

Instruções:

- Você pode consultar a internet e o cartão de referência do R. Você não pode consultar o caderno nem trocar informações com colegas;
- Todas as questões devem ser respondidas na folha de prova, com os códigos empregados e os resultados obtidos;
- É altamente recomendável que você faça a prova no computador do laboratório (linux). Isso porque a importação de dados pode gerar problemas devido a codificação de caracteres. Caso faça a prova no seu computador pessoal, a responsabilidade de fazer a importação total e correta dos conjuntos de dados é sua.

1. Dez indivíduos participaram de um programa de modificação alimentar para estimular a perda de peso. Seus pesos antes e depois da participação no programa são mostrados na tabela abaixo.

indivíduo	antes	depois
1	195	187
2	213	195
3	247	221
4	201	190
5	187	175
6	210	197
7	215	199
8	246	221
9	294	278
10	310	285

- passar os dados para o R;
- obtenha a média, a mediana e a variância amostral para o peso dos indivíduos antes e depois do programa de modificação alimentar;
- calcule a média e a variância amostral para a diferença de peso dos indivíduos;
- há evidência para confirmar a afirmação de que esse programa particular de modificação alimentar seja efetivo na redução do peso médio? Defina a hipótese nula e alternativa, cite o teste e a função R a serem empregados, aplique o teste de hipótese considerando $\alpha = 0,05$.

```
> #-----  
> # a) passando os dados para o R  
> # a.1) digitando em um data.frame  
> dados <- data.frame(antes=c(195, 213, 247, 201, 187,  
+ 210, 215, 246, 294, 310),  
+ depois=c(187, 195, 221, 190, 175,  
+ 197, 199, 221, 278, 285))  
> # a.2) digitando em uma matriz  
> dados <- c(195, 187, 213, 195, 247, 221, 201,  
+ 190, 187, 175, 210, 197, 215, 199,  
+ 246, 221, 294, 278, 310, 285)  
> dados <- matrix(dados, byrow=TRUE, ncol=2,  
+ dimnames=list(NULL, c("antes", "depois")))  
> # a.3) digitando na janela gráfica de edição  
> # edit(data.frame()) # então digitar os dados  
>  
> # a.4) usando textConnection()  
> lines <- "antes depois  
+ 195 187  
+ 213 195  
+ 247 221  
+ 201 190  
+ 187 175  
+ 210 197  
+ 215 199  
+ 246 221  
+ 294 278  
+ 310 285"  
> #dados <- read.table(textConnection(lines), header=TRUE)  
> #closeAllConnections()  
> str(dados)
```

```

num [1:10, 1:2] 195 213 247 201 187 210 215 246 294 310 ...
- attr(*, "dimnames")=List of 2
 ..$ : NULL
 ..$ : chr [1:2] "antes" "depois"
> #-----
> # b) média, mediana, variância para antes e depois
> # b.1) fazendo separado
> apply(dados, 2, mean)

antes depois
231.8 214.8
> apply(dados, 2, median)

antes depois
214 198
> apply(dados, 2, var)

antes depois
1759.733 1436.622
> # b.2) fazendo junto
> apply(dados, 2, function(x){
+ c(media=mean(x), mediana=median(x), variancia=var(x))
+ })

media antes depois
mediana 231.800 214.800
198.000
variancia 1759.733 1436.622
> #-----
> # c) média e variância para a diferença de peso
> pdif <- dados[,2]-dados[,1] # diferença = depois-antes
> c(media=mean(pdif), variancia=var(pdif))

media variancia
-17.00000 41.11111
> #-----
> # d) testar hipótese se  $m_{depois} < m_{antes}$  ou  $pdif < 0$ 
> #  $h_0: m_{depois} = m_{antes}$  ou  $pdif = 0$ 
> #  $h_1: m_{depois} < m_{antes}$  ou  $pdif < 0$ 
> # teste t pareado unilateal, t.test() ou
> # teste t unilateral para uma amostra usando pdif
> t.test(dados[,2], dados[,1],
+ paired=TRUE, alternative="less") # t pareado

Paired t-test

data: dados[, 2] and dados[, 1]
t = -8.3843, df = 9, p-value = 7.593e-06
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -13.2832
sample estimates:
mean of the differences
 -17
> t.test(pdif, alternative="less") # t 1 amostra

One Sample t-test

data: pdif
t = -8.3843, df = 9, p-value = 7.593e-06
alternative hypothesis: true mean is less than 0
95 percent confidence interval:
 -Inf -13.2832
sample estimates:
mean of x
 -17
>
> #-----
>

```

2. O tempo para que um sistema computacional execute determinada tarefa é uma variável aleatória com distribuição normal, com média 320 segundos e desvio padrão de 7 segundos.

- Qual a probabilidade da tarefa ser executada entre 310 e 330 segundos?
- Se a tarefa é colocada para execução 200 vezes, qual a probabilidade de ela demorar mais que 325 segundos em pelo menos 50 vezes?

```

> #-----
> # a)  $P(310 < X < 330 \mid \mu=320, \sigma=7)$ 
> mu <- 320; sigma <- 7
> pnorm(330, mu, sigma) - pnorm(310, mu, sigma)
[1] 0.8468725
> #-----
> # b)  $P(X > 50, n=200, p=p^*), p^*=P(X > 325 \mid \mu=320, \sigma=7)$ 
> n <- 200
> p <- pnorm(325, mu, sigma, lower.tail=FALSE); p
[1] 0.2375253
> 1-pbinom(50, size=n, prob=p)
[1] 0.305581
>
> #-----

```

3. Considere os dados sobre forma de agregados do solo coletados em duas profundidades e obtenha:

```

ag <- read.table(
  "http://www.leg.ufpr.br/~walmes/cursoR/agreg.txt",
  header=TRUE, sep="\t")

```

- faça a importação dos dados e descreva o objeto `ag`;
- a tabela de distribuição de frequência para a variável `aspecto` em cada profundidade. Use 6 classes de mesma amplitude. Relate as informações obtidas;
- a tabela com a profundidade na coluna e as estatísticas descritivas para `aspecto` nas linhas: média, mediana, desvio-padrão, mínimo, máximo e número de elementos na amostra. Comente esses resultados;
- faça um gráfico de caixas para a variável `aspecto` em função da profundidade. Comente o que você observa?
- faça o gráfico dos valores ordenados de `aspecto` da profundidade 5 contra os valores ordenados de `aspecto` da profundidade 20. Adicione a linha da função identidade. O que você pode concluir desse gráfico?
- aplique um teste de hipótese para saber se existe diferença entre os níveis de profundidade para o valor médio de `aspecto`. O que você conclui?

```

> #-----
> # a) importar e descrever o objeto
> ag <- read.table(
+   "http://www.leg.ufpr.br/~walmes/cursoR/agreg.txt",
+   header=TRUE, sep="\t")
> str(ag) # isso dá a descrição do objeto
'data.frame':   600 obs. of  7 variables:
 $ profundidade: int  5 5 5 5 5 5 5 5 5 ...
 $ area        : num  0.0038 0.005 0.008 0.0052 0.0055 0.0046 0.0055 0.0039 0.0051 0.0079 ...
 $ perimetro   : num  0.2 0.3 0.4 0.3 0.3 0.2 0.3 0.2 0.3 0.4 ...
 $ aspecto     : num  0.947 0.913 0.806 0.955 0.792 ...
 $ roundness   : num  0.968 0.871 0.795 0.83 0.897 ...
 $ maioreixo   : int  18 22 30 21 23 20 22 23 23 28 ...
 $ menoreixo   : int  17 20 24 20 18 20 19 15 17 27 ...
> #-----
> # b) tabela de distribuição de frequências com 6 classes
>
> class <- seq(0.4,1,0.1) # 6 classes de 0.4 à 1
> asp.class <- cut(ag$aspecto, class) # classifica
> table(asp.class, ag$profundidade) # conta
asp.class      5 20
 (0.4,0.5]    0  1
 (0.5,0.6]    7  4
 (0.6,0.7]   22 23
 (0.7,0.8]   67 60
 (0.8,0.9]   76 86
 (0.9,1]    128 126
> #-----
> # c) tabela com profundidade e estatísticas descritivas
> # c.1) fazendo subconjunto e depois calculando
> p05 <- subset(ag, profundidade==5)$aspecto
> p20 <- subset(ag, profundidade==20)$aspecto
> estat <- function(x){ # função que calcula estatísticas
+   c(me=mean(x), md=median(x), dp=sd(x),
+     min=min(x), max=max(x), n=length(x))}
> data.frame(em05=estat(p05), em20=estat(p20))

```

```

      em05      em20
me  0.8528470  0.8571670
md  0.8723000  0.8750000
dp  0.1114107  0.1037468
min 0.5143000  0.4615000
max 1.0000000  1.0000000
n   300.0000000 300.0000000

> # c.2) fazendo usando tapply+do.call ou +as.data.frame
> aux <- with(ag, tapply(aspecto, profundidade, estat))
> do.call(cbind, aux)

      5      20
me  0.8528470  0.8571670
md  0.8723000  0.8750000
dp  0.1114107  0.1037468
min 0.5143000  0.4615000
max 1.0000000  1.0000000
n   300.0000000 300.0000000

> #-----
> # d) gráfico de caixas
> boxplot(aspecto~profundidade, data=ag)
> # observa mesma amplitude, mesmo centro,
> # mesma assimetria à esquerda
>
> #-----
> # e) gráfico dos valores ordenados
> plot(sort(p05)~sort(p20))
> abline(a=0, b=1)
> # observa uma leve tendência de aspecto ser maior na 20
>
> #-----
> # e) teste de hipótese, t para duas médias bilateral
> # e.1) passando para a t.test() os vetores
> t.test(p05, p20, var=TRUE)

      Two Sample t-test

data:  p05 and p20
t = -0.4915, df = 598, p-value = 0.6233
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02158173  0.01294173
sample estimates:
mean of x mean of y
 0.852847  0.857167

> # e.2) passando para a t.test() a fórmula
> with(ag, t.test(aspecto~profundidade, var=TRUE))

      Two Sample t-test

data:  aspecto by profundidade
t = -0.4915, df = 598, p-value = 0.6233
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02158173  0.01294173
sample estimates:
mean in group 5 mean in group 20
 0.852847      0.857167

>
> #-----

```

4. Considere os dados sobre a população do Brasil por municípios ao longo dos anos e obtenha:

```

poptot <- read.table(
"http://www.leg.ufpr.br/~walmes/ensino/ce223-2011-01/dadospoptot2010.txt",
header=TRUE, sep="\t", quote="")

```

- faça a importação dos dados, descreva o objeto poptot e remova as observações perdidas;
- de quanto foi aumento populacional brasileiro de 2007 para 2010?
- no mesmo período de tempo, qual foi o aumento populacional do estado do Paraná?
- entre todos os estados, qual o que apresentou o maior crescimento populacional relativo (cpr) entre 2007 e 2010? Use

$$cpr = (\text{pop}_{2010} - \text{pop}_{2007}) / \text{pop}_{2007}$$

em que pop_{ano} é a população do estado em cada ano;

e) faça um gráfico de barras do cpr em função dos estados com os estados em ordem crescente de cpr.

```
> #-----
> # a) importa e descreve e remove os NA dos dados
> poptot <- read.table(
+ "http://www.leg.ufpr.br/~walmes/ensino/ce223-2011-01/dadospoptot2010.txt",
+ header=TRUE, sep="\t", quote="")
> poptot <- poptot[complete.cases(poptot),]
> str(poptot)

'data.frame':      4491 obs. of  8 variables:
 $ Sigla   : Factor w/ 27 levels "AC","AL","AM",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ Codigo  : int  1200054 1200104 1200203 1200302 1200336 1200344 1200385 1200401 1200450 1200500 ...
 $ Município: Factor w/ 5310 levels "Abadia de Goiás",...: 396 706 1412 1677 2768 2782 3710 4014 4740 4749 ...
 $ a1991   : int  2917 20263 66603 17769 10217 5327 15535 197376 17489 24197 ...
 $ a1996   : int  2918 13955 56705 22142 8699 5520 12101 228857 14280 23330 ...
 $ a2000   : int  3490 17013 67441 26722 11095 6374 15172 253059 19761 29420 ...
 $ a2007   : int  5351 19065 73948 31288 13785 7148 17258 290639 18863 34230 ...
 $ a2010   : int  6075 21438 78444 32311 15246 7989 17203 335796 20153 37993 ...

> #-----
> # b) aumento de 2007 para 2010 em número de habitantes
> au <- sum(poptot$a2010)-sum(poptot$a2007); au
[1] 6200641

> #-----
> # c) aumento populacional do PR
> pr <- subset(poptot, Sigla=="PR")
> sum(pr$a2010)-sum(pr$a2007)
[1] 131449

> #-----
> # d) maior crescimento populacional relativo
> a10 <- with(poptot, tapply(a2010, Sigla, sum)) # em 2010
> a07 <- with(poptot, tapply(a2007, Sigla, sum)) # em 2007
> cpr <- (a10-a07)/a07
> cbind(a10, a07, cpr)
      a10      a07      cpr
AC  624190  558060  0.118499803
AL  3063439 2983029  0.026955822
AM  3480937 3221939  0.080385755
AP   611553  540368  0.131734300
BA 13947064 14023207 -0.005429785
CE  8347060  8090718  0.031683467
DF  2562963  2455903  0.043592927
ES  3357108  3207213  0.046736840
GO  5428568  5147165  0.054671455
MA  5496530  5136571  0.070077684
MG 18642361 18363110  0.015207174
MS  2406329  2227527  0.080269285
MT  2732420  2570404  0.063031337
PA  6706310  6266075  0.070256899
PB  3541504  3418769  0.035900349
PE  8554254  8253286  0.036466445
PI  2595202  2513466  0.032519238
PR  9784322  9652873  0.013617604
RJ 14635068 14106036  0.037503945
RN  3093253  2940542  0.051932943
RO  1256544  1158316  0.084802420
RR   367147  323562  0.134703704
RS 10125458 10014957  0.011033597
SC  5936328  5562506  0.067203883
SE  2060993  1932830  0.066308470
SP 40587504 39199669  0.035404253
TO  1175145  1050812  0.118320879

> which.max(cpr)
RR
22

> #-----
> # e) gráfico de barras com estados ordenados por cpr
> barplot(sort(cpr))
>
> #-----
```


