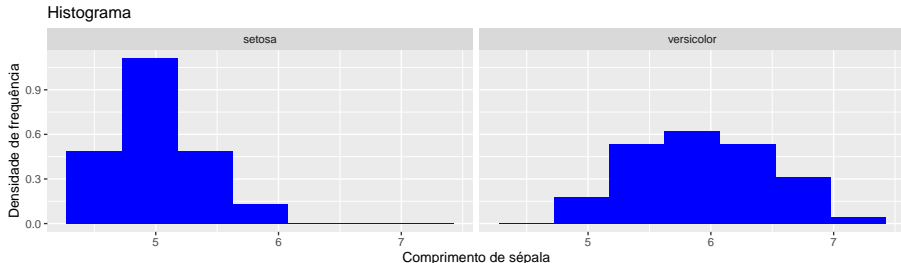
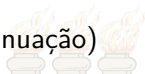
```
dados <- read_xlsx("dados.xlsx", sheet = "Iris")
dados_1 <- dados %>% filter(especie %in% c("setosa", "versicolosa"))

n_especie <- dados_1 %>% count(especie)
n_int <- round(1 + log2(n_especie$n))

ggplot(data = dados_1) +
  geom_histogram(aes(x = sepala_comp, y = ..density..),
                 bins = 7, fill = 'blue') +
  labs(x = 'Comprimento de sépala',
       y = 'Densidade de frequência',
       title = 'Histograma')+
  facet_grid(~especie)
```

# Testes para a diferença de médias de duas populações normais

## Exemplo: (continuação)





# Testes para a diferença de médias de duas populações normais

**Exemplo:** (continuação)

```
with(dados_1,  
      var.test(sepala_comp[especie %in% "setosa"],  
                sepala_comp[especie %in% "versicolor"],  
                alternative = "two.sided",  
                ratio = 1, conf.level = 0.95))
```

```
with(dados_1,  
      t.test(sepala_comp[especie %in% "setosa"],  
              sepala_comp[especie %in% "versicolor"],  
              alternative = "two.sided",  
              mu = 0, var.equal = F,  
              conf.level = 0.95))
```

# Testes para a diferença de médias de duas populações normais

## Exemplo: (continuação)

```
##  
## F test to compare two variances  
##  
## data:  sepala_comp[especie %in% "setosa"] and sepal  
## F = 0.46634, num df = 49, denom df = 49, p-value =  
## alternative hypothesis: true ratio of variances is  
## 95 percent confidence interval:  
## 0.2646385 0.8217841  
## sample estimates:  
## ratio of variances  
## 0.4663429
```

# Testes para a diferença de médias de duas populações normais

## Exemplo: (continuação)

```
##  
## Welch Two Sample t-test  
##  
## data:  sepala_comp[especie %in% "setosa"] and sepala_comp[especie %in% "versicolora"]  
## t = -10.521, df = 86.538, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.1057074 -0.7542926  
## sample estimates:  
## mean of x mean of y  
## 5.006 5.936
```

# Testes para a diferença de médias de duas populações normais

## Exemplo:

```
dados <- read_xlsx("dados.xlsx",  
                  sheet = "Repetibilidade")
```

```
with(dados,  
      var.test(tecnicol1, tecnico2,  
                alternative = "two.sided",  
                ratio = 1, conf.level = 0.99))
```

```
with(dados,  
      t.test(tecnicol1, tecnico2,  
              alternative = "two.sided",  
              mu = 0, var.equal = T,  
              conf.level = 0.99))
```

# Testes para a diferença de médias de duas populações normais

## Exemplo: (continuação)

```
##  
## F test to compare two variances  
##  
## data:  tecnico1 and tecnico2  
## F = 0.84278, num df = 5, denom df = 6, p-value = 0.  
## alternative hypothesis: true ratio of variances is  
## 99 percent confidence interval:  
##    0.07351729 12.23148465  
## sample estimates:  
## ratio of variances  
##          0.8427798
```

# Testes para a diferença de médias de duas populações normais

## Exemplo: (continuação)

```
##  
## Two Sample t-test  
##  
## data: tecnico1 and tecnico2  
## t = 0.28456, df = 11, p-value = 0.7813  
## alternative hypothesis: true difference in means is  
## 99 percent confidence interval:  
## -0.1345530 0.1616959  
## sample estimates:  
## mean of x mean of y  
## 24.99500 24.98143
```

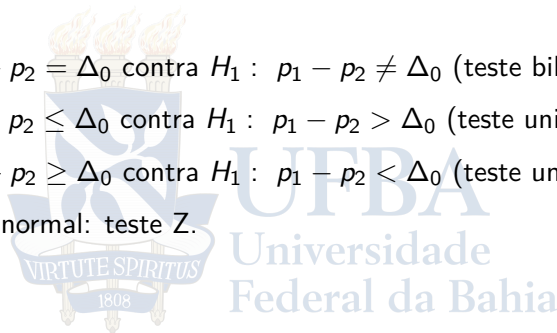
# Testes para a diferença de duas proporções populacionais

1 -  $H_0 : p_1 - p_2 = \Delta_0$  contra  $H_1 : p_1 - p_2 \neq \Delta_0$  (teste bilateral)

2 -  $H_0 : p_1 - p_2 \leq \Delta_0$  contra  $H_1 : p_1 - p_2 > \Delta_0$  (teste unilateral superior)

3 -  $H_0 : p_1 - p_2 \geq \Delta_0$  contra  $H_1 : p_1 - p_2 < \Delta_0$  (teste unilateral inferior)

Aproximação normal: teste Z.



# Testes para a diferença de duas proporções populacionais

## Exemplo:

Vamos utilizar os dados do artigo:

RUTLEDGE, Robert et al. The cost of not wearing seat belts. A comparison of outcome in 3396 patients. **Annals of surgery**, v. 217, n. 2, p. 122, 1993.

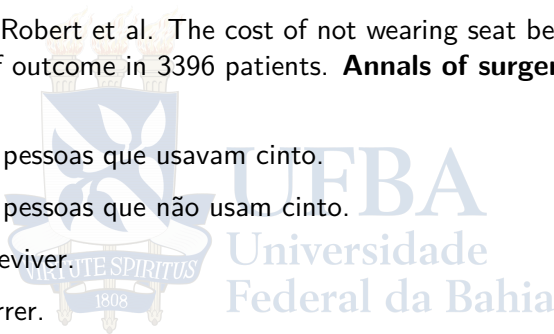
População 1: pessoas que usavam cinto.

População 2: pessoas que não usam cinto.

Sucesso: sobreviver.

Fracasso: morrer.

Usando o procedimento de teste de hipóteses, podemos verificar se a proporção de sobreviventes entre os que usavam cinto de segurança é maior do que a proporção de mortos entre os que não usavam cinto de segurança.





# Testes para a diferença de duas proporções populacionais

## Exemplo: (continuação)

```
sobreviventes <- matrix(c(1443,1781,47,135), ncol=2)
colnames(sobreviventes) <- c('Sobreviveram','Morerram')
rownames(sobreviventes) <- c('Com cinto','Sem cinto')

prop.test(sobreviventes, alternative = "greater", conf.level = 0.9)
```

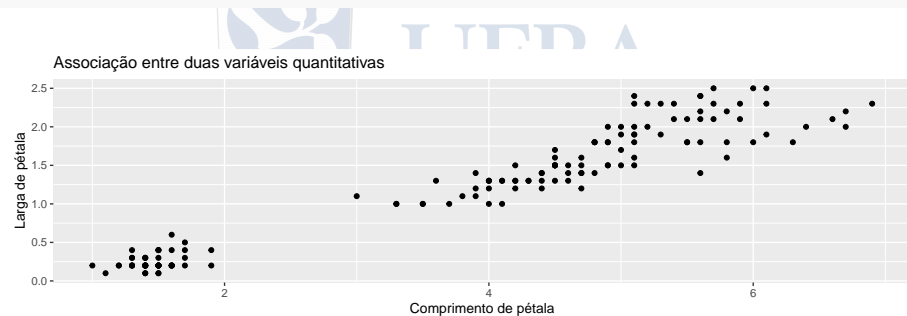
```
##
## 2-sample test for equality of proportions with continuity
## correction
##
## data: sobreviventes
## X-squared = 24.333, df = 1, p-value = 4.052e-07
## alternative hypothesis: greater
## 90 percent confidence interval:
## 0.02884211 1.00000000
## sample estimates:
## prop 1 prop 2
## 0.9684564 0.9295407
```

# Teste de correlação

$H_0 : \rho = 0$  contra  $H_1 : \rho \neq 0$  (teste bilateral)

```
dados <- read_xlsx("dados.xlsx", sheet = "Iris")
```

```
ggplot(data = dados) +  
  geom_point(aes(x = petala_comp, y = petala_larg)) +  
  labs(x = 'Comprimento de pétala', y = 'Larga de pétala',  
       title = 'Associação entre duas variáveis quantitativas')
```



# Teste de correlação

```
with(dados,  
      cor.test(petala_comp, petala_larg,  
                alternative = "two.sided",  
                conf.level = 0.95))
```

```
##  
## Pearson's product-moment correlation  
##  
## data: petala_comp and petala_larg  
## t = 43.387, df = 148, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equ  
## 95 percent confidence interval:  
## 0.9490525 0.9729853  
## sample estimates:  
## cor  
## 0.9628654
```