
Folha 3: Sumarização de Dados

1. Carregue o conjunto de dados `Aids2` da package `MASS`. Responda às seguintes questões.

- (a) Qual a categoria de transmissão do vírus mais frequente nestes dados?
E a categoria menos frequente?
(Sugestão: use as funções `table()`, `names()`, `which.max()` e `which.min()`).
- (b) Obtenha a tabela de contingência em função do sexo e categoria de transmissão do vírus.
- (c) Obtenha um sumário das propriedades estatísticas básicas da idade em função da categoria de transmissão do vírus.
- (d) Obtenha uma tabela com a categoria de transmissão do vírus ordenada por ordem decrescente da idade média.
- (e) Obtenha o número de indivíduos com idades entre 0-25, 25-50, 50-75 e 75-100. Verifique como eles se distribuem pelas diferentes categorias de transmissão do vírus.
- (f) Assuma que uma frequência inferior ou igual 1.5% de uma categoria numa variável discreta, representa um *outlier* nessa variável. Identifique as categorias de transmissão do vírus que podem ser consideradas *outliers*.
- (g) Liste o conjunto de exemplos cuja categoria de transmissão está nas condições anteriores. (Sugestão: use o operador `%in%`).
- (h) Inspeccione se existem *outliers* extremos, isto é, valores anormalmente altos e/ou baixos em cada uma das variáveis contínuas.
- (i) Faça uma função que, dado um qualquer data frame como argumento, imprime para cada atributo o seu nome e o conjunto de exemplos com valores considerados *outliers* nesse atributo.