

BOA PROVA!

Considere os dados do arquivo `dadosav2b.csv` disponível em
<http://leg.ufpr.br/~elias/ensino/ce083/dadosav2b.csv>

1. Importe esses dados para o **R**

```
> d <- read.csv2('http://leg.ufpr.br/~elias/ensino/ce083/dadosav2b.csv', fileEncoding='latin1')
```

(a) Quantas variáveis e quantas linhas tem esses dados. Qual a classe de cada variável

```
> str(d)

'data.frame':      150 obs. of  5 variables:
 $ Sexo           : Factor w/ 2 levels "Feminino","Masculino": 1 1 2 2 1 2 2 2 1 ...
 $ Idade          : int  53 46 60 61 44 54 59 72 45 56 ...
 $ Escolaridade   : Factor w/ 3 levels "Básico","Médio",...: 3 2 2 1 1 2 2 3 2 1 ...
 $ IMC            : num  26.2 23.6 29.2 26.7 26.1 28.1 27 28.8 24.4 25.9 ...
 $ PAS           : int  146 128 132 121 122 139 142 118 124 141 ...

> ## 150 linhas de 5 variáveis, três numéricas: uma contínua (IMC)
> ## e duas inteiras (Idade e PAS); e duas categóricas: uma
> ## nominal (Sexo) e uma ordinal (Escolaridade).
```

(b) Para as variáveis categóricas, indique quantas categorias tem cada variável

```
> sapply(d[,c(1,3)], nlevels)

      Sexo  Escolaridade
      2         3
```

(c) Para as variáveis numéricas, indique o maior e o menor valores

```
> sapply(d[,c(2,4:5)], range)

      Idade  IMC  PAS
[1,]    31 20.3 107
[2,]    85 30.2 172
```

2. Indique os comandos em **R** para fazer um resumo descritivo da variável

(a) Escolaridade. Qual a categoria com o maior percentual e qual é esse percentual.

```
> ### Resumo descritivo
> t.esc <- with(d, table(Escolaridade))
> (t.esc <- cbind(Frequencia=t.esc,
+               Percentual=round(100*prop.table(t.esc), digits=2)))

      Frequencia Percentual
Básico          40      26.67
Médio           60      40.00
Superior        50      33.33

> ### Categoria com maior percentual
> rownames(t.esc)[which.max(t.esc[,2])]

[1] "Médio"

> ### Percentual da categoria com maior percentual
> paste(t.esc[which.max(t.esc[,2]),2], "%")

[1] "40 %"
```

(b) Idade. Qual é a idade da pessoa mais nova e a idade média de todas essas pessoas

```
> ### Resumo descritivo
> with(d, c(n=length(Idade), summary(Idade), SD=sd(Idade)))
```

```

      n      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.     SD
150.00000 31.00000 46.00000 55.00000 54.91000 61.00000 85.00000 10.99966
> ### A pessoa mais nova tem
> paste(min(d$Idade), "anos de idade")
[1] "31 anos de idade"
> ### A média de idade é de
> paste(round(mean(d$Idade),2), "anos de idade")
[1] "54.91 anos de idade"

```

3. Faça uma descrição de Escolaridade por Sexo

(a) Mostre uma tabela de frequências.

```
> addmargins(tab <- with(d, table(Sexo, Escolaridade)))
```

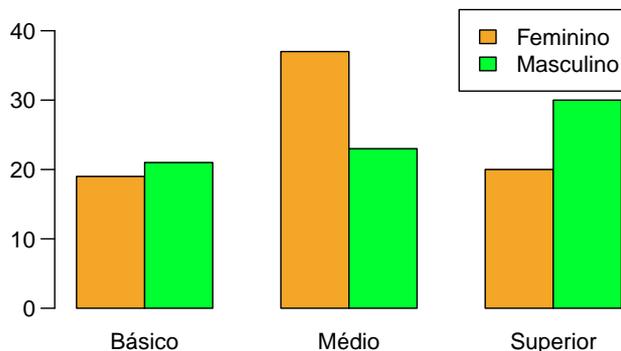
Sexo	Escolaridade			Sum
	Básico	Médio	Superior	
Feminino	19	37	20	76
Masculino	21	23	30	74
Sum	40	60	50	150

(b) Visualize um gráfico descritivo e comente

```

> par(mar=c(3,3,.5,.5), mgp=c(2,0.7,0), las=1)
> barplot(tab, beside=TRUE, legend=TRUE,
+         ylim=c(0,45), col=c('orange','green'))
> ## O número de mulheres e homens com escolaridade básica é próximo
> ## (19 e 21, respectivamente) e é próximo ao número de mulheres com
> ## ensino superior, 20 mulheres. As maiores frequências são de
> ## mulheres com ensino médio e de homens com ensino superior.

```



(c) Calcule e comente sobre o percentual de cada nível de escolaridade em relação a cada sexo.

```

> 100*prop.table(tab, margin=1)
      Escolaridade
Sexo      Básico  Médio Superior
Feminino 25.00000 48.68421 26.31579
Masculino 28.37838 31.08108 40.54054
> ## Um quarto das mulheres tem ensino básico, quase a metade tem ensino
> ## médio e 26,32% tem ensino superior. Entre os homens, 28,38% tem
> ## ensino básico, 31,08% tem ensino médio e 40,54% tem ensino superior.

```

4. Faça uma descrição de PAS por

(a) Sexo. Qual a média de PAS para cada sexo.

```

> ### Descrição de PAS por Sexo
> with(d, tapply(PAS, Sexo, summary))

```

```

$Feminino
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 112.0  127.8   134.0   134.6  141.0   158.0

$Masculino
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 107.0  131.0   137.0   137.2  143.0   172.0

> ## Para as mulheres, PAS varia de 112 a 158, com metade
> ## delas apresentando PAS menor que 134. Para os homens, PAS
> ## varia entre 107 a 172 e metade deles tem PAS menor que 137.
>
> ## calculando o desvio-padrão
> with(d, tapply(PAS, Sexo, sd))

Feminino Masculino
9.665602 11.896848

> ## O desvio padrão de PAS dos homens é maior das mulheres.
>
> ### média de PAS para cada sexo (ja calculadas)
> ### A média de PAS das mulheres é menor que a dos homens.

```

(b) Escolaridade. Visualize um gráfico descritivo e comente

```

> ### Descrição de PAS por Escolaridade
> tapply(d$PAS, d$Escolaridade, summary)

$Básico
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 112.0  128.0   136.0   135.0  140.2   157.0

$Médio
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 114.0  126.8   135.5   136.8  143.2   172.0

$Superior
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 107.0  130.2   135.5   135.3  143.0   158.0

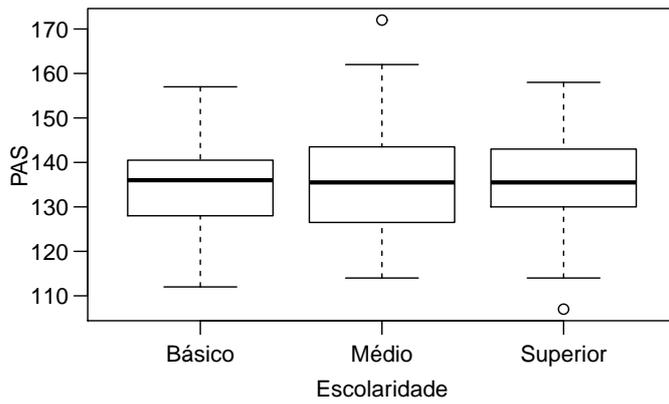
> ## A média e a mediana de PAS são aproximadamente iguais em todos os
> ## níveis de ensino. O menor valor de PAS é de uma pessoa com ensino
> ## superior, enquanto que o maior valor, de uma pessoa com ensino Médio.
> tapply(d$PAS, d$Escolaridade, sd)

Básico      Médio Superior
9.988423 11.496204 10.890194

> ## A variabilidade de PAS é menor para as pessoas com ensino
> ## básico, e a maior para as pessoas com nível médio.

> ### visualizando o gráfico
> par(mar=c(3,3,.5,.5), mgp=c(2,0.7,0), las=1)
> with(d, plot(PAS~Escolaridade))
> ## Os dados de PAS para cada nível de ensino distribuem-se de forma
> ## aproximadamente simétrica. Notamos um valor discrepante (muito grande)
> ## entre as pessoas com ensino médio. Também há um valor discrepante
> ## (muito pequeno) entre as pessoas com ensino superior.

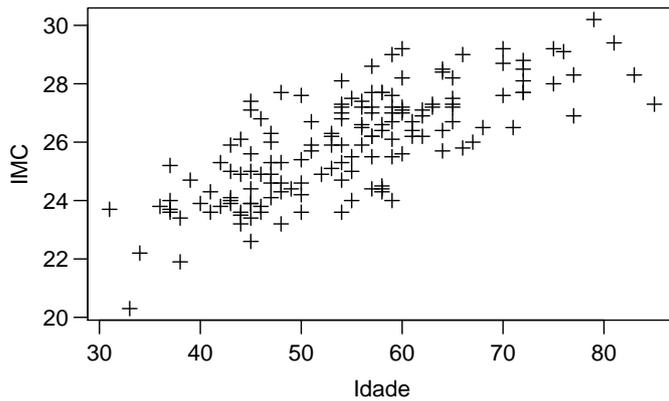
```



5. Faça um gráfico de

(a) IMC por Idade. Comente.

```
> par(mar=c(3,3,.5,.5), mgp=c(2,0.7,0), las=1)
> with(d, plot(Idade, IMC, pch=3))
> ### Notamos que quanto há uma tendência linear de
> ### crescimento de IMC com a Idade, isto é, quanto
> ### maior a idade, maior o IMC. Para pessoas com
> ### idade entre 30 a 40 anos, o IMC fica em torno
> ### de 10 a 25, enquanto que para pessoas com mais
> ### de 70 anos, o IMC fica maior que 26.
```



(b) IMC por PAS. Comente.

```
> par(mar=c(3,3,.5,.5), mgp=c(2,0.7,0), las=1)
> with(d, plot(IMC, PAS, pch=3))
> ### Não há relação clara entre IMC e PAS. A maioria dos
> ### pontos formam uma nuvem próxima do centro do gráfico,
> ### um pouco deslocada para baixo e para a direita. Em
> ### termos numéricos, a maioria das pessoas apresentam
> ### IMC entre 23 a 29 e PAS entre 115 a 155.
```

