

Trabalho_R

Pergunta 1 – Classificação das Variáveis

Nesta etapa, procedeu-se à classificação técnica de todas as variáveis presentes na base de dados, identificando a sua natureza e a respetiva escala de medição. Esta classificação é fundamental para selecionar os métodos estatísticos e as representações gráficas mais adequadas.

Variável	Tipo de Variável	Escala de Medição
Name	Qualitativa Nominal	Nominal
Age	Quantitativa Contínua	Razão
Gender	Qualitativa Nominal	Nominal
Blood Type	Qualitativa Nominal	Nominal
Medical Condition	Qualitativa Nominal	Nominal
Date of Admission	Quantitativa Contínua	Intervalo
Doctor / Hospital	Qualitativa Nominal	Nominal
Insurance Provider	Qualitativa Nominal	Nominal
Billing Amount	Quantitativa Contínua	Razão
Room Number	Qualitativa Nominal	Nominal
Admission Type	Qualitativa Nominal	Nominal
Discharge Date	Quantitativa Contínua	Intervalo
Medication	Qualitativa Nominal	Nominal
Test Results	Qualitativa Nominal	Nominal

Pergunta 2 -Caracterização de variáveis

Vamos usar o R Base para os gráficos da **Age** (Histograma e Boxplot) para evitar que o R tente instalar o **patchwork** ou o **ggplot2** a meio do PDF, o que “suja” o documento.

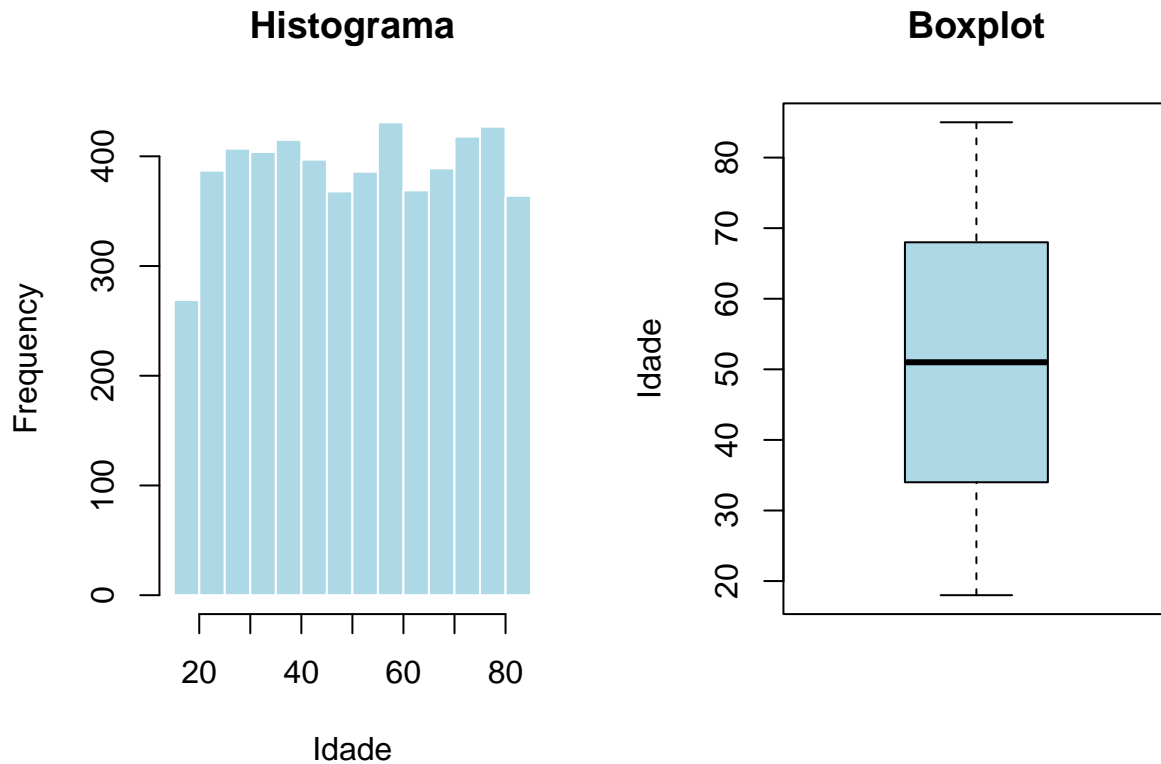
2.1 “Age”

Classificação: Quantitativa Contínua (Escala de Rácio).

Justificação: A variável **Age** é do tipo quantitativa contínua. Para a sua caracterização, foram calculadas as estatísticas descritivas de tendência central (média e mediana) e de dispersão (desvio padrão). Complementarmente, apresenta-se um histograma para observar a forma da distribuição etária e um boxplot para verificar a existência de valores atípicos (outliers). A idade média dos pacientes na amostra é de 51.25 anos. A proximidade entre a média e a mediana, aliada à observação do histograma, sugere uma distribuição relativamente uniforme/simétrica. O boxplot confirma a ausência de outliers, indicando que não existem idades extremas que possam distorcer a análise.

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	18.00	34.00	51.00	51.25	68.00	85.00

```
## [1] 19.67971
```



2.2 “Gender”

Classificação: Qualitativa Nominal

Justificação: Sendo uma variável categórica sem ordem, a análise baseia-se em frequências absolutas e relativas. Utilizou-se um Gráfico de Barras, pois é mais eficaz que o gráfico de setores para comparar proporções de forma rigorosa. Incluiu-se uma linha de referência nos 50% que serve para avaliar o balanço da amostra.

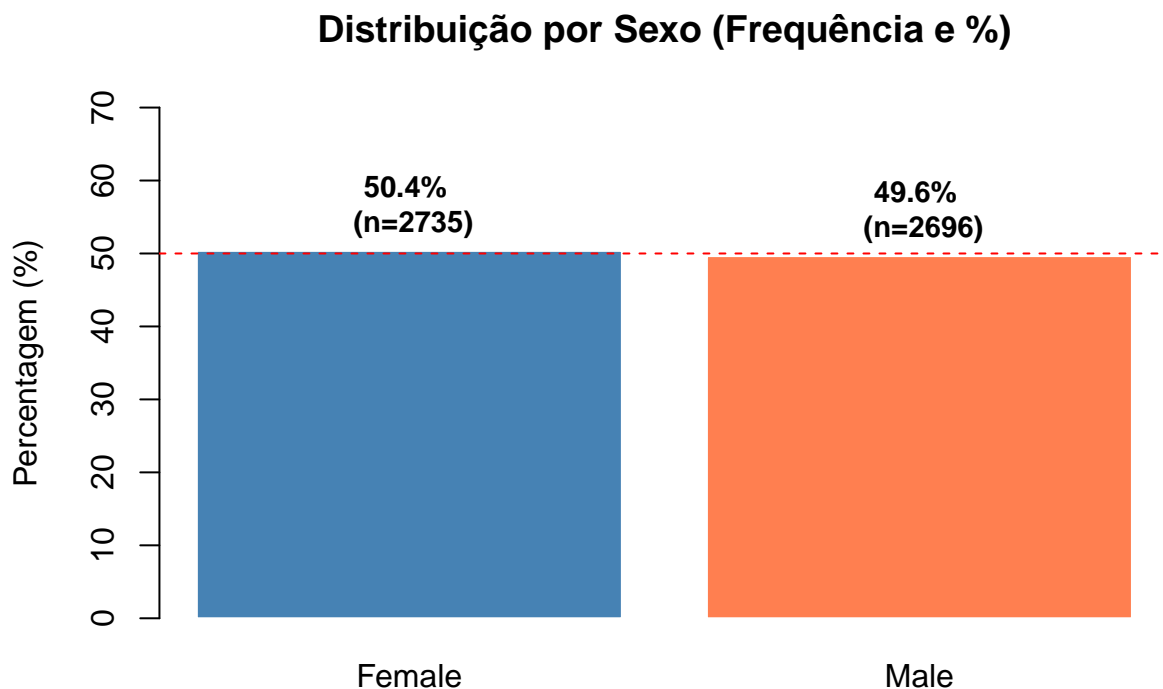
```
# 1. Obter frequências absolutas e relativas
genero_abs <- table(amostra$Gender)
genero_rel <- prop.table(genero_abs) * 100

# 2. Criar o gráfico de barras
# Ajustamos o ylim para 70 para garantir que o texto não corte no topo
bar_pos <- barplot(genero_rel,
  main = "Distribuição por Sexo (Frequência e %)",
  ylab = "Percentagem (%)",
  col = c("steelblue", "coral"),
  ylim = c(0, 70),
  border = "white")
```

```
# 3. Adicionar a linha de equilíbrio nos 50%
abline(h = 50, col = "red", lty = 2)

# 4. Criar as etiquetas: "Percentagem% (n=Valor)"
etiquetas <- paste0(round(genero_rel, 1), "% \n (n=", genero_abs, ")")

# 5. Colocar as etiquetas no topo das barras
text(x = bar_pos,
     y = genero_rel,
     labels = etiquetas,
     pos = 3,      # Coloca acima da coordenada Y
     cex = 0.9,    # Tamanho da fonte
     font = 2)     # Negrito
```



2.3 “Blood.Type”

Classificação: Variável qualitativa nominal com 8 categorias.

Justificação: Utilizou-se um gráfico de barras ordenado de forma decrescente, o que facilita a identificação imediata da “Moda” (o grupo mais frequente).

```
# 1. Preparar dados
blood_tab <- table(amostra$Blood.Type)
blood_tab_sorted <- sort(blood_tab, decreasing = TRUE)
```

```

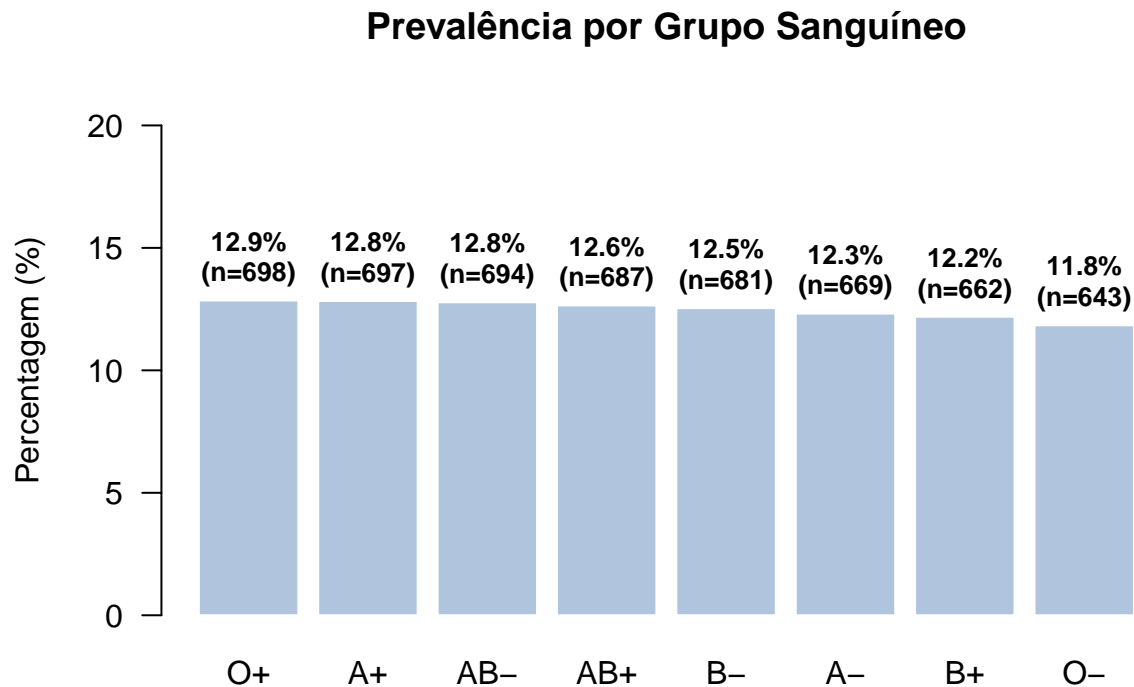
blood_perc <- prop.table(blood_tab_sorted) * 100

# 2. Criar o gráfico
# Aumentamos o ylim para o texto não ser cortado
bp_blood <- barplot(blood_perc,
  main = "Prevalência por Grupo Sanguíneo",
  col = "lightsteelblue",
  ylim = c(0, max(blood_perc) + 8),
  las = 1,
  border = "white",
  ylab = "Porcentagem (%)")

# 3. Criar etiquetas formatadas
txt_blood <- paste0(round(blood_perc, 1), "%\n(n=", blood_tab_sorted, ")")

# 4. Adicionar texto por cima das barras
text(x = bp_blood,
  y = blood_perc,
  labels = txt_blood,
  pos = 3,
  cex = 0.8,
  font = 2)

```



A análise da distribuição dos grupos sanguíneos revela que o tipo O+ é o mais frequente na amostra, registrando-se em 698 pacientes ($r = \text{round}(\text{blood_perc}[1], 1)\%$). Seguem-se os grupos A+ e AB-. Esta diversidade é característica de bases de dados hospitalares amplas, não se verificando uma predominância

absoluta que comprometa a representatividade de um grupo específico.

2.3 “Medical.Condition”

Classificação: Qualitativa Nominal

Justificação: Para a variável Medical.Condition, optou-se por um gráfico de barras horizontal. Esta escolha permite uma leitura clara das patologias, evitando a sobreposição de texto, e facilita a comparação da carga de doença entre as diferentes categorias.

```
# 1. Preparar dados
med_tab <- sort(table(amostra$Medical.Condition), decreasing = TRUE)
med_perc <- prop.table(med_tab) * 100

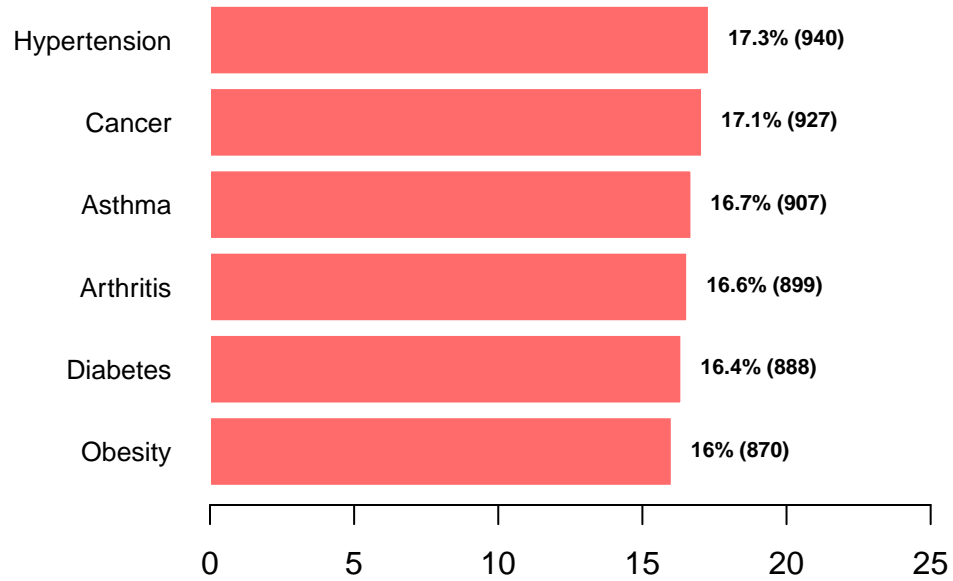
# 2. Inverter para o gráfico horizontal (maiores em cima)
med_tab_plot <- rev(med_tab)
med_perc_plot <- rev(med_perc)

# 3. Ajustar margens (Aumentar a margem esquerda para o texto das doenças)
par(mar = c(5, 10, 4, 2))

# 4. Gráfico Horizontal
bp_med <- barplot(med_perc_plot,
                  horiz = TRUE,
                  las = 1,
                  main = "Distribuição de Condições Médicas",
                  col = "indianred1",
                  xlim = c(0, max(med_perc_plot) + 10),
                  border = "white",
                  cex.names = 0.8)

# 5. Adicionar etiquetas à direita
txt_med <- paste0(round(med_perc_plot, 1), "% (", med_tab_plot, ")")
text(x = med_perc_plot, y = bp_med, labels = txt_med, pos = 4, cex = 0.7, font = 2)
```

Distribuição de Condições Médicas



```
# Reset das margens para o próximo gráfico não ficar estranho  
par(mar = c(5, 4, 4, 2))
```