

RAD1507 – Estatística Aplicada à Administração I
Prof. Dr. Evandro Marcos Saidel Ribeiro

RESUMO - 

O SOFTWARE R

O R é um software livre.

Informações e download do software:

<http://www.r-project.org/>

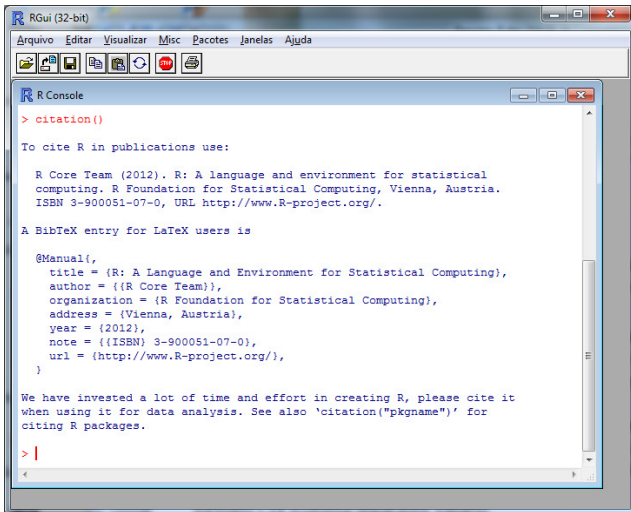


Figura 1 Área de trabalho do R com a janela de comandos.

LEITURA DE DADOS NO R

Existem várias formas para inserir dados no R. Podemos gravar os dados de uma planilha Excel no formato CSV (comma-separated values).

Para este resumo será utilizado o arquivo, disponibilizado no EAD:

Cap01_Corrar_etal_2007.csv

Para ler os dados temos que mudar o diretório do R: Na área de trabalho do R (Figura 1) utilize o menu: Arquivo > Mudar dir...

Em seguida selecione o diretório que contém o arquivo com os dados para análise

Para ler os dados digite na janela de comandos.

```
dados <- read.table("Cap01_Corrar_etal_2007.csv",  
header=TRUE,sep=";")
```

Observe as opções:

header=TRUE (Primeira linha com nomes)
sep=";" (separação entre os valores)

O comando "<- read.table" passa para "dados" o conteúdo do arquivo. Para verificar o conteúdo basta digitar "dados" na linha de comando.

Nesta forma criamos um **data.frame**. Esta é uma das classes que um objeto pode ter. No caso o objeto "dados" é um data.frame. As variáveis são as colunas do Dataframe.

Importante: Vamos fazer os exemplos utilizando o Dataframe como um todo, mas existem outras formas de realizar a análise.

REVISÃO DE INTRODUÇÃO À ESTATÍSTICA

Para obter um resumo das colunas do dataframe basta digitarmos o comando summary():

```
summary(dados)
```

É apresentado um resumo com os valores de mínimo, primeiro quartil, mediana, média, terceiro quartil e máximo. Mas num dataframe nem todas as colunas precisam ser da mesma classe e no caso as três primeiras são variáveis categóricas. As estatísticas apresentadas são válidas apenas para as variáveis métricas, PL, AC, PC, AP, ARLP, PELP e VLL.

Vamos transformar as variáveis categóricas para fatores assim elas serão utilizadas de forma correta.

Transformar CAP e TAM:

```
dados$CAP <- factor(dados$CAP, labels=
c("Capital Aberto", "Capital Fechado"))
```

```
dados$TAM <- factor(dados$TAM, labels=
c("Pequena", "Média", "Grande"))
```

Observe o resultado utilizando o comando `summary(dados)`

```
> summary(dados)
  Empresa      CAP      TAM      PL
Min.   : 1.00  Capital Aberto:60 Pequena:34 Min.   : 33875
1st Qu.: 25.75 Capital Fechado:40 Média :32  1st Qu.: 61991
Median : 50.50                                     Median : 67750
Mean   : 50.50                                     Mean   : 71246
3rd Qu.: 75.25                                     3rd Qu.: 81300
Max.   :100.00                                     Max.   :111110

  AC      PC      AP      ARLP
Min.   :14575 Min.   :12075 Min.   : 56425 Min.   : 1668
1st Qu.:29150 1st Qu.:41400 1st Qu.: 88450 1st Qu.:12302
Median :34450 Median :51750 Median :109038 Median :17931
Mean   :35311 Mean   :50249 Mean   :106094 Mean   :19716
3rd Qu.:39750 3rd Qu.:59081 3rd Qu.:126956 3rd Qu.:26897
Max.   :60950 Max.   :79350 Max.   :152500 Max.   :45036

  PELP      VLL
Min.   : 5868 Min.   :-0.117300
1st Qu.:25184 1st Qu.: 0.005575
Median :33252 Median : 0.015400
Mean   :34735 Mean   : 0.016946
3rd Qu.:44988 3rd Qu.: 0.030125
Max.   :59658 Max.   : 0.096500
> |
```

A partir deste ponto, para informar a escrita do comando vou considerar variáveis indicadas por x (para mais de uma variável utilizarei x1, x2, ...). No caso de exemplo serão utilizadas as variáveis do dataframe "dados".

Medidas de centro

Média: `mean(x)`

Exemplos

```
with(dados, mean(PL))
```

```
[1] 71245.9
```

Uma outra forma de acessar os dados é através da notação \$: `dataframe$variável`:

Exemplo:

```
mean(dados$PL)
```

```
[1] 71245.9
```

attach e detach

Para que as variáveis fique disponíveis de forma mais direta utilizamos o comando `attach(dataframe)` `attach(dados)`. Desta forma não precisamos utilizar a notação \$. Depois utilizamos `detach(dados)`

A partir daqui utilizamos: **attach(dados)**

Mediana: `median(x)`

A mediana é o valor que separa 50% dos valores inferiores dos 50% dos valores superiores, em outras palavras é o percentil 50.

Exemplos: `median(PL)` `median(AC)`

```
[1] 67750 [1] 34450
```

Percentil: `quantile(x,p)`
`quantile(x, probs = c(0.1, 0.5, 1, 2, 5, 10, 50)/100)`

Sendo p o percentil, por exemplo se p = 0.50 temos a mediana.

Exemplos: `quantile(PL, 0.50)`

```
50%
67750
```

`quantile(PL, probs = c(0.1, 0.5, 1, 2, 5, 10, 50)/100)`

| Percentil | 0.1% | 0.5% | 1% | 2% |
|-----------|----------|----------|----------|----|
| 34411.58 | 36557.90 | 39240.80 | 40622.90 | |
| 46002.25 | 51490.00 | 67750.00 | | |

Medidas de dispersão

Desvio Padrão Amostral `sd(x)`

Exemplos

```
sd(PL)          sd(AC)
[1] 15312.14    [1] 10213.83
```

Variância s^2 `var(x)`

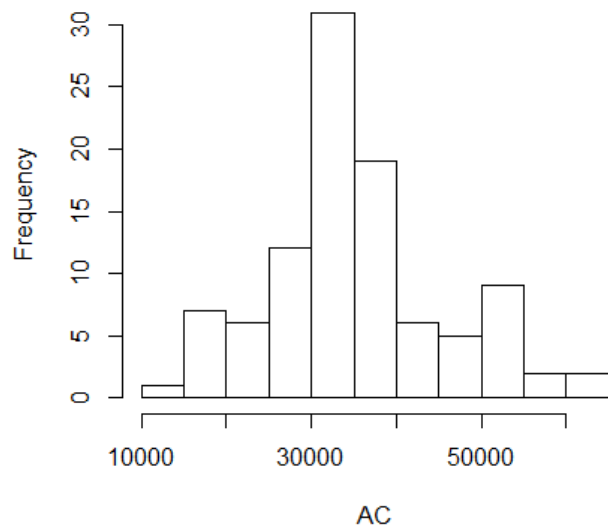
Exemplo: `var(PL)`

```
[1] 234461505
```

Gráficos

Histograma `hist(x)` `hist(AC)`

Histogram of AC



Boxplot `boxplot(x)`
`boxplot(x,range=0)` não mostra outliers

`boxplot(PL~TAM,ylab="PL($)",xlab="Tamanho")`

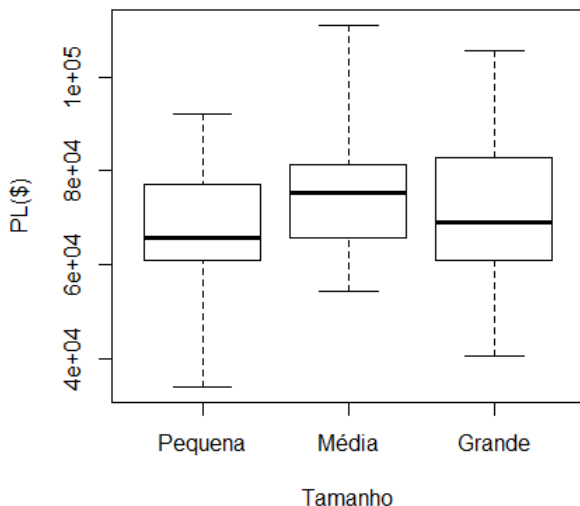


Gráfico de frequência para dados categóricos

Gráfico de Barras `barplot(table(x))`
`barplot(table(TAM),xlab="Tamanho",ylab="Contagem")`

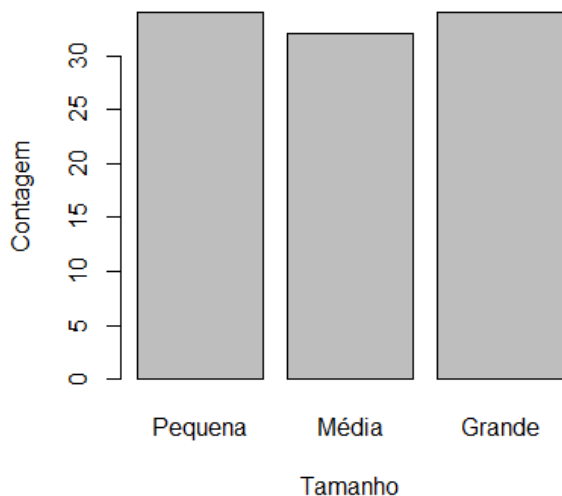
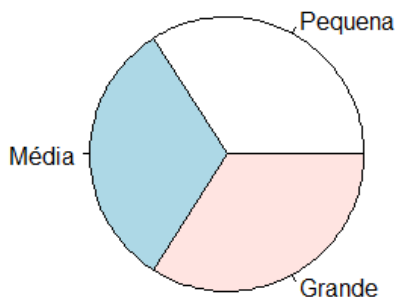


Gráfico de Pizza `pie(table(x))` `pie(table(TAM))`



Distribuição Binomial

A probabilidade de se obter exatamente x sucessos em n tentativas, sendo a p a probabilidade de sucesso em uma tentativa, é dada por

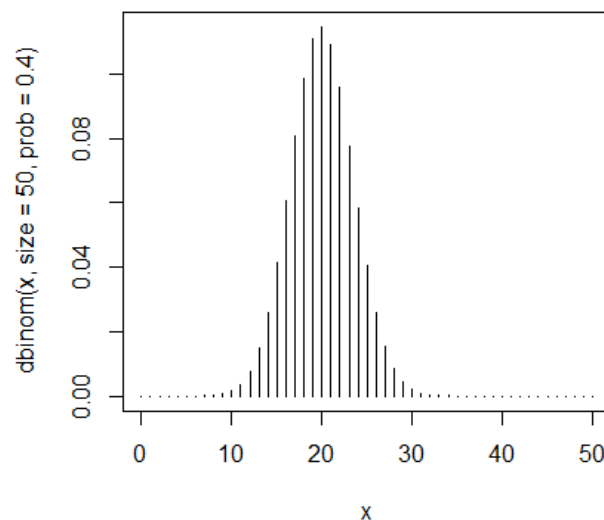
$$P(x) = \frac{n!}{(n-x)!x!} p^x q^{n-x} = \text{dbinom}(x,n,p)$$

e $q = 1 - p$

Gráfico

`x<-0:50`

`plot(x,dbinom(x,size=50,prob=.4),type="h")`



Distribuição Normal

Densidade de probabilidade = `dnorm(x, μ, σ)`

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Área à esquerda de x : = `pnorm(x, μ, σ)`

Score-z ou padronização dos dados `zscore(x)`

$$z = \frac{(x - \bar{x})}{s}$$

Instalar e carregar o pacote R.basic:

`Xz <- zscore(dados$PL)`

Intervalo de Confiança para média `t.test(x)`

`t.test(AC, conf.level=.90)`

One Sample t-test

data: AC

$t = 34.572$, $df = 99$, $p\text{-value} < 2.2e-16$

alternative hypothesis: true mean is not equal to 0

90 percent confidence interval:

33615.36 37007.14

sample estimates:

mean of x

35311.25

Intervalo de Confiança para proporção prop.test(x)

Contagem de itens válidos: $n = \text{length}(\text{CAP})$

Contagem de itens iguais à Capital Aberto:

$k = \text{sum}(\text{CAP} == \text{"Capital Aberto"})$

Teste de proporção $\text{prop.test}(k, n)$

1-sample proportions test with continuity correction

data: k out of n, null probability 0.5

X-squared = 3.61, df = 1, p-value = 0.05743

alternative hypothesis: true p is not equal to 0.5

95 percent confidence interval:

0.4970036 0.6952199

sample estimates:

p

0.6

ESTATÍSTICA APLICADA ADM 1

1 TESTE DE HIPÓTESE: uma amostra

Teste de Proporção

$\text{prop.test}(x, n, \text{conf.level}=\text{NC})$

$\text{prop.test}(x, n, p = \text{NULL},$

$\text{alternative} = \text{c}(\text{"two.sided"}, \text{"less"}, \text{"greater"}),$

$\text{conf.level} = 0.95, \text{correct} = \text{TRUE})$

x = número de sucessos

n = tamanho da amostra

p (opcional) proporção indicada em H_0

alternative (opcional, assume-se duas caudas)

NC = Nível de confiança (Ex.: NC = 0.99)

correct (correção de continuidade, para o cálculo sem esta correção, visto em aula, colocar correct = FALSE)

Exemplo 1: Testar se a proporção da população é maior do que 0.85 sendo que numa amostra de tamanho 100 foram observados 94 sucessos.

```
> prop.test(94, 100, p = 0.85, alternative = "greater",  
conf.level = 0.95, correct = FALSE)
```

Exemplo 2: Testar se a proporção da população é diferente de 0.5 sendo que numa amostra de tamanho 200 foram observados 105 sucessos.

```
> prop.test(105, 200)
```

```
> prop.test(105, 200, correct = FALSE)
```

Compare os resultados obtidos.

Cálculo da área à esquerda da estatística teste (z_{teste})

```
> pnorm( $z_{\text{teste}}$ )
```

Teste de média: teste-t

$\text{t.test}(x, \text{conf.level}=\text{NC})$ neste caso testa média = 0.

$\text{t.test}(x, \text{mu} = \text{valor}, \text{conf.level}=\text{NC})$ testa se a média é igual a "valor".

Exemplo 1: Teste se a média da população é igual a 368.0.

```
x <- c(380.8, 372.0, 379.8, 356.0, 349.3, 375.5,  
375.7, 370.4, 372.6, 380.9, 374.6, 372.8, 358.1,  
382.2, 381.9, 375.9, 359.7, 367.6, 362.1, 358.1,  
349.7, 383.4, 369.8, 387.6, 359.3)
```

```
t.test(x, conf.level=0.95, mu=368)
```

Exemplo 2: Teste se a média da população é maior do que 368.0.

```
t.test(x, conf.level=0.95, mu=368,  
alternative = "greater")
```

Cálculo da área à esquerda da estatística teste (t_{teste})

```
> pt( $t_{\text{teste}}$ )
```

2 TESTE DE HIPÓTESE: duas amostras

Teste de média: teste-t

$\text{t.test}(x, y, \text{conf.level}=\text{NC})$

$\text{t.test}(x \sim y, \text{conf.level}=\text{NC})$

3 TESTE DE HIPÓTESE: mais de duas amostras

4 CORRELAÇÃO

```
x <- c(4, 6, 3, 5, 1, 2)
y <- c(2, 1, 3, 1, 4, 3)
z <- c(1, 3, 2, 7, 5, 4)
```

Gráficos de Dispersão

```
plot(x,y)
plot(x,z)
plot(y,z)
```

Coeficiente de correlação de Pearson

```
cor(x,y)
cor(x,z)
cor(y,z)
```

Teste de hipótese, correlação:

```
cor.test(x,y)
cor.test(x, y,
  alternative = c("two.sided", "less", "greater"),
  method = c("pearson", "kendall", "spearman"),
  exact = NULL, conf.level = 0.95, continuity =
FALSE, ...)
```

5 REGRESSÃO LINEAR SIMPLES

```
Ys <- glm( y ~ x)
summary(Ys)
```

Resultado no R:

Call:

```
glm(formula = y ~ x)
```

Deviance Residuals:

```
 1    2    3    4    5
-0.01905  0.23810  0.35238 -0.39048  0.09524
 6
-0.27619
```

Coefficients:

| | Estim. | Std.E | t value | Pr(> t) |
|-------------|---------|-------|---------|----------|
| (Intercept) | 4.53 | 0.301 | 15.045 | 0.0001 |
| x | -0.6286 | 0.077 | -8.124 | 0.001 |

(Intercept) ***
x **

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

6 MATRIZ DE COVARIÂNCIA

```
cov(x,y)
cov(x,z)
cov(y,z)
```

7 REGRESSÃO LINEAR MÚLTIPLA

```
Ym <- glm( y ~ x + z)
summary(Ym)
```