

TAREFA 08

Bernardo Chrispim Baron

c) O conjunto de dados não tem comportamento de uma normal multivariada: o Teste de Shapiro-Wilk multivariado, proposto por Villasenor-Alva e Gonzalez-Estrada (2009) e implementado em R no pacote “mvShapiroTest”, retorna para o conjunto um p-valor de $4,9.10^{-14}$, corroborando a hipótese alternativa de que os dados não são normais multivariados.

d) Tendo em vista os gráficos “qqnorm” gerados e reproduzidos abaixo, bem como os p-valores do teste de Shapiro-Wilk para cada variável, é possível afirmar (com nível de significância de 0,05) que não são normais as variáveis idade (Idade), cintura (CINT), pulso (TXPUL), pressão sistólica (SIST), colesterol (COL), Índice de Massa Corpórea (IMC) e diâmetro da perna (Perna).

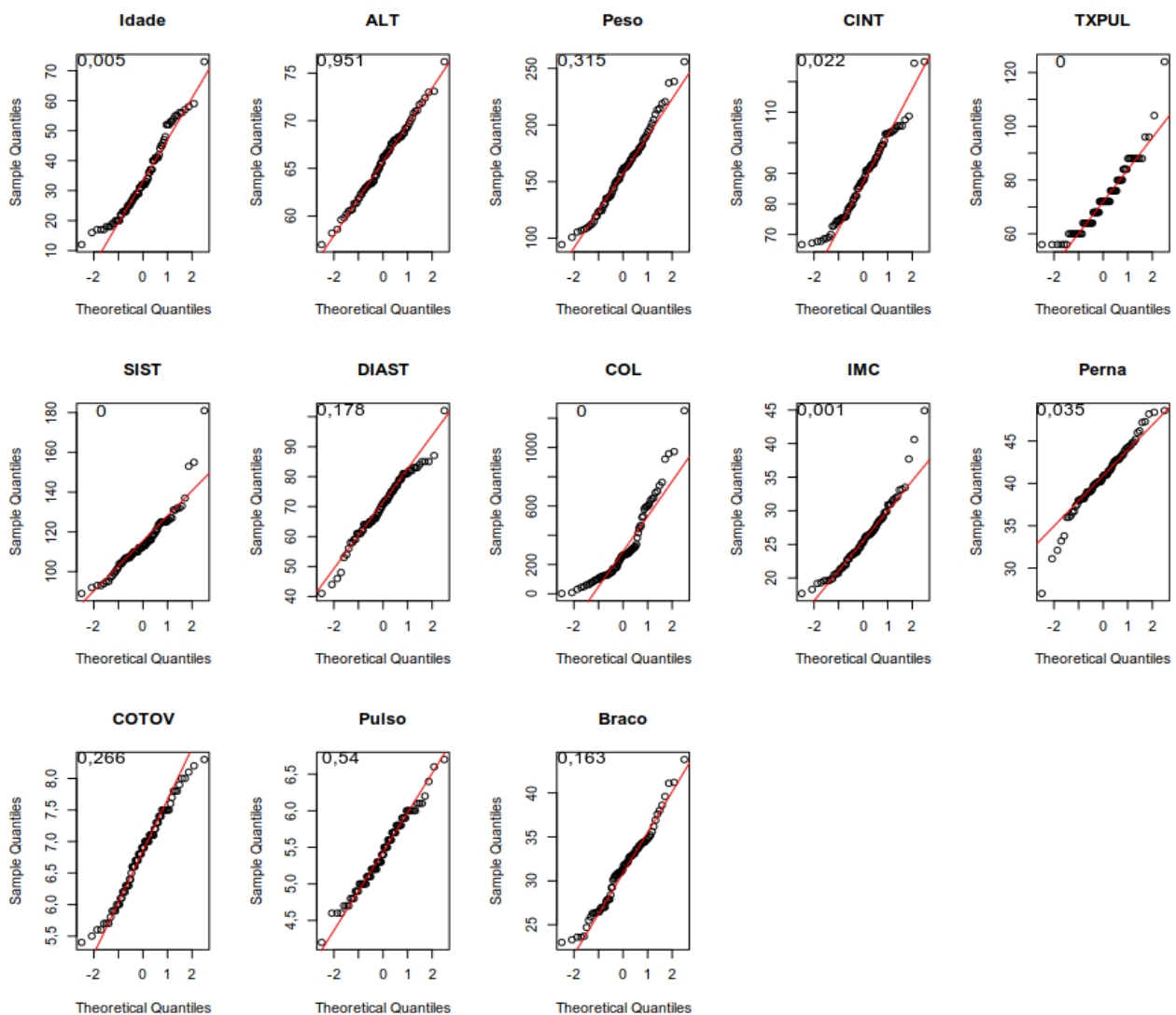


Figure 1: Gráficos "qqnorm" para as variáveis numéricas dos dados FSAÚDE.xls e MSAÚDE.xls.

e) O p-valor do Teste de Hotelling é de $1,1 \cdot 10^{-16}$, corroborando a hipótese alternativa de que os vetores de médias das variáveis medidas são distintos entre a população de mulheres e a população de homens das quais foram retiradas as amostras.

f) O p-valor do Teste BoxM é de 0,054, não permitindo descartar, com nível de significância de 0,05, a hipótese nula de que as matrizes de variância e covariância são iguais entre a população de homens e a população de mulheres das quais foram retiradas as amostras.

e) A análise discriminante com o total da amostra foi capaz de identificar corretamente 36 das quarenta mulheres (90%) e 37 dos 40 homens (92,5%). Isso constitui um erro de treinamento de 8,75%.

Sexo classificado	Sexo verdadeiro	
	Feminino	Masculino
Feminino	36	3
Masculino	4	37