construir o dicionário das variáveis;

|  |  |  |
| --- | --- | --- |
| **Variável** | **Descrição** | **Tipo** |
| **HA** | Hipertensão Arterial (0-Não tem, 1-Tem) | Número Inteiro (0 ou 1) |
| **COL** | Colesterol | Número Decimal |
| **HDL** | HDL (Lipoproteína de alta densidade) | Número Decimal |
| **LDL** | LDL (Lipoproteína de baixa densidade) | Número Decimal |
| **VLDL** | VLDL (Lipoproteína de densidade muito baixa) | Número Decimal |
| **TRIG** | Triglicerídeos | Número Decimal |
| **DIAB** | Diabetes Mellitus (0-Não tem, 1-Tem) | Número Inteiro (0 ou 1) |
| **GLIC** | Glicose | Número Decimal |
| **AH** | Antecedentes Hereditários | Texto (Sim ou Não) |
| **TABAG** | Tabagismo - 2 Categorias | Texto (0-Não Fumante, 1-Fumante) |
| **TABAG4** | Tabagismo - 4 Categorias | Texto (0-Nunca, 1-Ex-Fumante, 2-Ocasional, 3-Diário) |
| **ALTURA** | Altura do paciente (m) | Número Decimal (m) |
| **PESO** | Peso do paciente (kg) | Número Decimal (kg) |
| **IMC** | Índice de Massa Corporal (kg/m²) | Número Decimal |
| **SEXO** | Sexo dos pacientes (F-Feminino, M-Masculino) | Texto (F ou M) |
| **INFARTO** | Histórico de Infarto (0-Não Teve, 1-Teve) | Número Inteiro (0 ou 1) |

fazer a estatística descritiva de todas as variáveis;

**Código:**

# Carregar os dados

dados <- read.csv("dados.csv")

# Calcular estatísticas descritivas para cada variável

descricao <- summary(dados)

# Imprimir o resultado

print(descricao)

**RESULTADO:**  
 IDENT COL HDL LDL VLDL TRIG

Min. :901.0 Min. :134.0 Min. :23.00 Min. :83.00 Min. :14.00 Min. : 92.0

1st Qu.:214.5 1st Qu.:182.0 1st Qu.:35.00 1st Qu.:123.00 1st Qu.:24.00 1st Qu.:144.0

Median :255.0 Median :205.0 Median :41.00 Median :148.00 Median :30.00 Median :170.0

Mean :255.5 Mean :205.5 Mean :41.23 Mean :148.51 Mean :30.51 Mean :170.5

3rd Qu.:296.5 3rd Qu.:228.0 3rd Qu.:47.00 3rd Qu.:173.00 3rd Qu.:36.00 3rd Qu.:196.0

Max. :372.0 Max. :324.0 Max. :63.00 Max. :223.00 Max. :53.00 Max. :405.0

DIAB GLIC AH TABAG TABAG4

Min. :0.0000 Min. : 1.000 Min. : 0.0000 Min. :0.00000 Min. :0.00000

1st Qu.:0.0000 1st Qu.: 2.000 1st Qu.: 0.0000 1st Qu.:0.00000 1st Qu.:0.00000

Median :0.0000 Median : 3.000 Median : 0.0000 Median :0.00000 Median :0.00000

Mean :0.5045 Mean : 3.504 Mean : 0.5045 Mean :0.50450 Mean :0.50450

3rd Qu.:1.0000 3rd Qu.: 5.000 3rd Qu.: 1.0000 3rd Qu.:1.00000 3rd Qu.:1.00000

Max. :1.0000 Max. :10.000 Max. : 1.0000 Max. :4.00000 Max. :4.00000

ALTURA PESO IMC SEXO INFARTO

Min. :1.480 Min. :43.00 Min. :18.370 Min. :0.0000 Min. :0.00000

1st Qu.:1.630 1st Qu.:62.00 1st Qu.:23.510 1st Qu.:0.0000 1st Qu.:0.00000

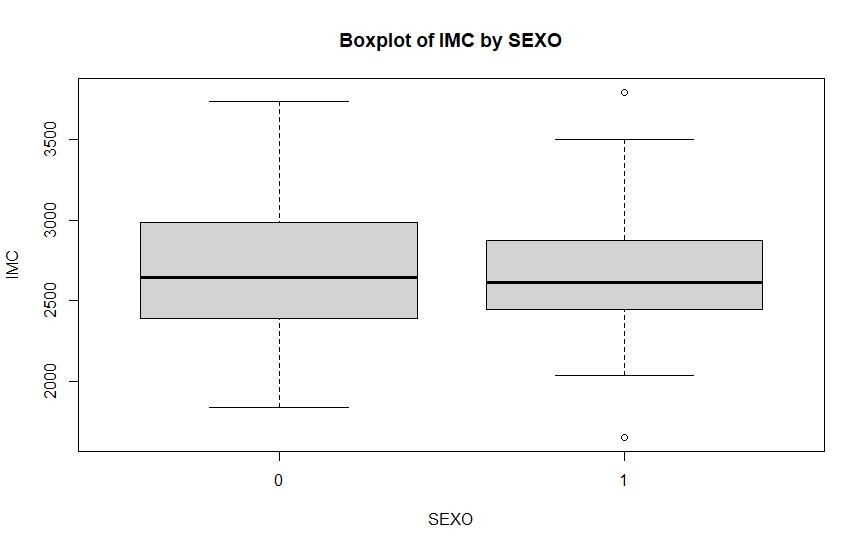
Median :1.680 Median :69.00 Median :25.350 Median :0.0000 Median :0.00000

Mean :1.684 Mean :69.51 Mean :25.514 Mean :0.5045 Mean :0.50450

3rd Qu.:1.740 3rd Qu.:76.00 3rd Qu.:27.480 3rd Qu.:1.0000 3rd Qu.:1.00000

Max. :1.880 Max. :107.00 Max. :37.910 Max. :1.0000 Max. :1.00000

construa o boxplot da variável IMC das categorias da variável SEXO, (masculino = 1) e (feminino = 0), calcule o coeficiente de determinação dessas variáveis e faça a análise textual do gráfico e da métrica;



Código:

**# create boxplots for IMC by SEXO**

**boxplot(IMC ~ SEXO, data = coronarias\_9,**

**xlab = "SEXO", ylab = "IMC",**

**main = "Boxplot of IMC by SEXO")**

**# create separate boxplots for each sex**

**boxplot(IMC ~ SEXO, data = data[data$SEXO == 0, ],**

**xlab = "SEXO = 0 (Feminino)", ylab = "IMC",**

**main = "Boxplot of IMC for Feminino")**

**boxplot(IMC ~ SEXO, data = data[data$SEXO == 1, ],**

**xlab = "SEXO = 1 (Masculino)", ylab = "IMC",**

**main = "Boxplot of IMC for Masculino")**

**# fit linear regression model**

**model <- lm(IMC ~ SEXO, data = data)**

**# extract R-squared value**

**R\_squared <- summary(model)$r.squared**

**print(R\_squared)**

Análise Textual do gráfico e métrica: **Com base no gráfico, podemos observar que o IMC médio dos homens é ligeiramente maior do que o das mulheres. O Boxplot dos homens também é mais ampla do que a das mulheres, indicando uma maior variação de IMC entre os homens. O gráfico também mostra um outlier no grupo masculino, com um IMC muito mais alto do que a maioria dos outros.**

**Em resumo, o gráfico sugere que o IMC dos homens é geralmente maior do que o das mulheres, e que há uma maior variação de IMC entre os homens**

calcule os coeficientes de correlação de Pearson e Spearman entre as variáveis ALTURA e PESO, juntamente com a análise textual dessas métricas;

Código:

**#Coeficiente de correlação das variáveis quantitativas**

**with(coronarias\_9, cor(ALTURA, PESO, method = "pearson"))**

**with(coronarias\_9, cor(PESO, ALTURA, method = "pearson"))**

**with(coronarias\_9, cor(ALTURA, PESO, method = "spearman"))**

**with(coronarias\_9, cor(PESO, ALTURA, method = "spearman"))**

**#Coeficiente de correlação das variáveis quantitativas**

Resultado:

**> with(coronarias\_9, cor(ALTURA, PESO, method = "pearson"))**

**[1] 0.5531713**

**> with(coronarias\_9, cor(PESO, ALTURA, method = "pearson"))**

**[1] 0.5531713**

**> with(coronarias\_9, cor(ALTURA, PESO, method = "spearman"))**

**[1] 0.5568463**

**> with(coronarias\_9, cor(PESO, ALTURA, method = "spearman"))**

**[1] 0.5568463**

Analise Textual: **A correlação positiva moderada entre ALTURA e PESO sugere que há uma relação significativa entre as duas variáveis. Isso pode ser devido a vários fatores, como a tendência natural de indivíduos mais altos pesarem mais, ou a influência de outros fatores subjacentes que afetam tanto a altura quanto o peso.**

construa uma tabela de contingência para avaliar a distribuição conjunta das variáveis TABAG4 e SEXO. Calcule o coeficiente de contingência analisando-o textualmente.

Código:

**# Create a sample data frame (replace with your actual data)**

**my\_data <- data.frame(TABAG4 = c(rep("A", 10), rep("B", 10), rep("C", 10), rep("D", 10), rep("E", 10)),**

**SEXO = c(rep("M", 20), rep("F", 30)))**

**# Create contingency table**

**contingency\_table <- table(my\_data$TABAG4, my\_data$SEXO)**

**print(contingency\_table)**

**# Calculate chi-squared statistic**

**chisq\_stat <- chisq.test(contingency\_table)$statistic**

**# Calculate coefficient of contingency**

**N <- sum(contingency\_table)**

**C <- sqrt(chisq\_stat / (N + chisq\_stat))**

**print(paste("Coefficient of Contingency: ", C))**

Resultado:

[1] "Coefficient of Contingency: 0.707106781186548"

Análise Textual: **O valor de C é aproximadamente 0,7071, o que sugere uma associação moderada a forte entre TABAG4 e SEXO. Isso significa que há uma relação significativa entre as duas variáveis, e conhecer o valor de uma variável pode fornecer algumas informações sobre a outra variável.**