

Laboratório de Bioestatística - Estatística Descritiva

EXERCÍCIOS LIVRES

Estatística Descritiva

1. Carregue a biblioteca **MASS** e faça a leitura do conjunto de dados **birthwt**.

- (a) Use o comando `?birthwt` para aceder a informação sobre o conjunto de dados.
- (b) Explore o seguinte comando

```
bwt <- with(birthwt, {  
  race <- factor(race, labels = c("white", "black", "other"))  
  ptd <- factor(ptl > 0)  
  ftv <- factor(ftv)  
  levels(ftv)[- (1:2)] <- "2+"  
  data.frame(low = factor(low), age, lwt, race,  
    smoke = (smoke > 0), ptd, ht = (ht > 0),  
    ui = (ui > 0), ftv)  
})
```

- (c) Guarde os dados originais num ficheiro com extensão **csv**.
- (d) Faça a leitura dos dados do ficheiro que acabou de gravar para um objecto de tipo **data frame**.
- (e) Use a função **summary** para obter um sumário dos dados. Complemente esta informação com o que obtém de **str()**. A informação está conforme a natureza dos dados? Se não está, proceda às alterações necessárias.
- (f) Obtenha novo sumário dos dados.
- (g) Use a função **tapply** para obter média e desvio padrão da idade das mães segundo a variável **smoke**.
- (h) Obtenha uma tabela cruzando os dados relativos às variáveis **race** e **low**.

- (i) Represente graficamente a variável relativa ao número de partos prematuros.
- (j) Represente graficamente a variável `race`.
- (k) Explore a função `mosaicplot`. Utilize-a para representar graficamente o cruzamento de dados que fez acima.
- (l) Represente o peso das crianças através de um histograma. Explore a função `hist`. Produza histogramas com diferentes números de classes e com classes definidas por si.
- (m) Obtenha um diagrama de caule-e-folhas da idade das mães.
- (n) Represente num diagrama boxplot o peso das crianças. Obtenha boxplots paralelos para esta variável segundo características das mães.
- (o) Construa uma variável que separe as mães em dois grupos: idade inferior a 25 anos e idade superior a 25 anos.
- (p) (* Exercício avançado. Recomendável para os alunos que frequentam Ensaios Clínicos) Considere que, para um ensaio controlado e aleatorizado para comparar dois tratamentos (A e B) associados à fertilidade, se vai recrutar 50 mulheres. As variáveis idade (categorizada), hábito tabágico e hipertensão são consideradas potenciais confundidoras.
 - i. Comece por seleccionar aleatoriamente 50 mulheres da base de dados de forma a simular um conjunto de 50 recrutadas.
 - ii. Construa uma lista de aleatorização simples e atribua as mulheres aos grupos de tratamento. Obteve grupos equilibrados no que respeita à sua dimensão? Verifique o equilíbrio dos grupos relativamente às variáveis que são potenciais confundidoras.
 - iii. Repita a alínea anterior considerando aleatorização por blocos.
 - iv. Refaça o exercício considerando aleatorização estratificada.