Instituto Superior de Agronomia
Modelos Matemáticos e Aplicações – 2017/2018

Test for Module I– duração: 2h30 3 de May 2018

1. Consider the data `rice` available in the R package `agricolae` that correspond to the yield of rice variety IR8 (g/m2) for land uniformity studies. The growing area is 18×36 meters. The data are given in a *data frame* with 36 observations on the following 18 variables `V1-V18`. To answer the following questions consider only six variables `V13-V18`.

Consider the *output* that is in the Appendix in which the chosen variables are studied through some commands executed in the R. Whenever possible use the results to answer the following questions:

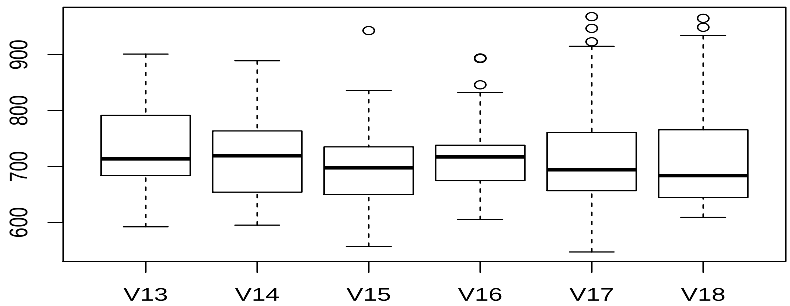
a) Sketch the **histogram**, indicated in the *output* but not plotted. What changes should you enter in the command for it to be plotted?

b) Interpret the results obtained with the command basicStats (V18).

c) Show the necessary calculations for plotting the *boxplot* for the variable `V18`. Clearly state the *boxplot* limits.

d) Compare the *boxplots* shown below.

e) Given the results presented in the *output*, it can be said that the variable `V13` has, on average, larger values than the `V15` variable? Justify properly.



2. a) It is intended to study the weight of rabbits reared on a property. For this a sample of 80 rabbits was collected on that property. Consider some results presented in the **Appendix**.

i) Give a point estimate of the mean weight of the rabbits on that property. With 95% confidence, what is the maximum limit of the error made when estimating the true mean weight of the rabbits by that indicated estimate?

ii) Can it be said that more than 20% of rabbits weigh 5.5 kg or more? Justify properly.

b) Suppose that the weight of rabbits reared on this property can be considered to have a normal distribution with a mean value of 5 kg and a standard deviation of 0.8 kg.

i) What is the percentage of rabbits weighing more than 7 kg?

ii) A dealer will buy 5000 rabbits and intends to sort them according to the weight according to the following rule: 20% of the lightest as "small", the following 55% as the average, the following 15% as "big"and 10% heavier as "extras". What are the weight limits for each ranking?

3. Let X be a r.v. representing the weight, in kg, of grapes produced by vines of a given quality, which is assumed to be approximately normal, $X \frown \mathcal{N}(\mu, \sigma)$, with $\mu = 60$ kg and $\sigma = 10$ kg.

a) Suppose you want to get 50 values ??of that model by simulation. What commands in R are required to get these values?

b) Write the necessary commands to calculate $P[55 < X < 70]$ using the R.

c) Let Z be a standard normal r.v. associated with X. What are the commands in R to get the density function graphs of Z and Z^2?

4. Consider you have a random sample with size n, $(X_1, X_2, .., X_n)$, obtained from a population X, with density function defined as, where $\theta > -1$ is an unknown parameter:

$$f(x;\theta) = \begin{cases} \dfrac{\theta+1}{\mathrm{e}^{\theta+1}} x^{\theta} & \text{se } \ 0 \le x \le \mathrm{e} \\ 0 & \text{other values of } x \end{cases}$$

Nota: It is known that $E[X] = \dfrac{(\theta+1)\mathrm{e}}{\theta+2}$.

a) Get the θ estimator by the Method of Moments.

b) Get the maximum likelihood estimator for θ.

c) Consider the observed values of the following sample, with size 30, extracted from that population, with which the calculations below were performed:

```
> dados
 [1] 2.13 2.26 1.32 1.88 1.63 2.27 2.65 2.60 1.80 2.00
[11] 1.85 1.54 0.63 2.47 1.67 2.14 1.47 2.30 2.02 2.57
[21] 1.77 2.40 1.96 2.33 2.12 1.62 0.82 2.39 2.57 2.59

> sum(dados)                > sum(log(dados))
[1] 59.77                    [1] 19.39522
```

i) Determine estimates for θ.

ii) Determine a maximum likelihood estimate for $E[X]$.

5. Indicate, justifying, whether the following statements are **True** or **False**. Please **correct the False**.

a) Seja $(X_1, ..., X_n)$ uma amostra aleatória de tamanho n, proveniente de uma população com valor médio μ e variância $\sigma^2 < +\infty$. Então $E[X_i X_j] = \mu^2$ quaisquer que sejam $i \neq j$.

b) μ, $\overline{X}$ e $\overline{x}$ são formas diferentes de designar a mesma quantidade: a média da população em estudo.

c) Seja $X \frown Normal(0,1)$ então $5X \frown Normal(5,5)$.

d) Seja$(X_1, ..., X_n)$ uma amostra aleatória de tamanho n e $X_{(n)} = max(X_i)$. Tem-se $P[X_{(n)} \le x] = nP[X_1 \le x]$, qualquer que seja x.

ANEXO I

```
########  Pergunta 1   ########

> library(agricolae)
> data(rice)
> rice
> attach(rice)

> hist(V15,breaks=c(500,560,620,680,740,800,950),plot=F)

$breaks
[1] 500 560 620 680 740 800 950

$counts
[1]  1  4 10 13  5  3

$density
[1] 0.0004629630 0.0018518519 0.0046296296 0.0060185185 0.0023148148
[6] 0.0005555556

$mids
[1] 530 590 650 710 770 875


> library(fBasics)
> basicStats(V18)
                     V18
nobs           36.000000
NAs             0.000000
Minimum       609.000000
Maximum       965.000000
1. Quartile   644.750000
3. Quartile   764.750000
Mean          724.277778
Median        683.500000
Sum         26074.000000
SE Mean        17.855665
LCL Mean      688.028850
UCL Mean      760.526705
Variance    11477.692063
Stdev         107.133991
Skewness        1.030894

> boxplot(V13,V14,V15,V16,V17,V18,names=c("V13","V14","V15","V16","V17","V18"))
> shapiro.test(V13);shapiro.test(V15)

   Shapiro-Wilk normality test
data:  V13
W = 0.9685, p-value = 0.3865

   Shapiro-Wilk normality test
data:  V15
W = 0.9568, p-value = 0.1705

> var.test(V13,V15)

        F test to compare two variances

data:  V13 and V15
F = 1.0022, num df = 35, denom df = 35, p-value = 0.9949
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5110389 1.9653902
sample estimates:
ratio of variances
          1.002193
```

```
> t.test(V13,V15,var.equal=TRUE)

        Two Sample t-test

data:  V13 and V15
t = 1.8916, df = 70, p-value = 0.06269
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.932271 72.987826
sample estimates:
mean of x mean of y
 733.5833  698.0556

> t.test(V13,V15,var.equal=TRUE,alternative="greater")

        Two Sample t-test

data:  V13 and V15
t = 1.8916, df = 70, p-value = 0.03134
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 4.219346      Inf
sample estimates:
mean of x mean of y
 733.5833  698.0556

> t.test(V13,V15,paired=TRUE,alternative="greater")

        Paired t-test

data:  V13 and V15
t = 2.7517, df = 35, p-value = 0.004664
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 13.71336      Inf
sample estimates:
mean of the differences
               35.52778

> detach(rice)


########  Pergunta 2   ########

> mean(peso)            > var(peso)
[1] 4.96125             [1] 0.8973402

> sort(peso)
 [1] 1.9 3.1 3.2 3.3 3.5 3.6 3.7 3.7 3.8 4.0 4.0 4.1 4.1 4.2 4.2
[16] 4.2 4.3 4.3 4.4 4.4 4.4 4.4 4.4 4.5 4.5 4.5 4.6 4.6 4.7 4.7
[31] 4.7 4.8 4.8 4.8 4.8 4.8 4.8 4.9 4.9 4.9 4.9 5.0 5.0 5.1 5.1
[46] 5.1 5.2 5.2 5.3 5.3 5.3 5.3 5.3 5.3 5.4 5.4 5.4 5.4 5.4 5.4
[61] 5.5 5.5 5.5 5.6 5.6 5.7 5.7 5.7 5.9 6.0 6.0 6.1 6.5 6.5 6.5
[76] 6.7 6.8 6.8 7.0 7.0


>
```