

**BOA PROVA!**

Considere os dados dos arquivos em  
<http://leg.ufpr.br/~elias/ensino/ce083/dad3a.csv>  
<http://leg.ufpr.br/~elias/ensino/ce083/dad3b.csv>

1. Importe esses dados para o **R**. Indique o comando em R para juntar ambos num unico data.frame, considerando

```
> ### os dados poderiam ser carregados com  
> d1 <- read.csv2("http://leg.ufpr.br/~elias/ensino/ce083/dad3a.csv")  
> d2 <- read.csv2("http://leg.ufpr.br/~elias/ensino/ce083/dad3b.csv")
```

(a) apenas registros com 'ids' comuns. Quantas linhas resultou?

```
> m1 <- merge(d1, d2)  
> nrow(m1)  
[1] 12
```

(b) apenas os registros com 'ids' disponíveis no primeiro arquivo. Quantas linhas resultou?

```
> m2 <- merge(d1, d2, all.x=TRUE)  
> nrow(m2)  
[1] 17
```

(c) apenas os registros com 'ids' disponíveis no segundo arquivo. Quantas linhas resultou?

```
> m3 <- merge(d1, d2, all.y=TRUE)  
> nrow(m3)  
[1] 37
```

(d) os registros de todos os 'ids'. Quantas linhas resultou?

```
> m4 <- merge(d1, d2, all=TRUE)  
> nrow(m4)  
[1] 42
```

2. Considere que 'conceito' é uma variável ordinal. Indique os comandos para

(a) colocar essa variável como 'ordered factor' considerando a ordem adequada das categorias (níveis)

```
> oconc <- factor(as.character(m4$conceito), ordered=TRUE,  
+               levels=c("Regular", "Bom", "Excelente"))
```

(b) visualizar um gráfico com legenda

```
> t.oc <- table(oconc)  
> barplot(t.oc, legend=TRUE, col=2:4,  
+        args.legend=list(x="topleft", bty="n"))
```

(c) fazer uma tabela de frequências absolutas e relativas

```
> tab.oc <- data.frame(t.oc)  
> tab.oc$FreqRelativa <- 100*prop.table(t.oc)  
> colnames(tab.oc)[1:2] <- c("Conceito", "FreqAbsoluta")
```

3. Indique o comando para fazer

(a) um sumário descritivo da variável 'x1' para cada nível de 'conceito'

```
> with(m4, tapply(x1, oconc, summary))
```

(b) Indique o comando para visualizar um histograma da variável 'x2' para cada nível de 'conceito'

```
> require(lattice)  
> histogram(~x2|oconc, data=m4)
```

4. Indique o comando para fazer um gráfico considerando as variáveis 'x1', 'x2' e 'conceito' ao mesmo tempo

```
> xyplot(x2~x1|oconc, data=m4)
```

5. Indique o comando para

(a) selecionar um sub-conjunto dos dados apenas com os dados quando conceito='Bom'

```
> sbom <- subset(m4, conceito=="Bom")
```

(b) selecionar um sub-conjunto dos dados apenas com os dados quando x1>70

```
> sx1g <- subset(m4, x1>70)
```

(c) selecionar um sub-conjunto dos dados apenas com os 50% dos dados com os maiores valores 'x2'

```
> sx2m <- subset(m4, x2>median(x2, na.rm=TRUE))
```

6. Considere a variável 'data'. Indique o comando para

(a) converte-la para um objeto de classe adequada para representar datas

```
> dats <- as.POSIXlt(as.character(m4$data), format="%d/%m/%Y")
```

(b) visualizar um gráfico de 'data' e 'x1'

```
> with(m4, plot(dats, x1))
```

(c) visualizar um gráfico de 'data' e 'x1', indicando a escala do eixo de 'data' em semanas (**dica:** par(las=2))

```
> par(las=2)
```

```
> with(m4, plot(dats, x1, axes=FALSE, xlab="Data"))
```

```
> box(); axis(2)
```

```
> w <- seq(min(dats, na.rm=TRUE), max(dats, na.rm=TRUE), by="week")
```

```
> axis.POSIXct(1, at=w, format="%d/%m")
```

7. Indique um comando para indicar

(a) as linhas com 'ids' iniciando com 'A'

```
> grep("^A", m4$ids)
```

(b) os 'ids' que iniciam com 'B'

```
> grep("^B", m4$ids, value=TRUE)
```

(c) calcular a média de 'x1' quando 'ids' não inicia com 'C'

```
> mean(m4$x1[!grep("^C", m4$ids)], na.rm=TRUE)
```

(d) as linhas cujo segundo caracter de 'ids' é igual a 1

```
> grep("^[1]", m4$ids)
```

```
> ### ou
```

```
> which(substr(m4$ids,2,2)=='1')
```

(e) as linhas com 'ids' possuindo mais que três caracteres

```
> grep("^...$", m4$ids, invert=TRUE)
```

```
> ### ou
```

```
> which(nchar(as.character(m4$ids))>3)
```