UNIVERSIDADE FEDERAL DA FRONTEIRA SUL
Campus CERRO LARGO

PROJETO DE EXTENSÃO
Software R:
Capacitação em análise estatística de dados utilizando um software livre.

Fonte: `https://www.r-project.org/`

Módulo III
Estatística inferencial

Ministrante: Tatiane Chassot

Blog do projeto: `https://softwarelivrer.wordpress.com/equipe/`

Equipe:
Coordenadora:
Profe. Iara Endruweit Battisti (iara.battisti@uffs.edu.br)
Colaboradores:
Profa. Denize Reis
Prof. Erikson Kaszubowski
Prof. Reneo Prediger
Profa. Tatiane Chassot
Mestrando Felipe Smolski
Bolsista:
Djaina Rieger - aluna de Engenharia Ambiental (djaina.rieger@outlook.com)
Voluntárias:
Jaíne Frank
Jaqueline Caye

Sumário

ESTATÍSTICA DESCRITIVA

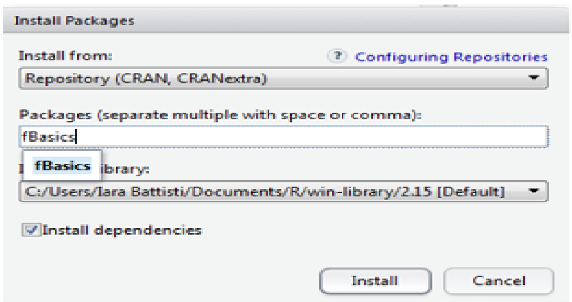
1 Importar arquivo de dados

```
library(readxl)
arvores <- read_excel("~/SOFTWARE R/arvores.xlsx")
View(arvores)
```

2 Intervalo de confiança para a média

Para calcular o intervalo de confiança para média de uma população utilizamos o comando ‘basicStats’ do pacote fBasics. Assim teremos que instalar o pacote fBasics clicando no botão Install na ficha ‘Packages’, conforme segue:

E digitar o nome no pacote fBasics no campo Packages, conforme segue:



Após ativar o pacote na linha de comando do painel console: **library(fBasics)**. Este comando para calcular o intervalo de confiança para a média de uma população, supondo **normalidade dos dados** e que os dados provem de uma **amostra aleatória simples**:

```
> basicStats(altura_m,ci=0.95)
                  altura_m
nobs            679.000000
NAs               0.000000
Minimum           2.000000
Maximum          26.100000
1. Quartile      11.950000
3. Quartile      18.050000
Mean             14.955376
Median           14.700000
Sum           10154.700000
SE Mean           0.172382
LCL Mean         14.616910
UCL Mean         15.293841
Variance         20.176752
Stdev             4.491854
Skewness         -0.157291
Kurtosis         -0.055489
```

Em que:

altura_m: a variável que estamos analisando, isto é, que queremos calcular o intervalo de confiança;

ci: identifica o nível de confiança considerado para o cálculo do intervalo de confiança.

O default do R é nível de confiança = 95%.

Os valores estimados para os limites inferior (LCL Mean) e superior do intervalo de confiança (UCL Mean) são 14,62 e 15,29, respectivamente. Indicando que a média da altura (em metros) das árvores da população com 95% de confiança deve estar neste intervalo.

Este comando também pode ser utilizado para obter estatísticas, semelhante ao comando 'summary' (já visto em outro encontro).

3 Intervalo de confiança para proporção

Por exemplo, vamos supor uma pesquisa de satisfação quanto o produto comprado, na qual participaram 485 consumidores aleatoriamente selecionados, destes 320 responderam estarem satisfeito ou muito satisfeito com o produto comprado.

Assim, o comando para calcular o intervalo de confiança para a proporção da população que neste caso é a proporção de consumidores satisfeitos com o produto comprado é:

```
prop.test(320,485)

##
##  1-sample proportions test with continuity correction
##
## data:  320 out of 485, null probability 0.5
## X-squared = 48.899, df = 1, p-value = 2.695e-12
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.6154670 0.7015481
## sample estimates:
##         p
## 0.6597938
```

4 Teste de hipótese para verificar a normalidade dos dados

Para verificar se os dados seguem uma distribuição normal, podemos, inicialmente usar o histograma (já visto em outro encontro) e depois confirmar com um teste estatístico para testar normalidade como Shapiro-Wilk ou Kolmogorov-Smirnov.

```
attach(arvores)
shapiro.test(altura_m)

##
##  Shapiro-Wilk normality test
##
## data:  altura_m
## W = 0.99052, p-value = 0.0002336
```

5 Teste de hipótese para a comparação de duas médias entre duas amostras

O **valor p** reflete a plausibilidade de se obter tais resultados no caso de Ho ser de fato verdadeira.

Regra de decisão do valor p

$p \leq 0{,}01$. Rejeita-se H_0 ao nível de 1% de significância.

$0{,}01 < p \leq 0{,}05$. Rejeita-se H_0 ao nível de 5% de significância.

$P > 0{,}05$. Não rejeita-se H_0.

5.1 Comparar médias entre duas amostras dependentes:

Para exemplificar, vamos digitar os dados diretamente na linha de comando:

EXEMPLO: É obtido o peso de seis indívíduos antes e após um treinamento de exercício físico. teste a hipótese de que a média antes do treinamento é diferente da média após o treinamento. Utilizando o nível de significância de 0,05.

Indivíduo	A	B	C	D	E	F
Peso antes do treinamento (kg)	99	62	74	59	70	73
Peso depois do treinamento (kg)	94	62	66	58	70	76

```
antes=c(99, 62, 74, 59, 70, 73)
depois=c(94, 62, 66, 58, 70, 76)
t.test(antes, depois, paired = TRUE)

##
##  Paired t-test
##
## data:  antes and depois
## t = 1.131, df = 5, p-value = 0.3094
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.333688  6.000355
## sample estimates:
## mean of the differences
##                1.833333
```

1ºpasso) Definir hipóteses

H0: média antes = média depois
H1: média de antes diferente da média depois

2º passo) Aplicar o teste estatístico adequado

Como são amostras dependentes, utiliza-se o teste t para comparar amostras dependentes.

5.2 Comparar médias entre duas amostras independentes

Primeiramente precisamos saber se existe homogeneidade de variâncias populacionais, a qual poderá ser verificada através de um teste de homogeneidade de variâncias utilizando os dados das duas amostras.

5.2.1 Teste para verificar a homogeneidade de variância

Se o resultado do teste (p menor que 0,01) for altamente significativo ou (valor de o entre 0,01 e 0,05) significativo então concluímos que as variâncias são diferentes, desta forma precisamos indicar no comando (VAR.EQUAL = FALSE).

No caso do teste não se significativo (p maior que 0,05) então concluímos que as variâncias não são diferentes, desta forma precisamos indicar no comando (VAR.EQUAL=TRUE)

```
attach(arvores)

## The following objects are masked from arvores (pos = 3):
##
##     altura_m, diametro_cm, Nomecientifico

var.test(diametro_cm,altura_m)

##
##  F test to compare two variances
##
## data:  diametro_cm and altura_m
## F = 112.07, num df = 678, denom df = 678, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##   96.39133 130.28931
## sample estimates:
## ratio of variances
##           112.0659
```

5.2.2 Teste para comparar médias entre duas amostras independentes

1º passo: Definir hipóteses:

H0 = média do diâmetro = média da altura
H1 = média do diâmetro ¡¿ média da altura

2º passo: Aplicar o teste estatístico adequado

Teste t para duas amostras independentes (paired=FALSE) e pequenas considerando variâncias populacionais diferentes (var.equal=FALSE).

```
attach(arvores)

## The following objects are masked from arvores (pos = 3):
##
##     altura_m, diametro_cm, Nomecientifico
## The following objects are masked from arvores (pos = 4):
##
##     altura_m, diametro_cm, Nomecientifico

t.test(diametro_cm,altura_m,var.equal=FALSE,paired=FALSE)

##
##  Welch Two Sample t-test
##
## data:  diametro_cm and altura_m
## t = 31.73, df = 690.1, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  54.56064 61.75839
## sample estimates:
## mean of x mean of y
##  73.11489  14.95538
```

6 REFERÊNCIAS BIBLIOGRÁFICAS

TRIOLA, M. F. **Introdução à estatística**. 2. ed. Rio de Janeiro: LTC, 1999.

FERREIRA, D. F. **Recursos Computacionais Utilizando o R**. UFLA, 2013.