

Exercícios 01

Gustavo H. Carvalho

2016-09-18

Dados para os exercícios

Neste exercício, usaremos um conjunto de dados com medidas de tamanho corporal de pardais. Os dados estão no objeto `bumpus` do pacote `semeco`:

```
library(semeco)
data(pugesek)
str(bumpus)
```

```
## 'data.frame':    136 obs. of  13 variables:
## $ id            : int   1 1 2 2 3 3 4 4 5 5 ...
## $ sex           : Factor w/ 2 levels "f","m": 2 2 2 2 2 2 2 2 2 2 ...
## $ age           : Factor w/ 3 levels "", "a", "y": 2 2 2 2 2 2 2 2 2 2 ...
## $ survival      : logi  TRUE FALSE FALSE TRUE TRUE FALSE ...
## $ total.length  : int   154 165 160 160 155 161 154 162 156 163 ...
## $ alar.extent   : int   241 240 245 252 243 249 245 246 247 250 ...
## $ weight        : num   24.5 26.5 26.1 26.9 26.9 25.6 24.3 25.9 24.1 25.5 ...
## $ length.beak.head : num   31.2 31 32 30.8 30.6 32.3 31.7 32.3 31.5 32.5 ...
## $ length.humerus : num   0.687 0.738 0.736 0.736 0.733 0.743 0.741 0.738 0.715 0.752 ...
## $ length.femur   : num   0.668 0.704 0.709 0.709 0.704 0.718 0.688 0.709 0.706 0.731 ...
## $ length.tibiotarsus: num   1.02 1.09 1.11 1.18 1.15 ...
## $ width.skull     : num   0.587 0.606 0.611 0.602 0.602 0.602 0.584 0.607 0.575 0.623 ...
## $ length.sternum  : num   0.83 0.847 0.842 0.841 0.846 0.828 0.839 0.869 0.821 0.888 ...
```

Nos dados, cada linha corresponde à um indivíduo. As medidas de comprimento estão em mm e o peso está em gramas. Crie um novo conjunto de dados chamado `bumpus_females` apenas com as fêmeas (utilize a coluna `sex`).

```
str(bumpus_females)
```

```
## 'data.frame':    49 obs. of  13 variables:
## $ id            : int   37 38 39 40 41 42 43 44 45 46 ...
## $ sex           : Factor w/ 2 levels "f","m": 1 1 1 1 1 1 1 1 1 1 ...
## $ age           : Factor w/ 3 levels "", "a", "y": 1 1 1 1 1 1 1 1 1 1 ...
## $ survival      : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
## $ total.length  : int   155 156 160 152 160 155 157 165 153 162 ...
## $ alar.extent   : int   240 240 242 232 250 237 245 245 231 239 ...
## $ weight        : num   26.3 25.8 26 23.2 26.5 24.2 26.9 27.7 23.9 26.1 ...
## $ length.beak.head : num   31.4 31.5 32.6 30.3 31.7 31 32.2 33.1 30.1 30.3 ...
## $ length.humerus : num   0.709 0.715 0.74 0.676 0.741 0.727 0.766 0.78 0.68 0.709 ...
```

```
## $ length.femur      : num  0.71 0.678 0.732 0.683 0.731 0.723 0.751 0.757 0.662 0.685 ...
## $ length.tibiotarsus: num  1.12 1.13 1.16 1.05 1.19 ...
## $ width.skull       : num  0.614 0.597 0.597 0.59 0.615 0.61 0.62 0.633 0.592 0.587 ...
## $ length.sternum    : num  0.815 0.812 0.854 0.78 0.886 0.787 0.841 0.895 0.781 0.911 ...
```

Selecione apenas as colunas com as variáveis relacionadas às medidas corporais:

```
bumpus_females <- bumpus_females[, -c(1:4)]
str(bumpus_females)
```

```
## 'data.frame': 49 obs. of 9 variables:
## $ total.length      : int  155 156 160 152 160 155 157 165 153 162 ...
## $ alar.extent       : int  240 240 242 232 250 237 245 245 231 239 ...
## $ weight            : num  26.3 25.8 26 23.2 26.5 24.2 26.9 27.7 23.9 26.1 ...
## $ length.beak.head  : num  31.4 31.5 32.6 30.3 31.7 31 32.2 33.1 30.1 30.3 ...
## $ length.humerus    : num  0.709 0.715 0.74 0.676 0.741 0.727 0.766 0.78 0.68 0.709 ...
## $ length.femur      : num  0.71 0.678 0.732 0.683 0.731 0.723 0.751 0.757 0.662 0.685 ...
## $ length.tibiotarsus: num  1.12 1.13 1.16 1.05 1.19 ...
## $ width.skull       : num  0.614 0.597 0.597 0.59 0.615 0.61 0.62 0.633 0.592 0.587 ...
## $ length.sternum    : num  0.815 0.812 0.854 0.78 0.886 0.787 0.841 0.895 0.781 0.911 ...
```

Exercícios

1. Crie uma matriz de covariâncias entre as variáveis e chame-a de `bumpus_cov`.
2. Crie uma matriz de correlações entre as variáveis e chame-a de `bumpus_cor`.
3. Crie diagramas de dispersão par à par com os dados e compare-os com as matrizes de covariâncias e correlações:

```
# pch = "." plota pontos (.) com um fator de expansão (cex) igual a 3.
pairs(bumpus_females, pch = ".", cex = 3)
```

4. Quais são as médias e os desvios-padrão das variáveis? Crie um objeto chamado `bumpus_female_std` com as variáveis padronizadas. Recalcule as médias e os desvios-padrão, agora com os dados padronizados.
5. Crie uma matriz de covariâncias entre as variáveis padronizadas. Algo mudou? E as correlações, foram alteradas?
6. O peso (`weight`) dos pardais está em gramas. Transforme-os para kilogramas e recalcule as covariâncias e correlações, prestando atenção aquelas que envolvem a variável `weight`. Algo mudou? Por quê?
7. Calcule o coeficiente de regressão não-padronizado e padronizado entre `weight` e `total.length` (`weight ~ total.length`). Analise as matrizes de covariâncias e correlações e entenda como os coeficientes foram gerados.