

Exercícios Módulo 2

Laboratório de Análise de Dados e Economia da Educação - educLAB

Fevereiro de 2021

Sobre o Dplyr

Considere a base “dados” e “dados_” para resolver as questões de 1 a 6.

```
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

1 - 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”.

2 - Como já sabemos o resultado final dos alunos, agora precisamos saber a média e o desvio padrão da sala.

3 - No computo final das notas terá sido aprovado o aluno que obtiver em resultado final maior ou igual a 7 pontos. A partir da base construída no item 1, mostre apenas os alunos aprovados.

4 - Considerando a base produzida no item 1, ordene de forma decrescente os alunos por resultado final. Agora, você conseguiu acesso a uma base complementar chamada “dados_”. Esta base possui informações adicionais sobre os alunos dessa turma referentes a identificação do aluno na escola, idade, sexo, se o aluno faz aulas de reforço, e qual o turno que estuda.

```
dados_ <- data.frame(ID = c(1:5, 8:13),  
IDADE = c(17, 17, 19, 18, 17, 18, 19, 18, 17, 17, 18),  
SEXO = c("F", "F", "M", "M", "F", "M", "M", "F", "M", "F", "F"),  
REFORCO = c("S", "N", "N", "N", "S", "S", "N", "S", "N", "N", "N"))
```

```
TURNO = c("MANHA")
```

```
dados_
```

```
##      ID IDADE SEXO REFORCO TURNO
## 1     1     17   F         S MANHA
## 2     2     17   F         N MANHA
## 3     3     19   M         N MANHA
## 4     4     18   M         N MANHA
## 5     5     17   F         S MANHA
## 6     8     18   M         S MANHA
## 7     9     19   M         N MANHA
## 8    10     18   F         S MANHA
## 9    11     17   M         N MANHA
## 10   12     17   F         N MANHA
## 11   13     18   F         N MANHA
```

5 - A escola solicitou uma nova planilha que contenha essas informações adicionais, ou seja, ela quer as demais informações sobre todos os alunos que foram apresentados na base “dados”. Construa essa nova base.

6 - Na construção de sua nova planilha você percebeu que alguns alunos não possuíam todas as informações. Apresente em uma planilha os alunos que não possuíam as informações complementares da base “dados_” e em outra planilha apresente os alunos que possuíam informações na base complementar, mas não possuíam correspondentes na base “dados”.

Sobre a Tibble

7 - Quais as principais vantagens em utilizar tibbles ao invés de data frames?

8 - Utilizando a função `tibble()`, crie uma tabela com no mínimo 30 linhas e 5 variáveis utilizando informações verdadeiras (pesquise na internet).

9 - Refaça o exercício anterior utilizando a função `tribble()`.

10 - Mostre na tela do console os valores das variáveis um por um, de maneiras diferentes, determinando o número de linhas que deseja observar.

Sobre a Tidy

11 - Porque é necessária uma boa organização dos dados?

12 - Mostre explique os argumentos não padrões das funções `gather()` e `separate()` (Utilize o `?help`).

13 - Use as funções `gather()` e `separate()` para organizar os dados que você criou utilizando o pacote `tibble`.

14 - Converta o formato do objeto anterior para data frame e calcule as médias das variáveis numéricas.