Instituto Superior de Agronomia
Mathematical Models and Applications – 2020/2021

Module I Test – duration: 2h30 9 March 2021

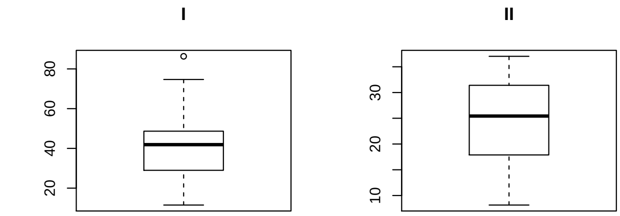
Note: Some calculations to support the resolution of this test can be found in the Appendix

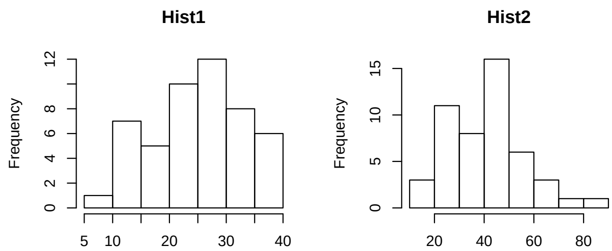
1. A buyer wants to make sure that the weight of pine nuts (in grams) he will get when buying pine cones is at least 35 grams on average per pine cone. In the negotiation phase, an owner affirms that its pine cone gives pine nuts that has, on average, weight greater than that value. To show that he is right, the owner makes a random sampling, in each of two plantations of stone pine in two different locations. You get a sample of 49 pine cones in each of the plantations. For the sampled pine cones the weight of each pine cone and the weight of the pine nuts are calculated, which are found in the *data frame* of the R `pinha.pinhao`. Refer to the Appendix to answer the following questions:

a) Identify the variables contained in the *data frame* and classify them, justifying.

b) Explain the result of executing the command identified by **A**.

c) Associate each boxplot with the corresponding histogram, justifying.





d) For *boxplot* I explain the calculations that led to its construction.

e) Consider the results for plantation A.

i) Provide a descriptive analysis of the results recorded for pine nuts weight by pine cone.

ii) Indicate a 95 % confidence interval for the average weight of pine nuts per pine cone from this plantation. Justify it properly.

iii) Based on the previous paragraph, what can you say about the owner's statement regarding this plantation? Justify.

iv) Indicate, justifying, what happens to the amplitude of the confidence interval obtained in i) for the average pinion weight of this plantation if you have collected a sample of 100 pine cones. Note: suppose there is no change in the observed variance value.

f) Consider the data collected in the two plantations, A and B.

i) Interpret the command identified by B and its result.

ii) Is the average weight of pine nuts in plantation A higher than the average weight of pine nuts in plantation B? Justify it properly.

2. Assume that the weight of pine cones from plantation A follows a normal distribution of an average value of 264 g and standard deviation of 68 grams and that the owner sells them at 0.90 euros / kg.

a) Determine the limits of the weight range that contains 90 % of the pine cones from that plantation.

b) How likely is it that a pinecone, chosen at random, has more than 400 g?

c) The pine cones are placed, at random, in bags that are considered to be filled with 50 pine cones. Answer the following questions, justifying it properly.

   i) How likely is it that in a bag there will be a maximum of 3 pine cones over 400 grams?

   ii) How likely is it that the price of a full bag is less than 12.5 euros?

3. Say, conveniently justifying, whether **true** or **false** are the statements in each of the following paragraphs. Correct the ones you think are false.

a) Consider the random pair (X, Y).

   i) If $Cov(X, Y) = 3$ then X and Y are not independent.

   ii) $Var(X - Y) = Var(X) - Var(Y)$.

b) Let X be a random variable.

   i) If $X \frown Poisson(3)$ then $P[2 < X < 3] = 0$.

   ii) If $X \frown calN(0.1)$ then $P[X = 0] = 1/2$.

   iii) If $X \frown \mathcal{N}(5.2)$ the command in R needed to obtain 50 values with this law is `rnorm (50)`.

c) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n, from a population X with an average value of μ and let $\overline{X}$ be the average of the random sample .

   i) If $X \frown Bernoulli(p)$ then $n\overline{X} \frown B(n, p)$.

   ii) $E[\mu] = \overline{X}$.

4. Consider the v.a. X that characterizes the time until a reaction is observed in the application of a treatment in a given culture, characterized by the following density function, dependent on the parameter, $\beta > 0$, unknown:

$$f(x|\beta) = \sqrt{\frac{2}{\beta\pi}}\, \mathrm{e}^{-\dfrac{x^2}{2\beta}}, \qquad x > 0,\ \beta > 0.$$

It is known that $E[X] = \sqrt{\dfrac{2\beta}{\pi}}$, e $E[X^2] = \beta$.

Consider that you have a random sample of size n, $(X_1, X_2, \ldots, X_n)$, associated with X.

a) Determine the β estimator calculated using the moment method.

b) Determine the maximum likelihood estimator for β.

c) The following sample of 12 values of X was observed:

0.9 0.6 1.2 0.4 1.3 1.7 1.5 2.2 0.7 0.8 1.8 0.3

Calculate estimates of β .

APPENDIX

```
> pinha.pinhao<-read.table(file="pinhas-pinhao.csv", sep=";",header=TRUE, as.is=TRUE)
>
> names(pinha.pinhao)
[1] "Plantacao"        "ppinha"           "ppinhao"

> dim(pinha.pinhao)
[1] 98  3

> head(pinha.pinhao)
  Plantacao    ppinha   ppinhao
1         A    301.89    56.39
2         A    274.55    45.10
3         A    300.89    44.52
4         A    307.01    41.50
5         A    205.45    23.62
6         A    310.84    61.22
> attach(pinha.pinhao)

> ppinhaoA<-ppinhao[Plantacao=="A"]    #Comando A
> ppinhaoB<-ppinhao[Plantacao=="B"]

> library(fBasics)
> basicStats(ppinhaoA)
              ppinhaoA
nobs         49.000000
NAs           0.000000
Minimum      11.500000
Maximum      86.330000
1. Quartile  28.970000
3. Quartile  48.660000
Mean         40.754898
Median       41.890000
Sum        1996.990000
SE Mean       2.257321
LCL Mean     36.216249
UCL Mean     45.293547
Variance    249.679476
Stdev        15.801249
Skewness      0.454293
Kurtosis      0.030192

>  summary(ppinhaoB)
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
8.15   17.88   25.43   24.68   31.40   37.03

> shapiro.test(ppinhaoA)
    Shapiro-Wilk normality test

data:  ppinhaoA
W = 0.9776, p-value = 0.4704

> shapiro.test(ppinhaoB)
    Shapiro-Wilk normality test

data:  ppinhaoB
W = 0.9582, p-value = 0.07973

> shapiro.test(ppinhaoA-ppinhaoB)
        Shapiro-Wilk normality test

data:  ppinhaoA - ppinhaoB
W = 0.9749, p-value = 0.3745
```

```
> var.test(ppinhaoA,ppinhaoB)   # Comando B

        F test to compare two variances

data:  ppinhaoA and ppinhaoB
F = 3.9276, num df = 48, denom df = 48, p-value = 5.402e-06
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 2.215437 6.962814
sample estimates:
ratio of variances
          3.927553


> t.test(ppinhaoA,ppinhaoB,paired=TRUE)

        Paired t-test

data:  ppinhaoA and ppinhaoB
t = 6.4205, df = 48, p-value = 5.695e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 11.03766 21.10275
sample estimates:
mean of the differences
                16.0702


> t.test(ppinhaoA,ppinhaoB,alternative="greater")

        Welch Two Sample t-test

data:  ppinhaoA and ppinhaoB
t = 6.3558, df = 70.955, p-value = 8.724e-09
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 11.85632      Inf
sample estimates:
mean of x mean of y
 40.75490  24.68469


> t.test(ppinhaoA,ppinhaoB,paired=TRUE,alternative="greater")

        Paired t-test

data:  ppinhaoA and ppinhaoB
t = 6.4205, df = 48, p-value = 2.848e-08
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 11.87217      Inf
sample estimates:
mean of the differences
                16.0702
```