Serviço Público Federal
Universidade Federal do Pará
Instituto de Ciências Exatas e Naturais
Faculdade de Estatística

Dionisio Alves da Silva Neto
Matrícula: 202007840008

Atividade 1 de Análise Multivariada II:
Análise Exploratória de Dados (AED)

Belém, PA
2022

1. Introdução

1.1 Propósitos e justificativa

O presente trabalho visa aplicar as técnicas de Análise exploratória de Dados (AED), a qual visa resumir, organizar e descrever as informações presentes no banco de dados em gráficos e tabelas. A aplicação da síntese feita pela AED retrata as características de determinada população ou fenômeno, determinando relações entre as variáveis analisadas.

1.2 Banco de dados

O banco de dados de dados utilizado é referente às medições dos níveis de glicose no sangue de pacientes do sexo feminino. Desse modo, na metodologia de coleta, foram utilizados dois recortes temporais: as semanas e o momento de ingestão de açúcar. Para as semanas houveram 3 períodos (1ª semana, 2ª semana e 3ª semana) e, em relação à ingestão de açúcar, houveram 2 instantes (jejum e após a ingestão da sacarose).

1.3 Pré-processamento dos dados

Ao decorrer deste projeto, foi utilizada a linguagem de programação R (R Core Team, 2022), a qual é fundamental para a aplicação de métodos, modelos e testes estatísticos. Nesta perspectiva, durante o pré-processamento dos dados, percebeu-se que o banco de dados disponibilizado encontrava-se no formato *wide*, o qual resume as informações numéricas em relação a duas ou três variáveis categóricas de modo simplificado. Para adequar ao formato *long*, o qual é o adequado para o tratamento de dados em linguagens de programação, foi utilizado o código (disponibilizado no **Anexo I)** para ajustar ao formato que a linguagem R opera. Logo, com o intuito de unir as informações de glicose juntamente com a variável de ingestão de açúcar, foi criado uma nova coluna no banco de dados para a categorização dos momentos (0: jejum e 1: após a ingestão de açúcar), a qual está presente no **Anexo II** deste trabalho.

2. Medidas descritivas

2.1 Vetor de Médias geral

Para o vetor de médias, considerando ambos os momentos de jejum e uma hora após, tem-se que as médias para a primeira, semana e terceira semana são 90,39, 89,09, 89,09, respectivamente. Destarte, temos a ideia que, em média, o nível de glicose nas três semanas são bastantes parecidos.

y1	y2	y3
90,39	89,09	89,09

2.2 Vetor de médias do grupo em jejum

Com o vetor de médias para o grupo de jejum, tem-se que 70,10, 73,60 e 75,20 representam os níveis médios de glicose para a primeira, segunda e terceira semana, respectivamente. Podemos observar que para este grupo existe uma pequena diferença entre os valores médios obtidos para as semanas 1 e 3. Por outro lado, o valor da semana 2 encontra-se bastante próximo para as outras semanas. Logo, conclui-se que, para o grupo de jejum, ocorre um aumento no nível médio de glicose ao longo das 3 semanas de observação das pacientes.

y1	y2	y3
70,10	73,60	75,20

2.3 Vetor de médias do grupo que consumiu a sacarose uma hora após

Para o grupo em jejum, nota-se que os valores médios do nível de glicose para as semanas 1, 2 e 3 são 111,00, 105,00 e 111,00, nesta respectiva ordem. Desse modo, tem-se que os valores da primeira e terceira semana são iguais, resultando em um valor médio de 111,00. Em contraste, na segunda semana, é perceptível uma queda no nível médio de glicose para 105,00. Em suma, o valor médio do nível de glicose nas 3 semanas de estudo apresenta um declínio durante a segunda semana e uma igualdade na primeira e terceira semana.

y1	y2	y3
111,00	105,00	111,00

2.4 Matriz de variâncias e covariâncias geral

Na análise da matriz de variâncias e covariâncias do nível de glicose para os dois momentos de ingestão conjuntamente, temos que as variâncias para a primeira, segunda e terceira semana são 847,81, 532,04 e 597,82, respectivamente. Percebe-se que a maior variabilidade está presente para os dados de glicose da semana 1 e, para as semanas 2 e 3, a dispersão dos dados bem próximos.

Em relação às covariâncias do nível de glicose entre as semanas, observamos que para as semanas 1 e 2 a covariância vale 479,32; para as semanas 1 e 3, a covariância vale 464,95 e; para as semanas 2 e 3, a covariâncias vale 362,84. Nesta visão, é correto afirmar que a menor covariância entre variáveis ocorre entre as semanas 2 e 3 e, para as demais medidas de variância conjunta, os valores são bem próximos.

	y1	y2	y3
y1	847,81	479,32	464,95
y2	479,32	532,04	362,84
y3	464,95	362,84	597,85

2.5 Matriz de variâncias e covariâncias para o grupo em jejum

Para o grupo em jejum, temos que as variâncias da primeira, segunda e terceira semana são 97,30, 74,60 e 77,00, nesta respectiva ordem. Os valores para as covariâncias são 17,80, para as semanas 1 e 2; 12,00, para as semanas 1 e 3 e; 14,20, para as semanas 2 e 3. Em análise, um primeiro fato observado mostra que em comparação à matriz de variâncias e covariâncias geral, a matriz especificada para o grupo em jejum apresenta valores consideravelmente menores para as medidas de variabilidade individual e conjunta.

	y1	y2	y3
y1	97,30	17,80	12,00
y2	17,80	74,60	14,20
y3	12,00	14,20	77,00

2.6 Matriz de Variâncias e Covariâncias para o grupo que consumiu sacarose uma hora após

Para o grupo que ingeriu glicose uma hora após, os valores 779,00, 510,00 e 485,00 representam as variâncias das semanas 1, 2 e 3 para tal conjunto de indivíduos. O valor das variabilidade conjunta entre as semanas 1 e 2 é 310.00; entre as semanas 1 e 3 é 192,00 e; entre as semanas 2 e 3 é 156,00. Logo, em comparação ao grupo em jejum, percebe-se que os valores de variância e covariância para o grupo de que ingeriu açúcar uma hora após são bastante elevados, tal fato explica que, nos dados de glicose, a variabilidade nas três semanas é altamente influenciada pelo grupo segundo momento de medição.

	y1	y2	y3
y1	779,00	310,00	192,00
y2	310,00	510,00	156,00
y3	192,00	156,00	485,00

.

2.7 Matriz de Correlações geral

A matriz de variâncias de variâncias e covariâncias é de grande interpretação, mas, normalmente, é preciso da matriz de correlações lineares para uma real mensuração da interação entre as variáveis no estudo. Por isso, na matriz de correlações geral, temos que a correlação para as semanas 1 e 2, vale 0,7137; para as semanas 1 e 3, vale 0,6531. Logo, conclui-se que a maior correlação do nível de glicose está entre as semanas 1 e 2, a qual é classificada como positiva e forte. Para a primeira e terceira semana, assim como a segunda e terceira semana, as correlações são próximas e classificadas como positiva e moderada.

	y1	y2	y3
y1	1,0000	0,7137	0,6531
y2	0,7137	1,0000	0,6434
y3	0,6531	0,6434	1,0000

2.8 Matriz de Correlações no momento jejum

Para o grupo de pacientes mulheres em jejum temos que a correlação entre a primeira e segunda semana vale 0,2090; para a primeira e terceira semana vale 0,1390 e para a segunda e terceira semana vale 0,1890. Nota-se que para esta parcela da população os valores da correlação linear de Pearson são classificados como positivos e fracos, diferente da situação geral.

	y1	y2	y3
y1	1,0000	0,2090	0,1390
y2	0,2090	1,0000	0,1880
y3	0,1390	0,1880	1,0000

2.9 Matriz de Correlações no momento após a ingestão de açúcar

Em relação ao grupo que ingeriu açúcar após 1 hora, temos que os valores para a correlação linear para as semanas 1 e 2 é de 0,4920; para as semanas 1 e 3, a correlação vale 0,3130 e; para as semanas 2 e 3, a correlação vale 0,3140. Do mesmo modo ao grupo anterior, percebe-se correlações positivas, porém fracas para os níveis de glicose.

	y1	y2	y3
y1	1,0000	0,4920	0,3130
y2	0,4920	1,0000	0,3140
y3	0,3130	0,3140	1,0000

2.10 Estatística Descritiva do banco de dados

Na **Tabela 1**, é apresentado o resumo descritivo completo sobre o banco de dados do nível de glicose para os momentos de medição e nas três semanas de coleta. Em análise, percebe-se que os valores para a mediana nas 3 semanas apresentam valores mais distantes em comparação ao já calculado para o vetor de médias.

Em segundo plano, para os valores de mínimo e máximo, percebe-se que o menor nível de glicose coletado foi na primeira semana do estudo(48,00), como também, o maior nível de glicose também foi coletado na semana 1. Em relação aos

quartis calculados para os dados, nota-se que o maior valor para o quartil 1 pertence à terceira semana, enquanto que o maior valor do quartil 3 pertence à primeira semana.

Tabela 1: Valores das estatísticas descritivas do nível de glicose para cada semana.

Estatística	y1	y2	y3
Mínimo	48,00	53,00	59,00
Q1	69,25	71,00	74,00
Mediana (Q2)	77,00	82,00	86,00
Média	90,39	89,09	93,01
Q3	112,00	101,00	109,00
Máximo	170,00	158,00	153,00

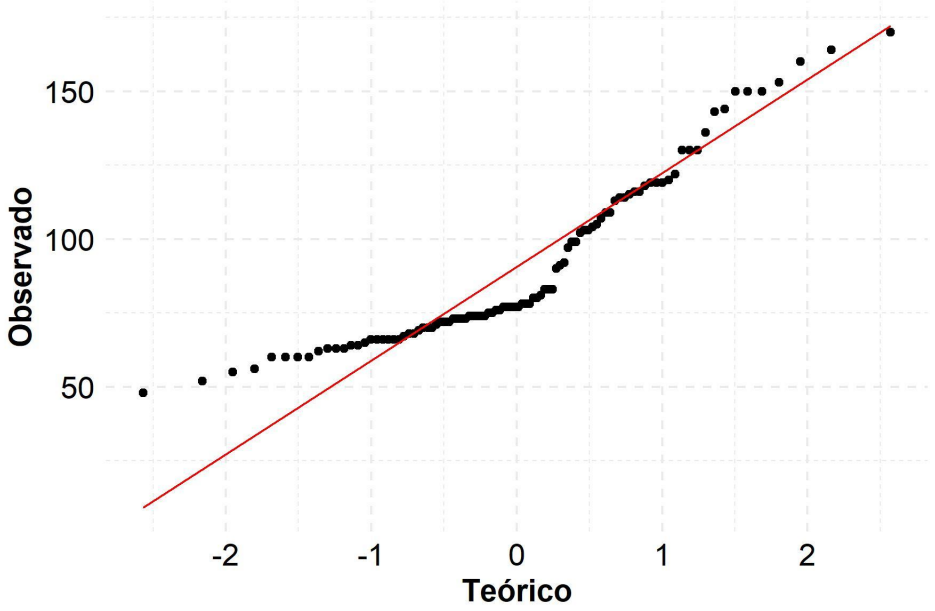
Fonte: Construído pelo autor, 2022.

3. Gráficos Univariados

3.1 Q-Q plot

Na **Figura 1**, nota-se o gráfico Q-Q plot do nível de glicose para a primeira semana do estudo. Podemos notar que a dispersão entre os valores teórico e observado é bastante nítida, pois o padrão dos dados foge da reta vermelha a qual afirma um modelo probabilístico de normalidade para os dados. Portanto, existem evidências de que esta variável não segue um modelo normal.

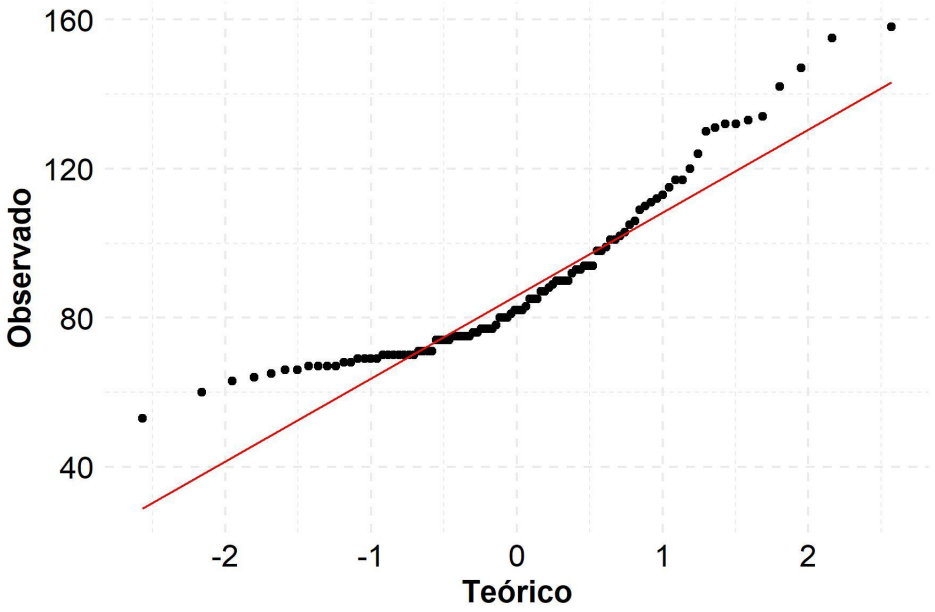
Figura 1: Q-Q plot para o nível de glicose na primeira semana.



Fonte: Construído pelo autor, 2022.

A **Figura 2** mostra o gráfico de Q-Q plot para os dados de glicose para a semana 2 do estudo. Através da imagem, percebe-se que existe uma grande distorção da linha que afirma uma padrão de normalidade, com destaque para as caudas do gráfico de dispersão. Por isso, existem poucas evidências de que esta variável segue um modelo normal de probabilidade.

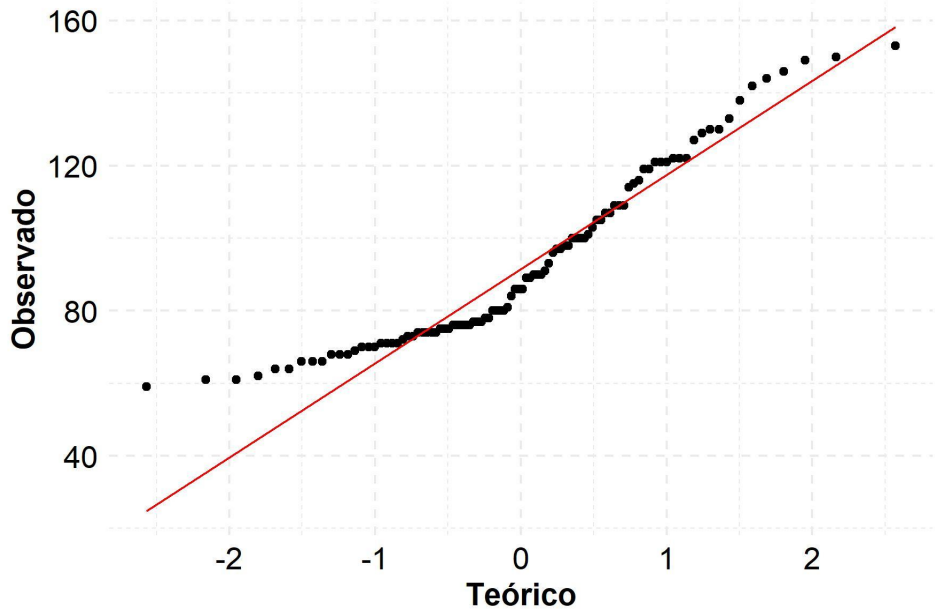
Figura 2: Q-Q plot para o nível de glicose na segunda semana.



Fonte: Construído pelo autor, 2022.

Na **Figura 3**, temos o gráfico Q-Q plot para a medição de glicose na terceira semana do estudo. Desse modo, percebe-se que a dispersão entre o valor observado e o valor teórico esperado para os dados não seguem a reta de tendência que poderia evidenciar indícios de normalidade nos dados desta variável. Desse modo, os níveis de glicose na terceira semana podem não apresentar um modelo normal de probabilidade.

Figura 3: Q-Q plot para o nível de glicose na terceira semana.

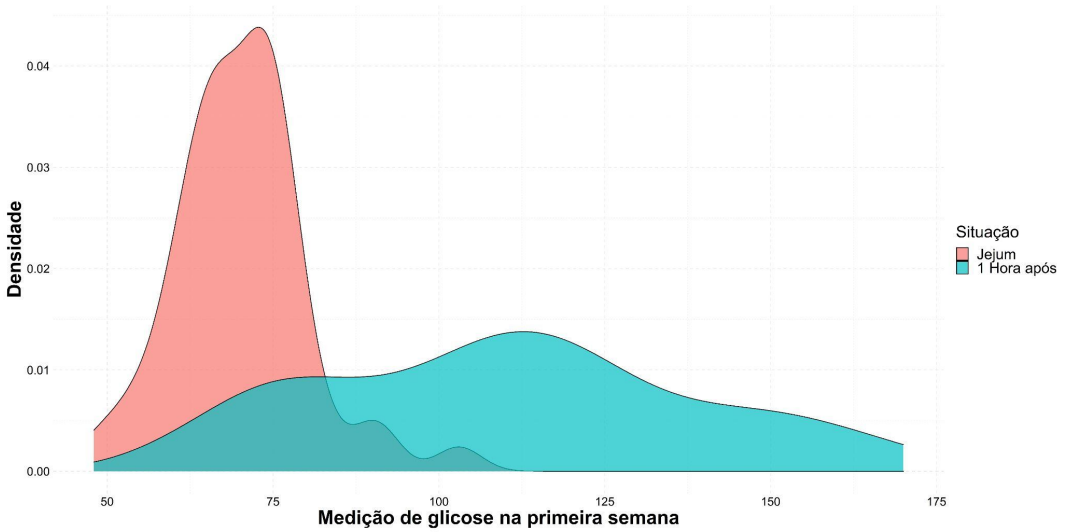


Fonte: Construído pelo autor, 2022.

3.2 Gráficos de Densidade

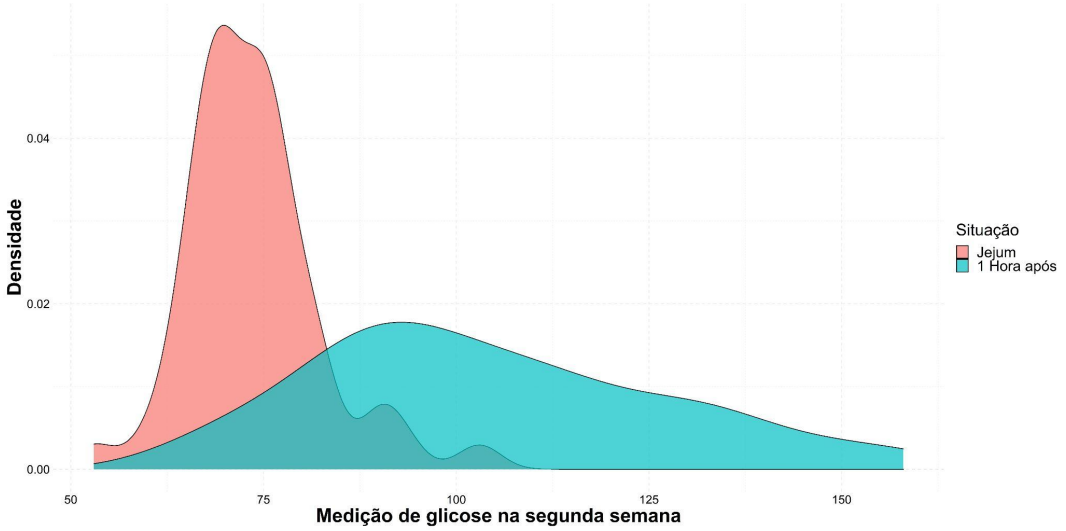
As **Figuras 4, 5** e **6** mostram o gráfico de densidade para as semanas 1, 2 e 3, respectivamente, do nível de glicose nas pacientes mulheres, estratificado pela situação. Portanto, pode-se concluir que, nas três situações, ocorre uma certa distinção entre os dois grupos, com o grupo mulheres em jejum apresentando valores bem concentrado em torno de 75 e o grupo de mulheres que ingeriram açúcar com valores bastantes dispersos, o que mostra uma alta variabilidade.

Figura 4: Gráfico de densidade para o nível de glicose na primeira semana, por momento.



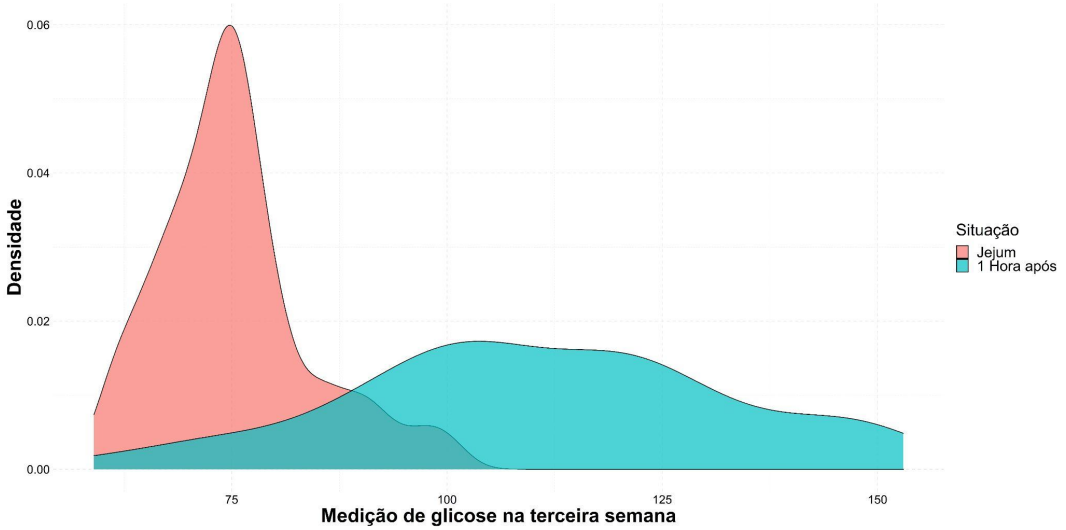
Fonte: Construído pelo autor, 2022.

Figura 5: Gráfico de densidade para o nível de glicose na segunda semana, por momento.



Fonte: Construído pelo autor, 2022.

Figura 6: Gráfico de densidade para o nível de glicose na terceira semana, por momento.

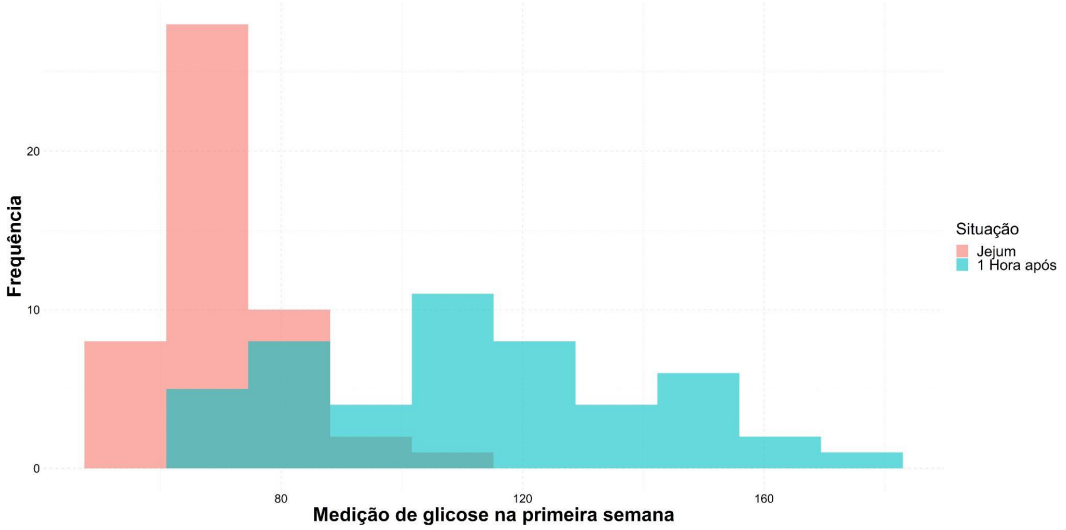


Fonte: Construído pelo autor, 2022.

3.3 Histogramas

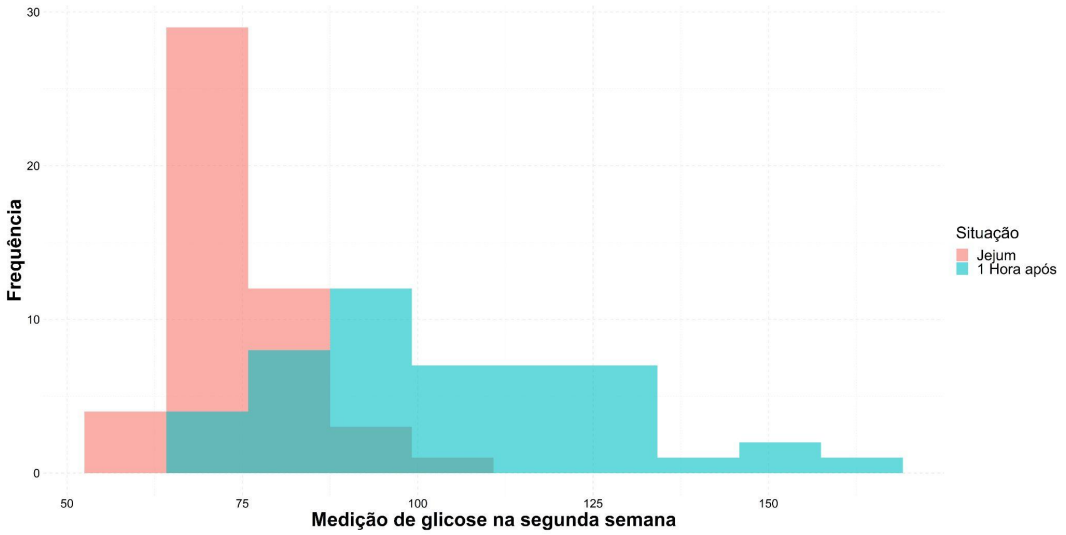
Os histogramas nas **Figuras 7, 8 e 9** mostram a frequência absoluta das medições do nível de glicose na primeira, segunda e terceira semana do estudo, estratificados pela situação das pacientes. Nos três casos, percebe-se que o grupo de mulheres em jejum apresenta níveis de glicose bem menores em relação ao grupo de mulheres que ingeriram açúcar uma hora após o experimento. Ademais, nota-se que os níveis de glicose das mulheres em jejum são bem mais concentrados do que as mulheres que ingeriram açúcar uma hora após o experimento.

Figura 7: Histograma para o nível de glicose na primeira semana, por momento.



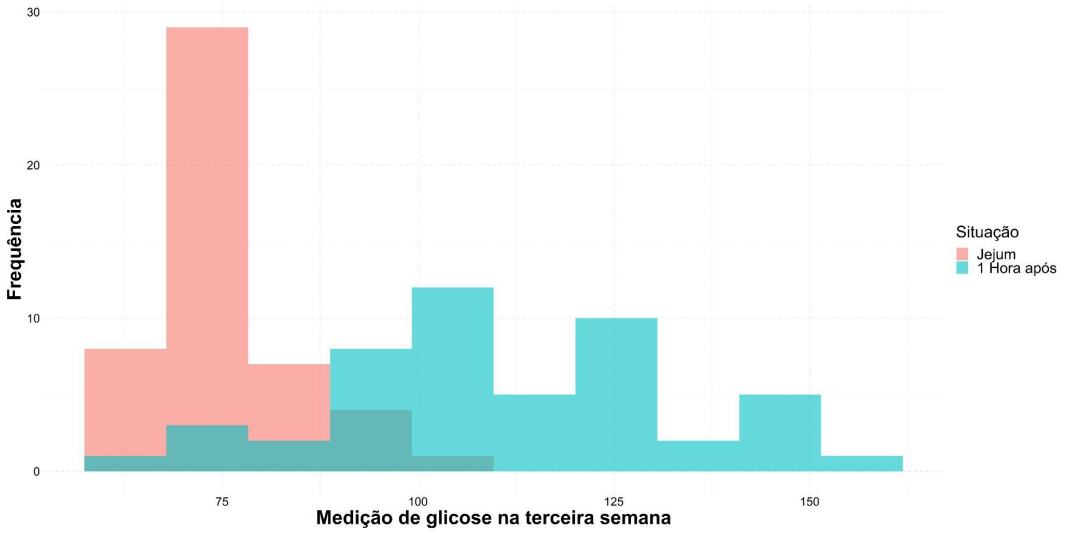
Fonte: Construído pelo autor, 2022.

Figura 8: Histograma para o nível de glicose na segunda semana, por momento.



Fonte: Construído pelo autor, 2022.

Figura 9: Histograma para o nível de glicose na terceira semana, por momento.

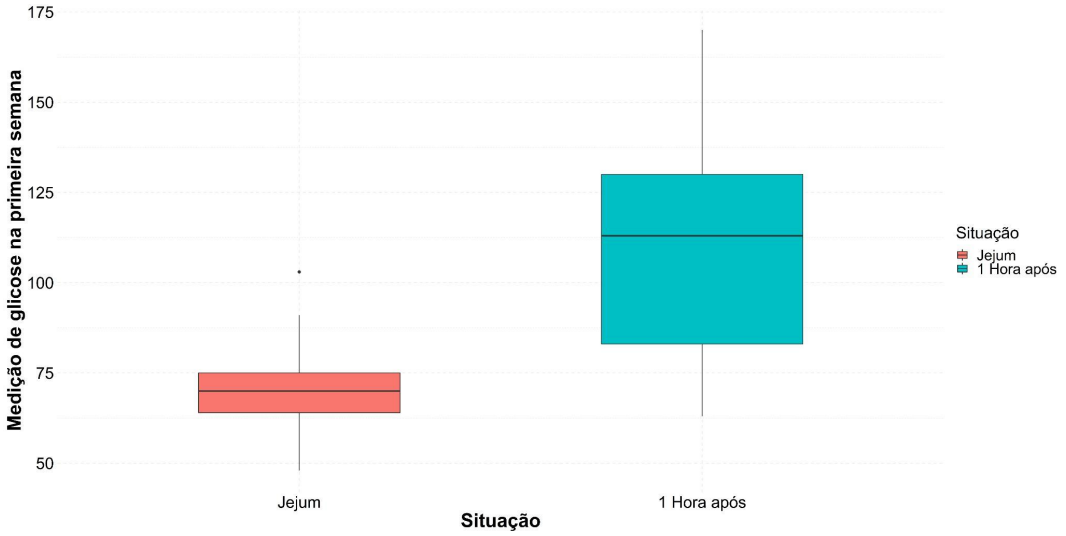


Fonte: Construído pelo autor, 2022.

3.4 Boxplots

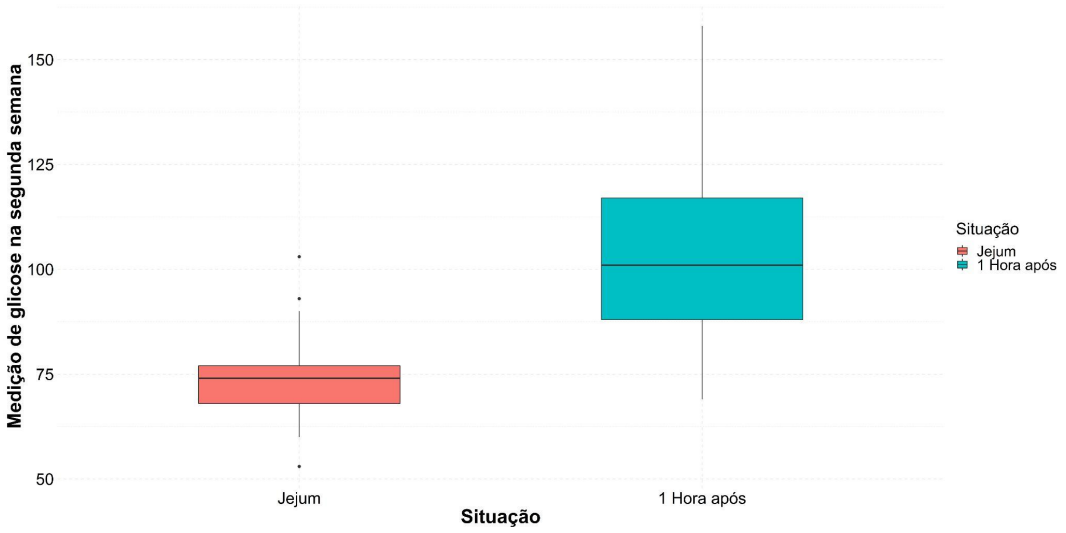
Nas **Figuras 10, 11 e 12** são abordados os diagramas de caixa para as três medições do nível de glicose (primeira semana, segunda semana e terceira semana), categorizado pela situação. Nos três casos, é perceptível que a variabilidade do grupo de ingeriu açúcar uma hora após é bem maior do que a do grupo que não ingeriu.

Figura 10: Boxplot para o nível de glicose na primeira semana, por momento.



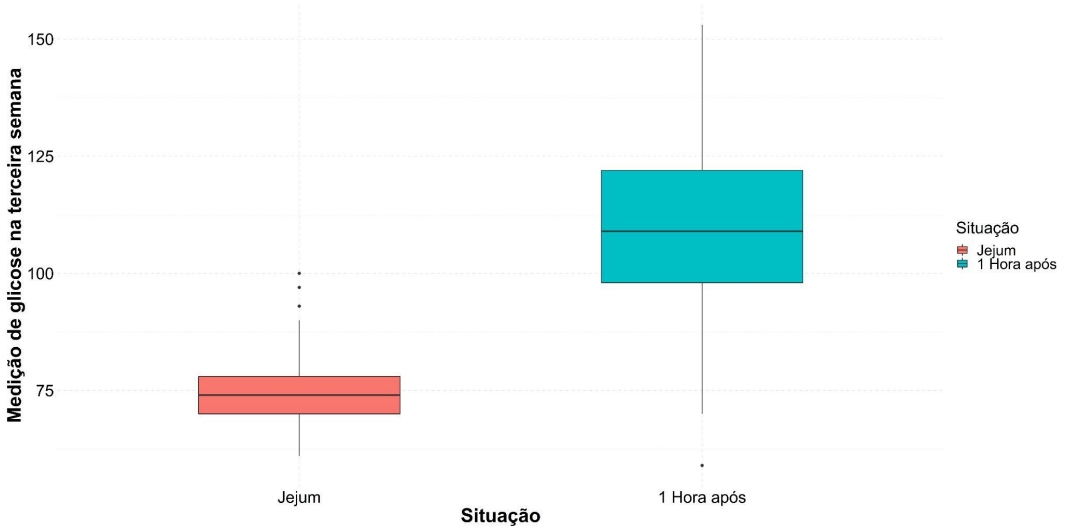
Fonte: Construído pelo autor, 2022.

Figura 11: Boxplot para o nível de glicose na segunda semana, por momento.



Fonte: Construído pelo autor, 2022.

Figura 12: Boxplot para o nível de glicose na terceira semana, por momento.



Fonte: Construído pelo autor, 2022.

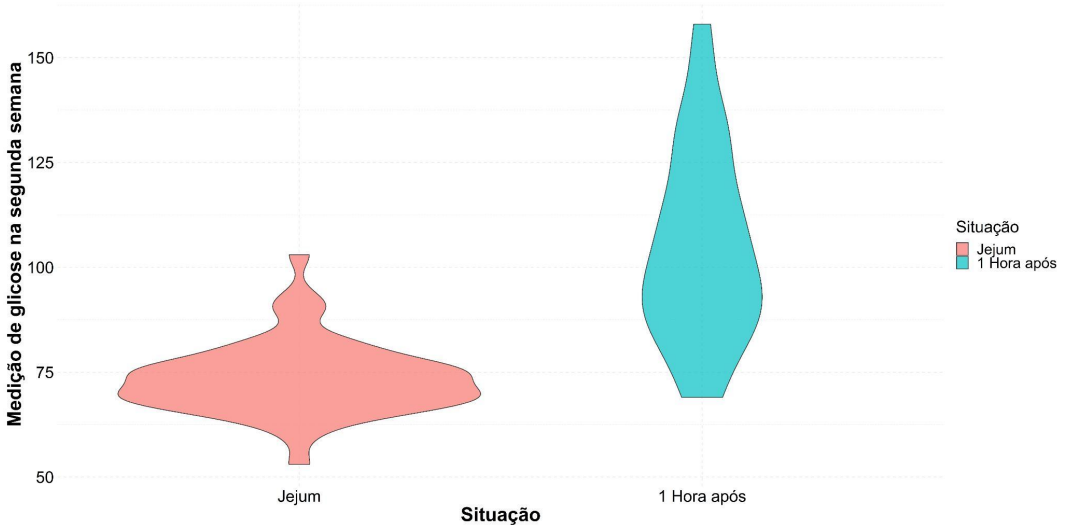
3.5 Gráfico de violino:

Uma outra forma de se avaliar a variabilidade dos dados é utilizando o gráfico de violino. Nesta visão, as **Figuras 13, 14 e 15** abordam a distribuição do nível de glicose para as semanas 1, 2 e 3, nesta respectiva ordem. Logo, é possível concluir que, para o grupo em jejum, as três imagens mostram que existem muitos valores próximos de 75 e, para o grupo e, para o grupo que ingeriu açúcar uma hora após, os valores estão bastante distribuídos, com concentração acima de 100 na primeira semana e concentração em torno de 80 na terceira semana.

Figura 13: Gráfico de violino para o nível de glicose na primeira semana, por momento.

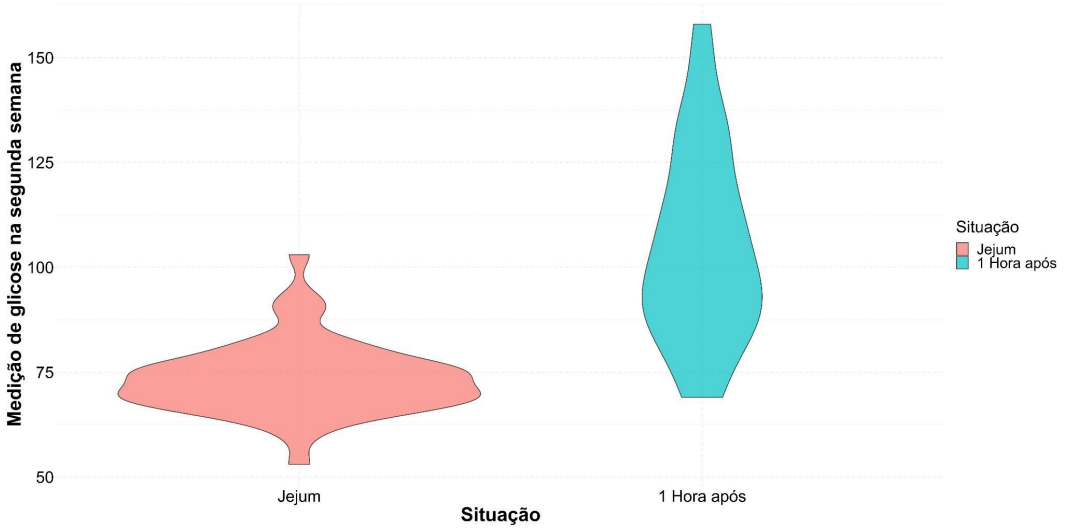
Fonte: Construído pelo autor, 2022.

Figura 14: Gráfico de violino para o nível de glicose na segunda semana, por momento.



Fonte: Construído pelo autor, 2022.

Figura 15: Gráfico de violino para o nível de glicose na terceira semana, por momento.



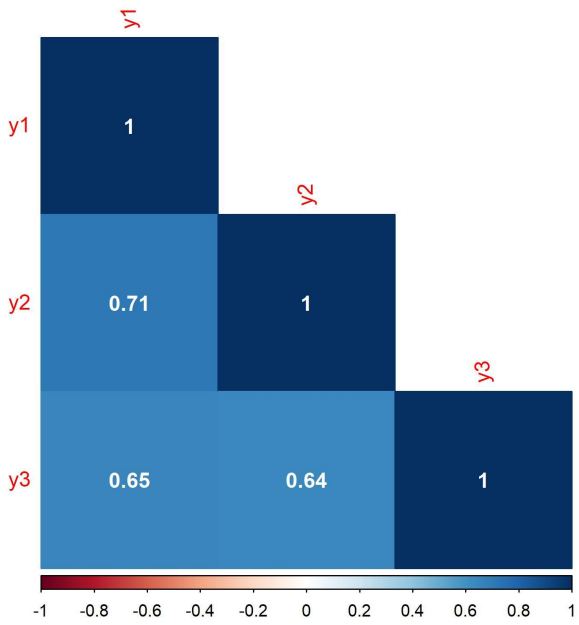
Fonte: Construído pelo autor, 2022.

4. Gráficos Multivariados

4.1 Gráfico de correlação

Por meio da matriz de correlações elaborada na **Seção 2.7**, foi criado um mapa de calor para se ter uma ideia mais clara do grau de intensidade entre as variáveis. Desse modo, na **Figura 16**, é evidente que todas as variáveis apresentaram uma correlação positiva, com destaque para o par (y1,y2), o qual representa a primeira e segunda semana do estudo, com um valor de 0,71 para o coeficiente de Pearson.

Figura 16: Mapa de calor para as correlações entre as variáveis para o nível de glicose.

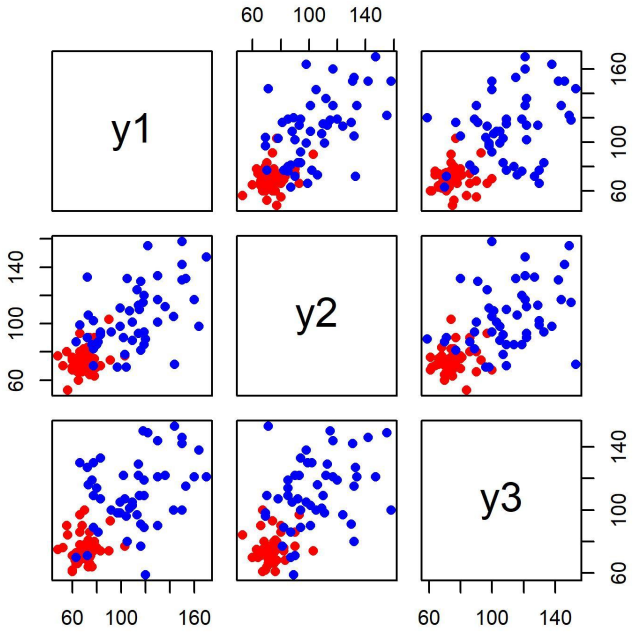


Fonte: Construído pelo autor, 2022.

4.2 Dispersão em pares

Na **Figura 17**, é abordado a dispersão entre os valores do nível de glicose entre as semanas do estudo, estratificado pelos grupos: jejum (vermelho) e ingestão de açúcar após uma hora (azul). Nesse viés, nota-se que para todas as semanas, no grupo das mulheres que ingeriam açúcar no experimento, ocorreu uma maior elevação no valor da variável em relação ao grupo de mulheres no jejum.

Figura 17: Dispersão entre as variáveis do nível de glicose, por par.

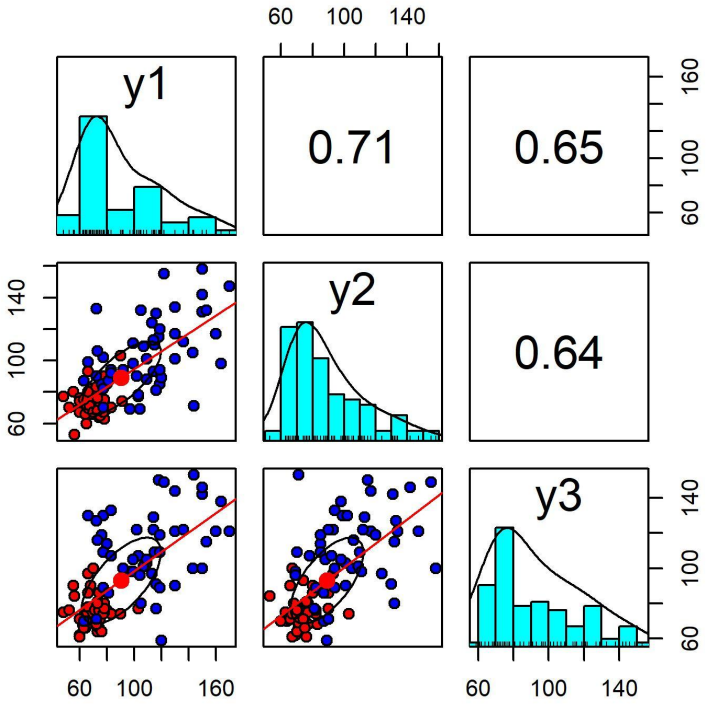


Fonte: Construído pelo autor, 2022.

4.3 Painéis, ajuste com elipse de confiança e histograma em pares.

Na **Figura 18**, é possível notar novamente a dispersão dos dados, só que agora ajustado por um modelo linear, dos valores para o coeficiente de pearson e dos histogramas individuais de cada variável. Dessa maneira, conclui-se que a dispersão entre os dados é elevada nos valores mais altos de glicose, o que mostra a inconsistência de ajustar um modelo linear em qualquer combinação entre pares do nível de glicose. Também, percebe-se que os níveis de glicose estão concentrados entre os valores de 60 e 80 nas três semanas.

Figura 18: Dispersão, ajuste, histogramas e correlações entre as variáveis do nível de glicose, por par.

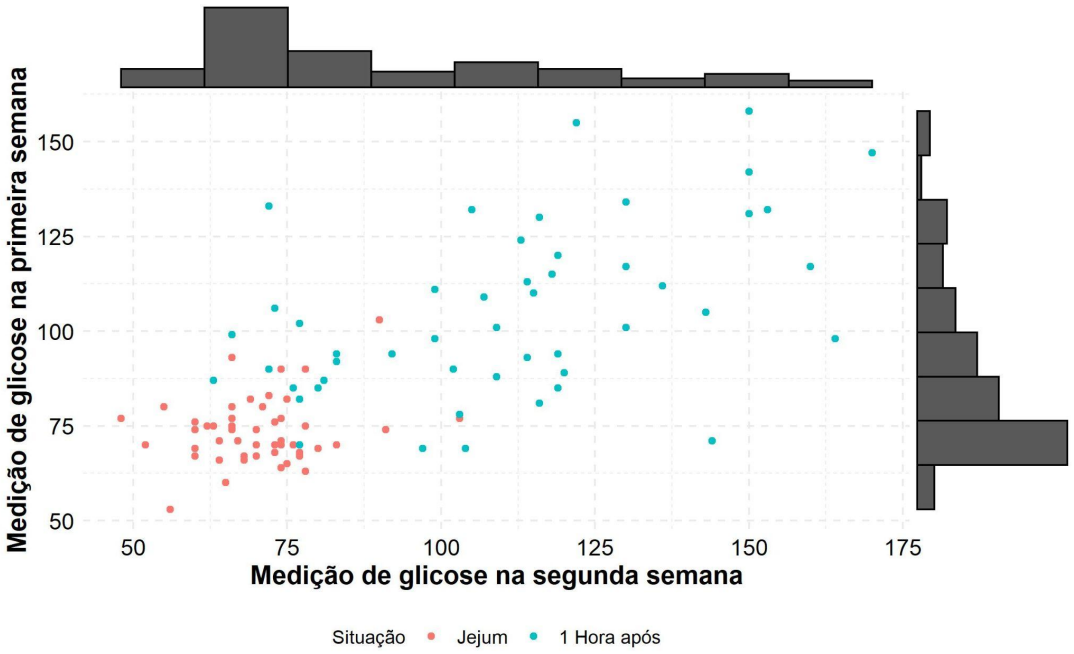


Fonte: Construído pelo autor, 2022.

4.4 Dispersão com histograma entre pares

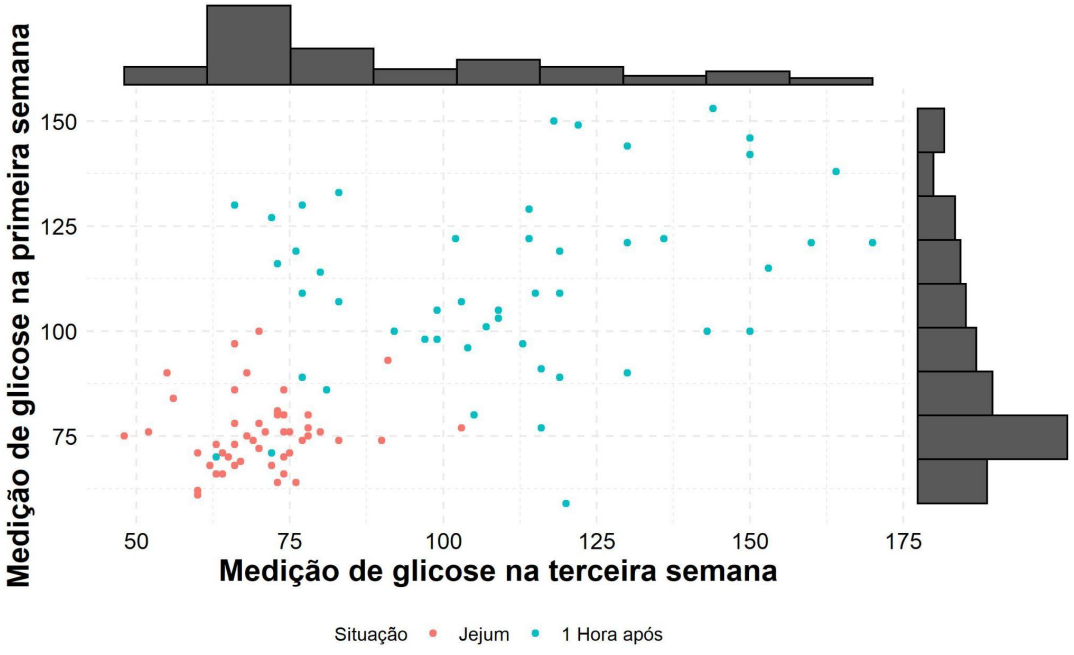
As **Figuras 19, 20 e 21** mostram a dispersão entre os pares do nível de glicose: primeira e segunda semana, primeira e terceira semana e segunda e terceira semana, respectivamente. Em todos os três gráficos, percebe-se uma concentração no início para o grupo de mulheres que estavam em jejum e uma alta dispersão para as mulheres que consumiram açúcar 1 hora depois. Em adição, os histogramas laterais nos três casos mostram uma alta concentração para o grupo em jejum em qualquer par de variáveis entre as semanas de medição.

Figura 19: Histogramas e dispersão entre as variáveis do nível de glicose para a primeira e segunda semana, por situação.



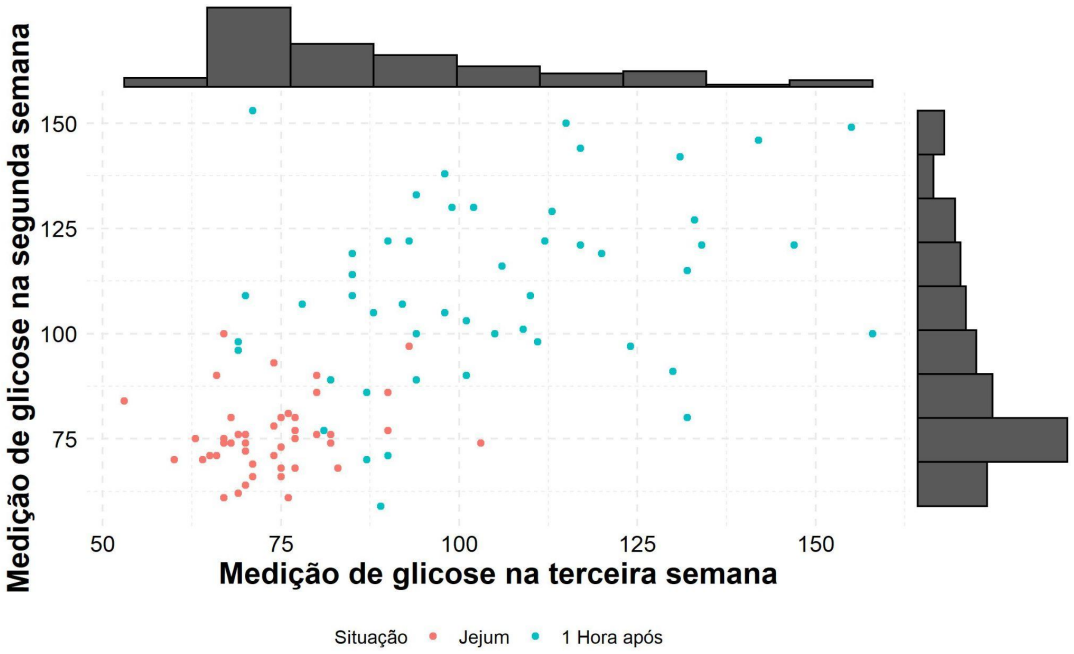
Fonte: Construído pelo autor, 2022.

Figura 20: Histogramas e dispersão entre as variáveis do nível de glicose para a primeira e terceira semana, por situação.



Fonte: Construído pelo autor, 2022.

Figura 21: Histogramas e dispersão entre as variáveis do nível de glicose para a segunda e terceira semana, por situação.



Fonte: Construído pelo autor, 2022.

5. Testes de Normalidade Bidimensional

O **Quadro 1** mostra o resumo para os testes de Normalidade Bidimensional para todos os pares da variável acerca do nível de glicose. Por meio do quadro abaixo, consegue-se perceber que nenhuma combinação de par para as variáveis do estudo conseguiu atender ao padrão de normalidade bidimensional em questão. Indícios sobre estes resultados já foram discutidos nas Figuras 1, 2 e 3, nas quais os gráficos de Q-Q plot indicavam um padrão fora da normalidade unidimensional para cada variável. Para maiores detalhes dos resultados exibidos no quadro abaixo, é recomendável conferir o **Anexo III** deste relatório.

Quadro 1: Conclusões dos testes de Normalidade Bidimensional, por par de variáveis.

Par	Royston	Mardia	Doornik-Hansen	Henze-Zirkler	Energy
y1 e y2	Não	Não	Não	Não	Não
y2 e y3	Não	Não	Não	Não	Não
y1 e y3	Não	Não	Não	Não	Não

Fonte: Construído pelo autor, 2022.

Anexo I: Script comentado e desenvolvido na linguagem R.

```
## ------
## Script da atividade 1 de multivariada
## ------

## ---
## Pacotes
## ---

if(!require(pacman)) install.packages("pacman")

p_load(readxl, dplyr, rstatix, corrplot, GGally,
      psych, PerformanceAnalytics, lattice, graphics, sm, gmodels,
      MVN, ggplot2, ggExtra,openxlsx)

## ---
## Banco de dados
## ---

setwd("D:/")

dir()

## ---
## leitura dos dados
## ---

## dados com os cabecalhos e as duas medicoes
data <- read.xlsx(xlsxFile = "dados_long_wide.xlsx",
                  fillMergedCells = TRUE, colNames = T)

## tamanho dos dados com as medicoes
n = dim(data)[1]
head(data)  # visualizando o cabecalho
tail(data)  # visualizando as ultimas linhas
summary(dados) # estatisticas descritivas

jejum = data[-1,1:3] ## tirando a primeira linha com o nome das variaveis (y1, y2, e y3)
apos = data[-1,4:6]  ## tirando a primeira linha com o nome das variaveis (x1, x2, e x3)

## colocando o mesmo cabecalho para juntar as linhas
colnames(jejum) = c("y1", "y2", "y3")
colnames(apos) = c("y1", "y2", "y3")

## colocando os labels para identificar as medicoes
## 0: no jejum
## 1: apos 1 hora

jejum["ingestao"] = rep(0,n - 1) ## retirando uma linha pois a primeira, pois identifica
```

```
apos["ingestao"] = rep(1, n - 1) ## as variaveis

## juntando as linhas
dados = rbind(jejum, apos)
head(dados)

## ajustando o index
rownames(dados) = 1:(dim(dados)[1])


dados

## Transformando os dados
dados$y1 = as.numeric(dados$y1)
dados$y2 = as.numeric(dados$y2)
dados$y3 = as.numeric(dados$y3)
dados$ingestao = as.factor(dados$ingestao)

## ---
## Analise descritiva
## ---

colMeans(dados[,1:3]) # vetor de medias
cov(dados[,1:3])      # matriz de covariancias
cor(dados[,1:3])      # matriz de correlacoes

## Vetores de medias por ingestao
dados %>% group_by(ingestao) %>%
  summarise(n = n(), mean_y1 = mean(y1),
            mean_y2 = mean(y2), mean_y3 = mean(y3))

## Matrizes de covariancias por ingestao

dados %>% group_by(ingestao) %>%
  do(data.frame(cov = t(cov(.[,1:3]))))

## Matrizes de correlacoes por ingestao
dados %>% group_by((ingestao)) %>%
  do(data.frame(cov = t(cor(.[,1:3]))))

## ---
## Construcao de graficos
## ---

# grafico qqplot
ggplot(data = dados, aes(sample= y1)) +
  stat_qq() +
  stat_qq_line() +
  xlab("Teórico") +
  ylab("Observado") +
```

```
  theme_minimal()

ggplot(data = dados, aes(sample= y2)) +
  stat_qq() +
  stat_qq_line() +
  xlab("Teórico") +
  ylab("Observado") +
  theme_minimal()

ggplot(data = dados, aes(sample= y3)) +
  stat_qq() +
  stat_qq_line() +
  xlab("Teórico") +
  ylab("Observado") +
  theme_minimal()

## grafico de correlacoes

corrplot(cor(dados[,1:3]), method = 'color',type = 'lower',
         cl.pos = 'b', addCoef.col = 'white')

## graficos de dispersao multivariados
pairs(dados[,1:3], pch = 16, col = c('red', 'blue')[dados$ingestao])

ggpairs(dados[,1:3],
        ggplot2::aes(color = factor(dados$ingestao), alpha = 0.7)) +
  theme_minimal()

pairs.panels(dados[,1:3],
             pch = 21,
             bg = c("red","blue")[dados$ingestao],
             lm = T)


chart.Correlation(dados[,1:3]) ## mlgs, histogramas e correlacoes
chart.Boxplot(dados[,1:3])

## grafico de estrela
stars(dados[,1:3],key.loc = c(11,2))



## boxplots
boxplot(dados[,1:3])

boxplot(dados$y1~factor(dados$ingestao))
boxplot(dados$y2~factor(dados$ingestao))
boxplot(dados$y3~factor(dados$ingestao))

## y1: primeira semana
```

```
ggplot() + aes(dados,
            x = dados$ingestao,
            y = dados$y1,
            fill = dados$ingestao)  +
  geom_boxplot(width = 0.5) +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_x_discrete("Situação", labels = c("Jejum", "1 Hora após")) +
  scale_y_continuous("Medição de glicose na primeira semana") +
  theme_minimal()

## y2: segunda semana
ggplot() + aes(dados,
            x = dados$ingestao,
            y = dados$y2,
            fill = dados$ingestao)  +
  geom_boxplot(width = 0.5) +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_x_discrete("Situação", labels = c("Jejum", "1 Hora após")) +
  scale_y_continuous("Medição de glicose na segunda semana") +
  theme_minimal()

## y3: terceira semana
ggplot() + aes(dados,
            x = dados$ingestao,
            y = dados$y3,
            fill = dados$ingestao)  +
  geom_boxplot(width = 0.5) +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_x_discrete("Situação", labels = c("Jejum", "1 Hora após")) +
  scale_y_continuous("Medição de glicose na terceira semana") +
  theme_minimal()

## grafico de violino

# primeira semana
ggplot() + aes(y = dados$y1,
            x = dados$ingestao,
            fill = dados$ingestao) +
  geom_violin(alpha = 0.7) +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_y_continuous("Medição de glicose na primeira semana") +
  scale_x_discrete("Situação", labels = c("Jejum", "1 Hora após")) +
  theme_minimal()

# segunda semana
ggplot() + aes(y = dados$y2,
```

```
            x = dados$ingestao,
            fill = dados$ingestao) +
  geom_violin(alpha = 0.7) +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_y_continuous("Medição de glicose na segunda semana") +
  scale_x_discrete("Situação", labels = c("Jejum", "1 Hora após")) +
  theme_minimal()

# terceira semana
ggplot() + aes(y = dados$y3,
            x = dados$ingestao,
            fill = dados$ingestao) +
  geom_violin(alpha = 0.7) +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_y_continuous("Medição de glicose na terceira semana") +
  scale_x_discrete("Situação", labels = c("Jejum", "1 Hora após")) +
  theme_minimal()

## grafico de densidade

## y1: primeira semana

dados %>% ggplot() + aes(x = y1, fill = ingestao) +
  geom_density(alpha =0.7) +
  scale_x_continuous("Medição de glicose na primeira semana") +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_y_continuous("Densidade") +
  theme_minimal()

## y2: segunda semana
dados %>% ggplot() + aes(x = y2, fill = ingestao) +
  geom_density(alpha =0.7) +
  scale_x_continuous("Medição de glicose na segunda semana") +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_y_continuous("Densidade") +
  theme_minimal()


## y3: terceira semana
dados %>% ggplot() + aes(x = y3, fill = ingestao) +
  geom_density(alpha =0.7) +
  scale_x_continuous("Medição de glicose na terceira semana") +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_y_continuous("Densidade") +
  theme_minimal()
```

```
## dipersao com densidades marginais

# primeira e segunda semana
p = ggplot() + aes(dados, x = dados$y1,
            y = dados$y2,
            colour = dados$ingestao)

p = p + geom_point(size = 1.5)  +
  scale_color_discrete(name = "Situação",
                       labels = c("Jejum", '1 Hora após')) +
  scale_y_continuous("Medição de glicose na primeira semana") +
  scale_x_continuous("Medição de glicose na segunda semana") +
  theme(legend.position = 'bottom')


ggMarginal(p, type = 'histogram')
#ggMarginal(p , type = 'density')
#ggMarginal(p , type = 'boxplot')

# primeira e terceira semana
p = ggplot() + aes(dados, x = dados$y1,
                   y = dados$y3,
                   colour = dados$ingestao)

p = p + geom_point(size = 1.5)  +
  scale_color_discrete(name = "Situação",
                       labels = c("Jejum", '1 Hora após')) +
  scale_y_continuous("Medição de glicose na primeira semana") +
  scale_x_continuous("Medição de glicose na terceira semana") +
  theme(legend.position = 'bottom')
ggMarginal(p, type = 'histogram')


# segunda e terceira semana
p = ggplot() + aes(dados, x = dados$y2,
                   y = dados$y3,
                   colour = dados$ingestao)

p = p + geom_point(size = 1.5)  +
  scale_color_discrete(name = "Situação",
                       labels = c("Jejum", '1 Hora após')) +
  scale_y_continuous("Medição de glicose na segunda semana") +
  scale_x_continuous("Medição de glicose na terceira semana") +
  theme(legend.position = 'bottom')
ggMarginal(p, type = 'histogram')


GGally::ggpairs(dados, columns = 1:4,
```

```
            aes(colour = ingestao))


# histograma
dados %>%
  ggplot(aes(x = y1, fill = ingestao)) +
  geom_histogram(position = 'identity', bins = 10,
                 alpha = 0.6) +
  theme_minimal() +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                                 name = "Situação") +
  scale_x_continuous("Medição de glicose na primeira semana") +
  scale_y_continuous("Frequência")

dados %>%
  ggplot(aes(x = y2, fill = ingestao)) +
  geom_histogram(position = 'identity', bins = 10,
                 alpha = 0.6) +
  theme_minimal() +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_x_continuous("Medição de glicose na segunda semana") +
  scale_y_continuous("Frequência")

dados %>%
  ggplot(aes(x = y3, fill = ingestao)) +
  geom_histogram(position = 'identity', bins = 10,
                 alpha = 0.6) +
  theme_minimal() +
  scale_fill_discrete(labels = c("Jejum", '1 Hora após'),
                      name = "Situação") +
  scale_x_continuous("Medição de glicose na terceira semana") +
  scale_y_continuous("Frequência")




## ---
## Teste de normalidade bidimensional
## ---

colnames(dados)

## para y1 e y2

mvn(cbind(dados$y1, dados$y2), mvnTest = 'royston')# nao

mvn(cbind(dados$y1, dados$y2), mvnTest = 'mardia') # nao
```

```
mvn(cbind(dados$y1, dados$y2), mvnTest = 'dh')     # nao

mvn(cbind(dados$y1, dados$y2), mvnTest = 'hz')     # nao

mvn(cbind(dados$y1, dados$y2), mvnTest = 'energy') # nao

## para y2 e y3

mvn(cbind(dados$y3, dados$y2), mvnTest = 'royston')# nao

mvn(cbind(dados$y3, dados$y2), mvnTest = 'mardia') # nao

mvn(cbind(dados$y3, dados$y2), mvnTest = 'dh')     # nao

mvn(cbind(dados$y3, dados$y2), mvnTest = 'hz')     # nao

mvn(cbind(dados$y3, dados$y2), mvnTest = 'energy') # nao

## para y1 e y3

mvn(cbind(dados$y3, dados$y1), mvnTest = 'royston')# nao

mvn(cbind(dados$y3, dados$y1), mvnTest = 'mardia') # nao

mvn(cbind(dados$y3, dados$y1), mvnTest = 'dh')     # nao

mvn(cbind(dados$y3, dados$y1), mvnTest = 'hz')     # nao

mvn(cbind(dados$y3, dados$y1), mvnTest = 'energy') # nao
```

Anexo II: Tabela no formato *long* desenvolvida dentro do script.

Tabela 2: Tabela no formato *long* após o pré-processamento dos dados.

y1	y2	y3	ingestao
60	69	62	0
56	53	84	0
80	69	76	0
55	80	90	0
62	75	68	0
74	64	70	0
64	71	66	0
73	70	64	0
68	67	75	0
69	82	74	0
60	67	61	0
70	74	78	0
66	74	78	0
83	70	74	0
68	66	90	0
78	63	75	0
103	77	77	0
77	68	74	0
66	77	68	0
70	70	72	0
75	65	71	0
91	74	93	0
66	75	73	0
75	82	76	0
74	71	66	0
76	70	64	0
74	90	86	0
74	77	80	0
67	71	69	0
78	75	80	0
64	66	71	0
71	80	76	0
63	75	73	0
90	103	74	0
60	76	61	0
48	77	75	0
66	93	97	0
74	70	76	0
60	74	71	0

63	75	66	0
66	80	86	0
77	67	74	0
70	67	100	0
73	76	81	0
78	90	77	0
73	68	80	0
72	83	68	0
65	60	70	0
52	70	76	0
97	69	98	1
103	78	107	1
66	99	130	1
80	85	114	1
116	130	91	1
109	101	103	1
77	102	130	1
115	110	109	1
76	85	119	1
72	133	127	1
130	134	121	1
150	158	100	1
150	131	142	1
99	98	105	1
119	85	109	1
164	98	138	1
160	117	121	1
144	71	153	1
77	82	89	1
114	93	122	1
77	70	109	1
118	115	150	1
170	147	121	1
153	132	115	1
143	105	100	1
114	113	129	1
73	106	116	1
116	81	77	1
63	87	70	1
105	132	80	1
83	94	133	1
81	87	86	1
120	89	59	1
107	109	101	1

99	111	98	1
113	124	97	1
136	112	122	1
109	88	105	1
72	90	71	1
130	101	90	1
130	117	144	1
83	92	107	1
150	142	146	1
119	120	119	1
122	155	149	1
102	90	122	1
104	69	96	1
119	94	89	1
92	94	100	1

Fonte: Construído pelo autor, 2022.

Anexo III: Saída da linguagem R dos testes de Normalidade Multivariada para as variáveis.

Para y1 e y2:

Royston:

```
$multivariateNormality
    Test       H      p value MVN
1 Royston 47.23727 4.880246e-11  NO

$univariateNormality
          Test  Variable Statistic   p value Normality
1 Anderson-Darling  Column1    4.1524  <0.001      NO
2 Anderson-Darling  Column2    3.5698  <0.001      NO

$Descriptives
   n     Mean  Std.Dev Median Min Max  25th 75th      Skew   Kurtosis
1 98 90.38776 29.11712     77  48 170 69.25  112 0.9255969 -0.1367493
2 98 89.09184 23.06606     82  53 158 71.00  101 1.0656981  0.4101178
```

Mardia:

```
$multivariateNormality
          Test       Statistic           p value Result
```

1 Mardia Skewness 32.2947605633935 1.66525070381227e-06 NO

2 Mardia Kurtosis 2.72691029091733 0.00639304174060085 NO

3 MVN <NA> <NA> NO

$univariateNormality

Test Variable Statistic p value Normality

1 Anderson-Darling Column1 4.1524 <0.001 NO

2 Anderson-Darling Column2 3.5698 <0.001 NO

$Descriptives

n Mean Std.Dev Median Min Max 25th 75th Skew Kurtosis

1 98 90.38776 29.11712 77 48 170 69.25 112 0.9255969 -0.1367493

2 98 89.09184 23.06606 82 53 158 71.00 101 1.0656981 0.4101178

Doornik-Hansen:

$multivariateNormality

Test E df p value MVN

1 Doornik-Hansen 40.95695 4 2.743557e-08 NO

$univariateNormality

Test Variable Statistic p value Normality

1 Anderson-Darling Column1 4.1524 <0.001 NO

2 Anderson-Darling Column2 3.5698 <0.001 NO

$Descriptives

n Mean Std.Dev Median Min Max 25th 75th Skew Kurtosis

1 98 90.38776 29.11712 77 48 170 69.25 112 0.9255969 -0.1367493

2 98 89.09184 23.06606 82 53 158 71.00 101 1.0656981 0.4101178

Henze-Zirkler:

$multivariateNormality

Test HZ p value MVN

1 Henze-Zirkler 4.472083 7.210652e-10 NO

$univariateNormality

Test Variable Statistic p value Normality

1 Anderson-Darling Column1 4.1524 <0.001 NO

2 Anderson-Darling Column2 3.5698 <0.001 NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	90.38776	29.11712	77	48	170	69.25	112	0.9255969	-0.1367493
2	98	89.09184	23.06606	82	53	158	71.00	101	1.0656981	0.4101178

Energy:

$multivariateNormality

	Test	Statistic	p value	MVN
1	E-statistic	4.077517	0	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	4.1524	<0.001	NO
2	Anderson-Darling	Column2	3.5698	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	90.38776	29.11712	77	48	170	69.25	112	0.9255969	-0.1367493
2	98	89.09184	23.06606	82	53	158	71.00	101	1.0656981	0.4101178

Para y2 e y3:

Royston:

$multivariateNormality

	Test	H	p value	MVN
1	Royston	43.01494	4.484591e-10	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	3.5698	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74	109	0.7203791	-0.5426254
2	98	89.09184	23.06606	82	53	158	71	101	1.0656981	0.4101178

Mardia:

$multivariateNormality

	Test	Statistic	p value	Result
1	Mardia Skewness	33.031734814583	1.17674285535925e-06	NO
2	Mardia Kurtosis	2.90896433176704	0.00362628229749751	NO
3	MVN	<NA>	<NA>	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	3.5698	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74	109	0.7203791	-0.5426254
2	98	89.09184	23.06606	82	53	158	71	101	1.0656981	0.4101178

Doornik-Hansen:

$multivariateNormality

	Test	E	df	p value	MVN
1	Doornik-Hansen	44.93437	4	4.10286e-09	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	3.5698	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74	109	0.7203791	-0.5426254
2	98	89.09184	23.06606	82	53	158	71	101	1.0656981	0.4101178

Henze-Zirkler:

$multivariateNormality

	Test	HZ	p value	MVN
1	Henze-Zirkler	4.790731	2.076315e-10	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	3.5698	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74	109	0.7203791	-0.5426254
2	98	89.09184	23.06606	82	53	158	71	101	1.0656981	0.4101178

Energy:

$multivariateNormality

	Test	Statistic	p value	MVN
1	E-statistic	4.210457	0	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	3.5698	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74	109	0.7203791	-0.5426254
2	98	89.09184	23.06606	82	53	158	71	101	1.0656981	0.4101178

Para y1 e y3:

Royston:

$multivariateNormality

	Test	H	p value	MVN
1	Royston	43.86382	2.905399e-10	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	4.1524	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74.00	109	0.7203791	-0.5426254
2	98	90.38776	29.11712	77	48	170	69.25	112	0.9255969	-0.1367493

Mardia:

$multivariateNormality

	Test	Statistic	p value	Result
1	Mardia Skewness	24.7101707458878	5.75309131964422e-05	NO
2	Mardia Kurtosis	0.450986870964495	0.651999011490641	YES
3	MVN	<NA>	<NA>	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	4.1524	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74.00	109	0.7203791	-0.5426254
2	98	90.38776	29.11712	77	48	170	69.25	112	0.9255969	-0.1367493

Doornik-Hansen:

$multivariateNormality

	Test	E	df	p value	MVN
1	Doornik-Hansen	45.56368	4	3.035416e-09	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	4.1524	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74.00	109	0.7203791	-0.5426254
2	98	90.38776	29.11712	77	48	170	69.25	112	0.9255969	-0.1367493

Henze-Zirkler:

$multivariateNormality

	Test	HZ	p value	MVN
1	Henze-Zirkler	5.279394	3.358391e-11	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	4.1524	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74.00	109	0.7203791	-0.5426254
2	98	90.38776	29.11712	77	48	170	69.25	112	0.9255969	-0.1367493

Energy:

$multivariateNormality

	Test	Statistic	p value	MVN
1	E-statistic	4.575392	0	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Anderson-Darling	Column1	2.7831	<0.001	NO
2	Anderson-Darling	Column2	4.1524	<0.001	NO

$Descriptives

	n	Mean	Std.Dev	Median	Min	Max	25th	75th	Skew	Kurtosis
1	98	93.01020	24.45087	86	59	153	74.00	109	0.7203791	-0.5426254
2	98	90.38776	29.11712	77	48	170	69.25	112	0.9255969	-0.1367493