

Tabulação, importação e exportação de dados

Recomendações gerais na tabulação, leitura e escrita de dados usando R

Prof. Me. Lineu Alberto Cavazani de Freitas

Métodos Estatísticos em Pesquisa Científica
Apoio computacional em linguagem R

Departamento de Estatística
Laboratório de Estatística e Geoinformação



Importação e exportação

- ▶ Usando o R é possível **importar** e **exportar** dados em diferentes formatos.
- ▶ Os formatos comuns são **txt**, **csv** e **xls/xlsx**.
- ▶ É possível trabalhar com outros formatos e até mesmo bancos de dados relacionais.
- ▶ Mas para que a leitura seja bem sucedida, uma etapa anterior merece atenção: a **tabulação dos dados**.



Tabulação

Tabulação

- ▶ Em Estatística e análise de dados de modo geral trabalhamos com **dados estruturados** em um **formato tabular**.
 - ▶ Cada **linha** representa uma **observação**.
 - ▶ Cada **coluna** uma **variável**.
 - ▶ Cada **célula** o valor da variável para aquela **observação**.
- ▶ A ideia é simples, mas na prática... Vemos muitas inconsistências.
- ▶ Uma tabulação adequada economiza tratamentos posteriores para realizar a leitura dos dados e, conseqüentemente, adianta a análise.

Inconsistências comuns

1. Níveis de variáveis qualitativas tabulados de forma inconsistente.
2. Variáveis quantitativas representadas com caracteres.
3. Inconsistência nas escalas.
4. Inconsistência nos decimais.
5. Valores ausentes tabulados de forma equivocada.
6. Planilhas formatadas (células mescladas, subtabelas coloridas).
7. Variáveis de interesse em múltiplas abas de uma planilha.
8. Codificações que só fazem sentido para o pesquisador.
9. A mesma variável tabulada em diferentes colunas (comum em estudos longitudinais).

Dicas e boas práticas

- ▶ Na tabulação evite nomes de colunas longos ou caracteres especiais.
- ▶ Seja consistente na tabulação dos níveis de variáveis qualitativas (maiúsculas e minúsculas, espaços, abreviações).
- ▶ Se a variável é quantitativa, não coloque texto nas células.
- ▶ Defina qual será o representador de decimal e use-o em toda a tabulação.
- ▶ Se o valor for ausente, coloque NA na célula ou deixe-a em branco.
- ▶ Não use células mescladas se seu interesse é importar os dados.
- ▶ Gere uma tabela “final” com as variáveis que você tem interesse em analisar. Evite várias planilhas ou abas com diferentes variáveis que devem ser analisadas juntas.
- ▶ Evite códigos e procedimentos que só fazem sentido pra você. Você está habituado com os dados, quem vai analisar não.



Importação

Importação

- ▶ Chamamos de **importação** ou **leitura** o procedimento de entrar com um conjunto de dados externo no R.
- ▶ Podemos ter dados em diferentes formatos, por exemplo:
 - ▶ Arquivos simples de texto com extensão **txt**.
 - ▶ Arquivos separados por “,” ou “;” no formato **csv**.
 - ▶ Planilhas do Excel em formato **xlsx** e **xls**.
 - ▶ Arquivos de dados do R como RData, rda, rds.
 - ▶ Arquivos de outros softwares como SPSS, Stata e SAS.
 - ▶ Bancos de dados relacionais.

Importação

- ▶ Vamos explorar os comandos para leitura dos tipos mais comuns: arquivos txt, csv e xls/xlsx.
 - ▶ Para arquivos txt: `read.table()`.
 - ▶ Para arquivos csv: `read.csv()` e `read.csv2()`.
 - ▶ Para planilhas do Excel (arquivos xls e xlsx) precisamos do auxílio de bibliotecas para fazer a importação, como a `openxlsx` que possui a função `read.xlsx()`.
- ▶ Existem formas alternativas para entrar com conjuntos de dados de forma manual no R, como as funções `scan()` e `edit()`.



Exportação

Exportação

- ▶ Chamamos de **exportação** o procedimento de enviar um conjunto de dados para fora do R como um arquivo.
- ▶ Considerando os já mencionados tipos (txt, csv e xls), podemos usar as funções:
 - ▶ `write.table()` para exportar arquivos txt e csv.
 - ▶ `write.csv()` para arquivos csv.
 - ▶ `write.xlsx()` do pacote `openxlsx`.

O que foi visto

Tabulação de dados

- ▶ Inconsistências comuns.
- ▶ Boas práticas.

Importação

- ▶ Entrada de dados no R com as funções `read.table()`, `read.csv()`, `read.xlsx()`, `scan()` e `edit(data.frame())`.

Exportação

- ▶ Saída de dados do R com as funções `write.table()`, `write.csv()` e `write.xlsx()`