

Fatores de Risco para Alzheimer

Análise Descritiva do Banco de Dados

O banco é composto por dados de pacientes com e sem a doença de Alzheimer. A pesquisa aqui se objetiva a buscar compreender os fatores que podem aumentar o risco de desenvolvimento da doença. Para isso, começamos analisando descritivamente o banco de dados e suas variáveis de 1) idade, 2) gênero, 3) IMC, 4) consumo de álcool, 5) atividade física e 6) qualidade de sono. Em seguida avaliamos variáveis clínicas envolvendo doenças pré existentes dos pacientes.

Idades

A começar pela idade, é possível ver na Figura 1 e 2, temos uma distribuição simétrica das idades dos pacientes, tendo sua mediana em 75 anos de idade, sem outliers e com menor variabilidade. Seu coeficiente de variação é de 12%, indicando baixa dispersão das idades dos pacientes.

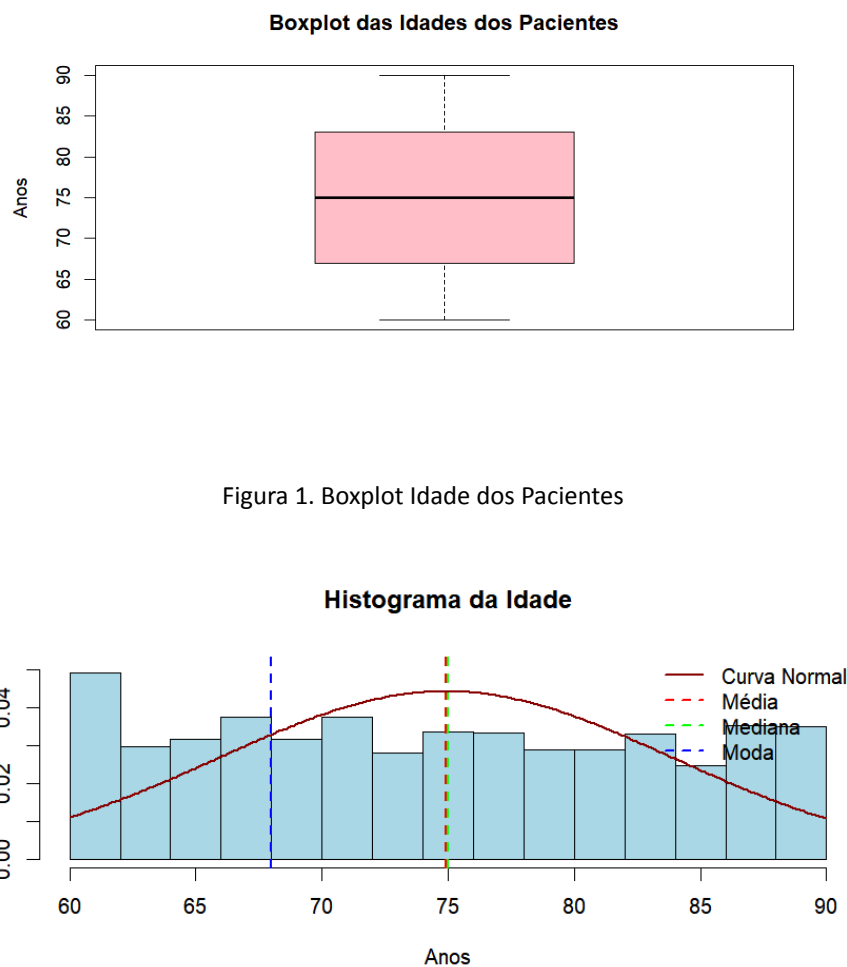


Figura 1. Boxplot Idade dos Pacientes

Figura 2. Histograma Idade dos Pacientes

Gênero

Como os dados a respeito de gênero são binários, utilizamos uma análise por Bernoulli para identificar qual era a proporção da dispersão. Tendo 1061 pacientes identificados como 0 - masculino e 1088 pacientes identificados como 1 - feminino. Concluímos, então, que a maioria dos pacientes é de mulheres, porém com pouca diferença representativa no total da população analisada.

Resultados 0	Resultados 1
1061	1088
Gênero masculino	Gênero feminino

IMC

A respeito do IMC, é possível observar o boxplot da Figura 3 e 4. Os dados apresentam pouca variabilidade e dispersão, sem conter outliers. Ao comparar a mediana e a média, observamos uma leve assimetria à esquerda. Ainda sobre a mediana, obteve-se um valor de 27,82 kg/m², o que coloca a maioria dos pacientes em sobrepeso, com um coeficiente de variação de 26,1%, indicando dispersão moderada do IMC dos pacientes.

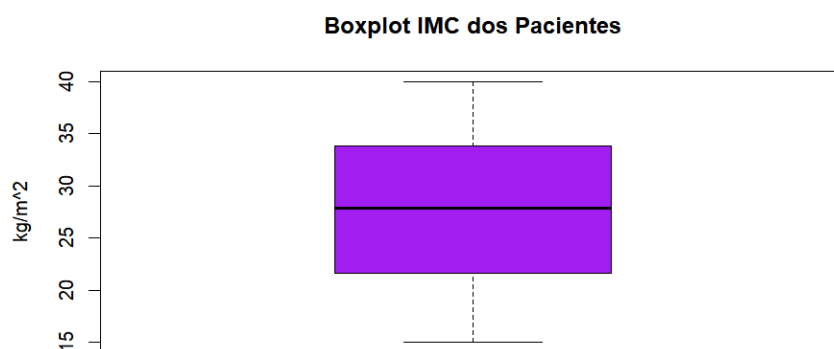


Figura 3. Boxplot IMC dos Pacientes

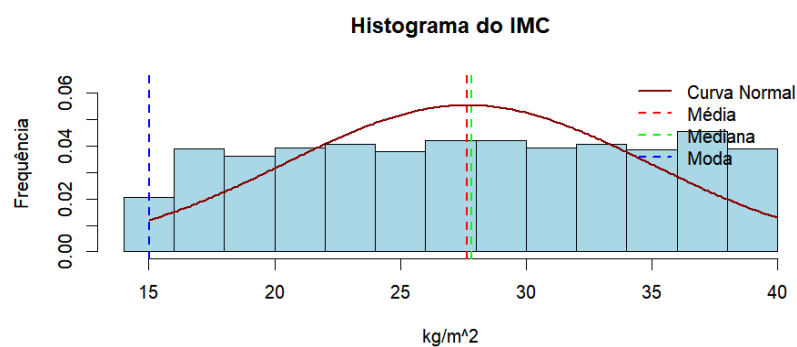


Figura 4. Histograma IMC dos Pacientes

Consumo de álcool

Sobre o consumo de álcool, identificou-se o boxplot na Figura 5 e 6. Identificou-se, também, uma leve assimetria à direita já que a mediana é menor do que a média, um coeficiente de variação alto de 57,83%, representando alta dispersão na quantidade de álcool consumido pelos pacientes. A mediana do consumo de álcool dos pacientes é de 10 unidades semanais.

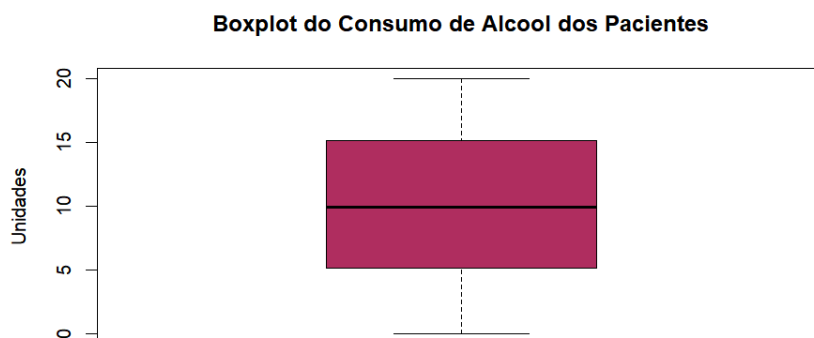


Figura 5. Boxplot Consumo de Álcool dos Pacientes

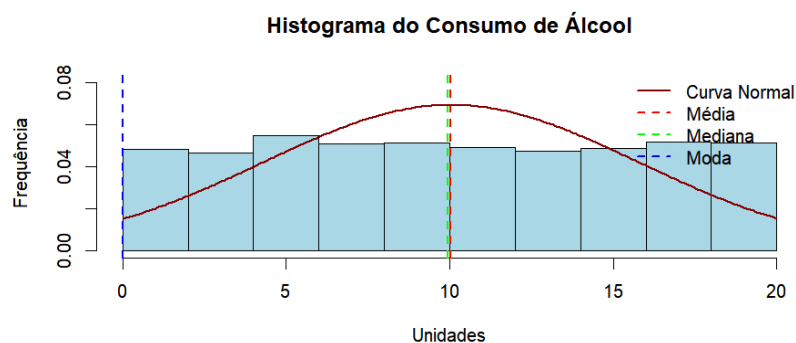


Figura 6. Histograma Consumo de Álcool dos Pacientes

Atividade Física

A respeito da quantidade de atividade física praticada, a Figura 7 e 8 apresenta o boxplot da medida. O boxplot mostra que a amostra apresenta uma leve assimetria à direita, tendo mediana correspondente à 4,7 horas semanais de atividade física com coeficiente de variação de 57,35%, o que sugere grande dispersão na quantidade de exercício entre os pacientes.

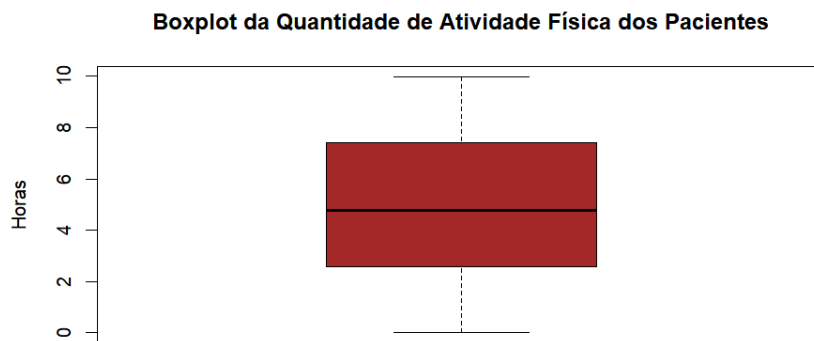


Figura 7. Boxplot Atividade Física dos Pacientes

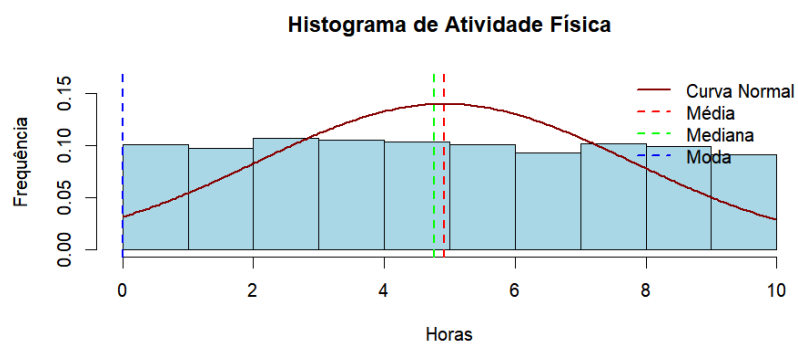


Figura 8. Histograma Atividade Física dos Pacientes

Qualidade do Sono

Por fim, a respeito da qualidade de sono dos pacientes, é possível identificar na Figura 9 e 10 uma simetria no score de sono dos pacientes observados com mediana de 7 pontos de sono diário com coeficiente de variação de 25%, o que indica uma dispersão moderada nos pontos de sono dos pacientes.

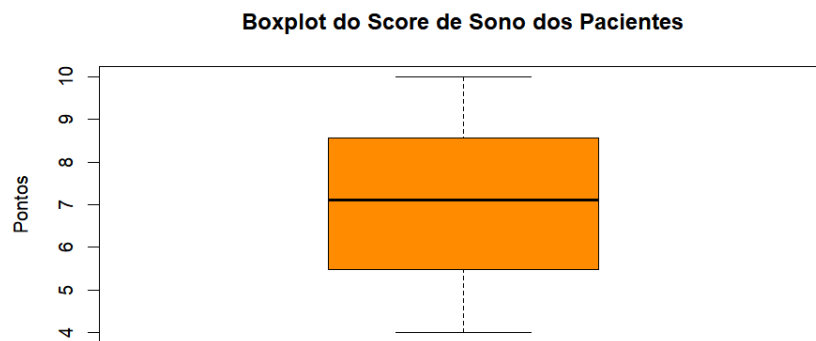


Figura 9. Boxplot Qualidade do Sono dos Pacientes

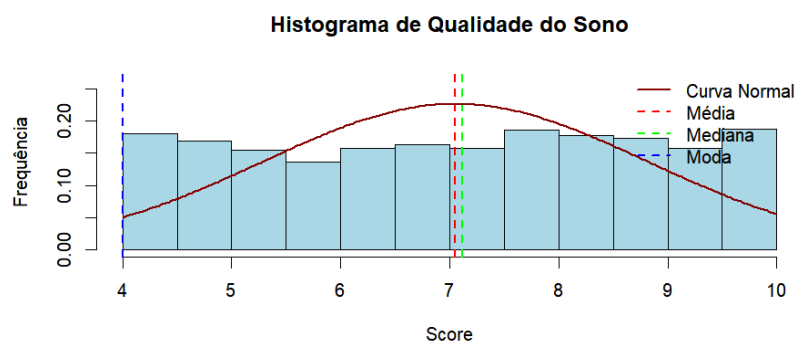


Figura 10. Histograma Qualidade do Sono dos Pacientes

Diagnóstico

Por fim, estudando o diagnóstico através da dispersão de Bernoulli podemos perceber uma maioria de pacientes diagnosticados com a Doença de Alzheimer, porém tendo pouca diferença na população total.

Resultados 0	Resultados 1
1389	760
Diagnóstico negativo para Alzheimer	Diagnóstico positivo para Alzheimer

Em seguida, passamos a avaliar dados mais complexos relacionados a doenças pré existentes nos pacientes, como doenças cardiovasculares, diabetes, depressão, hipertensão e problemas de colesterol.

Doenças Cardiovasculares

A respeito da presença de doenças cardiovasculares, a média dos dados aponta a um diagnóstico negativo na maioria dos pacientes pesquisados, porém com coeficiente de variação de 243%. Atribuímos esse valor de alta variação para o fato dos dados serem binários, sempre aproximados de 0 e 1.

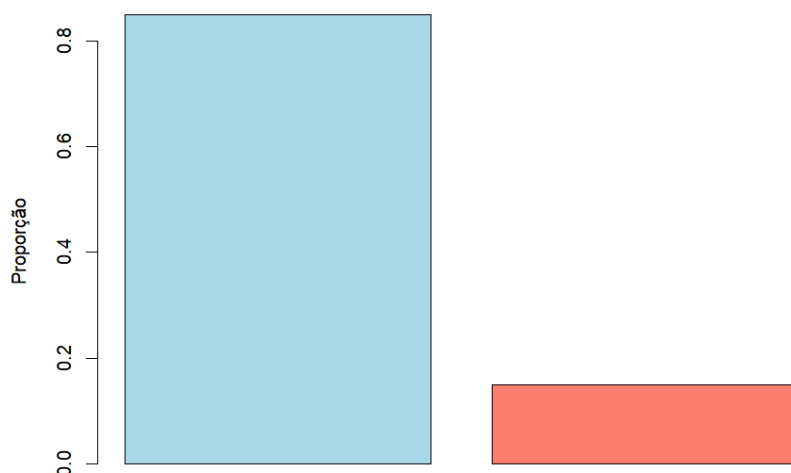


Figura 11. Distribuição de Doenças Cardiovasculares nos pacientes

Diabetes

A respeito da presença de diabetes, a média dos dados aponta a um diagnóstico negativo na maioria dos pacientes pesquisados, porém com coeficiente de variação de 237.39%. Atribuímos esse valor de alta variação para o fato dos dados serem binários, sempre aproximados de 0 e 1.

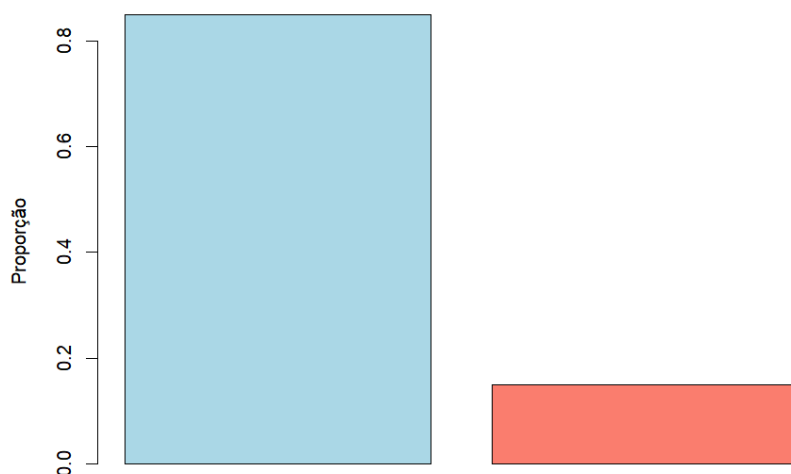


Figura 12. Distribuição de Diabetes nos pacientes

Depressão

A respeito da presença de depressão, a média dos dados aponta a um diagnóstico negativo na maioria dos pacientes pesquisados, porém com coeficiente de variação de 199.7%. Atribuímos esse valor de alta variação para o fato dos dados serem binários, sempre aproximados de 0 e 1.

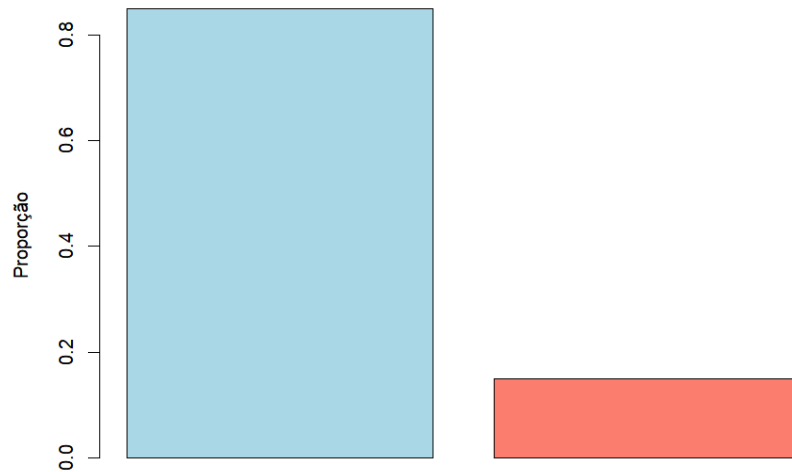


Figura 13. Distribuição de Depressão nos pacientes

Hipertensão

A respeito da presença de hipertensão, a média dos dados aponta a um diagnóstico negativo na maioria dos pacientes pesquisados, porém com coeficiente de variação de 239.13%. Atribuímos esse valor de alta variação para o fato dos dados serem binários, sempre aproximados de 0 e 1.

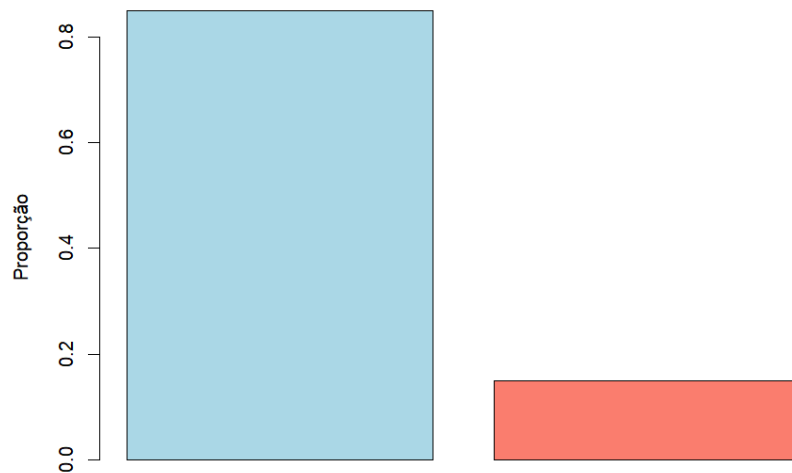


Figura 12. Distribuição de Diabetes nos pacientes

Colesterol Total

A respeito dos níveis de colesterol nos pacientes, identificamos a média de colesterol total em 225.2 mg/dL, mediana de 225.1 mg/dL e coeficiente de variação equivalente a 18.89%, o que representa uma dispersão mediana dos dados. Os valores de média e mediana indicam um valor de colesterol total **muito alto** na população, porém bem próximo ao valor que já seria considerado alto. A Figura 13 indica uma distribuição que se assemelha à normal.

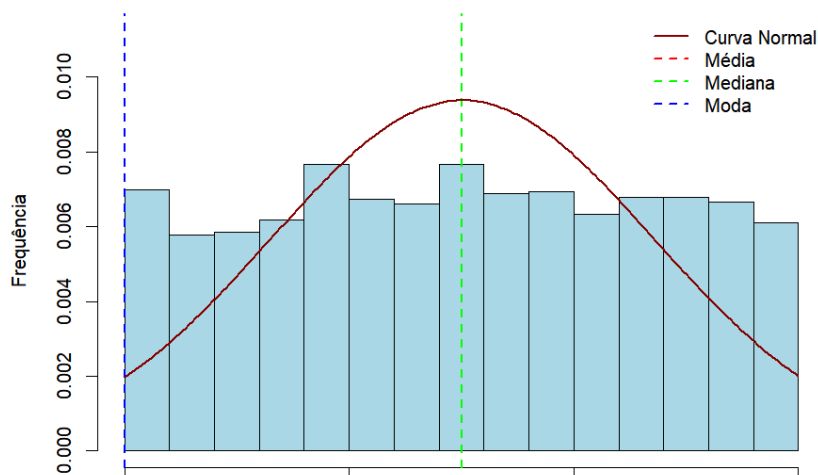


Figura 13. Histograma de Colesterol Total dos Pacientes

Colesterol LDL

A respeito dos níveis de colesterol do tipo LDL nos pacientes, identificamos a média de colesterol LDL em 124.34 mg/dL, mediana de 123.34 mg/dL e coeficiente de variação equivalente a 34.88%, o que representa uma alta dispersão dos dados. Os valores de média e mediana indicam um valor de LDL **desejável** na população, porém bem próximo ao valor que já seria considerado alto. A Figura 14 indica uma distribuição que é simétrica, porém com uma subpopulação com baixo LDL como indicado pela moda.

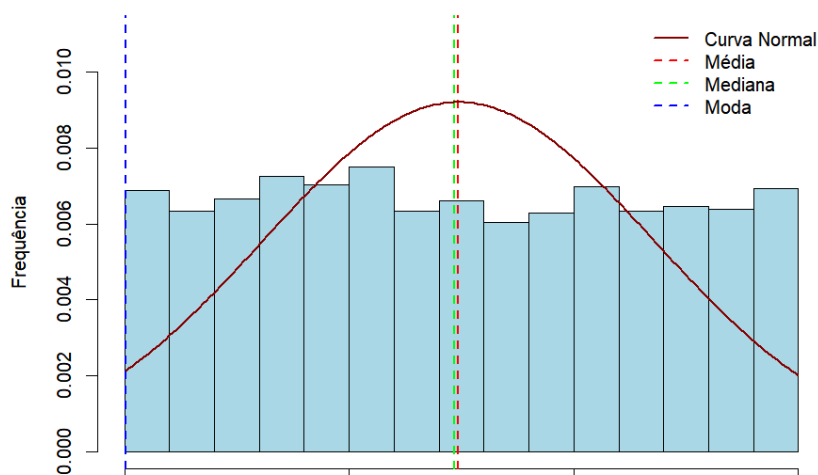


Figura 14. Histograma de Colesterol LDL dos Pacientes

Colesterol HDL

A respeito dos níveis de colesterol do tipo HDL nos pacientes, identificamos a média de colesterol HDL em 59.46 mg/dL, mediana de 59.77 mg/dL e coeficiente de variação equivalente a 38.91%, o que representa uma alta dispersão dos dados. Os valores de média e mediana indicam um valor de HDL **condizente** na população. A Figura 15 corresponde ao histograma dos dados que está mais proporcional ao uma curva normal do que os valores da Figura 16, porém podemos ver uma moda que indica uma subpopulação com baixo HDL.

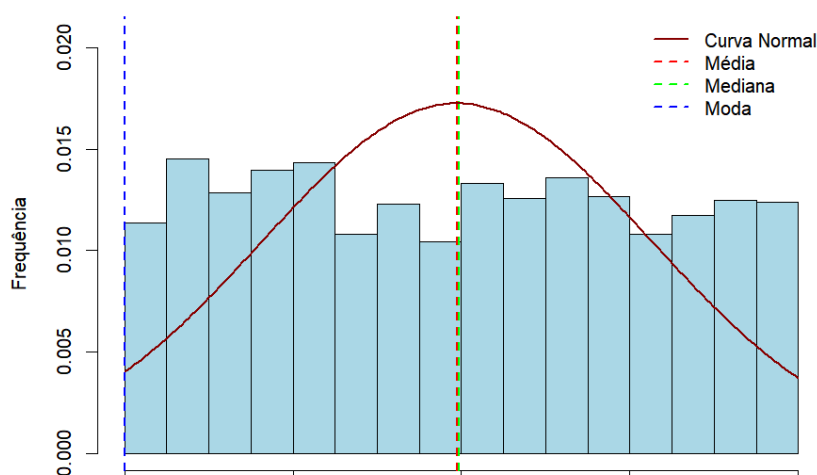


Figura 15. Histograma de Colesterol HDL dos Pacientes

Colesterol Triglicerídeos

A respeito dos níveis de colesterol do tipo triglicerídeos nos pacientes, identificamos a média de colesterol triglicerídeos em 228.26 mg/dL, mediana de 230.3 mg/dL e coeficiente de variação equivalente a 44.68%, o que representa uma dispersão mediana dos dados. Os valores de média e mediana indicam um valor de triglicerídeos **alto** na população. A Figura 16 corresponde ao histograma dos dados que pode indicar duas subpopulações em uma distribuição bimodal dos dados, mostrando que há sobrevalência de dois grandes grupos de valores de triglicerídeos.

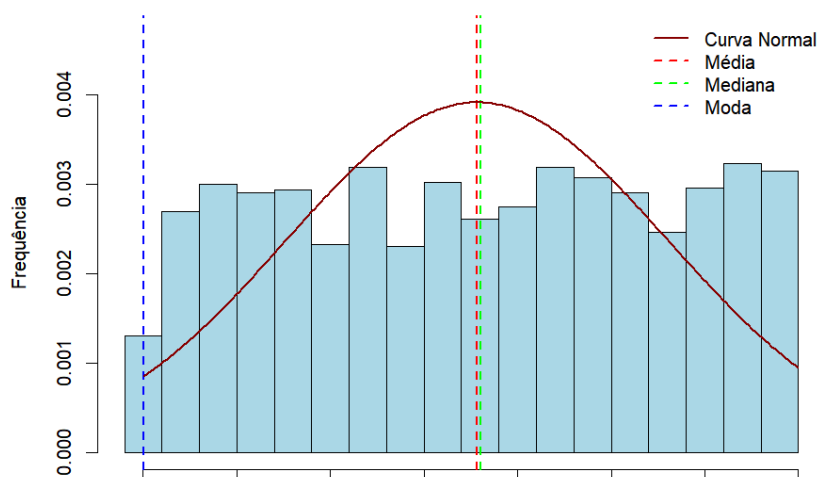


Figura 16. Histograma de Colesterol Triglicerídeos dos Pacientes

Correlatos de Variáveis

Entendendo a distribuição desses dados no nosso banco, passamos a buscar saber quais dessas variáveis mais se correlacionavam com o diagnóstico positivo para a Doença de Alzheimer. Para isso utilizamos a função cor para entender as correlações que são apresentadas na Tabela 1.

Tabela 1. Correlação de dados gerais dos pacientes

Qualidade de Sono	-0.0565
BMI	0.0263
Gênero	-0.0210
Consumo de Álcool	-0.0076
Atividade Física	0.0059
Idade	-0.0055

Aqui podemos identificar quais são as variáveis mais relevantes com os primeiros dados do estudo. Nesse caso, quanto mais próxima de 1, maior a correlação e quanto mais próxima de -1 menor era a correlação da variável com o diagnóstico positivo. Assim, podemos identificar a variável de IMC mais relevante para o estudo; podendo supor que quanto maior o IMC, maior o risco de um diagnóstico positivo.

No entanto, é preciso analisar mais dados para podermos afirmar quais são as variáveis determinantes para o desenvolvimento da doença, por isso realizamos o mesmo teste para as variáveis clínicas que estudamos; o resultado encontra-se na Tabela 2.

Tabela 2. Correlação de dados clínicos dos pacientes

Colesterol HDL	0.0426
Hipertensão	0.0351
Colesterol LDL	-0.0320
Doença Cardiovascular	0.0315
Diabetes	-0.0315
Colesterol Triglicerídeos	0.0227
Colesterol Total	0.0064
Depressão	-0.0059

Podemos compreender que, nesse caso, a variável mais importante para a correlação se mostra ser o nível de Colesterol do tipo HDL, seguida da presença de Hipertensão nos pacientes. Mostrando que maior o HDL no sangue, maior a chance de diagnóstico positivo para Alzheimer.

Testes de Hipótese

Entendendo a distribuição desses dados no nosso banco, passamos a buscar saber quais dessas variáveis mais se correlacionavam com o diagnóstico positivo para a Doença de Alzheimer. Para isso utilizamos o teste ANOVA para compreender se as diferenças entre as médias das variáveis e do diagnóstico ocorrem pelo acaso ou não.

Na Tabela 3 é possível encontrar as variáveis estudadas, seus devidos p-valor e o significado para a pergunta “há diferença significativa entre as médias da variável e dos diagnósticos do banco de dados?”.

Tabela 3. Resultados teste ANOVA para variáveis clínicas

Variável	p-valor	Significativo? (p-valor<0.05)
Idade	0.7993	Não
IMC	0.2222	Não
Atividade Física	0.7830	Não
Qualidade de Sono	0.0087	Sim
Consumo de Álcool	0.7241	Não
Doença Cardiovascular	0.1445	Não
Diabetes	0.1443	Não
Depressão	0.7848	Não
Hipertensão	0.1040	Não
Colesterol Total	0.7670	Não
Colesterol LDL	0.1384	Não
Colesterol HDL	0.0484	Sim
Colesterol Triglicérides	0.2935	Não

Para entender onde estava a variação nas variáveis Qualidade de Sono e Colesterol HDL, realizamos um Teste Tukey. Para o caso da Qualidade de sono, o resultado do teste mostrou que os pacientes com Alzheimer (grupo 1) têm, em média, 0.02 unidades a menos de sono do que os pacientes sem Alzheimer (grupo 2), como é possível observar na Figura 18.

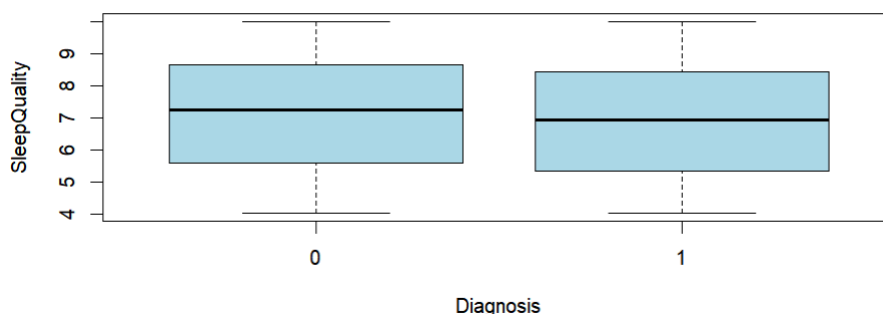


Figura 18. Boxplot do teste Tukey para Qualidade de Sono X Diagnóstico

Já no caso do Colesterol HDL, o resultado do teste apresentou que os pacientes com Alzheimer (grupo 1) têm, em média, 2.06 unidades a mais de HDL que os pacientes sem Alzheimer (grupo 0), como pode ser observado na Figura 18.

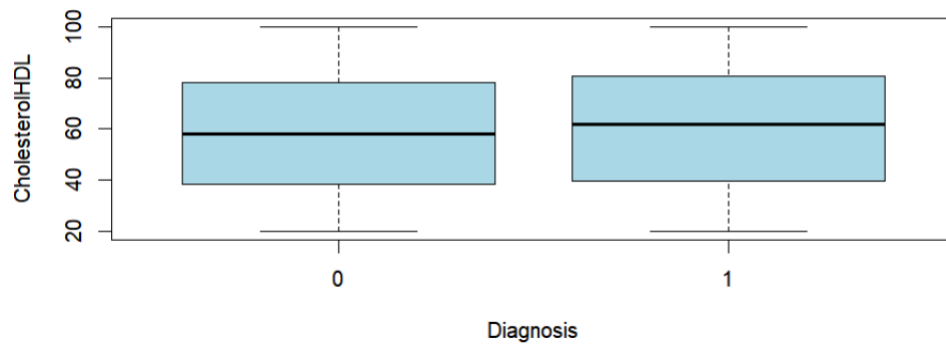


Figura 17. Boxplot do teste Tukey para Colesterol HDL X Diagnóstico

Conclusão

Conclui-se que, de todas as variáveis clínicas estudadas no banco de dados, apenas as variáveis a respeito da Qualidade de Sono e o Colesterol tipo HDL dos pacientes se mostraram estatisticamente significativas para o diagnóstico positivo da doença de Alzheimer. Esse resultado pode indicar que uma má rotina de sono e um aumento no colesterol HDL podem se tornarem fatores importantes para a presença da doença quando associados a fatores genéticos, ambientais e metabólicos que influenciam no desenvolvimento da patologia.