

Exercícios para treino referentes aos blocos 1 e 2 – resolver e enviar a Prof^a Manuela Neves (manela@isa.ulisboa.pt) até dia 6 de Maio

1. a) Realizou-se uma experiência em que se enxertaram genótipos de laranja doce em três tipos de árvores: limoeiro, tangerineira e laranjeira, a fim de estudar a incidência de cancro cítrico nas folhas de cada um daqueles tipos. Na época adequada seleccionaram-se folhas de cada uma daquelas árvores e classificou-se a incidência em três categorias: ‘não se revelou’, ‘revelou-se moderada’, ‘revelou-se grave’. Pretende-se verificar se a incidência de cancro cítrico difere significativamente consoante o tipo de árvore. Os valores observados encontram-se no quadro abaixo:

Árvore	‘não se revelou’	‘revelou-se moderada’	‘revelou-se grave’
Limoeiro	45	32	12
Tangerineira	26	47	28
Laranjeira	19	20	17

- i) Face ao modo descrito de realização da experiência como classifica esta tabela de contingência e que tipo de teste estatístico se deve realizar?
- ii) Complete o que deve estar nas letras em falta indicadas no *output* por A, B e C
- iii) Responda de forma completa à questão colocada
- b) A distribuição beta tem sido utilizada para modelar a taxa de incidência de cancro cítrico nas folhas. Vamos aqui considerar uma forma simples daquela distribuição, dependente apenas de um parâmetro, $\alpha > 0$, sendo a função densidade assim dada:

$$f(x|\alpha) = \alpha x^{\alpha-1}, \quad 0 < x < 1.$$

Considerando uma amostra aleatória de dimensão n , (X_1, X_2, \dots, X_n) , retirada de uma população X , com aquela função densidade

- i) Obtenha o estimador de α pelo método dos momentos.
- ii) Determine o estimador de máxima verosimilhança para α .
- iii) Observada a seguinte amostra de 10 valores de X
 0.51 0.92 0.74 0.10 0.41 0.54 0.11 0.92 0.90 0.89
 determine estimativas para α .

- c) Considere que se pretende estudar o coeficiente de variação da variável referida em b): $CV = \frac{\sigma_X}{\mu_X} \times 100\%$. Como o estudo inferencial deste parâmetro não é fácil com recurso a procedimentos de Estatística Clássica considerou-se o recurso à metodologia *bootstrap*, que se aplicou à amostra observada em b)iii).

- i) Determine uma estimativa *bootstrap* de CV .
- ii) Obtenha um intervalo *bootstrap* a 95% de confiança para CV .

ANEXO

```

> respostas<-matrix(c(45,32,12,26,47,28,19,20,17),byrow=T,nr=3,
+   dimnames = list(c("Limoeiro", "Tangerineira","Laranjeira"),
+     c("Não", "Moderado", "Grave")))
>
> respostas
      Não Moderado Grave
Limoeiro      45      32      12
Tangerineira  26      47      28
Laranjeira    19      20      17
> chisq.test(respostas)

      Pearson's Chi-squared test
data:  respostas
X-squared = 15.5473, df = A, p-value = 0.003691

> chisq.test(respostas)$expected
      Não Moderado  Grave
Limoeiro    32.56098 35.81707 20.62195
Tangerineira 36.95122 40.64634 23.40244
Laranjeira   20.48780 22.53659 12.97561

> chisq.test(respostas)$residuals
      Não  Moderado  Grave
Limoeiro    2.1799052 -0.6378014 -1.8986315
Tangerineira -1.8015574  0.9965822  B
Laranjeira   -0.3286989 -0.5343249  1.1172136
> pchisq(15.5473,4)
[1] C

> ## alínea b)

> amostra<-c(0.51, 0.92, 0.74,0.10,0.41,0.54, 0.11, 0.92,0.90,0.89)

> sum(amostra)          > sum(log(amostra))
[1] 6.04                [1] -7.380751

> ## alínea c)

> n<-length(amostra);n
[1] 10

> CV<-vector()
> for(i in 1:1000)
+ {   amostra_star<-sample(amostra,n,rep=TRUE)
+     CV[i]<-sd(amostra_star)/mean(amostra_star) }

> sum(CV)
[1] 525.7187

> quantile(CV,prob=c(0.01,0.025,0.05,0.10,0.90,0.95,0.975,0.99))
      1%      2.5%      5%      10%      90%      95%      97.5%      99%
0.2018559 0.2493638 0.2813498 0.3430363 0.7283245 0.7643532 0.8493279 0.8830583

```

2. Considere os seguintes comandos e o output associado:

```
> dna<-c("A","C","G","T")
  > seq2<-sample(dna,1000,replace=T,
+   prob=c(0.20,0.30,0.18,0.32))
```

```
> table(seq2)
seq2
  A   C   G   T
192 289 183 336
```

```
> pbinom(192,1000,0.20)
[1] 0.2783474
```

```
> 1-pbinom(207,1000,0.20)
[1] 0.2749125
```

Explique o que cada linha de comandos executa e diga se é admissível supor que o nucleótido “A” ocorre na proporção definida no estudo.