

- 1) Dado o arquivo txt de uma pesquisa em escola americana de ensino médio sobre o desempenho do exame nacional. É um data.frame com 11 variáveis a saber:

Id # índice amostrado  
 genero # female ou male  
 raca # raça entre White, african american, hispanic, asian  
 socioeco # nível sócio econômico: low, Middle, high  
 oriescol # origem da escolaridade: public, private  
 niveispro # níveis do programa: academic, vocational, general  
 literatura # pontuação em literatura  
 redação # pontuação em redação  
 matematica # pontuação em matemática  
 ciencia # pontuação em ciência  
 sociologia # pontuação em sociologia

O **data set** tem cerca de 200 observações e 11 itens.

Adapte e Execute o extract da ETL a seguir:

```
desempenho <- read.delim("C:/Users/Marise Miranda/Downloads/desempenho.txt")
Print o View(desempenho)
```

- 1) Qual a média e 1º quartil do seu conjunto de dados para a coluna matematica.

Média 52.65

1ºquartil 45.00

```
> summary(desempenho$matematica)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 33.00  45.00   52.00   52.65  59.00   75.00
```

- 2) Determine a quantidade de alunos por genero

```
> table(desempenho$genero)
```

```
female  male
  109    91
```

female male

109 9

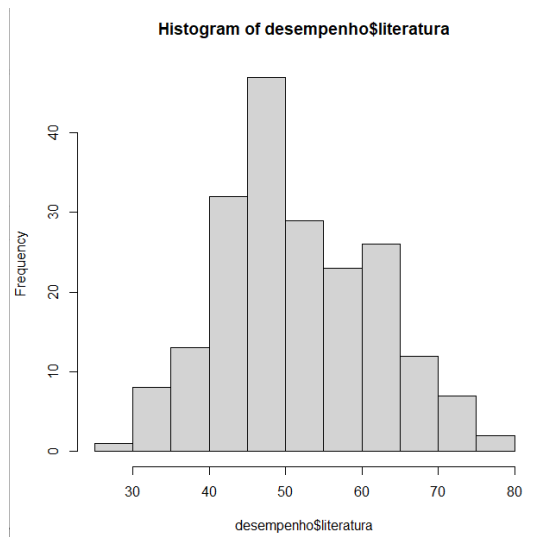
- 3) Qual é a faixa que está aprox 65% menor da maior faixa da coluna literatura

- a) 50 - 55
- b) 40 - 45
- c) 50 - 55
- d) 60 - 65
- e) 45 - 50

Resposta

```
hist(desempenho$literatura)
```

[https://datascience.csuchico.edu/event/r\\_import/](https://datascience.csuchico.edu/event/r_import/)



```
table(desempenho$literatura)
```

```
28 31 34 35 36 37 39 41 42 43 44 45 46 47 48 50 52 53 54
```

```
1 1 6 1 3 2 8 2 13 2 13 2 1 27 1 18 14 1 1
```

```
55 57 60 61 63 65 66 68 71 73 76
```

```
13 14 9 1 16 9 1 11 2 5 2
```

```
> cbind(table(desempenho$literatura))
```

```
 [,1]
```

```
28 1
```

```
31 1
```

```
34 6
```

```
35 1
```

```
36 3
```

```
37 2
```

```
39 8
```

```
41 2
```

```
42 13
```

```
43 2
```

```
44 13
```

```
45 2
```

```
46 1
```

```
47 27
```

```
48 1
```

```
50 18
```

```
52 14
```

```
53 1
```

```
54 1
```

```
55 13
```

```
57 14
```

```
60 9
```

```
61 1
```

```
63 16
```

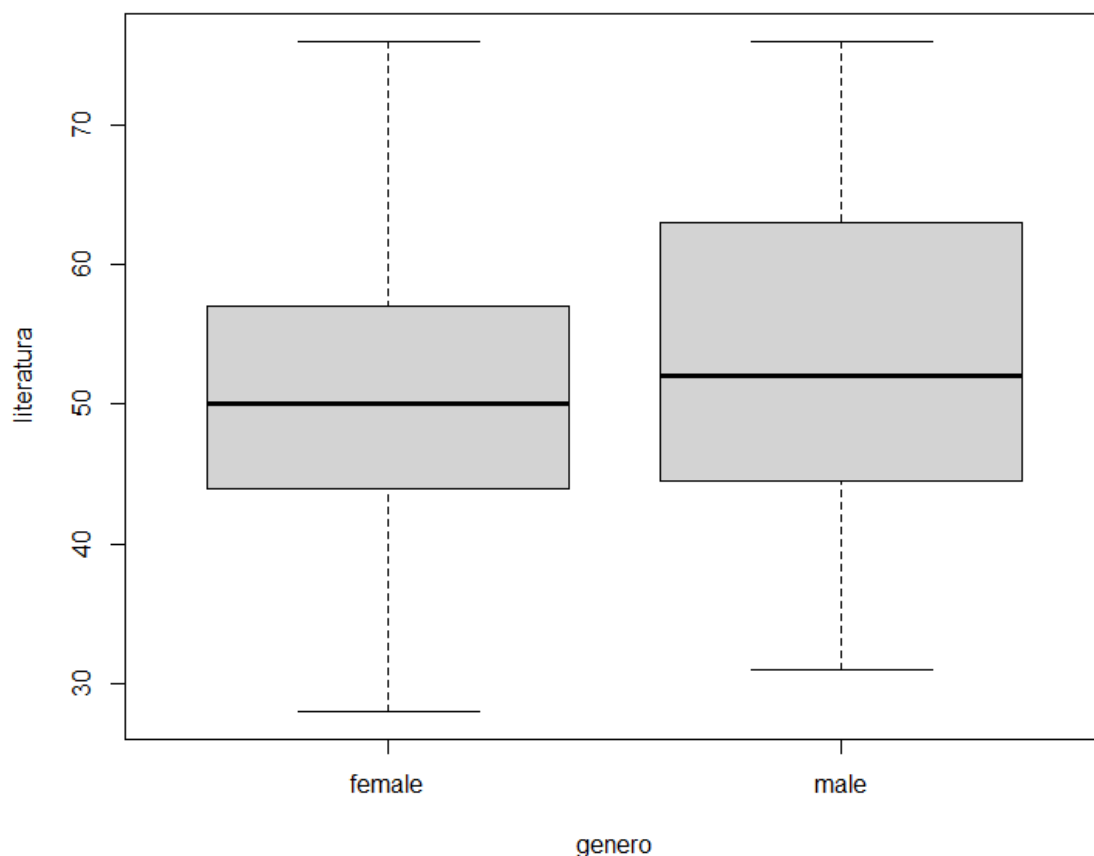
```
65 9
```

```
66 1
```

```
68 11
```

71 2  
73 5  
76 2

4) Com base no data set desempenho, obtivemos este resultado



Descreva a análise dos resultados.

5) A simbolização de uma conjunção é ...

Escolha uma:

- a.  $p \rightarrow q$
- b.  $p \vee q$
- c.  $p \wedge q$
- d.  $\sim p$

6) Aponte a alternativa correta que representa as mesmas combinações a seguir mas na forma de range (faixa)

- I. `x <- c(157, 158, 159, 160, 161, 162, 163, 164)`
- II. `x <- c(15, 16, 17, 18, 20, 21, 22, 23, 24)`

III. `x <- c(10, 9, 8, 7, 6, 5, 4, 3, 2, 1)`  
 IV. `x <- c(-1071, -1072, -1073, -1074, -1075, -1074, -1073, -1072, -1071)`  
 V. `x <- c(1.5, 2.5, 3.5, 4.5, 5.5)`

- a) I) `x <- c(157:164)`; II) `x <-c(15:24)`; III) `x<-c(10:1)`; IV) `x<-c(-1071:-1075,-1074:-1071)`; V) `x<-c(1.5:5.5)`  
 b) I) `x <- c(157;164)`; II) `x <-c(15;24)`; III) `x<-c(10;1)`; IV) `x<-c(-1071;-1071)`; V) `x<-c(1,5;5,5)`  
 c) I) `x <- c(164;157)`; II) `x <-c(24;15)`; III) `x<-c(1;10)`; IV) `x<-c(-1075;-1071,-1071;-1074)`; V) `x<-c(5,5;1,5)`  
 d) I) `x <- c(164:157)`; II) `x <-c(24:15)`; III) `x<-c(1:10)`; IV) `x<-c(-1075:-1071,-1071:-1074)`; V) `x<-c(1,5;5,5)`  
 e) `x <- c(157;164)`; II) `x <-c(15;24)`; III) `x<-c(10;1)`; IV) `x<-c(-1071;-1075,-1074;-1071)`; V) `x<-c(1.5;5.5)`

- 7) Qual o resultado da função de combinação a seguir  
`x<-c(10:20)*20`

- a) 200 220 240 260 280 300 320 340 360 380 400  
 b) 20 22 24 26 28 30 32 34 36 38 40  
 c) 2.0 2.2 2.4 2.6 2.8 3.0 3.2 3.4 3.6 3.8 4.0  
 d) 100 121 144 169 196 225 256 289 324 361 400  
 e) [1] 10000 9801 9604 9409 9216 9025 8836 8649  
 [9] 8464 8281 8100 7921 7744 7569 7396 7225  
 [17] 7056 6889 6724 6561 6400 6241 6084 5929  
 [25] 5776 5625 5476 5329 5184 5041 4900 4761  
 [33] 4624 4489 4356 4225 4096 3969 3844 3721  
 [41] 3600 3481 3364 3249 3136 3025 2916 2809  
 [49] 2704 2601 2500 2401 2304 2209 2116 2025  
 [57] 1936 1849 1764 1681 1600 1521 1444 1369  
 [65] 1296 1225 1156 1089 1024 961 900 841  
 [73] 784 729 676 625 576 529 484 441  
 [81] 400

- 8) Uma rede de hospitalar com foco em telemedicina, contratou um analista de TI para elaborar um plano de prevenção de erros na operação do software dos atendentes nas clínicas, hospitais e laboratórios. Com o objetivo de melhorar a tomada de decisão, o analista classificou os erros quanto ao tempo de emprego dos atendentes e quantificou os erros por setores aos quais estavam alocados. Os dados são apresentados na tabela seguinte.

	Setor da rede hospitalar telemedicina erros cometidos operação software/setor			
Tempo de emprego do atendente	Clínica	hospital	laboratório	Total de erros
< 2 anos	55	25	10	90
2 a 4 anos	20	6	4	30
> 4 anos	35	15	10	60
Total	110	46	24	180

Considerando as informações, conclui-se que a probabilidade aproximada de um atendente, com tempo de emprego menor ou igual a 4 anos pode ser calculada pela seguinte fórmula:

Probabilidade (%) = (tempo de emprego  $\geq$  4 anos)/total de erros

Escolha a alternativa com o resultado correta.

- a) 17%
- b) 33%
- c) 50%
- d) 67%
- e) 83%

**Resposta:**

Adaptado de enade 2021 SI.

Total de erros todos setores com tempo emprego  $\geq$  4 anos

$55+25+10+20+6+4=120$

Ou somar o totalizador de cada linha que contempla o requisito

Total de erros =  $90 + 30 = 120$

Probabilidade =  $120/180 = 0,6667 \rightarrow 0,6667 * 100 = \sim 67\%$