

Exame final

Estatística Computacional I - Estatística (2011)

(13 de janeiro de 2012)

Prof. Walmes M. Zeviani & Elías T. Krainski - Departamento de Estatística - UFPR

Acadêmico:

Turma:



1. (2 pontos) Faça a importação dos seguintes arquivos de dados para o R. Indique os comandos usados para fazer e dê a descrição de cada um dos arquivos importados: número de registros, número de colunas, tipos de valores em cada coluna.

<http://www.leg.ufpr.br/~walmes/ensino/ce083-2011-02/renda.csv>
<http://www.leg.ufpr.br/~walmes/ensino/ce083-2011-02/esperanca.csv>

```
> # opção 1
> ren <- read.table("http://www.leg.ufpr.br/~walmes/ensino/ce083-2011-02/renda.csv",
+                 header=TRUE, sep=";", stringsAsFactors=FALSE)
> # opção 2
> ren <- read.csv("http://www.leg.ufpr.br/~walmes/ensino/ce083-2011-02/renda.csv",
+               sep=";", stringsAsFactors=FALSE)
> # opção 3
> ren <- read.csv2("http://www.leg.ufpr.br/~walmes/ensino/ce083-2011-02/renda.csv",
+                dec=".", stringsAsFactors=FALSE)
> # estrutura: número de linhas, colunas, tipo
> # poderia usar, ncol(), nrow(), dim(), class(), typeof(), summary()...
> str(ren)
'data.frame':   5462 obs. of  4 variables:
 $ Sigla      : chr  "RJ" "MG" "PR" "CE" ...
 $ Codigo     : int  3302908 3170305 4119251 2304269 3517307 4123105 4314555 2909109 2505402 3132503 ...
 $ Município: chr  "Miguel Pereira" "Umburatiba" "Pinhal de São Bento" "Deputado Irapuan Pinheiro" ...
 $ X2000      : num  379.6 116.4 124.7 82.8 205 ...

> esp <- read.csv2("http://www.leg.ufpr.br/~walmes/ensino/ce083-2011-02/esperanca.csv",
+                dec=".", stringsAsFactors=FALSE)
> str(esp)
'data.frame':   5493 obs. of  4 variables:
 $ Sigla      : chr  "SC" "SC" "RJ" "SP" ...
 $ Codigo     : int  4217402 4209805 3399919 3546207 2603009 1101708 5107958 2304954 4300646 5201108 ...
 $ Município: chr  "Schroeder" "Leoberto Leal" "São João Marcos" "Santa Cruz da Conceição" ...
 $ X2000      : num  75.3 71.4 NA 72.5 69.9 ...
```

2. (4 pontos) O arquivo renda.csv contém dados de renda (R\$) per capita e o arquivo esperanca.csv contém os valores em esperança de vida (anos) ao nascer. Esses valores são para os municípios do Brasil no ano de 2000. Forneça os comandos usados para o que se pede em cada item e mostre os valores obtidos para os Estados da região sul.

- a) a renda média para cada Estado;
- b) a esperança de vida média para cada Estado;
- c) a amplitude (máximo-mínimo) da esperança de vida para cada Estado;
- d) o número de municípios com renda superior à R\$ 200.

```
> # média de renda por Estado
> tapply(ren$X2000, ren$Sigla, mean, na.rm=TRUE)
      AC      AL      AM      AP      BA      CE      DF      ES
108.21300  75.18612  82.85841 133.52094  93.35357  82.72151 605.40600 211.26705
      GO      MA      MG      MS      MT      PA      PB      PE
197.22854  69.79488 178.37354 211.62913 225.30141 116.66302  82.36328 100.86122
      PI      PR      RJ      RN      RO      RR      RS      SC
 73.27570 201.79219 261.33944  93.18647 187.12325 133.73853 245.75532 256.67645
      SE      SP      TO
 92.48186 276.60882 115.68634

> # média de esperança de vida por Estado
> tapply(esp$X2000, esp$Sigla, mean, na.rm=TRUE)
      AC      AL      AM      AP      BA      CE      DF      ES
66.21673 63.07342 64.86367 66.10750 63.18821 66.23153 70.37000 68.17863
      GO      MA      MG      MS      MT      PA      PB      PE
68.92987 60.12108 69.53665 69.38155 68.70219 67.63045 61.24837 65.13874
      PI      PR      RJ      RN      RO      RR      RS      SC
62.00547 69.12926 69.02441 65.59306 66.24437 65.11320 72.53075 73.42305
      SE      SP      TO
62.63834 71.65265 63.93718

> # amplitude de esperança de vida por Estado
> tapply(esp$X2000, esp$Sigla, max, na.rm=TRUE)-tapply(esp$X2000, esp$Sigla, min, na.rm=TRUE)
      AC      AL      AM      AP      BA      CE      DF      ES      GO      MA      MG
5.190 13.017 11.956  9.362 14.271 13.523  0.000 12.306 14.639 14.843 17.632
      MS      MT      PA      PB      PE      PI      PR      RJ      RN      RR
7.995 10.937 12.399 13.164 19.403 13.432 17.022 11.895 15.675  7.016 10.288
      RS      SC
13.687 12.000 13.784 17.189 16.849

> # número de municípios com renda superior à R$ 200 por Estado
> tapply(ren$X2000, ren$Sigla, function(x) sum(x>200, na.rm=TRUE))
      AC      AL      AM      AP      BA      CE      DF      ES      GO      MA      MG
1 1 2 6 1 1 41 97  1 287 40 62 12  3  8  1 174 74  2
      RO      RR      RS      SC      SE      SP      TO
20 1 308 220  1 534  5
```

3. (4 pontos) Faça o que se pede a seguir:

- a) obtenha dois objetos com as informações de renda e esperança de vida apenas para os municípios do estado do Paraná;
- b) faça a junção das informações em um único objeto mantendo apenas os municípios presentes em ambos;
- c) faça o gráfico de dispersão da esperança de vida em função o logaritmo da renda e descreva-o (associação linear/não linear, forte/fraca, positiva/negativa).
- d) classifique os municípios para as seguintes classes de renda: (0-150], (150-200], (200-250], (250-∞); qual o número de municípios em cada classe?

```

> # subconjunto com municípios do Paraná
> renPR <- subset(ren, Sigla=="PR")
> espPR <- subset(esp, Sigla=="PR")
> # juntando os dois data.frames, Código ou Municípios são ID
> erPR <- merge(renPR, espPR, by=c("Sigla","Codigo","Município"))
> str(erPR)
'data.frame':   390 obs. of  5 variables:
 $ Sigla      : chr  "PR" "PR" "PR" "PR" ...
 $ Codigo     : int  4100103 4100202 4100301 4100400 4100459 4100509 4100608 4100707 4100806 4100905 ...
 $ Município: chr  "Abatiá" "Adrianópolis" "Agudos do Sul" "Almirante Tamandaré" ...
 $ X2000.x    : num  149 116 154 198 123 ...
 $ X2000.y    : num  69.6 69.9 67.2 66.1 67.5 ...
> # gráfico de dispersão, ou usar xyplot(), associação linear fraca positiva
> plot(X2000.y~log(X2000.x), erPR)
> # classificando a renda nos intervalos dados
> rc <- cut(erPR$X2000.x, c(0,150,200,250,Inf))
> str(rc)
Factor w/ 4 levels "(0,150]", "(150,200]", "...: 1 1 2 2 1 2 2 3 3 2 ...
> # tabela de frequência das classes de renda
> table(rc)
rc
(0,150] (150,200] (200,250] (250,Inf]
      70       145        101         70

```
