

Problem Set 1

3 de outubro de 2023

Problema 1

Avalie a expressão seguinte em R.

$$\frac{\log_{10}(67000) * 92 + 36}{\sqrt{759}}$$

Arredonde a resposta a uma casa decimal.

Problema 2

Utilize o código seguinte para calcular sua resposta.

```
set.seed(42) # so everyone produces the same answer  
x <- round(rnorm(20, mean = 100, sd = 10), 2) # returns vector of 20 random numbers
```

- O que é o segundo elemento de `x`?
- O que é a “class” e o “type” de `x`?
- O que é o valor máximo de `x`? (Utilize uma função de R!)

Problema 3

Utilize o código seguinte para calcular sua resposta.

Primeiro, faça o download de “einstein_soro.rds” para sua pasta de R.

```
soro <- readRDS("einstein_soro.rds")
```

- Esta conjunto de dados tem quantos casos?
- Esta conjunto de dados tem quantas variáveis?
- O que é o “class” e o “type” de `soro`?
- O que é a data de coleção (`dt_collect`) do terceiro caso?
- Imprima os quinta até décimo anos de nascimento (`birth_yr`).
- Quantas cidades diferentes existem nesta amostra?

Problema 4

O arquivo `trplasma.csv` mostra todas as mutações que uma amostra dos pacientes com HIV em falha virológica tiveram no gene transcriptase reversa do vírus. Um “1” na célula da planilha indica que a mutação estava presente e “0” demonstra ausência da mutação naquele momento.

A primeira coluna é o número de código do paciente e todas as outras colunas representam as mutações. Os nomes das colunas podem ser interpretados como:

- “tr” para *transcriptase reversa*,
- o número seguinte como a posição do aminoácido (codon) com a mutação
- o código para semana de exame de sangue (“bl” = *baseline* ou “_12” = 12 semanas).

- a. Primeira tarefa: importar o arquivo usando `readr::read_csv()` para um tibble chamada `trplas`.
- b. Segunda tarefa: Transforme os dados para um conjunto *tidy*, usando as funções de `tidyr`.

Dica: planeje a transformação. Quais são as variáveis? Como vai dividir o código da mutação? Desenhe no papel como vai aparecer o tibble.

Problema 5

Carregue na memória o arquivo “`pac_demo.xlsx`” como `pac_data`, e responda à pergunta seguinte.

- a. Primeira tarefa: Quando falamos de carga viral de HIV, normalmente falamos de uma transformação das cópias do vírus em log10. Usando a função `log10()`, crie uma nova variável da valor logaritmico de carga viral, `log_cv` e salvar ele de volta para `pac_data`.
- b. Segunda tarefa: Com a variável de escolaridade (`escol`), mude a valor para ser um factor com 4 níveis e mostra uma tabela com os níveis na ordem correta.

Quando você consegue estas respostas (não se ...), você estará no caminho para o sucesso no curso MAD!