

Prova de Certificação - Curso de Introdução ao R

Laboratório de Análise de Dados e Economia da Educação - educLAB

Fevereiro de 2021

- 1 - Como podemos definir o diretório? Escreva o código e explique sua funcionalidade.
- 2 - Construa uma matriz que contenha 20 elementos e tenha 4 linha e 5 colunas.
- 3 - Qual a diferença de um data frame para uma matriz?
- 4 - Sabendo que a inicial "S" em atividade extra significa que o aluno entregou os exercícios e que eles valem 0.5 pontos na média e que "N" significa que o aluno não entregou os exercícios, logo não possui pontos a mais na média, construa uma nova coluna para o resultado final a qual será a soma de "RESULT_PROVA" e "ATIV_EXTRA".

Considere a base de dados a seguir.

```
dados <- data.frame(ID = c(1:10),  
NOME = c("A", "B", "C", "D", "E", "F", "G", "H", "I", "J"),  
RESULT_PROVA = c(8.05, 6.65, 7.77, 6, 8.95, 6.96, 5.36, 7.95, 8.22, 5),  
ATIV_EXTRA = c("N", "S", "S", "N", "N", "N", "S", "N", "S", "S"))
```

dados

##	ID	NOME	RESULT_PROVA	ATIV_EXTRA
## 1	1	A	8.05	N
## 2	2	B	6.65	S
## 3	3	C	7.77	S
## 4	4	D	6.00	N
## 5	5	E	8.95	N
## 6	6	F	6.96	N
## 7	7	G	5.36	S
## 8	8	H	7.95	N
## 9	9	I	8.22	S
## 10	10	J	5.00	S

- 5 - Utilizando a função `tibble()`, crie uma tabela com no mínimo 30 linhas e 5 variáveis utilizando informações verdadeiras (pesquise na internet).
- 6 - Mostre explique os argumentos não padrões das funções `gather()` e `separate()`.
- 7 - Utilizando a base `mtcars` crie um gráfico de dispersão que identifique o deslocamento (`disp`) por galão gasto por milha (`mpg`). A partir do resultado anterior faça uma clusterização pela quantidade de cilindros (`cyl`).
- 8 - Use a base do PIB Municipal do Ipeadata para plotar um mapa com o PIB dos municípios de maneira que se possa identificar regiões com alta concentração de maiores resultados.
- 9 - Com a mesma base anterior, encontre o PIB acumulado por estado e plot um gráfico com esses resultados.

10 - Usando a função `readline` (<https://www.datamentor.io/r-programming/examples/user-input/>) do R crie uma função que peça ao usuário que digite “par” a “impar” e que digite um número inteiro entre 0 a 9. Salve a primeira como “`aposta_do_desafiante`” e a segunda como “`jogada_do_desafiante`”.

11 - Adicione a esta função a geração pelo R de um número aleatório inteiro 0 ou 1. Se o R gerar 0 crie uma variável com o nome “par”. Se o R gerar 1 modifique tal variável com o nome “impar”. Chame essa variável de “`aposta_do_R`”.

12 - Crie outra variável chamada de “`jogada_do_R`” que também gere um número aleatório inteiro entre 0 e 9.

13 - Suponha que o tempo necessario para atendimento de clientes em uma central de atendimento telefonico siga uma distribuição normal de média de 8 minutos e desvio padrão de 2 minutos.

- Qual e a probabilidade de que um atendimento dure menos de 5 minutos?
- E mais do que 9.5 minutos?
- E entre 7 e 10 minutos?
- 75% das chamadas telefonicas requerem pelo menos quanto tempo de atendimento?

14 - Seja uma distribuição normal $X \sim N(3; 16)$

- Calcule $P(2 < X < 5)$
- $P(X > 4)$

Crie esses dois conjuntos de dados para resolver a questão 15:

```
x <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
```

```
y <- c(2, 5, 8, 5, 3, 9, 10, 12, 9, 10)
```

15 - Onde $(x \sim y)$ (*variável resposta* \sim *variável preditora*). Após esse processo, Ajuste um modelo de regressao linear aos dados, estime dos parâmetros, **Erro** e **R2** do modelo e por fim expresse graficamente a função que ajusta a reta do modelo aos dados.