Curso de Pós-Graduação em Ciências Veterinárias - UFRRJ

Métodos Estatísticos

Prof: Wagner Tassinari

wagner.tassinari@ini.fiocruz.br

Intervalos de Confiança e Testes de Hipótese

Teste para a média populacional

Teste para a média populacional

- Nosso objetivo agora é apresentar procedimentos estatísticos simples para verificar se um conjunto de dados amostrais dá ou não suporte à uma conjectura sobre o valor médio (desconhecido) de uma característica de interesse, observável em "indivíduo" de uma população.
- Mais precisamente, procedimentos para testar hipóteses sobre , tomando como base o valor médio dessa característica, observado em uma amostra casual simples de tamanho n desses "indivíduos".

Teste para a média populacional ( ) - One *Sample t-test*

- Intervalo de Confiança

$$P[\mu \in (\bar{x} \pm t_{(n-1;\alpha)} \times s/\sqrt{n}] = 1 - \alpha$$

- Hipóteses
  - $H_0 : \mu = \mu_0$
  - $H_1 : \mu \neq \mu_0$
  - $H_1 : \mu < \mu_0$
  - $H_1 : \mu > \mu_0$
- Estatística Teste

$$t_{calc} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

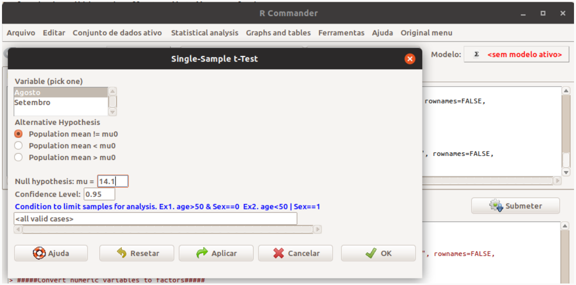
Exemplo:

Voltando ao exemplo dos pássaros migratórios, bibliografias recentes sobre estes passaros mostram que a massa média destes pássaros no mês de agosto é de $= 14.1$. Verifique se a média dessa amostra coletada corresponde com a massa média verificada na maioria das literaturas sobre o assunto (, média populacional).

- $H_0 : = 14.1$
- $H_1 : = 14.1$

Solução utilizando o plugin Rcommander.EZT

- Importar o arquivo “massapassaros0.xlsx”
  - Rcommander → Arquivo → Importar arquivos de dados → from Excel data set
- Testar a média
  - Rcommander → Estatísticas → Médias → Teste t para uma amostra



Solução:

Submeter

Output

```
> (res <- t.test(banco0$Agosto, alternative='two.sided', mu=14.1, conf.level=0.95))

        One Sample t-test

data:  banco0$Agosto
t = -10.288, df = 9, p-value = 0.000002823
alternative hypothesis: true mean is not equal to 14.1
95 percent confidence interval:
 10.89173 12.04827
sample estimates:
mean of x
    11.47


> cat(gettext(domain="R-RcmdrPlugin.EZR", "mean"), " = ", res$estimate, ", ", gettext(domain="R-RcmdrPlugin.EZR", "95% CI"), " ",
+   res$conf.int[1], "-", res$conf.int[2], ", ", gettext(domain="R-RcmdrPlugin.EZR", "p.value"), " = ", signif(res$p.value, digits=3),
+    "
+ ", sep="")
mean = 11.47, 95% CI 10.89173-12.04827, p.value = 0.00000282
```

como o $p - valor < 0.001$, rejeita-se H_0, ou seja, a massa média dos pássaros desta amostra não corresponde com o a média encontrada na literatura.

Teste para a comparação de duas médias populacionais ($_1 e_2$) - Grupos Independentes

- Intervalo de confiança

$$P[\mu_1 - \mu_2 \in (\bar{x}_1 - \bar{x}_2 \pm t_{(n_1+n_2-2;\alpha)} \cdot s_p\sqrt{1/n_1 + 1/n_2})] = 1 - \alpha$$

- Sendo s_p o desvio-padrão conjugado

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n2 - 2}}$$

- Hipóteses:
  - $H_0 : \mu_1 = \mu_2 \therefore \mu_1 - \mu_2 = 0$
  - $H_1 : \mu_1 \neq \mu_2 \therefore \mu_1 - \mu_2 \neq 0$
- Estatística de teste

$$t_{calc} = \frac{\bar{x}_1 - \bar{x}_2}{s_p/\sqrt{1/n_1 + 1/n_2}}$$

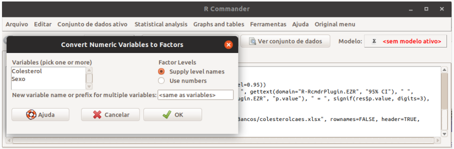
Exemplo:

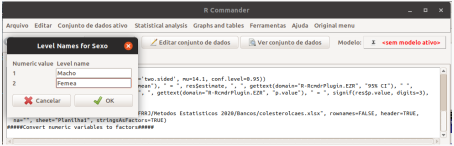
Comparando as distribuições do teor de colesterol sérico em cães machos e femêas normais, medidos em mg/100ml, pergunta-se: É possível responder se existe diferença significativa no teor de de colesterol entre cães machos e femêas ?

- $H_0 : \mu_{macho} = \mu_{femea} \therefore \mu_{macho} - \mu_{femea} = 0$
- $H_1 : \mu_{macho} \neq \mu_{femea} \therefore \mu_{macho} - \mu_{femea} \neq 0$

Solução utilizando o plugin Rcommander.EZT

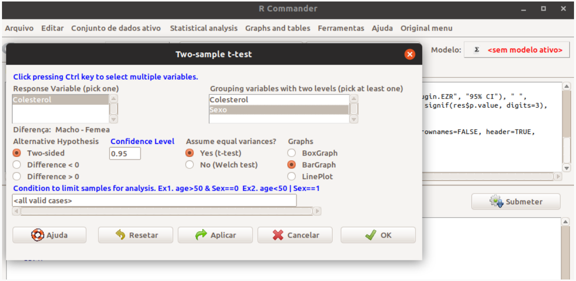
- Importar o arquivo "colesterolcaes.xlsx"
  - Rcommander → Arquivo → Importar arquivos de dados → from Excel data set
- Colocar os *labels* na variável categórica (qualitariva), sexo
- Rcommander → Conjunto de dados ativo → Variables → Convert numeric variables to factors



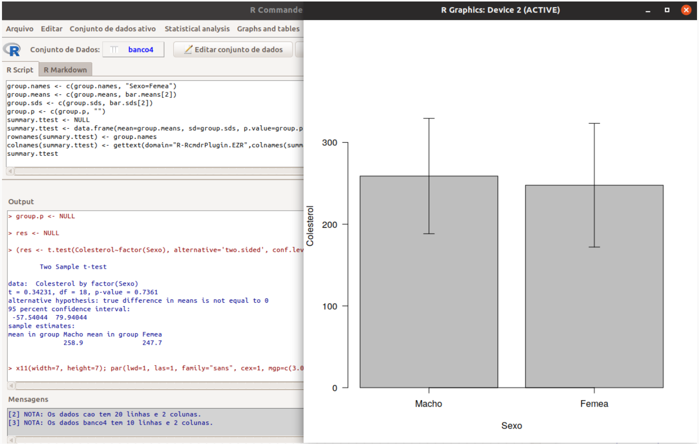


Solução utilizando o plugin Rcommander.EZT

- Testar os dois grupos
  - Rcommander → Statistical analysis → Continuous variables → Two-sample t-test

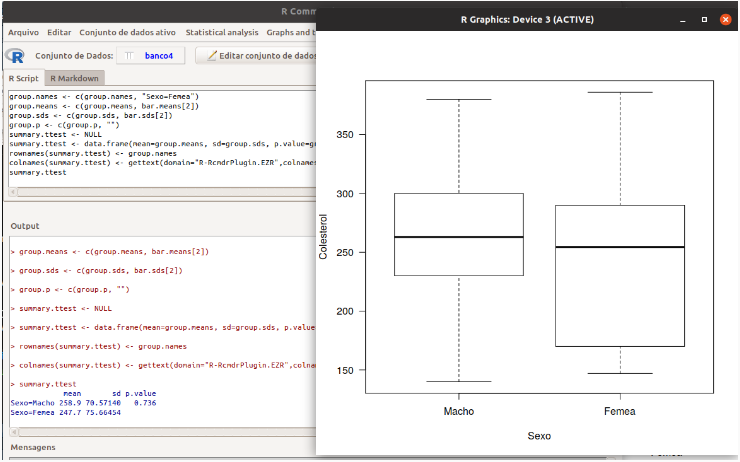


Solução utilizando o plugin Rcommander.EZT



como o $p - valor = 0,7361$, não rejeita-se H_0, ou seja, não existe diferença significativa entre o nível dos coleterois entre cães fêmeas e cães machos.

Solução utilizando o plugin Rcommander.EZT



Teste para a comparação de duas médias populacionais (μ_1 e μ_2) - Dados Pareados

- Intervalo de Confiança

$$P[\mu \in (\bar{d} \pm t_{(n-1;\alpha)} \cdot s_d/\sqrt{n}] = 1 - \alpha$$

- Hipóteses
    - $H_0 : \mu_{antes} = \mu_{depois} \therefore \mu_{antes} - \mu_{depois} = 0$
    - $H_1 : \mu_{antes} \neq \mu_{depois} \therefore \mu_{antes} - \mu_{depois} \neq 0$
- Estatística Teste

$$t_{calc} = t_d = \frac{\bar{d}}{s_d/\sqrt{n}}$$

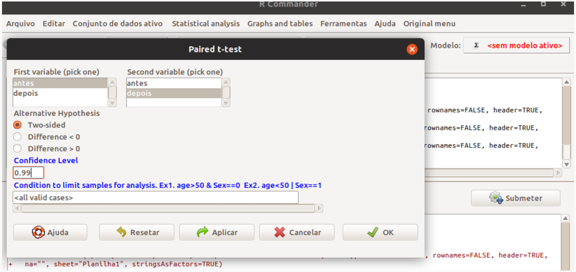
Exemplo:

Dez cobaias foram submetidas ao tratamento de engorda com certa ração. Os pesos em gramas, antes e após o teste são dados a seguir (supõe-se que provenham de distribuições normais). A 1% de significância, podemos concluir que o uso da ração contribuiu para o aumento do peso médio dos animais ?

- $H_0 : \mu_{antes} = \mu_{depois} \therefore \mu_{antes} - \mu_{depois} = 0$
- $H_1 : \mu_{antes} \neq \mu_{depois} \therefore \mu_{antes} - \mu_{depois} \neq 0$

Solução utilizando o plugin Rcommander.EZT

- Importar o arquivo "pesocobaias0.xlsx"
  - Rcommander → Arquivo → Importar arquivos de dados → from Excel data set
- Testar os dois grupos
  - Rcommander → Statistical analysis → Continuous variables → Paried t-test



Solução utilizando o plugin Rcommander.EZT

```
> #####Paired t-test#####

> (res <- t.test(dado6$antes, dado6$depois, alternative='two.sided', conf.level=0.99, paired=TRUE))

        Paired t-test

data:  dado6$antes and dado6$depois
t = -2.9635, df = 9, p-value = 0.01587
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 -13.8377274   0.6377274
sample estimates:
mean of the differences
                   -6.6


> mean1 <- mean(dado6$antes, na.rm=TRUE)

> mean2 <- mean(dado6$depois, na.rm=TRUE)

> sd1 <- sd(dado6$antes, na.rm=TRUE)

> sd2 <- sd(dado6$depois, na.rm=TRUE)
```

```
> sd1 <- sd(dado6$antes, na.rm=TRUE)

> sd2 <- sd(dado6$depois, na.rm=TRUE)

> summary.ttest <- NULL

> summary.ttest <- data.frame(mean=c(mean1, mean2), sd=c(sd1, sd2), p.value=c(signif(res$p.value, digit=3),""))

> rownames(summary.ttest) <- c("antes", "depois")

> colnames(summary.ttest) <- gettext(domain="R-RcmdrPlugin.EZR",colnames(summary.ttest))

> summary.ttest
        mean       sd p.value
antes  648.4 58.85236  0.0159
depois 655.0 59.20023

Mensagens
[10] NOTA: Os dados banco4 tem 10 linhas e 2 colunas.
```

Como $p - valor = 0,0158$, rejeita-se H_0, ou seja, existe diferença significativa entre as médias do antes e do depois.

Teste para a comparação para mais de duas médias populacionais - ANOVA

- Hipóteses:
  - $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$
  - $H_1 : \mu_i \neq \mu_j$, sendo $i \neq j$
- Estatística Teste: Teste F

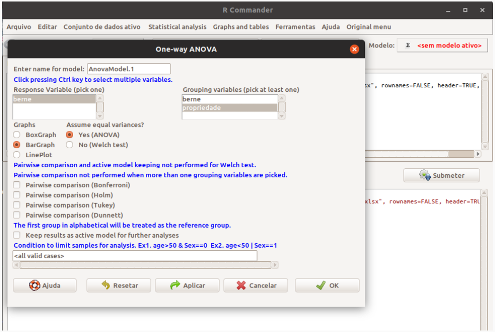
Exemplo:

Gostaríamos de verificar se o número de bernes por animal é a mesma nas três propriedades distintas. Para cada propriedade foram selecionados 5 animais.

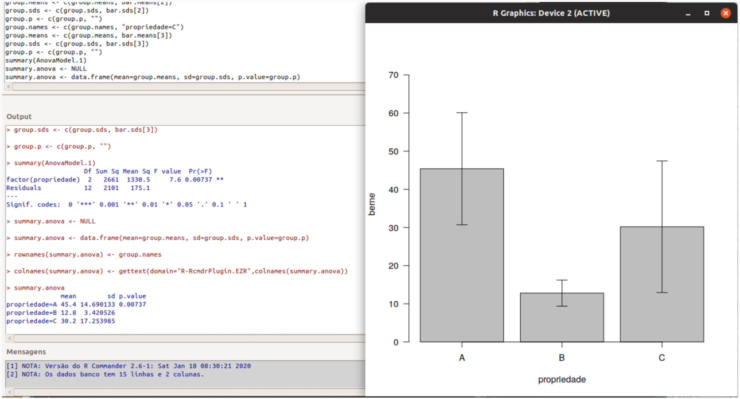
- $H_0 : \mu_a = \mu_b = \mu_c$
- $H_1 : \mu_i \neq \mu_j$, sendo $i \neq j$

Solução utilizando o plugin Rcommander.EZT

- Importar o arquivo "bernepropriedades.xlsx"
  - Rcommander → Arquivo → Importar arquivos de dados → from Excel data set
  - Testar os grupos
  - Rcommander → Statistical analysis → Continuous variables → one-Way ANOVA



Solução utilizando o plugin Rcommander.EZT



- Como $p - valor = 0,00104$, rejeita-se H_0
- Verificamos que existe diferença significativa entre as médias.

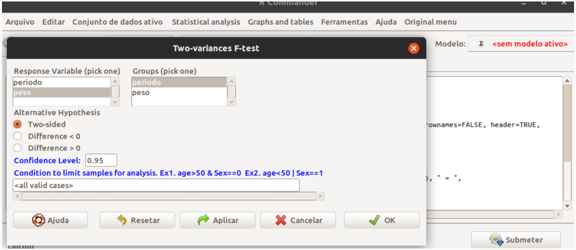
Testes para a variância

Teste para a comparação de duas variâncias populacionais

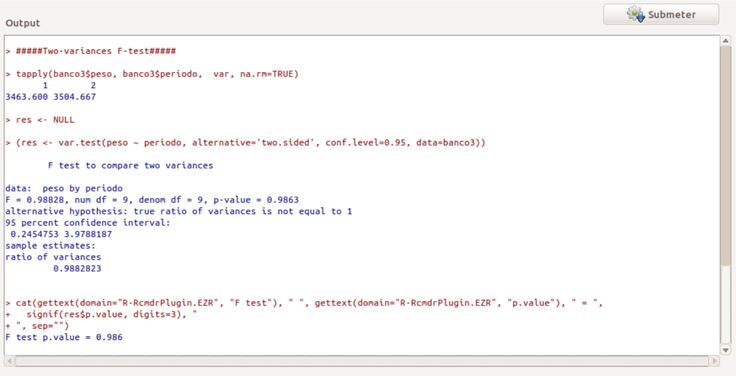
- Hipóteses
  - $H_0 : \sigma_1^2 = \sigma_2^2 (= \sigma)$ ou $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$
    - $H_1 : \sigma_1^2 \neq \sigma_2^2$ ou $\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$
- Estatística do teste
  - $F_{calc} = \dfrac{s_1^2}{s_2^2}$
- A utilização deste teste é necessária quando precisamos comparar diferenças entre médias. Alguns softwares já fazem isso, como é o caso do Bioestat.

Solução utilizando o plugin Rcommander.EZT

- Importar o arquivo "pesocobaias.xlsx"
  - Rcommander → Arquivo → Importar arquivos de dados → from Excel data set
- Testar a variância dos dois períodos
  - Rcommander → Statistical analysis → Continuous variables → Two-variences F-test



Solução utilizando o plugin Rcommander.EZT


Submeter
Output
```
> #####Two-variances F-test#####

> tapply(banco3$peso, banco3$periodo,  var, na.rm=TRUE)
       1        2 
3463.600 3504.667 

> res <- NULL

> (res <- var.test(peso ~ periodo, alternative='two.sided', conf.level=0.95, data=banco3))

	F test to compare two variances

data:  peso by periodo
F = 0.98828, num df = 9, denom df = 9, p-value = 0.9863
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2454753 3.9788187
sample estimates:
ratio of variances 
         0.9882823 


> cat(gettext(domain="R-RcmdrPlugin.EZR", "F test"), " ", gettext(domain="R-RcmdrPlugin.EZR", "p.value"), " = ",
+   signif(res$p.value, digits=3), "
+ ", sep="")
F test p.value = 0.986
```

- Como $p - valor = 0.986$, não rejeita-se H_0
- Verificamos que as variâncias entre os períodos são homogêneas (iguais).

Teste para a comparação de duas ou mais variâncias populacionais - Bartlett's Test

- Hipóteses
  - $H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$
  - $H_1 : \sigma_i^2 \neq \sigma_j^2$, sendo $i \neq j$
- Estatística de teste

$$T = \frac{(n-k)ln(s_p^2) - \sum_{i=1}^{k}(n_i - 1)ln(s_p^2)}{1 + (1/(3(k-1)))((\sum_{i=1}^{k} 1/(n_i - 1)) - 1/(n-k))}$$

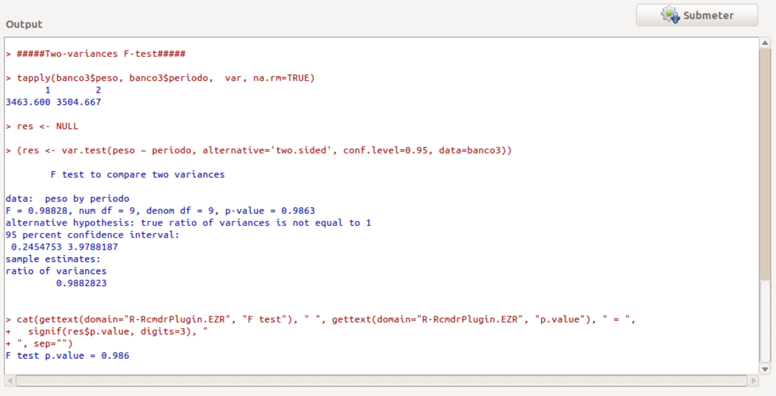
- sendo $n = n_1 + n_2 + \ldots + n_k$, k o número de grupos ou amostras e a variância conjunta sendo:

$$s_p^2 = \frac{\sum_{i=1}^{k}(n_i - 1)(s_i^2)}{n-k}$$

Solução utilizando o plugin Rcommander.EZT

- Importar o arquivo “bernepropriedades.xlsx”
  - Rcommander → Arquivo → Importar arquivos de dados → from Excel data set
  - Comparar as variâncias do número de berne entre as propriedades
  - Rcommander → Statistical analysis → Continuous variables → Bartlett’s test

Solução utilizando o plugin Rcommander.EZT



Submeter

Output

```
> #####Two-variances F-test#####

> tapply(banco3$peso, banco3$periodo,  var, na.rm=TRUE)
       1        2 
3463.600 3504.667 

> res <- NULL

> (res <- var.test(peso ~ periodo, alternative='two.sided', conf.level=0.95, data=banco3))

	F test to compare two variances

data:  peso by periodo
F = 0.98828, num df = 9, denom df = 9, p-value = 0.9863
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2454753 3.9788187
sample estimates:
ratio of variances 
         0.9882823 


> cat(gettext(domain="R-RcmdrPlugin.EZR", "F test"), " ", gettext(domain="R-RcmdrPlugin.EZR", "p.value"), " = ", 
+   signif(res$p.value, digits=3), "
+ ", sep="")
F test p.value = 0.986
```

- Como $p - valor = 0,02908$, rejeita-se H_0
- Verificamos que a variabilidade do número de bernes entre entre as propriedades é heterogênea.

Teste para proporções

Teste para Proporção Populacional (π)

- Intervalo de Confiança

$$P[\pi \in (p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n}})] = 1 - \alpha$$

- Hipóteses
    - $H_0 : \pi = \pi_0$
    - $H_1 : \pi \neq \pi_0$
    - $H_1 : \pi < \pi_0$
    - $H_1 : \pi > \pi_0$
- Estatística Teste

$$z_{calc} = \frac{p - \pi_0}{\sqrt{\frac{p(1-p)}{n}}}$$

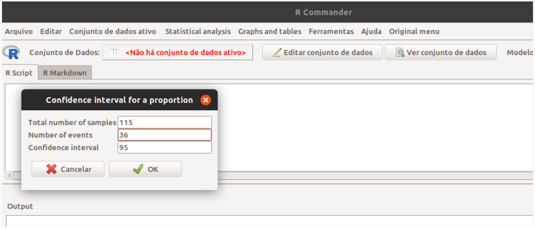
Exemplo:

Amostras de sangue de 115 cabeças de gado foram analisadas através de um teste sorológico para *Leptospira sp.* e, de acordo com o título, cada amostra foi classificada como positiva ou negativa. Na amostra, 36 animais apresentaram título positivo. Deseja-se saber se a prevalência de animais infectados foi de 40% em todo estabelecimento.

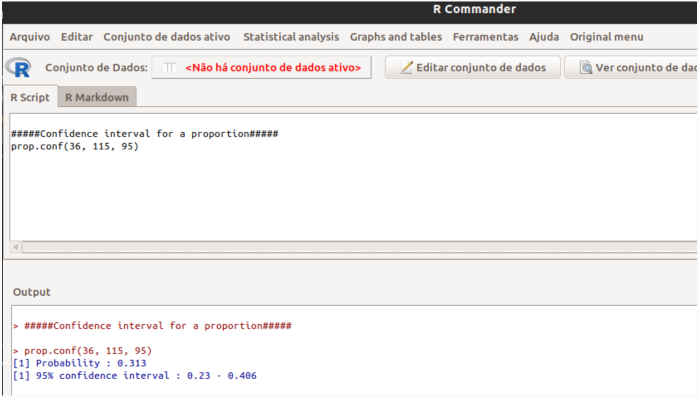
- $H_0 : \pi = 40\%$
- $H_1 : \pi \neq 40\%$

Solução utilizando o plugin Rcommander.EZT

- Não existe a necessidade de abrir um banco de dados (tabela)
- Estimando o intervalo de confiança para a proporção
    - Rcommander → Statistical analysis → Discrete variables → Confidence interval for a proportion



Solução utilizando o plugin Rcommander.EZT



- Como o valor 40% está incluido dentro do intervalo de confiança, podemos dizer que não rejeitamos H_0, ou seja, a prevalência de animais infectados pode ser considerada de 40% em todo estabelecimento.

Solução utilizando a linha de comando do R

```r
prop.test(36, 115, 0.4)
```

```

	1-sample proportions test with continuity correction

data:  36 out of 115, null probability 0.4
X-squared = 3.2699, df = 1, p-value = 0.07056
alternative hypothesis: true p is not equal to 0.4
95 percent confidence interval:
 0.2315949 0.4071869
sample estimates:
        p
0.3130435
```

- Como o $p - valor = 0,071$, não rejeitamos H_0.

Teste para a Comparação de duas Proporções Populacionais (π_1 e π_2)

- Intervalo de Confiaça

$$P[\pi_1 - \pi_2 \in (p_1 - p_2 \pm z_{\alpha/2} \cdot \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}] = 1 - \alpha$$

- Hipóteses
  - $H_0 : \pi_1 = \pi_2$
  - $H_1 : \pi_1 \neq \pi_2$
- Estatística Teste

$$z_{calc} = \frac{p_1 - p_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

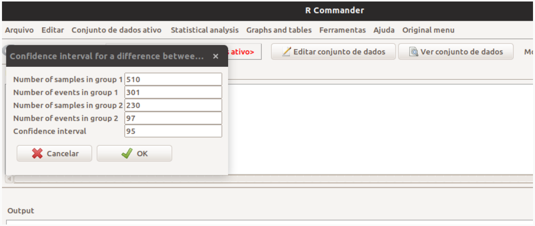
Exemplo:

Pesquisadores decidiram avaliar se a proporção de cães machos é idêntica em cães domiciliados e não-domiciliados. Fizeram um levantamento em uma determinada cidade, e observaram que, dos 510 cães domiciliados amostrados, 301 eram machos e, dentre os 230 não-domiciliados, 97 eram machos. Pergunta-se, existe diferença significativa entre as duas proporções ?

- $H_0 : \pi_1 = \pi_2$
- $H_1 : \pi_1 \neq \pi_2$

Solução utilizando o plugin Rcommander.EZT

- Não existe a necessidade de abrir um banco de dados (tabela)
- Estimando o intervalo de confiança para a proporção
    - Rcommander → Statistical analysis → Discrete variables → Confidence interval for a difference between two proportions



Solução utilizando o plugin Rcommander.EZT

- Como o 0 (zero) não está incluido dentro do intervalo de confiança, podemos dizer que que existe diferença significativa entre as proporções.

Solução utilizando a linha de comando do R

```r
prop.test(c(301,97), c(510,230))
```

```
    2-sample test for equality of proportions with continui

data:  c(301, 97) out of c(510, 230)
X-squared = 17.425, df = 1, p-value = 2.988e-05
alternative hypothesis: two.sided
95 percent confidence interval:
 0.08852378 0.24839011
sample estimates:
   prop 1    prop 2
0.5901961 0.4217391
```

- Como o $p - valor < 0,05$, rejeitamos H_0.