

Nome: Rodrigo Gregolin Giannotti

Ao entregar este trabalho final no e-class Você estará declarando:

"Estou ciente das razões pelas quais estamos fazendo este trabalho de forma remota. Assim, declaro que utilizei somente formas éticas de resolver as questões e não consultei ninguém, de nenhuma forma, para respondê-las "

Instruções:

- Este teste tem **5 páginas**, incluindo esta.
- São 4 questões
- Consulta a notas de aula, apontamentos, net, scripts etc. é permitida.
- **Trabalho é individual. A consulta a colegas ou outras pessoas, qualquer que seja o meio, é proibida.**
- **Todos os cálculos e gráficos deverão ser feitos com o software R**
- Responda as perguntas digitando as respostas neste documento, nos espaços indicados. Respostas fora desses espaços não serão consideradas.
- Por favor, só responda o que for pedido. Não perca tempo tentando explicar o que fez ou por que fez (mesmo porque eu não lerei).
- Se Você copiar algum output do R, formate-o como *Lucida Console*. Cuidado: se não formatar apresentação poderá ficar feia e sua nota será penalizada. É mais fácil copiar os comandos da caixinha de script do que de output
- Os gráficos tem que caber nos espaços indicados
- Ao terminar, **não esqueça de colocar seu nome**. Verifique cuidadosamente se não esqueceu nenhuma resposta nos espaços indicados.
- **Salve como .pdf** e faça upload no ECLASS na caixa de **entrega de atividades**, na pasta correspondente. Verifique se o arquivo salvo ficou em "boas condições" (ou seja se não foi desformatado. Pode acontecer) **Testes entregues no formato .doc não serão aceitos** .
- Veja que a data de entrega é **até as 10h28min do dia 22/09/2020**
- Não envie as respostas por e mail ou WhatsApp. Não serão aceitas

Escrevi para que Você leia, por favor!!!
O não cumprimento das instruções será penalizado

Questão 1: Considere a planilha **CAMBRIDGE** no arquivo **DADOS PAX DW 2020 09 08** em Excel postado no eclass.

- Não altere esta planilha a menos que seja solicitado.
- Não elimine observações, a menos que seja solicitado
- Para simplificar batize esse arquivo como **cam= CAMBRIDGE**

Questão 1: Elabore um boxplot com o arquivo original CAMBRIDGE para comparar a variável TEMPO entre as diferentes categorias de BANDEIRA. Coloque o título no gráfico e faça o grid em cor vermelha (red).

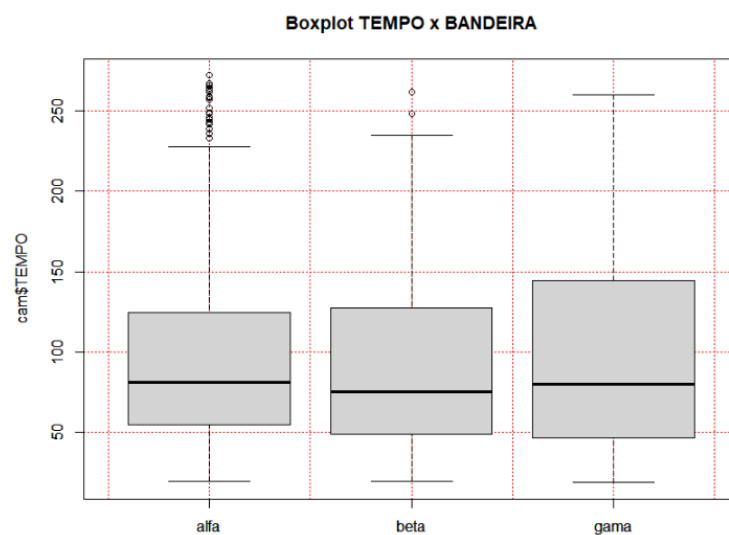
Cole os comandos do R e o boxplot nos espaços abaixo. O gráfico tem que caber na caixa. (para copiar um gráfico clique na caixinha *Export*, na saída de gráficos e *Copy to clipboard*)

Depois calcule com o R a mediana de TEMPO para cada categoria de BANDEIRA

Comandos do gráfico

```
boxplot(cam$TEMPO ~ cam$BANDEIRA)
grid(col = "red")
par(new=TRUE)
boxplot(cam$TEMPO ~ cam$BANDEIRA,
        main = "Boxplot TEMPO x BANDEIRA")
```

Boxplot



Cálculo das medianas (cole os comandos e o output)

```
cam %>%
  group_by(BANDEIRA) %>%
  summarise(median(TEMPO))

# A tibble: 3 x 2
  BANDEIRA `median(TEMPO)`
  <chr>          <dbl>
1 alfa           81
2 beta          75.5
3 gama           80
```

Questão 2: Construa uma tabela de dupla entrada colocando SEXO nas linhas e BANDEIRA nas colunas. Cole a tabela obtida com a proporção por linhas (ou seja a proporção de cada tipo de BANDEIRA dentro de cada SEXO) arredondando as proporções para 3 casas decimais.

Cole os Comandos

```
table(cam$SEXO, cam$BANDEIRA) %>%
  prop.table(margin = 1) %>%
  round(digits = 3)
```

Cole a Tabela

	alfa	beta	gama
fem	0.791	0.135	0.074
masc	0.805	0.119	0.076

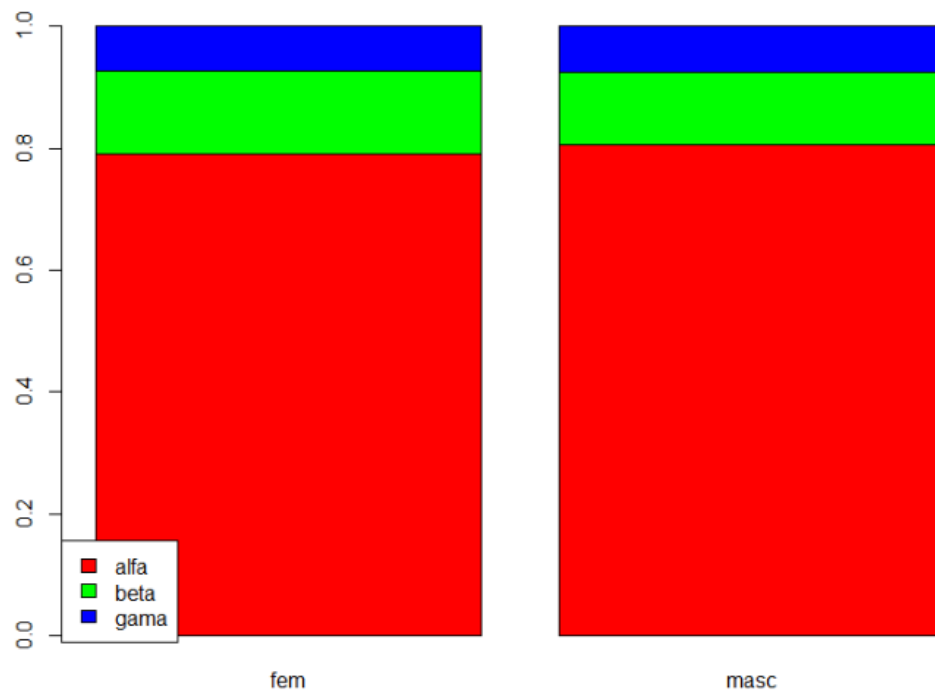
Questão 3: Construa um gráfico de barras comparando a porcentagem de cada categoria de BANDEIRA por categoria de SEXO (SEXO no eixo x e porcentagem de BANDEIRA na vertical). As duas colunas serão da mesma altura (=1,00). Utilize `col=rainbow(3)`. Inclua a legenda, assim Você aprende mais uma coisa!

```
legend('bottomleft', legend = c('alfa','beta','gama'), col=rainbow(3), fill=rainbow(3) )
```

Cole os comandos

```
tabela <- table(cam$BANDEIRA, cam$SEXO) %>%  
  prop.table(margin = 2) %>%  
  barplot(col = rainbow(3))  
legend('bottomleft', legend = c('alfa','beta','gama'), col=rainbow(3),  
fill=rainbow(3) )
```

Gráfico



Questão 4:

a) digite o comando que Você utilizaria no R para criar um novo arquivo contendo apenas clientes cuja BANDEIRA=beta. Denomine o novo arquivo como "betacard". Qual o número de linhas obtido?

Comando(s)

```
betacard <- cam %>%  
  filter(BANDEIRA == "beta")
```

Número de linhas= 248

b) utilizando os comandos do R, calcule o desvio padrão de TEMPO para esse novo arquivo, para cada um dos dois SEXOS.

Co!e o(s) comando(s) e output

```
betacard %>%  
  group_by(SEXO) %>%  
  summarise(sd(TEMPO))
```

```
# A tibble: 2 x 2  
  SEXO   `sd(TEMPO)`  
  <chr>     <dbl>  
1 fem       49.3  
2 masc      57.7
```