|  |  |
| --- | --- |
|  |  |



# 

# Pós-Graduação em Data Science

**Avaliação de “R com Data Science”**

# Prof. Gabrielle Gomes dos Santos Ribeiro

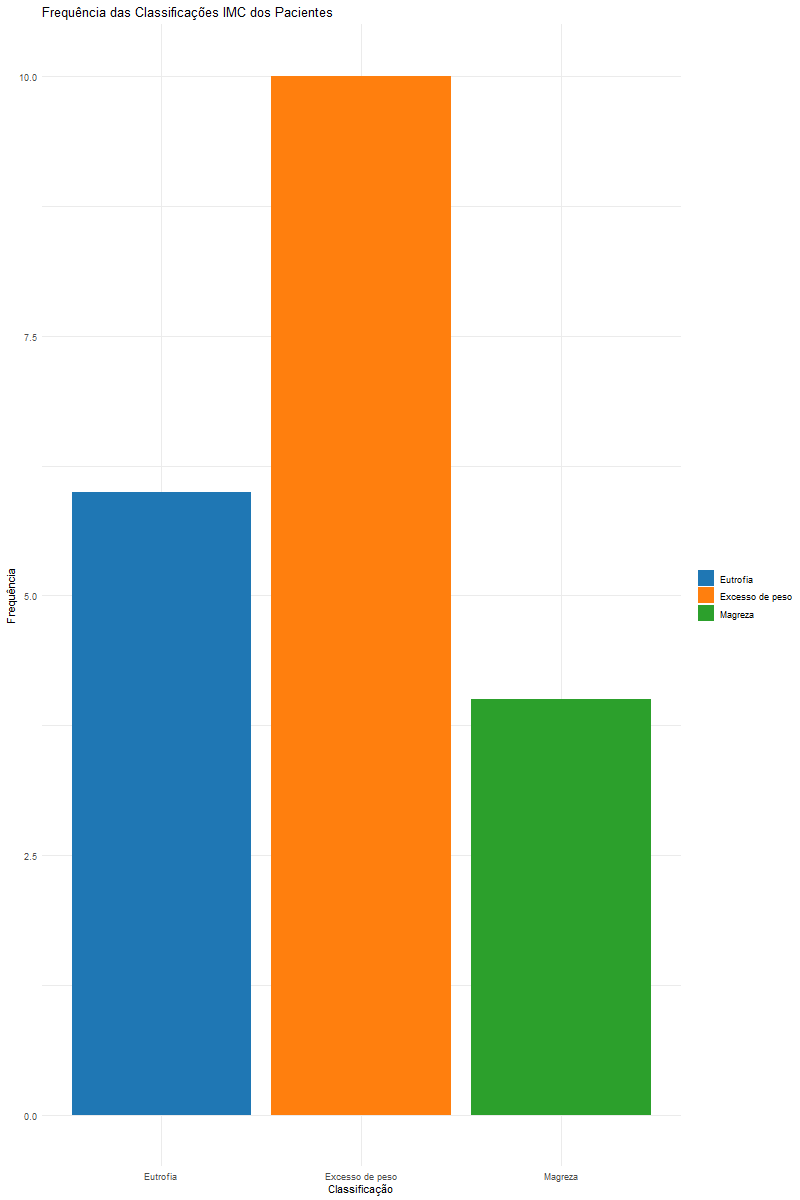
# Nome:\_Luciano da Silva Martins\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_RA:\_9422410584\_\_\_\_\_\_\_\_

**QUESTÕES**

**QUESTÃO 1**

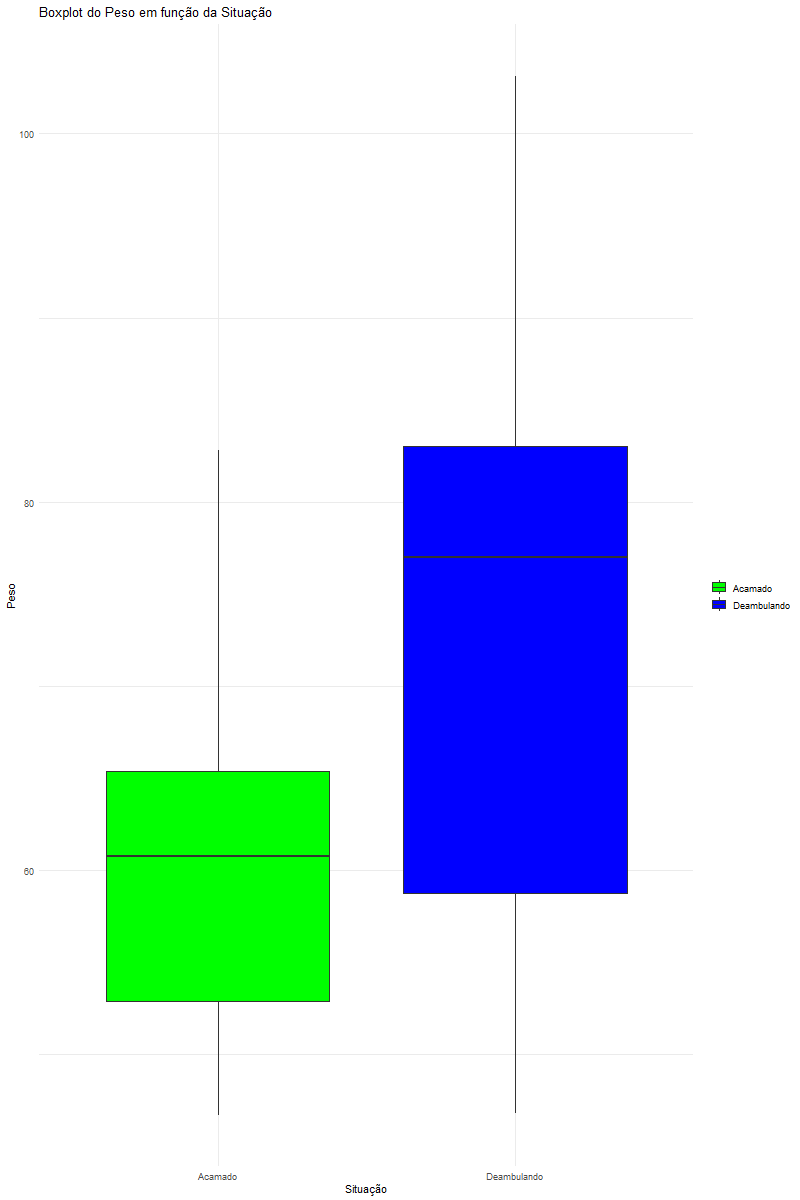
Considere o conjunto de dados “Nutrição.csv” e responda as questões abaixo através de comandos do R:

1. Construa uma tabela de frequências para a variável “Classificação” e interprete o resultado quanto às características da amostra. Construa um gráfico para representar a tabela, com título, cor, legenda e título nos eixos.



**A análise das classificações de IMC dos pacientes revela uma preocupação significativa com o "Excesso de peso", que representa 40% da amostra. Programas focados em controle de peso e prevenção de obesidade podem ser prioritários para melhorar a saúde dessa população. Ao mesmo tempo, a presença de pacientes com "Magreza" e "Eutrofia" indica a necessidade de uma abordagem equilibrada que aborde todas as categorias de IMC de forma eficaz.**

1. Faça um gráfico boxplot do Peso em função da variável Situação. Coloque a cor verde no gráfico da situação “acamado” e azul na situação “deambulando”. O que se pode concluir a partir dos gráficos?



**Essas observações sugerem que os pacientes acamados podem ter mais problemas relacionados ao peso, como ganho excessivo de peso devido à falta de mobilidade, o que é um ponto importante a ser considerado para intervenções médicas e nutricionais. Por outro lado, os pacientes deambulando têm uma distribuição de peso mais consistente, o que pode ser devido a um nível de atividade física mais elevado.**

1. Considere as variáveis Peso e Altura, aplique um teste adequado para verificar a existência de correlação entre as variáveis, especificando as hipóteses do teste e fazendo a conclusão do resultado. A partir da reta de regressão, estime o peso de uma pessoa de 1,80m.

**- Como a correlação é: 0.562058 o peso tende a aumentar com a altura**

**- Existe uma correlação estatisticamente significativa entre Peso e Altura (p-value = 0.0099).**

- **O peso estimado para uma pessoa com 1,80m de altura é 70 kg.**

**QUESTÃO 2**

Um farmacêutico comprou um material específico de dois diferentes fornecedores. Para comparar o nível de impurezas presentes nas compras feitas dos dois fornecedores, o farmacêutico mediu a porcentagem de impurezas presentes em cada um dos grupos, obtendo o que segue:

Fornecedor A: 1,8 2,5 1,5 1,2 1,0

Fornecedor B: 1,6 2,5 1,2 2,3 1,5

1. Calcule a média, mediana e desvio-padrão de cada conjunto.

media\_a # 1.6

* + 1. mediana\_a # 1.5

desvio\_padrao\_a # 0.587367

media\_b # 1.82

mediana\_b # 1.6

desvio\_padrao\_b # 0.5540758

1. Calcule a amplitude dos 2 conjuntos e explique o motivo pelo qual essa medida não é a mais indicada para medir a dispersão dos dados.

maior\_menor\_nro\_A <- range(fornecedor\_A)

maior\_menor\_nro\_A

amplitude\_A <- diff(maior\_menor\_nro\_A)

amplitude\_A # 1.5

maior\_menor\_nro\_B <- range(fornecedor\_B)

maior\_menor\_nro\_B

amplitude\_B <- diff(maior\_menor\_nro\_B)

amplitude\_B # 1.3

**A amplitude é uma medida simples e fácil de calcular, mas suas limitações em termos de sensibilidade a outliers, falta de informação sobre a distribuição dos dados e incapacidade de refletir a dispersão central tornam-na inadequada para uma análise completa da variabilidade dos dados.**

b) Cite as 3 condições que devem ser atendidas para poder comparar a variabilidade dos grupos apenas pelo desvio-padrão. Investigue se neste caso é possível e justifique sua resposta.

**1) Unidade de Medida Consistente**

**2) Escala Similar**

**3) Distribuições Similares**

**Sim é possível pois os dados atendentes as 3 critérios pedidos.**

c) Determine o grupo mais homogêneo.

**O fornecedor\_B é o grupo mais homogêneo**.

**QUESTÃO 3**

Suponha que uma amostra de tamanho n = 6 será extraída de uma população de 20 indivíduos, sendo a idade a variável de interesse. A população é mostrada na íntegra a seguir.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ordem** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| Idade | 43 | 49 | 32 | 64 | 28 | 17 | 41 | 37 | 21 | 14 |
|  |  |  |  |  |  |  |  |  |  |  |
| **Ordem** | **11** | **12** | **13** | **14** | **15** | **16** | **17** | **18** | **19** | **20** |
| Idade | 73 | 59 | 22 | 38 | 36 | 28 | 36 | 63 | 71 | 48 |

Selecione os elementos da amostra utilizando a técnica de amostragem sistemática. A partir dos dados da amostra coletada, estime a idade média da população.

idade <- c(14, 17, 21, 22, 28, 28, 32, 36, 36, 37, 38, 41, 43, 48, 49, 59, 63, 64, 71, 73)

N <- length(idade)

n <- 6

intervalo <- N / n

mean(idade)

41

**QUESTÃO 4**

Considere o conjunto Milsa.txt e responda os itens a seguir:

1. Selecione uma amostra aleatória simples de 20% do total de funcionários e a partir da amostra, verifique, a partir da medida numérica adequada, se os salários dos funcionários podem ser considerados homogêneos ou não.

dados <- read.table("Milsa.txt", header = T)

vinte\_por\_cento\_dos\_dados <- ceiling(0.2 \* salarios) # 8

set.seed(123)

amostra\_indices <- sample(1:N, vinte\_por\_cento\_dos\_dados) # 15 19 14  3 10  2  6 11

amostra <- dados$salario[amostra\_indices] # 9.13 10.53  8.95  5.25  7.44  4.56  6.66  8.12

desvio\_padrao\_amostra <- sd(amostra) # 2.024451

media\_amostra <- mean(amostra) # 7.58

cv\_amostra <- (desvio\_padrao\_amostra / media\_amostra) \* 100 # 26.70779

homogeneidade <- ifelse(cv\_amostra < 20, "Homogêneos", "Não homogêneos") # "Não homogêneos"

**À partir dos dados analisados foi concluído que os salários NÃO SÃO HOMOGÊNEOS**

1. Utilize a variável “Instrução” para dividir a população em estratos. Considerando uma amostra de tamanho 10, determine o tamanho dos estratos, a quantidade de funcionários que serão amostrados em cada um e os elementos selecionados

**# Tamanho dos Estratos**

**instr Tamanho Proporcao Amostra**

**1 1 16 0.40 4**

**2 2 17 0.42 4**

**3 3 7 0.18 2**

**# Amostra Selecionada**

**Func civil instr filhos salario ano mes regiao**

**2 2 2 1 1 4.56 32 10 2**

**3 3 2 1 2 5.25 36 5 2**

**12 12 1 1 NA 8.46 27 11 2**

**8 8 1 1 NA 7.39 43 4 2**

**14 14 2 1 3 8.95 44 2 3**

**15 15 2 2 0 9.13 30 5 1**

**16 16 1 2 NA 9.35**