Instituto Superior de Agronomia
Modelos Matemáticos – 2015/2016

Teste do Módulo I - second date 6 de Abril de 2016

1. After harvesting, pears harvested in the orchard are separated according to their size (in mm). Selecting randomly 1000 from one day harvest, the data obtained are now summarized in the following table:

Size classes (mm)	[30;50]	]50;55]	]55;60]	]60;70]	]70;90]
Number of pears	40	160	390	350	60

Note: Use the results that may be useful, presented in Annex , to answer the following questions.

a) What is the variable under study? What type is it? Justify.

b) Construct a table of relative frequencies for the data observed and sketch the corresponding graph.

c) Calculate the approximate values for mean and median of the size of pears, Comment.

d) calculate an estimate of the proportion of pears with a size over 60 mm. What is the precision associated with this estimate with a confidence of 95%?

e) Can the producer say that on average, the size of pears in this orchard is larger than 60 mm? Answer this question using an appropriate statistical analysis.

2. The Rayleigh distribution has a wide range of applications including life testing experiments and clinical studies. One major application of this model is used in analyzing wind speed data or even environmental situations where "severe"values can occur. Let us consider X a random variable following a model of Rayleigh in a very simple way, with only one unknown parameter $\theta > 0$, with density function defined as:

$$f(x|\theta) = \frac{x}{\theta}\exp[-x^2/2\theta], \quad \text{se} \ \ x > 0, \quad \text{and is zero in remaining values??of } x$$

Consider we have a random sample $(X_1, X_2, .., X_n)$ associated to X.

a) Obtain the estimate of θ by the method of moments.

Note: It is known that $E[X] = \sqrt{\frac{\theta\pi}{2}}$ and $Var[X] = \frac{4-\pi}{2}\theta$.

b) Obtain the maximum likelihood estimate of θ.

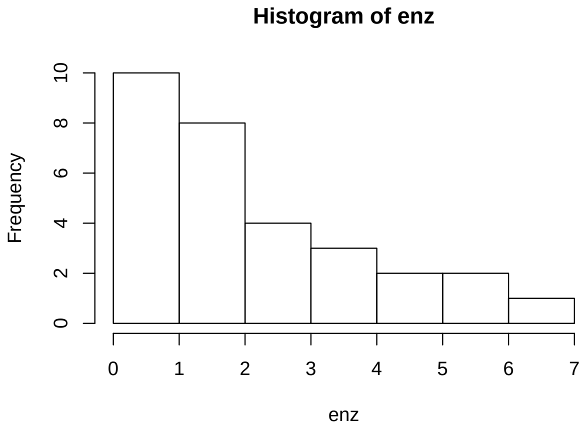
3. In a completely randomized design two different methods of growing corn resulted in various yields per acre on various plots of ground where the two methods were tried. The results obtained are in Appendix.

a) What is the estimated mean and variance of the yield in each method?

b) The collected data allow to say that one of the methods is the best? Justify giving a complete answer.

4. Suppose we are measuring the survival time of a given enzyme in solution and the values observed (in hours) are given in the Appendix.

It is intended to model the probability distribution of the lifetime of this enzyme. Initially we draw the histogram of the observations, which is presented below



a) I will tell you that is suspected to be an exponential model with parameter $\beta > 0$, i.e.

$f(x) = \frac{1}{\beta}\exp(-x/\beta)$ for $x > 0$ and $F(x) = 1 - \exp(-x/\beta)$ for $x > 0$.

Comment this my suspicion.

b) Now let's see if I will have reason!

i) Based on the sample data obtain an estimate of the parameter β.

ii) Consider the data grouped in the following classes

$[0, 1]$ $]1, 2]$ $]2, 3]$ $]3, 5]$ $]5, \infty[$

Now test the hypothesis of the survival time of these enzymes follow an exponential distribution. Justify adequately your answer.

Note: Use the results considered relevant, presented in Appendix.

5. :Let $(X_1, X_2, \ldots, X_n)$ and $(Y_1, Y_2, \ldots, Y_n)$ be two random samples of size $n \geq 2$ taken from independent normal populations, $X \frown \mathcal{N}(\mu_1, \sigma)$ and $Y \frown \mathcal{N}(\mu_2, \sigma)$. Suppose that σ is known.

It is intended to test the hypotheses $H_0 : \mu_1 = \mu_2 \quad vs \quad H_0 : \mu_1 \neq \mu_2$ and is decided to use the following decision rule:

- If the confidence intervals with level α $(0 < \alpha < 1)$ have at least a common value, i.e., if they overlap, the decision is *not reject H_0* ;
- If the intervals do not overlap *H_0 is rejected and H_1 is accepted*.

a) If H_0 is true show that the probability that the two intervals **overlap** is $2\Phi(\sqrt{2}\ z_{\alpha/2}) - 1$.

Note: Note that the two confidence intervals $(1 - \alpha) \times 100\%$ for the mean values of$X$ and Y do not overlap when the difference between the sample means is greater than the semi-sum of the amplitudes of each interval.

b) Let us consider $\alpha = 0.05$ and calculate the probability given in the previous question.

c) Using the decision rule establishe, what should the value of α be in order that the probability of Type I error is 0.05?

ANEXO

```
###----------------
####exercício 1
###---------------


> hist(amostra,breaks=c(30,50,55,60,70,90),plot=F)
$breaks
[1] 30 50 55 60 70 90

$counts
[1]  40 160 390 350  60

$density
[1] 0.002 0.032 0.078 0.035 0.003

$mids
[1] 40.0 52.5 57.5 65.0 80.0

$equidist
[1] FALSE


> t.test(amostra)

        One Sample t-test

data:  amostra
t = 232.2, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 59.87518 60.89582
**********************

> t.test(amostra, alternative="greater",mu=60)

        One Sample t-test

data:  amostra
t = 1.4824, df = 999, p-value = 0.06928
alternative hypothesis: true mean is greater than 60
95 percent confidence interval:
 59.95734      Inf
*****************

> t.test(amostra,alternative="greater")

        One Sample t-test

data:  amostra
t = 232.2, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 0
95 percent confidence interval:
 59.95734      Inf
**************
```

```
###------------
#### Exercício 3
#####-----------


> rend1
 [1] 83 91 94 89 89 96 91 92 90 91 90 81 83 84 83 88 91
> rend2
 [1]  89  84 101 100  91  93  96  95  94  78  82  81  77  79  81  80  81
> shapiro.test(rend1)

        Shapiro-Wilk normality test

data:  rend1
W = 0.9203, p-value = 0.1492

> shapiro.test(rend2)

        Shapiro-Wilk normality test

data:  rend2
W = 0.898, p-value = 0.06283

> shapiro.test(rend1-rend2)

        Shapiro-Wilk normality test

data:  rend1 - rend2
W = 0.9711, p-value = 0.8368

> var.test(rend1,rend2)

        F test to compare two variances

data:  rend1 and rend2
F = 0.2819, num df = 16, denom df = 16, p-value = 0.01561
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1020840 0.7784008
sample estimates:
ratio of variances
         0.2818905

> t.test(rend1,rend2)

        Welch Two Sample t-test

data:  rend1 and rend2
t = 0.6345, df = 24.356, p-value = 0.5317
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.176887  6.000416
sample estimates:
mean of x mean of y
 88.58824  87.17647
```

```
> t.test(rend1,rend2,paired=TRUE)

        Paired t-test

data:  rend1 and rend2
t = 0.8524, df = 16, p-value = 0.4066
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.099274  4.922804
sample estimates:
mean of the differences
               1.411765

> wilcox.test(rend1,rend2)

        Wilcoxon rank sum test with continuity correction

data:  rend1 and rend2
W = 166, p-value = 0.4681
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(rend1, rend2) :
  cannot compute exact p-value with ties
```

```
##---------
## Exercício 4
##---------


> enz<-c(4.75, 3.4, 1.8, 0.9, 2.2, 0.4, 5.8, 1.6, 2.4, 2.25,
+ 1.36, 0.93, 3.97, 6.95, 1.24, 0.80, 0.30, 0.02, 1.04, 0.89,
+ 3.82,1.25,0.04,2.65,4.3,0.52,5.8, 0.12, 1.5, 1.01)

> hist(enz)

> sum(enz)
[1] 64.01

> beta.est<-mean(enz)
> hist(enz,breaks=c(0,1,2,3,5,7),plot=F)
$breaks
[1] 0 1 2 3 5 7

$counts
[1] 10  8  4  5  3



$mids
[1] 0.5 1.5 2.5 4.0 6.0

$xname
[1] "enz"

$equidist
[1] FALSE


> ni<-hist(enz,breaks=c(0,1,2,3,5,7),plot=F)$counts



> p1<-pexp(1,1/beta.est)
> p2<-pexp(2,1/beta.est)-pexp(1,1/beta.est)
> p3<-pexp(3,1/beta.est)-pexp(2,1/beta.est)
> p4<-pexp(5,1/beta.est)-pexp(3,1/beta.est)
> p5<-1-pexp(5,1/beta.est)
> prob<-c(p1,p2,p3,p4,p5)
> prob
[1] 0.37417016 0.23416685 0.14654860 0.14911215 0.09600223
> prob*length(enz)
[1] 11.225105  7.025006  4.396458  4.473365  2.880067
>
```