

**Universidade de São Paulo
Instituto de Matemática e Estatística**

Centro de Estatística Aplicada

Relatório de Análise Estatística

RAE-CEA–19P02

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

**“Análise automatizada de Hiperintensidades de Substância Branca em pacientes
com CCL e Doença de Alzheimer”**

Carmen Diva Saldiva de André

Cleudiana Carvalho

Geraldo Busatto Filho

Luíza Baratojo

Milene Vaiano Farhat

São Paulo, junho de 2019

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “Análise automatizada de Hiperintensidades de Substância Branca em pacientes com CCL e Doença de Alzheimer”.

PESQUISADORA: Cleudiana Carvalho

ORIENTADOR: Prof. Dr. Geraldo Busatto Filho

INSTITUIÇÃO: Faculdade de Medicina da Universidade de São Paulo

FINALIDADE DO PROJETO: Mestrado

RESPONSÁVEIS PELA ANÁLISE: Carmen Diva Saldiva de André
Luíza de Mattos Baratojo
Milene Vaiano Farhat

REFERÊNCIA DESTE TRABALHO: ANDRÉ, C. D. S.; BARATOJO, L. M., FARHAT, M. V. Relatório de análise estatística sobre o projeto: “Análise automatizada de Hiperintensidades de Substância Branca em pacientes com CCL e Doença de Alzheimer”. São Paulo, IME-USP, 2019. (RAE–CEA-19P02)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

BRANT, R. (1990). Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. **Biometrics**, **46**, 1171 - 1178.

CARVALHO, C.L. (2019). **Análise automatizada de Hiperintensidades de Substância Branca em pacientes com CCL e Doença de Alzheimer**. São Paulo. 30p. Dissertação (Mestrado). Faculdade de Medicina - USP.

DAVID, J.S.E.; FERRARI, S.L.P. (2000). **Modelos de Superdispersão para Dados de Contagem**. São Paulo. Dissertação (Iniciação Científica). Instituto de Matemática e Estatística - USP.

DAVID, J.S.E.; FERRARI, S.L.P. (1999). **Regressão Logística, Regressão de Poisson e Modelos Lineares Generalizados**. São Paulo. 71p. Dissertação (Iniciação Científica). Instituto de Matemática e Estatística - USP.

FERRARI, S.L.P.; CRIBARI-NETO, F. (2004). Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, **31**, 799-815.

NETER, J.; WASSERMAN, W.; KUTNER, M.H. (1989). **Applied linear regression models**. 5ª ed. McGraw-Hill/Irwin.

MORAL, R.A.; HINDE, J.; DEMETRIO, C.G.B. (2017) Half-normal Plots and Overdispersed models in R: The hn Package. **Journal of Statistical Software**, **81** - doi 10.18637/jss.v081.i10.

PAULA, G.A. (2004). **Modelos de regressão: com apoio computacional**. São Paulo: IME-USP.

R Core Team (2018), **R: A Language and Enviroment for Statistical Computing**, Vienna, Austria: The R Foundation. www.R-project.org/.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Microsoft Word for Windows (versão Online)

Microsoft Excel for Windows (versão Online)

RStudio Desktop Open Source License (versão 1.2.1335)

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Unidimensional (03:010)

Análise Descritiva Multidimensional (03:020)

Testes de Hipóteses Paramétricas (05:010)

Análise de Associação e Dependência de Dados Quantitativos (06:010)

Associação e Dependência de Dados Qualitativos (06:020)

Análise de Regressão Clássica (07:020)

Regressão Logística (07:090)

ÁREA DE APLICAÇÃO

Bioestatística (14:030)

Resumo

Observou-se nas últimas décadas o aumento significativo de pessoas com doenças crônicas degenerativas, entre elas a doença de Alzheimer (DA) e comprometimento cognitivo leve (CCL). Pouco se sabe sobre as causas desses distúrbios e seus diagnósticos ainda são muito subjetivos. A fim de colaborar para uma descrição mais precisa de padrões típicos de alterações observadas, o seguinte projeto investiga correlação de variáveis avaliadas em imagens de ressonância magnética (RM) de pacientes em 3 grupos distintos: com DA, com CCL e saudáveis (Carvalho, 2019).

O estudo utilizou imagens de ressonância magnética de 120 idosos classificados como saudáveis, com DA ou com CCL. Primeiramente foi feita uma análise descritiva dos dados com o objetivo de compreender melhor seu comportamento e estrutura. Em seguida, realizou-se testes de correlação de Spearman entre as variáveis. Depois, com o teste de Kruskal Wallis, verificou-se que entre as variáveis de interesse, volume de lesão normalizado de substância branca, volume de lesão normalizado intracraniano e número de lesões, e as demais variáveis sociodemográficas, apenas idade apresentou, a um nível de significância de 0.05, associação positiva. Por fim, foram ajustados modelos de regressão nos quais cada variável da RM foi considerada uma variável resposta e as demais variáveis explicativas. Em ambos os casos foi comprovado por meio do teste de razão de verossimilhanças que não havia efeito de interação. Para chegar no melhor modelo, quando a variável resposta se tratava de um volume, testou-se a regressão de mínimos quadrados, a beta e por último uma regressão categórica. Quando a variável resposta era o número de lesões, testou-se primeiro a regressão de mínimos quadrados, depois a de Poisson, em seguida a binomial negativa e por fim a categórica. Independentemente da variável resposta, idade apresentou significativa associação e quando se tratando apenas de volume da lesão normalizado, diagnóstico também foi significativo.

Sumário

1. Introdução.....	9
2. Objetivo.....	10
3. Descrição do estudo.....	10
4. Descrição das variáveis.....	11
5. Análise descritiva.....	12
6. Análise inferencial.....	15
6.1 Associação das variáveis da ressonância magnética.....	15
6.2 Ajuste dos modelos de regressão múltipla.....	16
6.2.1 Volume de lesão normalizado SB.....	16
6.2.2 Número de lesões.....	19
7. Conclusões.....	21
Apêndice A.....	22
Apêndice B.....	27

1. Introdução

O aumento da longevidade é um fenômeno que vem sendo verificado nas últimas décadas. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), cerca de 10% dos brasileiros são idosos, ou seja, têm 60 anos de idade ou mais. Junto com o envelhecimento da população foi possível observar também o crescimento do número de pessoas com doenças crônicas degenerativas. Dentre essas patologias, a doença de Alzheimer (DA) representa 75% dos casos de demência, sendo assim a de maior prevalência.

A doença de Alzheimer é uma doença neurodegenerativa caracterizada principalmente pela perda de memória progressiva. Essa deficiência é associada ou precedida por outros tipos de comprometimentos cognitivos e manifestações emotivas, gerando a incapacitação social e ocupacional do indivíduo.

As pessoas com comprometimento cognitivo leve (CCL) fazem parte de um grupo particular com níveis de cognição inferior ao esperado pela idade, porém que não tiveram suas atividades diárias comprometidas. Esse tipo de demência pode vir a ser um pré-estágio para a doença de Alzheimer.

Apesar da expressiva manifestação de DA em idosos com patologias, pouco se sabe sobre as causas do comprometimento cognitivo e ambos os diagnósticos, da DA e do CCL, são muito subjetivos. Por essas razões, cria-se a necessidade de novas técnicas e ferramentas para um diagnóstico mais preciso dessas doenças.

O cérebro humano é composto por substâncias cinzenta e branca. Já se notou que a presença de alterações na estrutura da substância branca (SB) como as hiperintensidades de SB, visualizadas por exames de ressonância magnética (RM), presentes no cérebro de pacientes com DA e CCL, podem desempenhar relevância na fisiopatologia destes transtornos e um estudo mais aprofundado dessa relação pode ser útil para o diagnóstico da doença.

2. Objetivo

O objetivo do projeto é amplo e se resume à tentativa de investigar padrões de correlação de variáveis avaliadas em imagens de ressonância magnética (RM) com o comprometimento cognitivo observado nos pacientes com DA e CCL e outras variáveis clínicas e sociodemográficas, a fim de contribuir para uma descrição mais acurada de padrões típicos de alterações observadas na RM.

3. Descrição do estudo

O estudo foi do tipo transversal e utilizou imagens de exames de RM de um banco de dados armazenado no Laboratório de Investigação Médica em Neuroimagem Psiquiátrica (LIM21) do Instituto de Psiquiatria (IPq) do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP) (Carvalho, C. 2019).

Para o presente trabalho, foram utilizados dados cognitivos e exames de RM de 120 voluntários idosos (60 anos ou mais), subdivididos pela pesquisadora em três grupos:

- DA: voluntários diagnosticados com Doença de Alzheimer;
- CCL: voluntários diagnosticados com Comprometimento Cognitivo Leve;
- Controle: voluntários saudáveis.

Inicialmente a base de dados apresentada pela pesquisadora contava com dados de 125 pacientes, que continham informações de exames de ressonância magnética, traduzidos por um software em dois momentos:

- Pré Upgrade, com 120 voluntários;
- Pós Upgrade, com 5 voluntários.

Devido ao reduzido tamanho da amostra de pacientes no pós upgrade, consideramos apenas as informações anteriores ao upgrade do software de RM.

4. Descrição das variáveis

As variáveis consideradas no estudo são:

- Número de lesões;
- Volume de lesão normalizado SB: volume da lesão (ml) / volume substância branca (ml);
- Volume de lesão normalizado intracraniano: volume da lesão (ml) / volume intracraniano (ml);
- Idade (anos);
- Gênero:
 - o Feminino
 - o Masculino;
- Escolaridade:
 - o Baixa: de 0 a 7 anos
 - o Média: de 8 a 12 anos
 - o Alta: maior que 12 anos;
- Diagnóstico
 - o Controle
 - o CCL
 - o DA;

- PIB: marcador de medicina nuclear que indica presença de lesão no cérebro:
 - o Positivo: com lesão
 - o Negativo: sem lesão;
- MMSE (escore de 0-30): teste psicológico em que quanto maior o valor, melhor o nível de cognição do paciente.

5. Análise descritiva

Nesta seção, apresentamos a análise descritiva dos dados, que consiste no resumo dos resultados observados na amostra.

Pela Tabela A.1 e Figura B.1 é possível notar que a média da idade observada no grupo com a doença de Alzheimer (73,2 anos), bem como sua mediana (75 anos), são maiores do que nos demais grupos, o que é esperado já que o número de doenças degenerativas cresce conforme a população envelhece.

Pode ser observado na Tabela A.2 e Figura B.2 que a porcentagem de indivíduos com escolaridade baixa é maior no diagnóstico de Doença de Alzheimer (38,5%) do que nas demais categorias do diagnóstico, enquanto que a porcentagens de indivíduos com escolaridade alta é menor (18,2%) que as observadas nas outras duas categorias de diagnóstico.

A Tabela A.3 e Figura B.3 apresentam a distribuição do gênero em cada categoria de diagnóstico. As distribuições são semelhantes nos três grupos, o que indica não associação entre as variáveis.

Conforme esperado, os indivíduos do controle apresentaram escore médio do MMSE superior aos outros dois tipos do diagnóstico (26,6), sendo que as pessoas com DA apresentaram o menor escore médio, de 22,7, como aponta a Tabela A.4. As distribuições do MMSE em cada categoria do diagnóstico estão representadas, de forma aproximada, na Figura B.4. O que chama a atenção é que mesmo dentro do grupo com DA, o escore máximo atingiu o valor 30 (máximo resultado do teste), o que

mostra como o diagnóstico da doença é subjetivo e evidencia a necessidade de formas mais precisas para identificar o nível de comprometimento cognitivo do paciente.

O PIB é outra variável que auxilia na detecção de lesões no cérebro. A Tabela A.5 e Figura B.5 mostram que nos grupos DA e CCL foram observados maiores percentuais de pessoas com PIB positivo (47,5% e 48,6%, respectivamente) comparativamente ao percentual de indivíduos do grupo controle.

A Figura B.6 mostra um diagrama de dispersão do volume lesão normalizado SB e volume lesão normalizado intracraniano e evidencia alta correlação entre essas variáveis. De fato, o coeficiente de correlação de Spearman entre elas na amostra é 0,99. Por esse motivo, apenas o volume lesão normalizado SB foi considerado nas análises a seguir.

A Figura B.7 mostra um diagrama de dispersão do volume lesão normalizado SB e número de lesões. O coeficiente de Spearman dessas duas variáveis observado na amostra é 0,81.

As médias e medianas amostrais do volume de lesão normalizado na substância branca são iguais nos grupos CCL e DA e maiores que no controle (Tabela A.6). Os *box plots* na Figura B.8 sugerem assimetria da distribuição dessa variável nos três grupos do diagnóstico.

Os diagramas de dispersão na Figura B.9 indicam que o volume de lesão normalizado na substância branca tende a aumentar com a idade, nos três grupos definidos pela variável diagnóstico. As correlações de Spearman são baixas: 0,36 no controle, 0,25 no CCL e 0,41 no DA.

A Figura B.10 sugere que a associação do volume de lesão normalizado SB com a escolaridade não parece ser a mesma para cada diagnóstico. Quando se compara os três diagnósticos para as categorias de escolaridade média e alta, percebe-se que o volume da lesão tende a aumentar com o comprometimento cognitivo.

A Figura B.11 mostra que pessoas do gênero feminino possuem a mediana do volume de lesão normalizado na substância branca maior do que as do gênero

masculino nas três categorias de diagnóstico. Em ambos os gêneros, o volume de lesão normalizado SB foi menor para pacientes do grupo controle e maior para indivíduos com DA. O comportamento do volume em cada sexo nos três grupos de diagnóstico é semelhante ao observado na Figura B.8, no qual não foi considerada a divisão por gênero.

A Figura B.12 sugere a não associação da variável MMSE com o volume lesão normalizado SB, o que é evidenciado pelas correlações de Spearman baixas de 0,24, -0,22 e 0,17 para os grupos Controle, CCL e DA respectivamente.

A Figura B.13 mostra os *box plots* do volume de lesão normalizado SB por categoria do PIB para cada tipo de diagnóstico. O volume da lesão tende a ser menor no grupo controle do que no CCL e no DA nas duas categorias de PIB, o que já havia sido observado na Figura B.8. Nota-se também que no grupo CCL a mediana do volume de lesão normalizado SB é maior para pacientes com PIB positivo. Nos demais grupos, não parece haver diferença entre as medianas para as duas categorias do PIB.

Quanto ao número de lesões, variável da ressonância magnética, não é observado aumento das médias e medianas com o crescimento do grau de comprometimento (Tabela A.7), o que também pode ser observado nos *box plots* na Figura B.14.

Os diagramas de dispersão na Figura B.15 indicam que o número de lesões tende a aumentar suavemente com a idade para os três diagnósticos. Os coeficientes de correlação amostral de Spearman entre o número de lesões e idade dos pacientes no grupo controle é 0,24, no grupo CCL é 0,25 e no diagnóstico DA é 0,49.

A Figura B.16 sugere que a associação do número de lesões com a escolaridade depende do diagnóstico do paciente. Apenas para o diagnóstico DA, percebe-se que o número de lesões tende a aumentar conforme a escolaridade, isto é, é menor nos indivíduos com baixa escolaridade e maior nos indivíduos com alta escolaridade.

Pela Figura B.17 verifica-se que à medida em que o diagnóstico do paciente se agrava, o número de lesões mediano é crescente para o gênero masculino, ou seja, o maior número de lesões mediano para pacientes do sexo masculino é no diagnóstico

DA. O mesmo não ocorre para o gênero feminino. Percebe-se maior número de lesões mediano para pacientes do sexo feminino do que do sexo masculino, nos diagnósticos Controle e CCL, ocorrendo o contrário no grupo DA.

A Figura B.18 sugere associação entre o número de lesões e o MMSE para cada diagnóstico, sendo os coeficientes de correlação amostral de Spearman iguais a 0,41 no grupo controle, 0,57 no grupo CCL e 0,29 no diagnóstico DA.

A Figura B.19 mostra o *box plot* do número de lesões por categoria do PIB para cada diagnóstico. Nos diagnósticos CCL e DA, o número de lesões é menor para PIB negativo e maior para PIB positivo. No grupo de diagnóstico controle ocorre o inverso, ou seja, o número de lesões mediano é maior para PIB negativo.

6. Análise inferencial

Na análise inferencial foi avaliada a associação das variáveis da RM: número de lesões e volume de lesão normalizado SB com o diagnóstico, idade, gênero, escolaridade, PIB e MMSE. Inicialmente foi avaliada a associação das variáveis da RM e cada uma das demais variáveis. Posteriormente foram ajustados modelos de regressão, nos quais as variáveis da RM foram consideradas as variáveis respostas e o diagnóstico, idade, gênero, escolaridade, PIB e MMSE, as variáveis explicativas. Nos testes de hipótese, foi fixado nível de significância de 0,05.

6.1 Associação das variáveis da ressonância magnética

Para avaliar a associação do número de lesões e volume de lesão normalizado SB e cada uma das variáveis qualitativas: diagnóstico, escolaridade, gênero e PIB, foi aplicado o teste de Kruskal-Wallis. Foi adotada uma técnica não paramétrica porque os gráficos de probabilidade normal das observações, em cada categoria das variáveis qualitativas, apontaram desvios grosseiros da distribuição Normal (Figuras B.22 e B.23). Na Tabela A.8 são encontrados os valores-p obtidos no teste da hipótese de

igualdade das distribuições das variáveis da RM nas categorias das variáveis qualitativas. Os valores-p apontam que não há diferença entre as distribuições das variáveis da RM nas categorias de cada uma das variáveis: gênero, diagnóstico, PIB e escolaridade.

A associação do número de lesões e do volume de lesão normalizado SB com a idade e MMSE foi avaliada por meio do cálculo do coeficiente de correlação de Spearman. Os valores do coeficiente e correspondentes valores-p apresentados na Tabela A.9 apontam que há correlação positiva das variáveis da RM com a idade, isto é, o número de lesões e o volume de lesão normalizado SB tendem a aumentar com o aumento da idade. Não há correlação das variáveis da RM com o MMSE.

6.2 Ajuste dos modelos de regressão múltipla

A seguir passamos a descrever as etapas do processo de ajuste dos modelos de regressão para cada uma das variáveis respostas: volume de lesão normalizado SB e número de lesões.

6.2.1 Volume de lesão normalizado SB

Nesta etapa da análise a variável resposta é o volume de lesão normalizado SB. As variáveis explicativas consideradas foram: diagnóstico, idade, gênero, escolaridade, PIB e MMSE. As categorias de referência para as variáveis qualitativas consideradas foram: diagnóstico controle, gênero feminino, escolaridade baixa e PIB negativo. Foram também adicionados ao modelo termos de interação do diagnóstico e as demais variáveis explicativas, de forma a permitir que possíveis diferenças do volume nos três diagnósticos pudessem depender da idade, gênero, escolaridade, PIB e MMSE.

Inicialmente o modelo de regressão foi ajustado pelo método de mínimos quadrados (Neter et al., 1989), considerando a função *lm* do Software R (R Core Team, 2018). Porém a análise dos resíduos (Figuras B.22 e B.23) apontaram desvios

grosseiros das suposições de normalidade e igualdade de variâncias dos erros do modelo.

Dessa forma, levando em consideração que o volume de lesão normalizado SB é uma proporção, ou seja, seus valores variam entre 0 e 1, ajustamos um modelo de regressão Beta (Ferrari e Cribari, 2004) com as mesmas variáveis explicativas, utilizando para isto a função *betareg* do software R. A função de ligação considerada foi a logito da média do volume, lembrando que o logito de uma proporção p é dado por

$\log \frac{p}{(1-p)}$. Assim, o logito do volume da substância branca é dado por

$$\log \frac{\text{volume de lesão norm. SB}}{(1 - \text{volume de lesão norm. SB})}.$$

Denominamos o quociente $\frac{\text{volume de lesão norm. SB}}{(1 - \text{volume de lesão norm. SB})}$, que também pode ser escrito como $\frac{\text{volume com lesão norm. SB}}{\text{volume sem lesão norm. SB}}$, por razão de volume da substância branca.

Foi também ajustado um modelo sem as interações do diagnóstico com as demais variáveis explicativas. A contribuição das interações na explicação da variável resposta foi testada por meio do teste da razão de verossimilhanças com o auxílio da função *lrtest* do R. O valor observado da estatística do teste foi $\chi^2_o = 11,9$, e o correspondente valor-p obtido de uma distribuição χ^2 com 12 graus de liberdade foi igual a 0,456, o que significa que não há efeito das interações na explicação da variável resposta. Assim, foi considerado o modelo sem as interações. As estimativas dos coeficientes das variáveis explicativas, obtidas pelo método de máxima verossimilhança, respectivos erros padrão, valores da estatística do teste de Wald para a significância dos coeficientes e valores-p são apresentados na Tabela A.10. Pode ser observado nessa tabela que há associação do volume com a idade e diagnóstico. A interpretação dos coeficientes é a seguinte:

- Com o aumento de um ano na idade, a razão de volume fica multiplicada por $e^{0,06}=1,06$, ou seja, a razão de volume da substância branca aumenta 6%,

mantidos constantes os valores das demais variáveis no modelo.

- A razão de volume da substância branca no grupo com Doença de Alzheimer é 1,65 vezes a razão de volume no Controle, ou seja, há um aumento de 65% no volume da lesão normalizado de substância branca nos sujeitos com DA em relação aos do grupo Controle, mantidos constantes os valores das demais variáveis no modelo.
- A razão de volume no grupo com Comprometimento Cognitivo Leve é igual à razão do grupo Controle, ou seja, não há um aumento no volume da lesão normalizado de substância branca nos sujeitos com CCL em relação aos do grupo Controle, mantidos constantes os valores das demais variáveis no modelo.

O valor do pseudo- R^2 foi de 0,14 e o coeficiente de precisão foi 71,3. A análise de resíduos é apresentada nas Figuras B.24 e B.25.

Embora a análise de resíduos não tenha apontado desvios acentuados das suposições do modelo, há 4 pontos localizados fora do envelope (Figura B.24) e pontos apresentando altos valores de resíduos do desvio (Figura B.25). Decidimos então aplicar uma forma alternativa de análise categorizando o volume de lesão normalizado SB de acordo com seus tercís: 1º tercís = 0,00079 e segundo tercís = 0,00510.

Foi então ajustado um modelo logístico ordinal (Brant, 1990) para a variável volume de lesão normalizado SB categorizado, com função de ligação logito, considerando as variáveis explicativas idade, diagnóstico, gênero, MMSE, PIB e escolaridade e as interações do diagnóstico com as demais variáveis explicativas. Foi também ajustado um modelo sem as interações e a contribuição destes termos na explicação da variável resposta foi testada a partir do teste da razão de verossimilhanças. O valor observado da estatística do teste foi $\chi^2_o = 13,9$ e o correspondente valor-p foi 0,309. Portanto, não há evidência de que as interações afetam o volume. As estimativas dos coeficientes das variáveis explicativas obtidas no modelo sem interação, seus respectivos erros padrão, valores da estatística do teste de

Wald para a significância dos coeficientes e valores-p são apresentados na Tabela A.11. A interpretação dos coeficientes é a seguinte:

- O coeficiente positivo da idade indica que maiores valores do volume tendem a estar associados a pessoas mais idosas. Especificamente, quando a idade aumenta 1 ano, a chance do volume ser maior que 0,0051 fica multiplicada por $e^{0,10}=1,11$, ou seja, é 11% maior, mantidas as demais variáveis constantes.
- A chance do volume ser maior que 0,0051 no grupo DA é 3,2 vezes a chance no grupo controle e a chance do volume ser maior que 0,0051 no CCL é a mesma do controle ($p=0,084$), mantidas as demais variáveis constantes.
- A chance do volume ser maior que 0,0051 no gênero masculino é 0,4 vezes a chance no feminino, mantidas as demais variáveis constantes.
- Não há contribuição da escolaridade, PIB e MMSE na explicação do volume.

O desvio residual foi de 214, com 204 graus de liberdade, e correspondente valor-p igual a 0,301, nos levando a não rejeitar a hipótese de bom ajuste do modelo. Ainda, o teste de *Brant*, que verifica a hipótese de haver chances proporcionais (Brant, 1990), indicou, através do valor-p apresentado de 0,82, que a suposição de proporcionalidade está satisfeita. A análise foi realizada com o auxílio da função *polr*, da versão 7.3-51.4 do pacote MASS, do Software R.

A suposição de que as chances são proporcionais implica que os resultados acima se mantêm quando é considerada a chance do volume ser maior que 0,00079.

6.2.2 Número de lesões

Para o número de lesões tentou-se, primeiramente, ajustar um modelo de regressão linear pelo método de mínimos quadrados, porém a análise dos resíduos apontou desvios grosseiros das suposições do modelo (Figura B.26, Figura B.27).

Por a variável resposta se tratar de uma contagem, decidimos ajustar um modelo de regressão de Poisson (Paula, 2004), com o auxílio da função *glm* do *software* R (R Core Team, 2018). A função de ligação adotada foi a logarítmica. Obtivemos que o desvio do modelo foi de 249,50 (93 graus de liberdade), o que leva à rejeição da hipótese de que o modelo está bem ajustado ($p < 0,001$). Além disso, o gráfico de probabilidade meio normal com envelopes simulados (Moral et al., 2007) apresentado na Figura B.28 sugere a ocorrência de sobredispersão.

Uma alternativa, então, foi ajustar um modelo de regressão Binomial Negativa (David e Ferrari, 2000). Adotamos na análise a função de ligação logarítmica e utilizamos a função *glm.nb*, da versão 7.3-51.4 do pacote *MASS*, do *Software* R. O desvio do modelo ajustado foi de 101 (95 graus de liberdade). O gráfico de probabilidade meio normal com envelopes simulados do modelo (Figura B.29) evidenciou pontos muito discrepantes ao que é esperado quando o modelo está bem especificado.

Assim como foi feito para a variável volume de lesão normalizado SB, decidiu-se categorizar a variável número de lesões segundo seus tercís, sendo o 1º tercíl igual a 4 e o 2º igual a 9, e ajustar um modelo logístico ordinal. A estratégia da análise foi a mesma da análise do volume.

Inicialmente foram consideradas as mesmas 6 variáveis explicativas (idade, escolaridade, diagnóstico, MMSE, PIB e gênero) e interações do diagnóstico com as demais. Foi também ajustado um modelo sem as interações. O teste de razão de verossimilhanças apontou que não há contribuição das interações na explicação da variável resposta ($p = 0,229$) e foi então considerado o modelo sem as interações.

Na verificação de que as chances são proporcionais, por meio da função *brant*, e pacote *brant*, do programa R, foi obtido o valor-p igual a 0,460, o que significa que não há evidências para rejeitar a hipótese de chances proporcionais. Depois aplicou-se o teste de desvios para verificar o ajuste do modelo. O teste apresentou valor-p igual a 0,145, ou seja, também não se rejeita a hipótese de bom ajuste.

No modelo final, como pode ser visto na Tabela A.12, apenas o coeficiente da idade foi considerado significativo, levando à seguinte interpretação:

- A idade está associada ao número de lesões e o coeficiente positivo indica que as pessoas mais idosas tendem a ter maior número de lesões. Quando a idade aumenta em 1 ano, a chance do número de lesões ser maior que 9 fica multiplicada por $e^{0,07}=1,07$, ou seja, é 7% maior, mantidas constantes as demais variáveis do modelo. A chance do número de lesões ser maior que 5 também fica multiplicada por 1,07 com o aumento de 1 ano na idade, devido à suposição de chances proporcionais.

7. Conclusões

Inicialmente, definiu-se as variáveis volume de lesão normalizado SB e número de lesões como variáveis de interesse, pois suspeitava-se que seriam variáveis associadas ao diagnóstico de problemas de cognição, principalmente doença de Alzheimer, e fáceis de se obter pelo exame de ressonância magnética.

Verificou-se que ambas as variáveis de interesse, volume de lesão normalizado SB e número de lesões, estão associadas positivamente com a idade do paciente. Ou seja, quando aumentado a idade da pessoa, provavelmente, maior será seu volume de lesão de substância branca e mais lesões o indivíduo tende a ter.

Apesar de número de lesões não ter apresentado associação com o diagnóstico, foi possível observar que o diagnóstico de Doença de Alzheimer está relacionado com maior volume de lesão normalizado SB, entre indivíduos de mesma idade e gênero.

APÊNDICE A

Tabelas

Tabela A.1 Medidas descritivas para a idade (anos), por diagnóstico do paciente.

Diagnóstico	Média	Desvio Padrão	Mínimo	Mediana	Máximo
Controle	71,9	5,9	60	73,0	84
CCL	72,4	5,9	62	72,5	86
DA	73,2	7,7	60	75,0	87
Total	72,5	6,4	60	73,0	87

Tabela A.2 Distribuição de frequências de escolaridade, por diagnóstico do paciente.

Diagnóstico	Escolaridade			Total
	Baixa	Média	Alta	
Controle	11 (28,2%)	13 (35,2%)	19 (43,2%)	43 (100%)
CCL	13 (33,3%)	12 (32,4%)	17 (38,6%)	42 (100%)
DA	15 (38,5%)	12 (32,4%)	08 (18,2%)	35 (100%)

Tabela A.3 Distribuição de frequências de gênero, por diagnóstico do paciente.

Diagnóstico	Feminino	Masculino	Total
Controle	32 (74,4%)	11 (25,5%)	43 (100%)
CCL	33 (78,6%)	09 (21,4%)	42 (100%)
DA	24 (68,6%)	11 (31,4%)	35 (100%)

Tabela A.4 Medidas descritivas para MMSE por diagnóstico do paciente.

Diagnóstico	Média	Desvio Padrão	Mediana	Mínimo	Máximo
Controle	26,6	3,1	27	18	30
CCL	26,2	2,4	26	20	30
DA	22,7	4,7	23	12	30
Total	25,4	3,8	26	12	30

Tabela A.5 Distribuição de frequências de PIB, por diagnóstico do paciente.

Diagnóstico	Negativo	Positivo	Total
Controle	23 (67,6%)	11 (32,4%)	34 (100%)
CCL	21 (52,5%)	19 (47,5%)	40 (100%)
DA	18 (51,4%)	17 (48,6%)	35 (100%)

Tabela A.6 Medidas descritivas para volume de lesão normalizado SB por diagnóstico do paciente.

Diagnóstico	Média	Desvio Padrão	Mediana	Mínimo	Máximo
Controle	0,006	0,012	0,001	0	0,056
CCL	0,007	0,010	0,004	0	0,038
DA	0,007	0,010	0,004	0	0,043
Total	0,007	0,012	0,003	0	0,056

Tabela A.7 Medidas descritivas para número de lesões por diagnóstico do paciente.

Diagnóstico	Média	Desvio Padrão	Mediana	Mínimo	Máximo
Controle	6,7	6,0	5,0	0	24
CCL	7,9	5,4	7,5	0	26
DA	6,9	4,3	6,0	0	16
Total	7,2	5,3	6,5	0	26

Tabela A.8 Valores-p obtidos na comparação das distribuições do número de lesões e volume de lesão normalizado SB nas categorias de diagnóstico, escolaridade, gênero e PIB.

Variáveis da RM	Variáveis qualitativas			
	Diagnóstico	escolaridade	gênero	PIB
Número de lesões	0,354	0,999	0,871	0,645
Volume de lesão normalizado SB	0,237	0,601	0,535	0,392

Tabela A.9 Coeficientes de correlação de Spearman do número de lesões e volume de lesão normalizado SB e as variáveis idade e MMSE.

Variáveis da RM	Variáveis qualitativa			
	Idade		MMSE	
	Coeficiente de correlação	Valor-p	Coeficiente de correlação	Valor-p
Número Lesões	0,282	0,002	-0,016	0,862
Volume de lesão normalizado SB	0,352	<0,001	0,016	0,858

Tabela A.10 Resultados obtidos no ajuste do modelo final de regressão beta com o volume de lesão normalizado SB como variável resposta e a idade, gênero, diagnóstico, escolaridade, PIB e MMSE como variáveis explicativas.

Coeficientes	Estimado	Erro Padrão	Valor teste	Valor-p
intercepto	-10,77	1,27	-8,41	<0,001
idade	0,06	1,01	4,54	<0,001
gênero (Masculino)	-0,23	0,21	-1,09	0,272
escolaridade (Média)	0,39	0,22	1,75	0,078

escolaridade (Alta)	0,08	0,23	0,36	0,718
diagnóstico (CCL)	0,35	0,21	1,62	0,103
diagnóstico (DA)	0,50	0,24	2,02	0,043
PIB (positivo)	-0,07	0,18	-0,43	0,663
MMSE	0,02	0,02	0,99	0,319

Tabela A.11 Resultados obtidos no ajuste do modelo de regressão logística ordinal com o volume de lesão normalizado SB como variável resposta e a idade, gênero, diagnóstico, PIB, escolaridade e MMSE como variáveis explicativas.

Coeficientes	Estimado	Erro Padrão	Valor teste	Valor-p
intercepto 1 2	9,29	2,78	3,34	0,001
intercepto 2 3	10,68	2,84	3,86	<0,001
idade	0,10	0,03	-3,34	0,001
gênero (Masculino)	-0,94	0,47	2,02	0,044
diagnóstico (CCL)	0,79	0,46	1,73	0,084
diagnóstico (DA)	1,15	0,54	2,14	0,033
MMSE	0,10	0,06	1,66	0,098
PIB (positivo)	0,19	0,38	0,50	0,616
escolaridade (Média)	-0,36	0,48	0,76	0,448
escolaridade (Alta)	-0,57	0,48	1,17	0,240

Tabela A.12 Resultados obtidos no ajuste do modelo de regressão logística ordinal com o número de lesões como variável resposta e a idade, gênero, diagnóstico, escolaridade, PIB e MMSE como variáveis explicativas.

Coeficientes	Estimado	Erro Padrão	Valor teste	Valor-p
intercepto 1 2	5,59	2,61	2,14	0,036
intercepto 2 3	7,07	2,64	2,66	0,009
idade	0,07	0,02	2,41	0,016
gênero (Masculino)	-0,80	0,44	-1,79	0,078
escolaridade (Média)	0,13	0,45	0,28	0,958
escolaridade (Alta)	0,10	0,46	0,22	0,821
diagnóstico (CCL)	0,70	0,46	1,53	0,115

diagnóstico (DA)	0,69	0,52	1,31	0,183
PIB (positivo)	0,01	0,38	0,02	0,984
MMSE	0,02	0,05	0,44	0,655

APÊNDICE B

Figuras

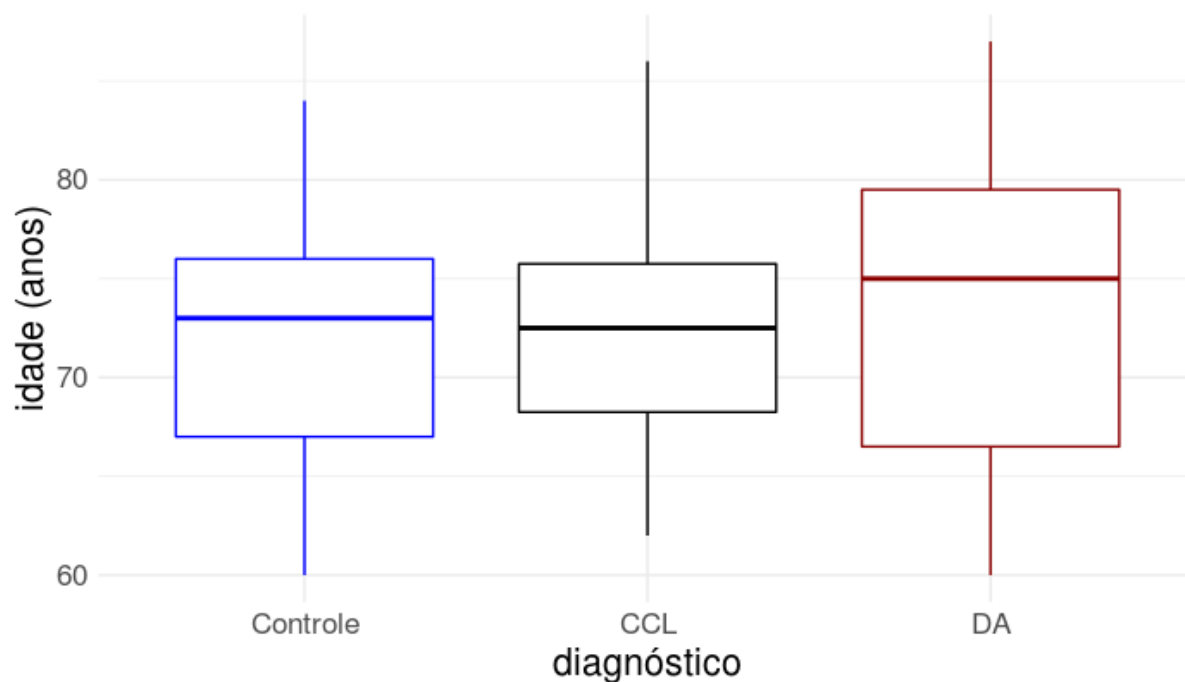


Figura B.1 Box plot da idade (anos) por diagnóstico do paciente.

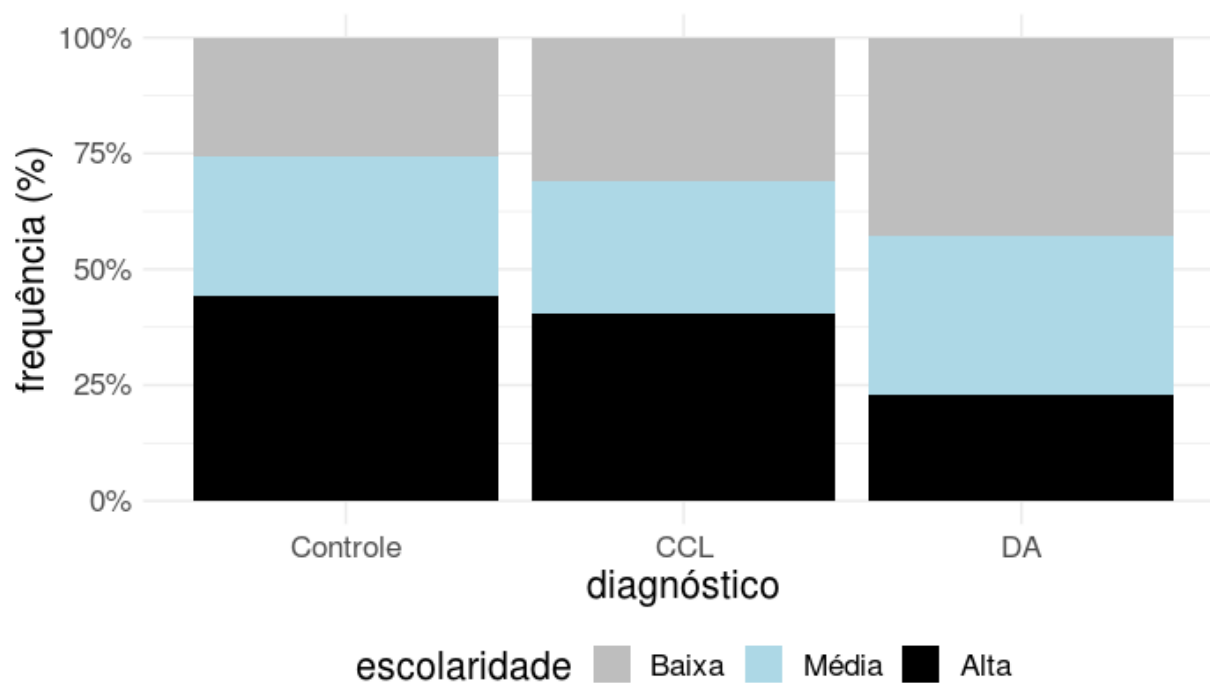


Figura B.2 Gráfico da distribuição de frequências da escolaridade por diagnóstico do paciente.

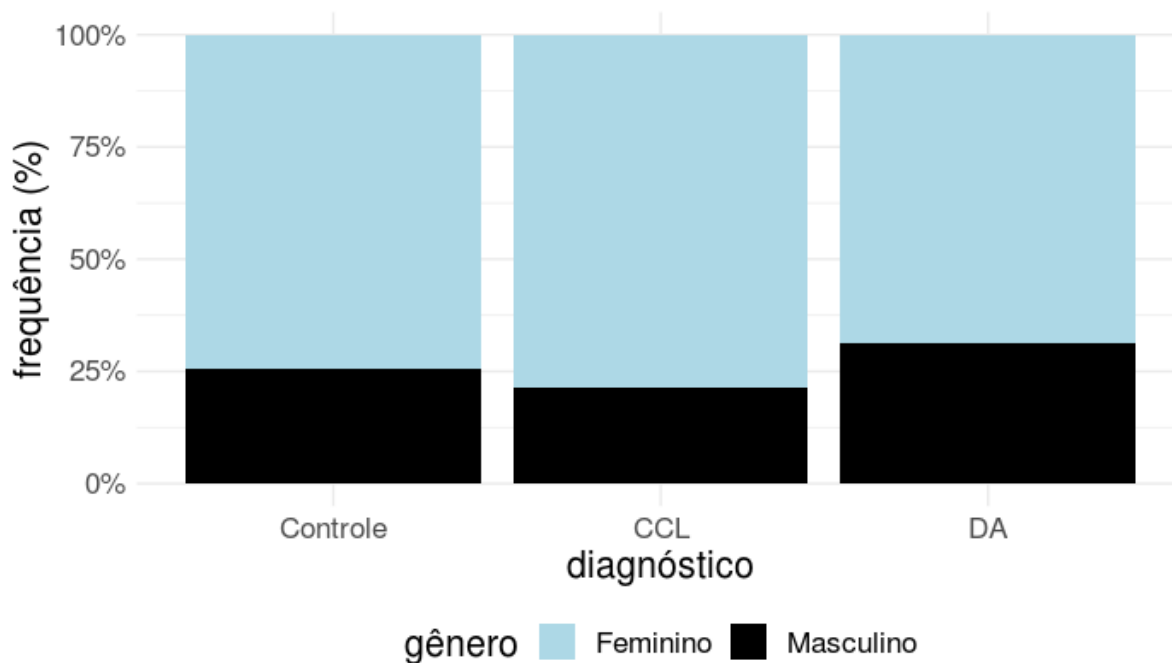


Figura B.3 Gráfico da distribuição de frequências do gênero por diagnóstico do paciente.

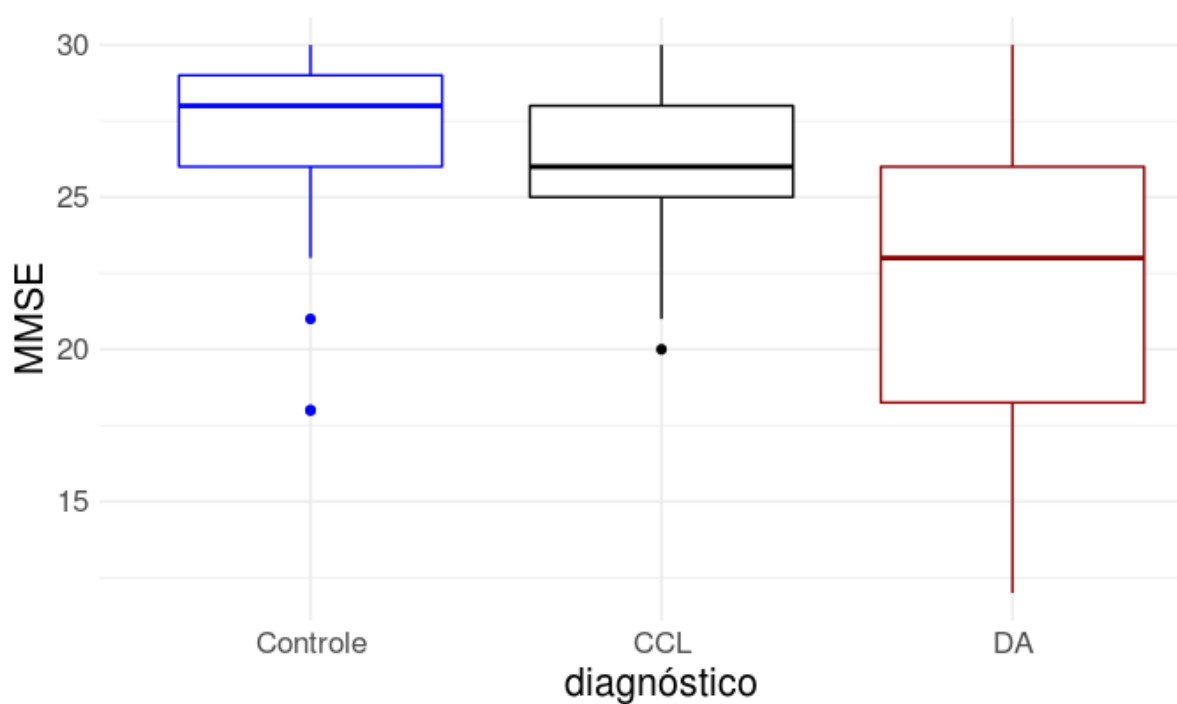


Figura B.4 Box plot do MMSE por diagnóstico do paciente.

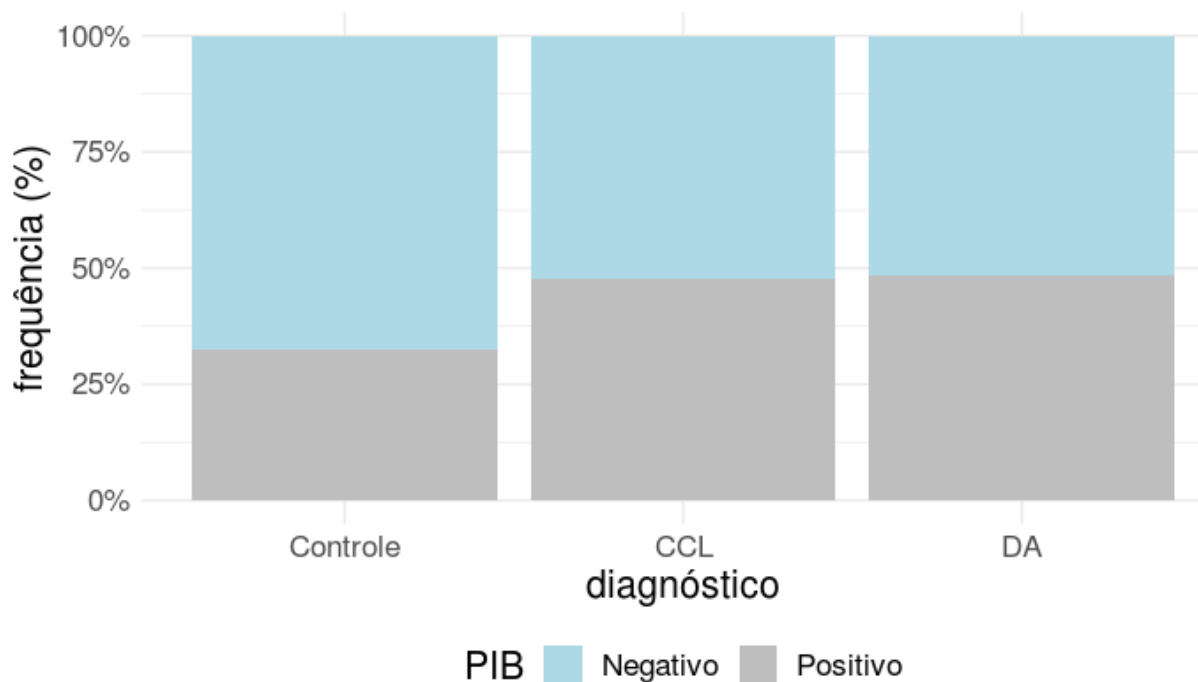


Figura B.5 Gráfico da distribuição de frequências do PIB por diagnóstico do paciente.

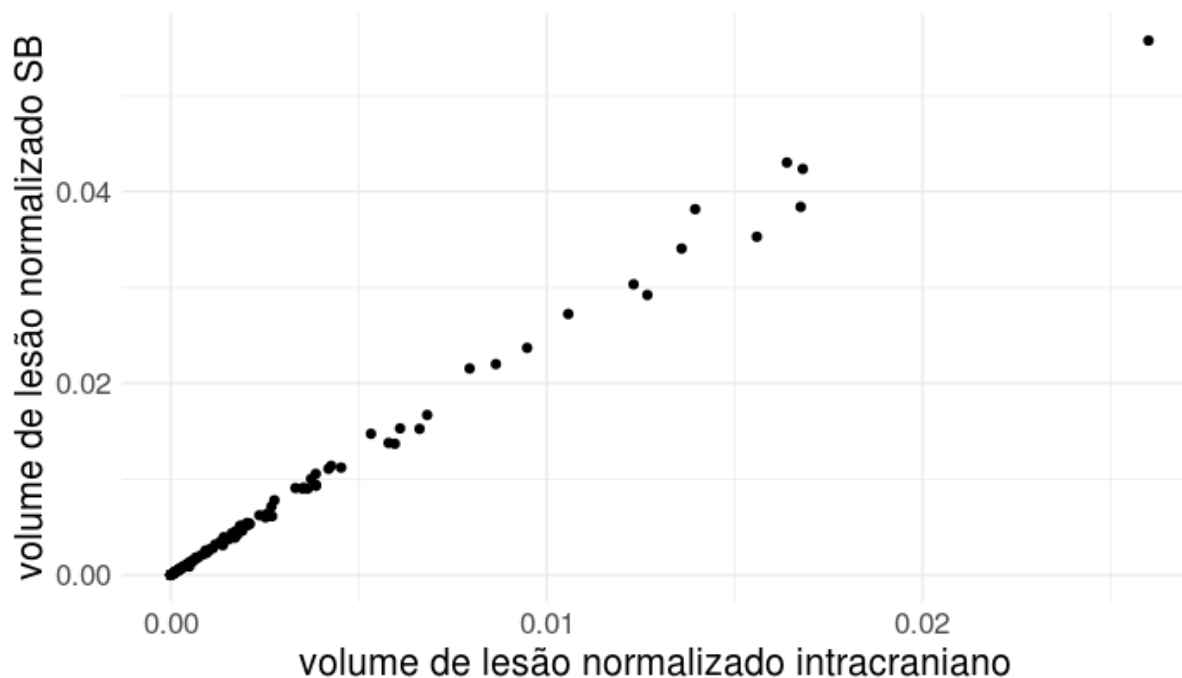


Figura B.6 Gráfico de dispersão do volume de lesão normalizado SB e volume de lesão normalizado intracraniano.

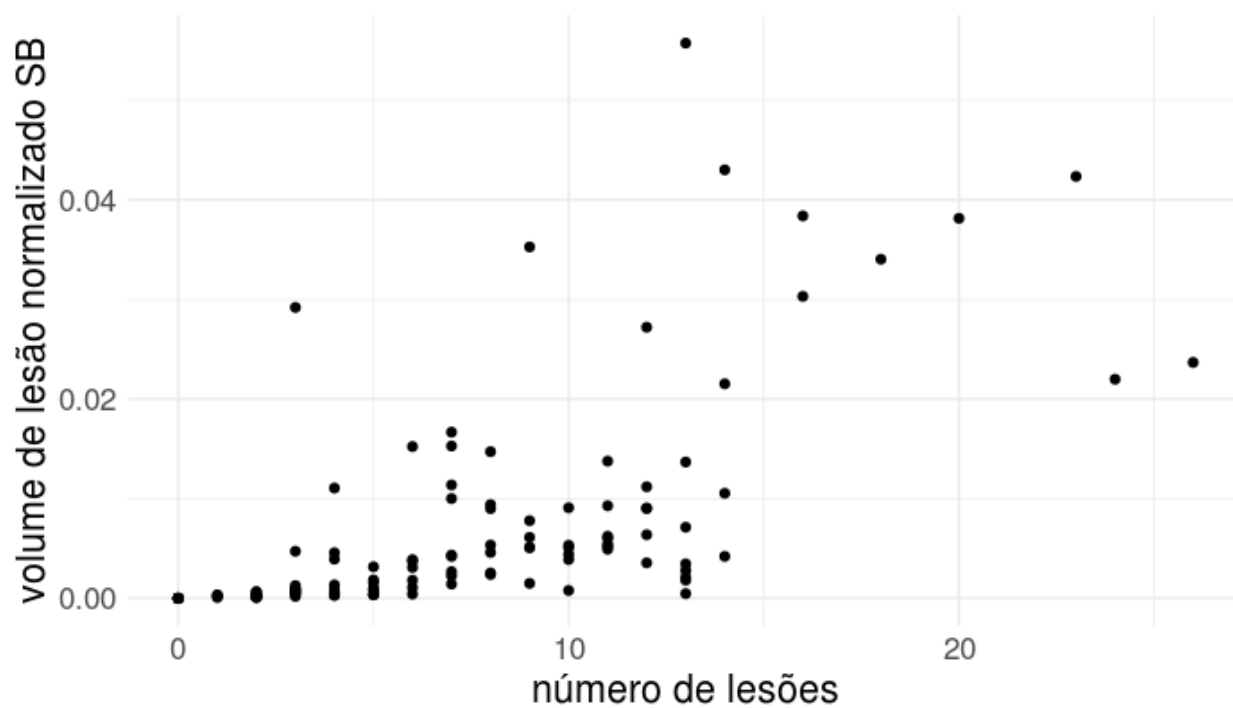


Figura B.7 Gráfico de dispersão do volume de lesão normalizado SB e número de lesões.

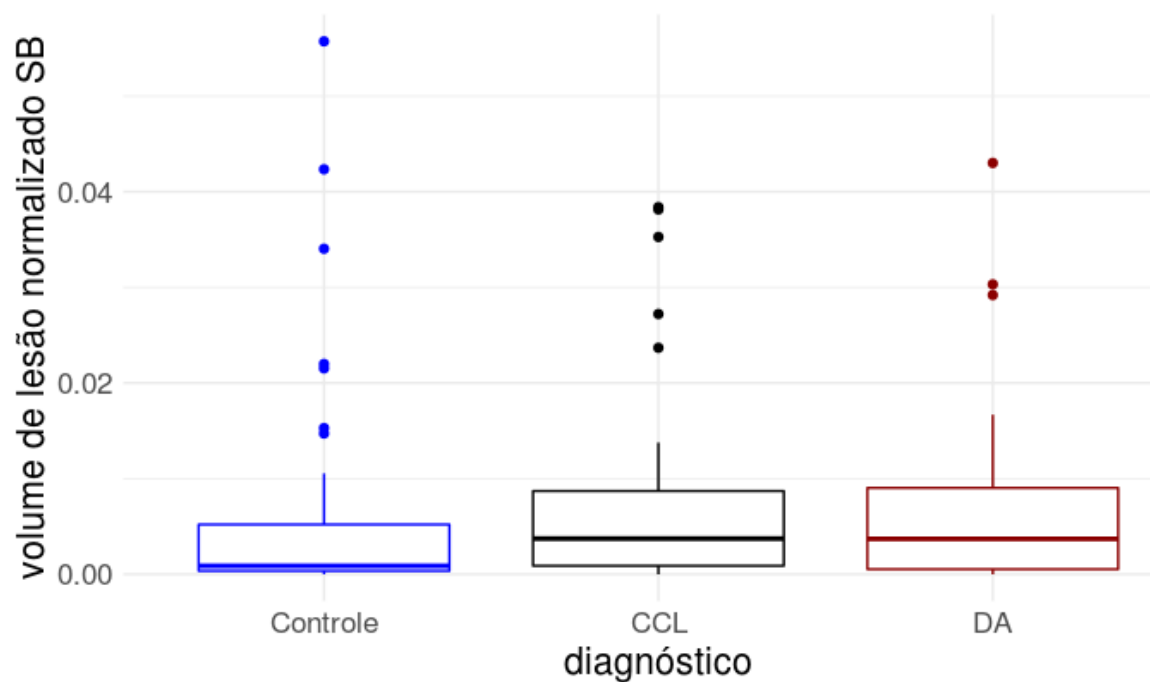


Figura B.8 Box plot do volume de lesão normalizado SB por diagnóstico do paciente.

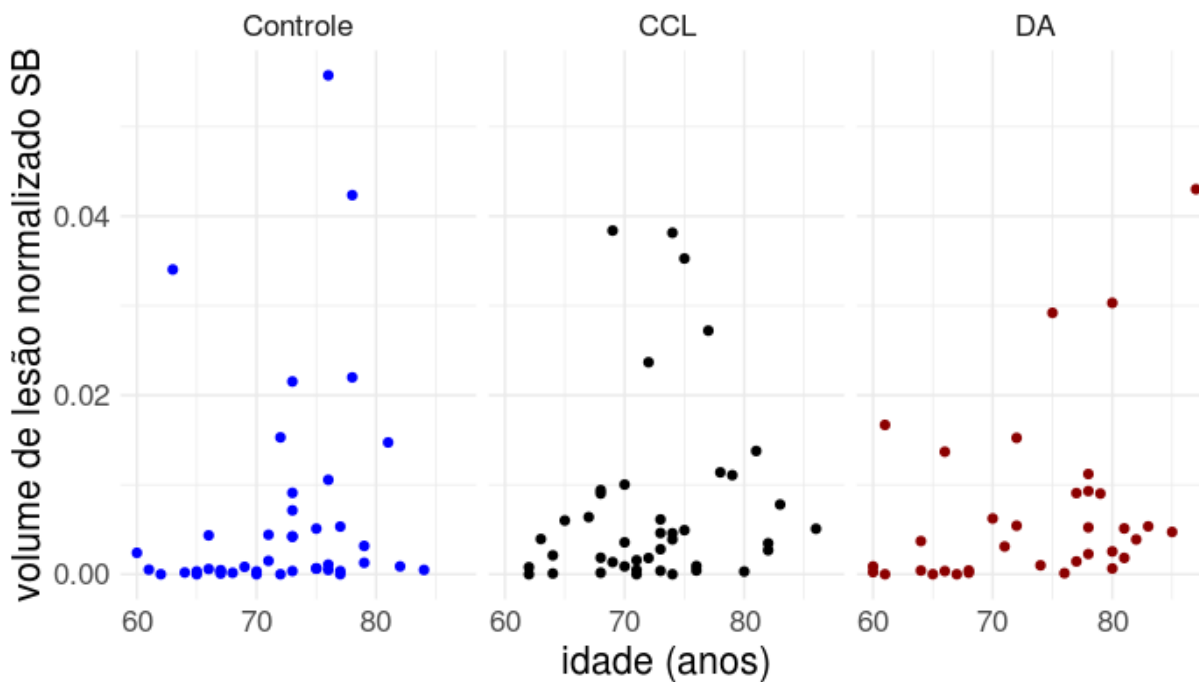


Figura B.9 Gráficos de dispersão do volume de lesão normalizado SB e idade por diagnóstico.

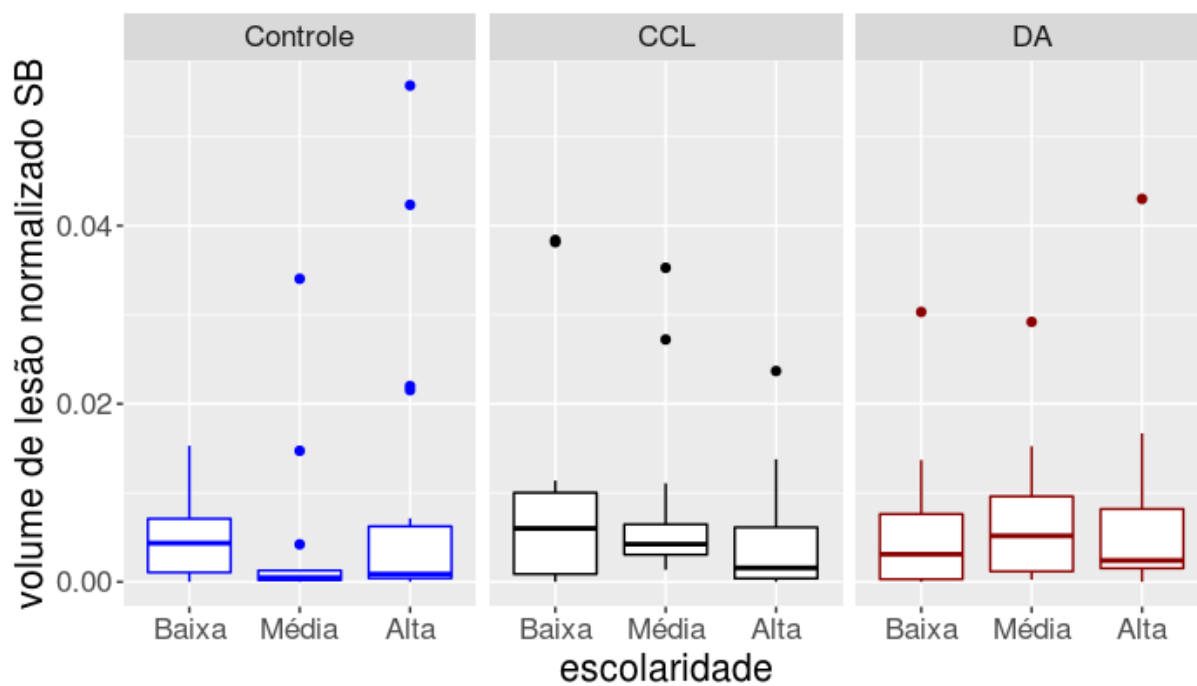


Figura B.10 *Box plot* do volume de lesão normalizado SB em cada categoria de escolaridade por diagnóstico do paciente.

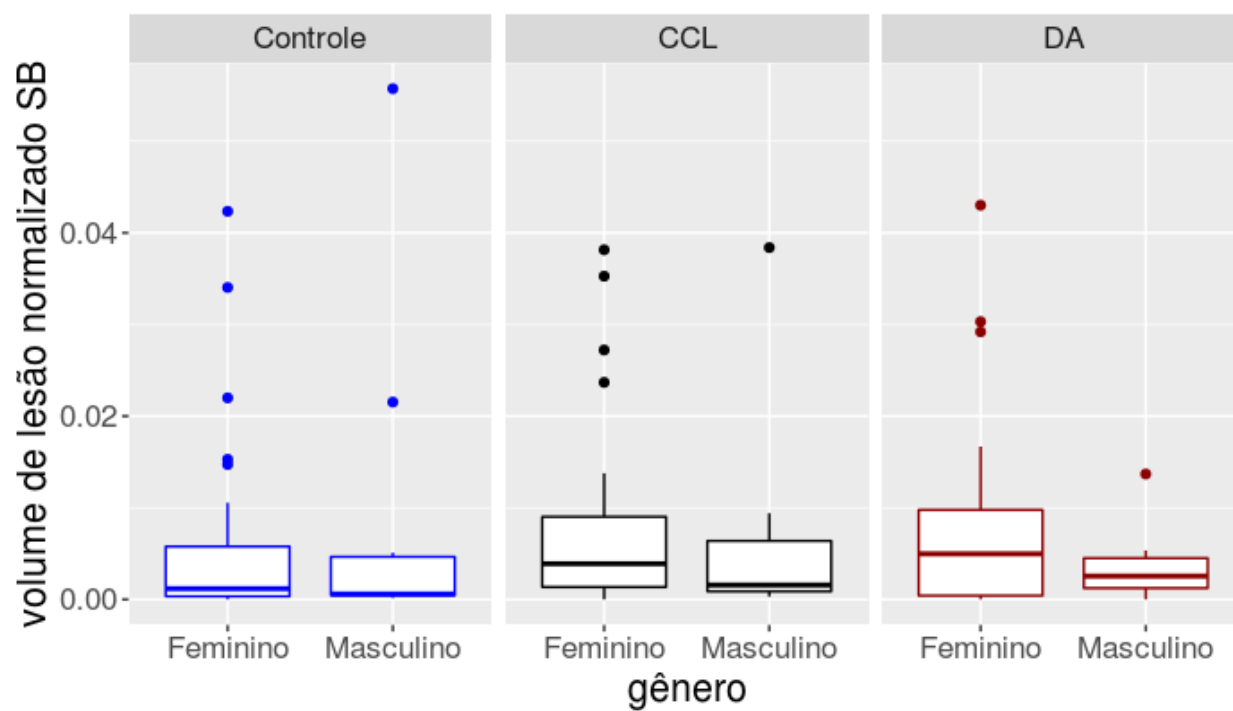


Figura B.11 *Box plot* do volume de lesão normalizado SB em cada gênero por diagnóstico do paciente.

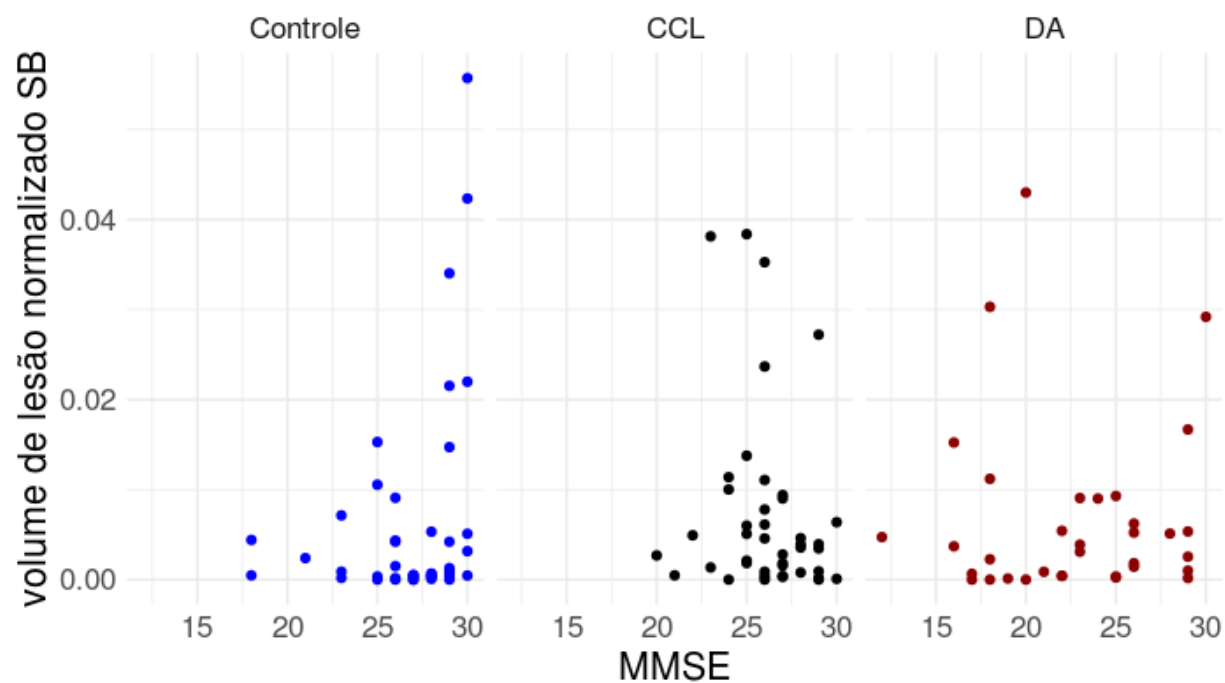


Figura B.12 Gráficos de dispersão do volume de lesão normalizado SB e MMSE por diagnóstico do paciente.

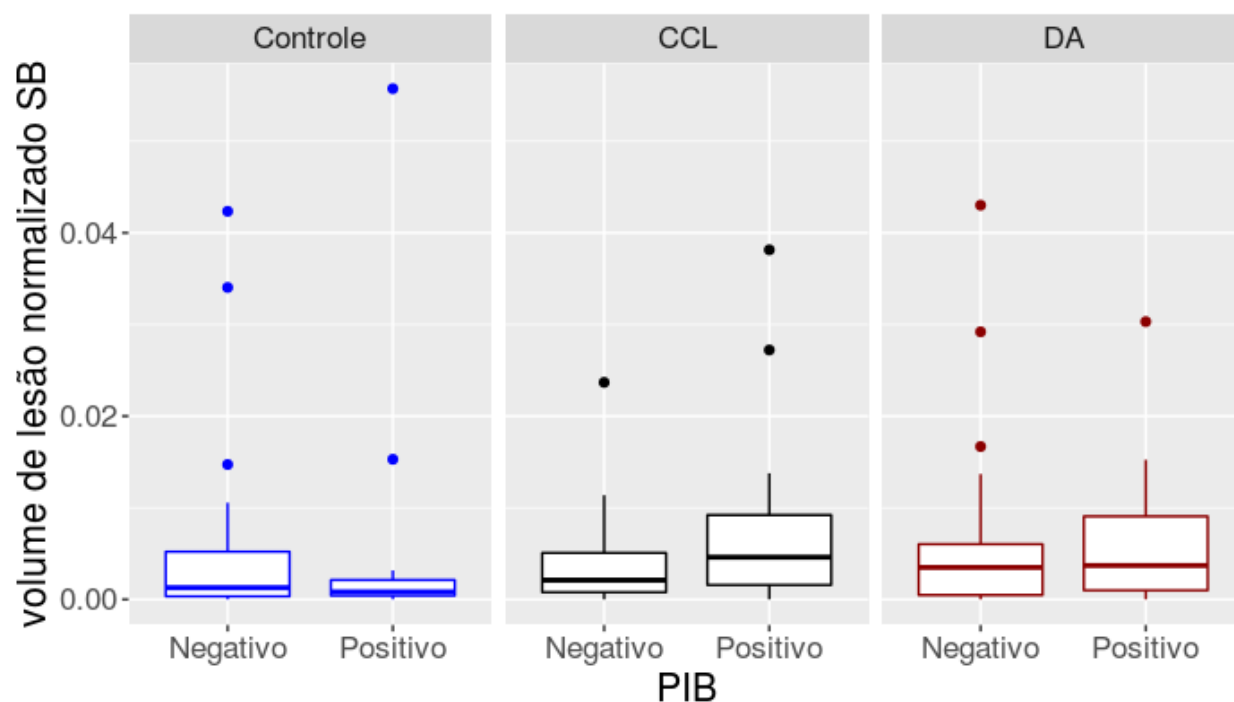


Figura B.13 Box plot do volume de lesão normalizado SB em cada categoria de PIB por diagnóstico do paciente.

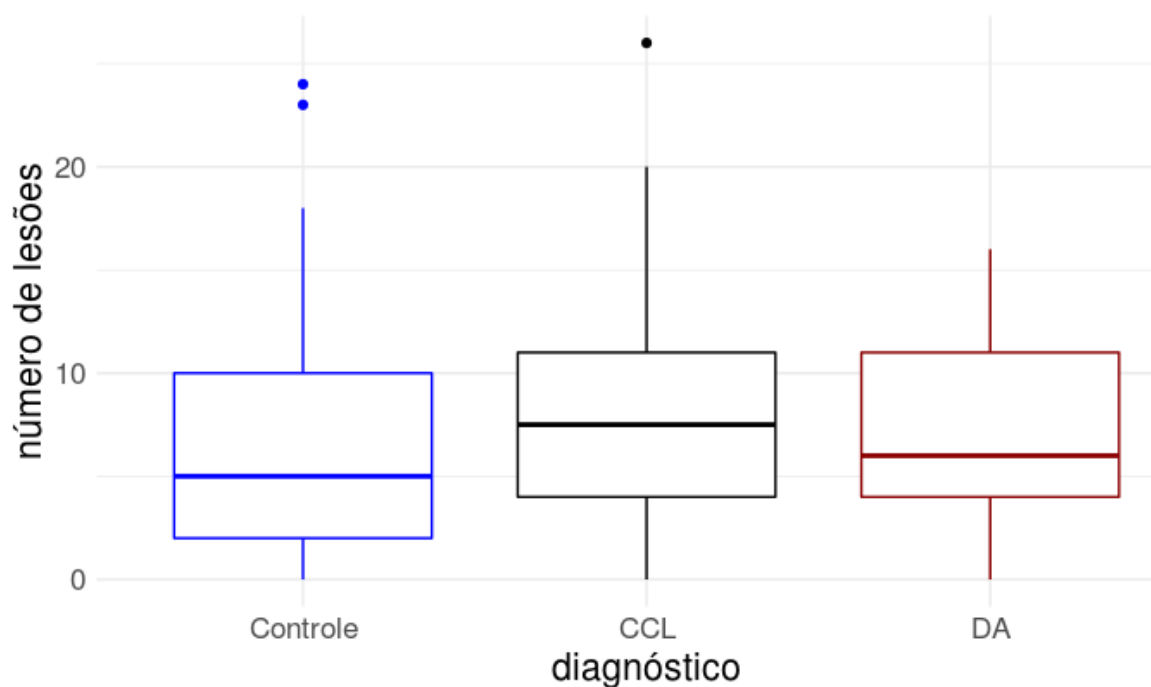


Figura B.14 Box plot do número de lesões por diagnóstico do paciente.

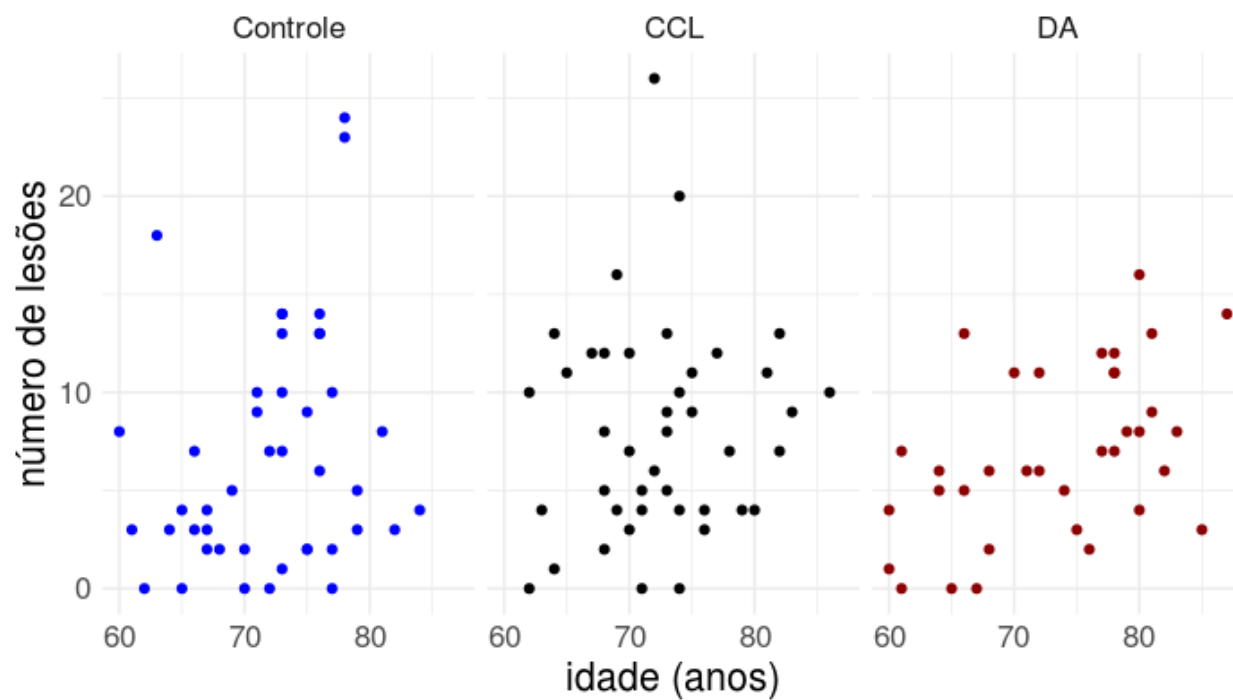


Figura B.15 Gráficos de dispersão do número de lesões e idade por diagnóstico do paciente.

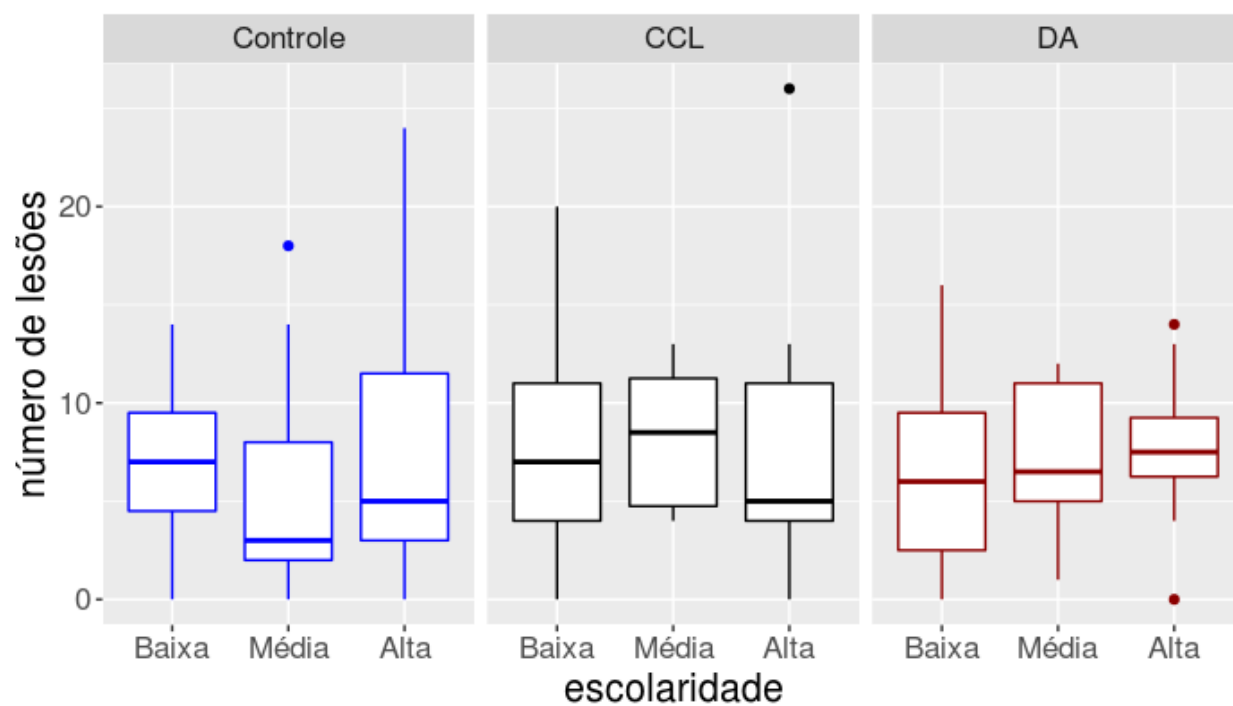


Figura B.16 Box plot do número de lesões em cada categoria de escolaridade por diagnóstico do paciente.

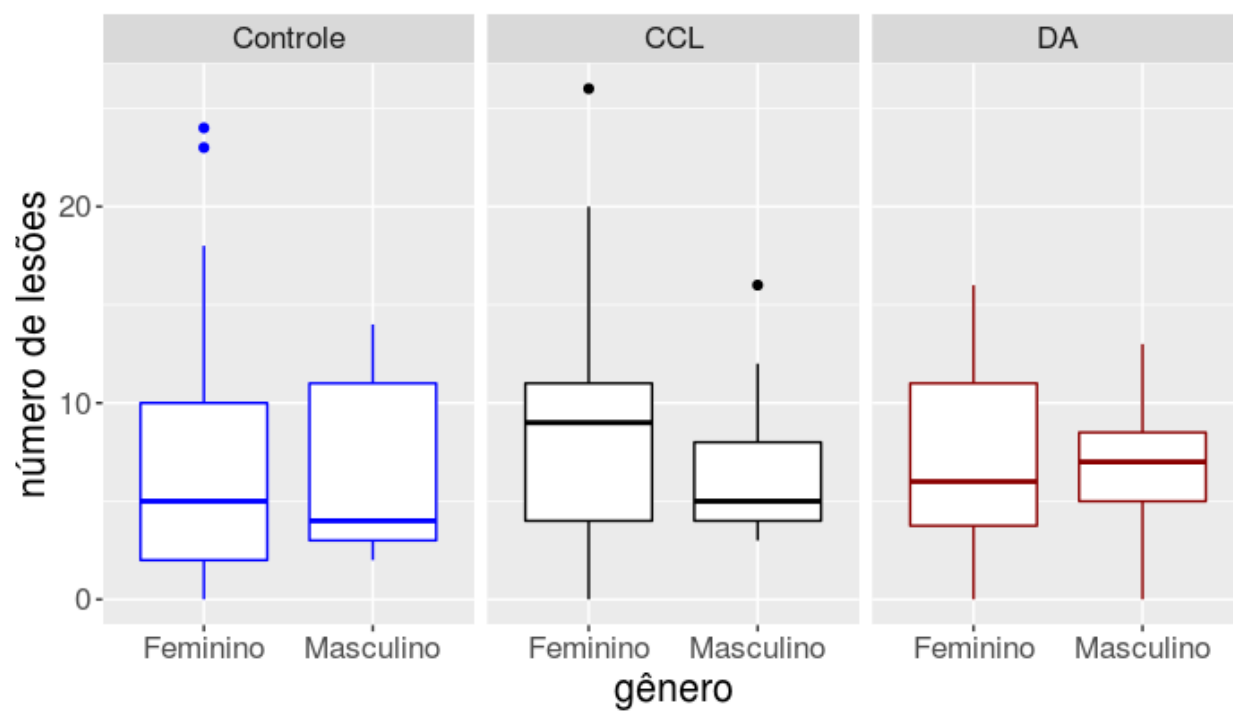


Figura B.17 Box plot do número de lesões em cada gênero por diagnóstico do paciente.

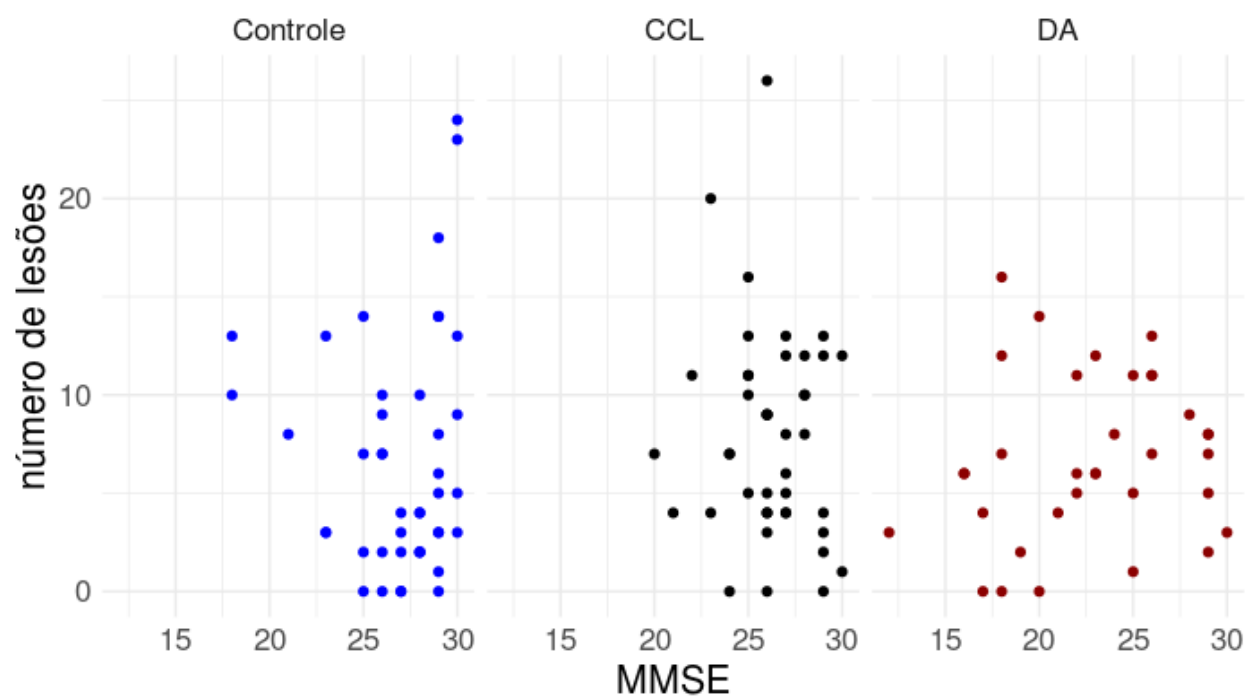


Figura B.18 Gráficos de dispersão do número de lesões e MMSE por diagnóstico do paciente.

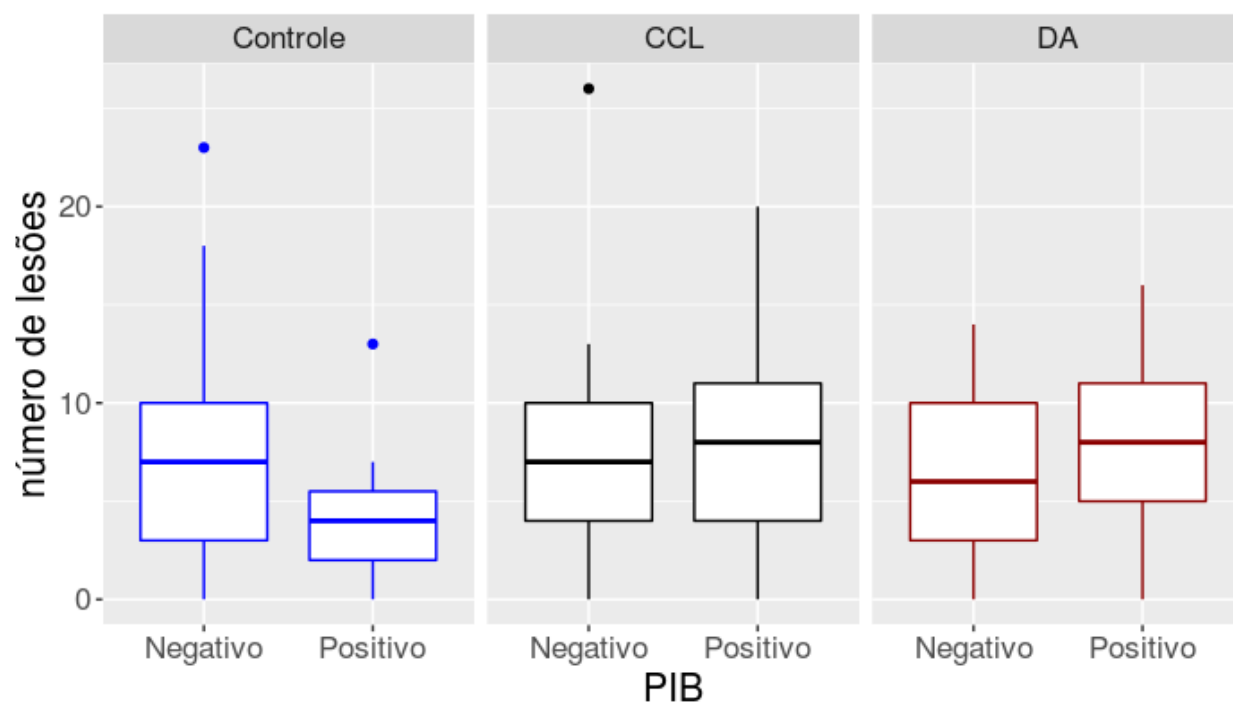


Figura B.19 Box plot do número de lesões por diagnóstico do paciente em cada categoria de PIB.

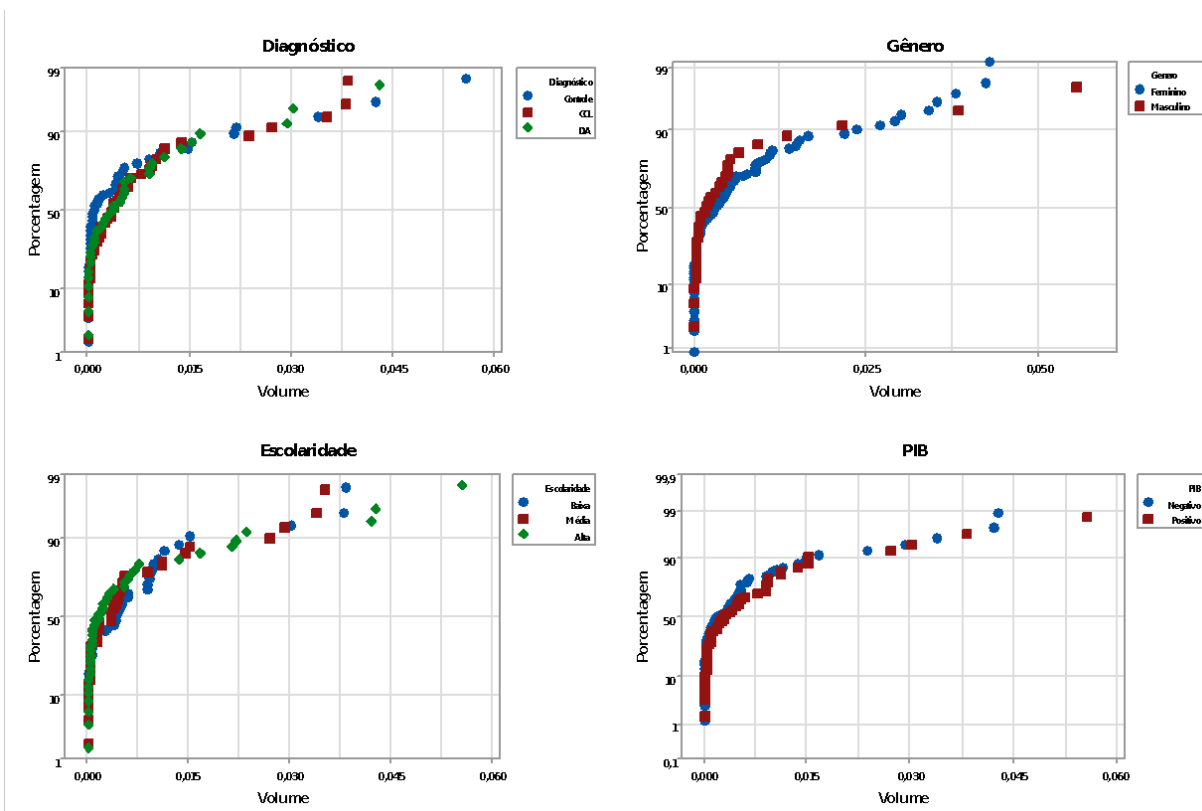


Figura B.20 Gráficos de probabilidade normal do volume de lesão normalizado SB em cada categoria de diagnóstico, gênero, escolaridade e PIB.

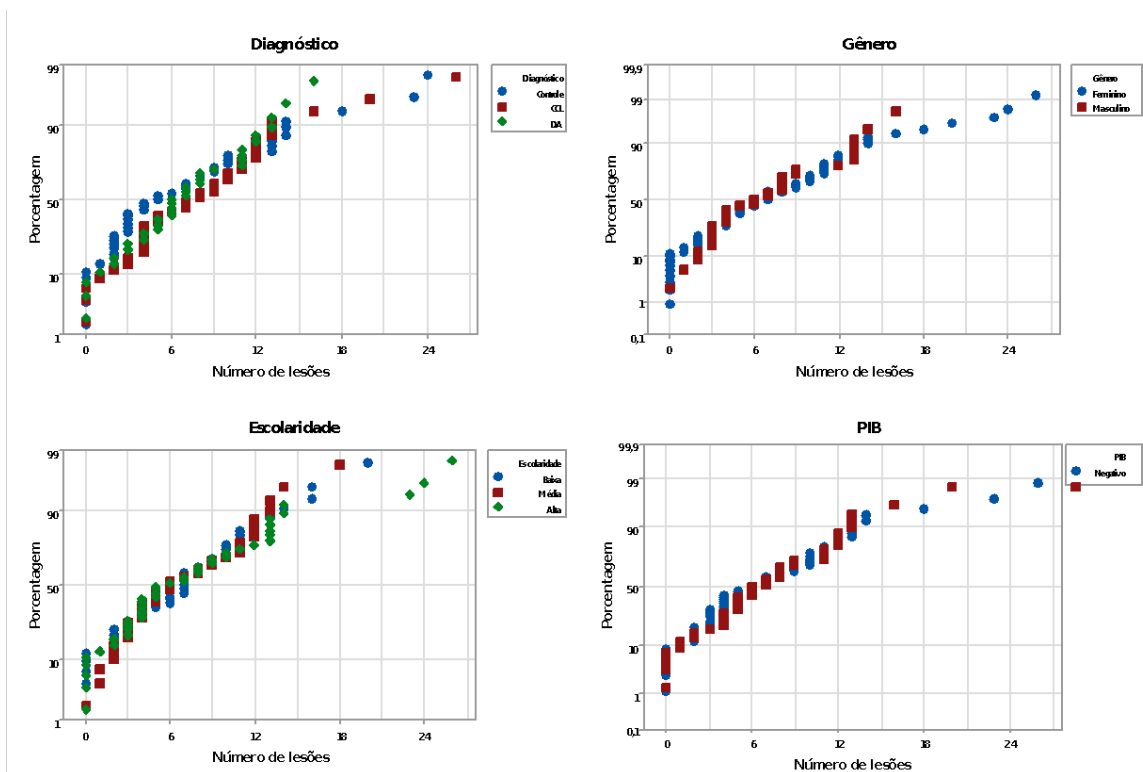


Figura B.21 Gráficos de probabilidade normal do número de lesões em cada categoria de diagnóstico, gênero, escolaridade e PIB.

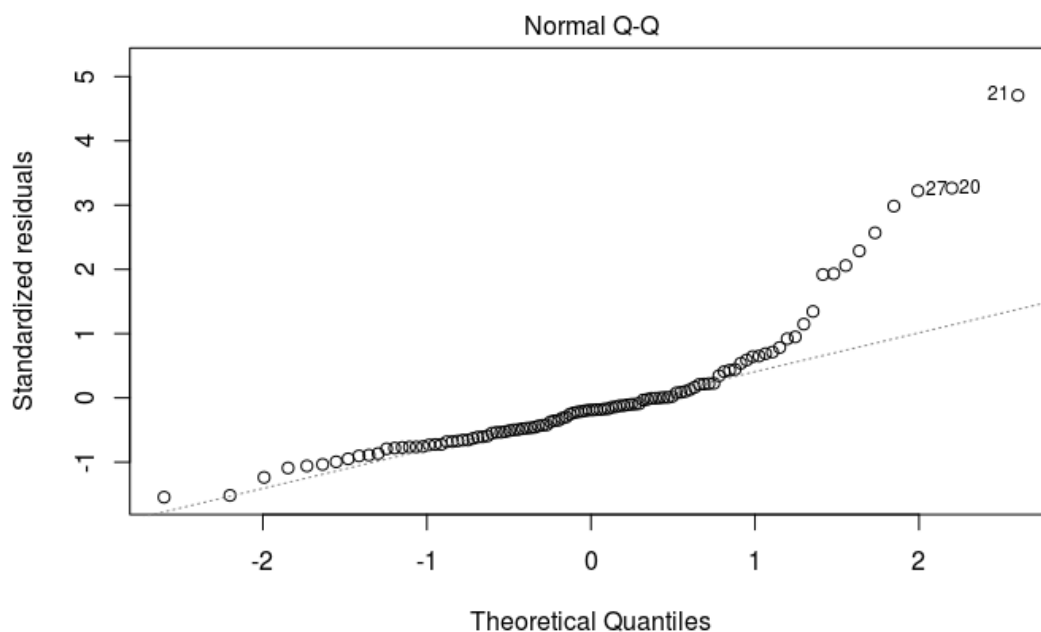


Figura B.22 Gráfico de probabilidade Normal dos resíduos do modelo linear de volume de lesão normalizado SB.

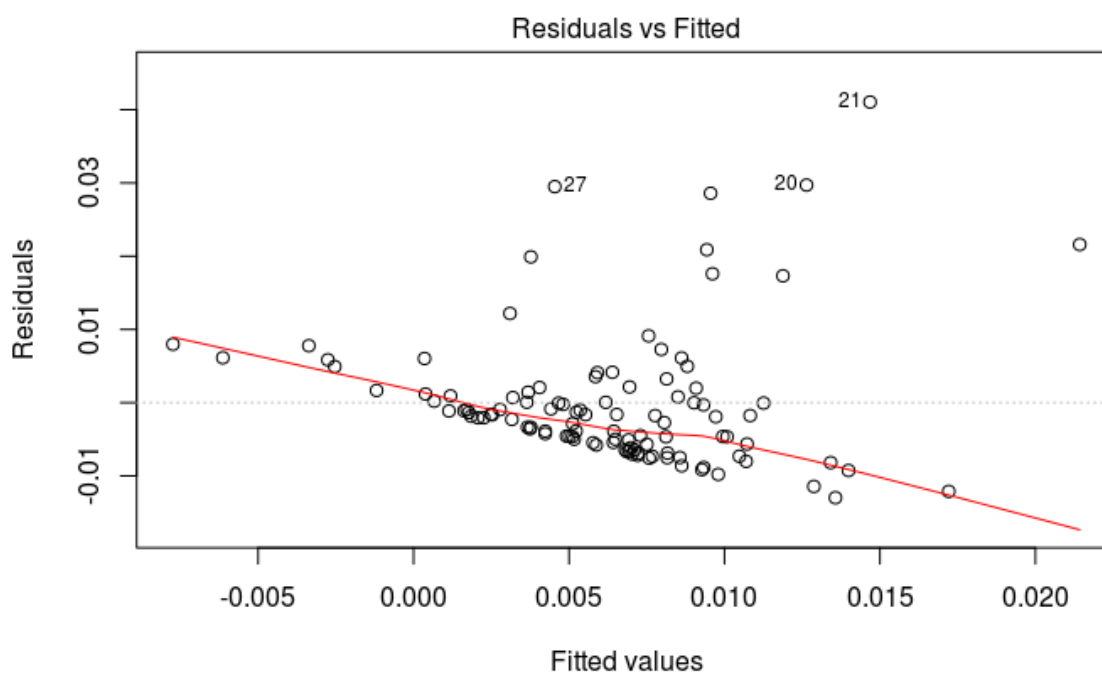


Figura B.23 Gráfico de resíduos pelos valores ajustados do modelo linear de volume de lesão normalizado SB.

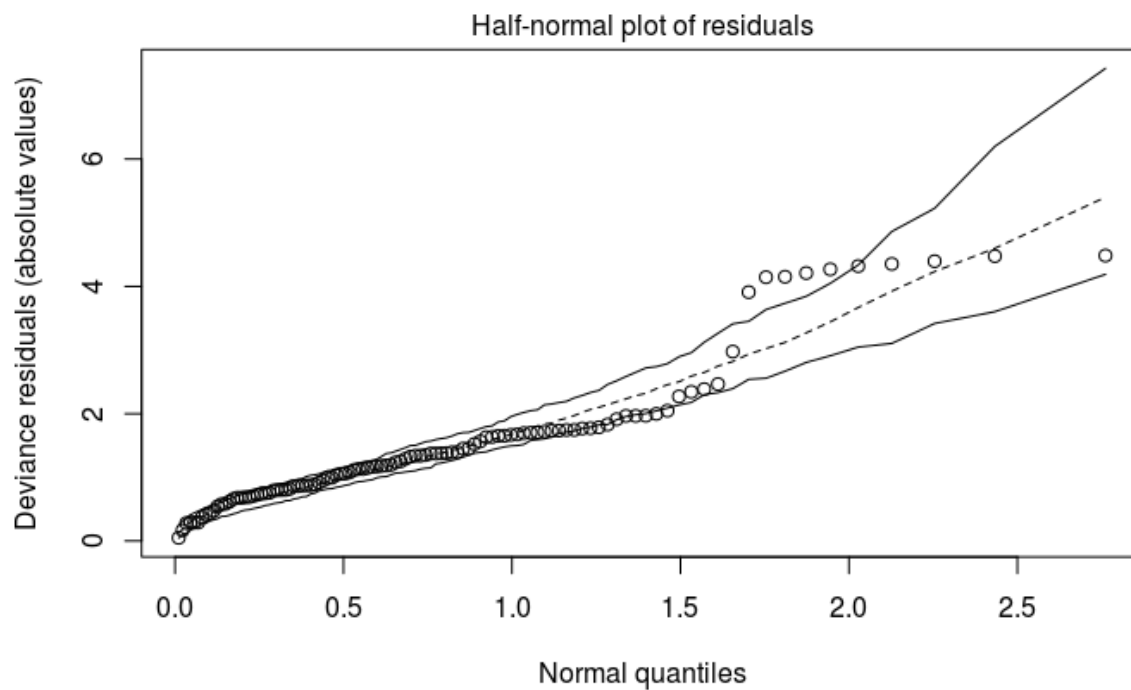


Figura B.24 Gráfico de probabilidade meio normal com envelopes simulados do resíduo do desvio do modelo Beta cuja variável resposta é volume de lesão normalizado SB.

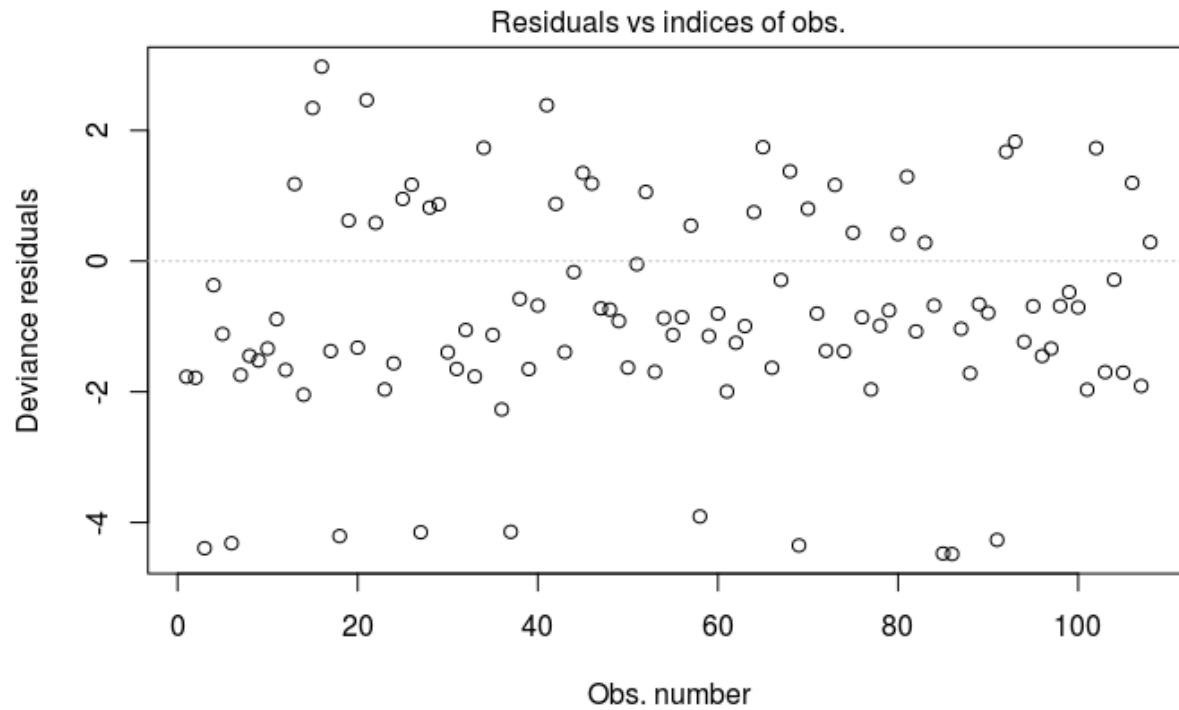


Figura B.25 Gráfico dos resíduos do desvio versus índice das observações do modelo beta cuja variável resposta é volume de lesão normalizado SB.

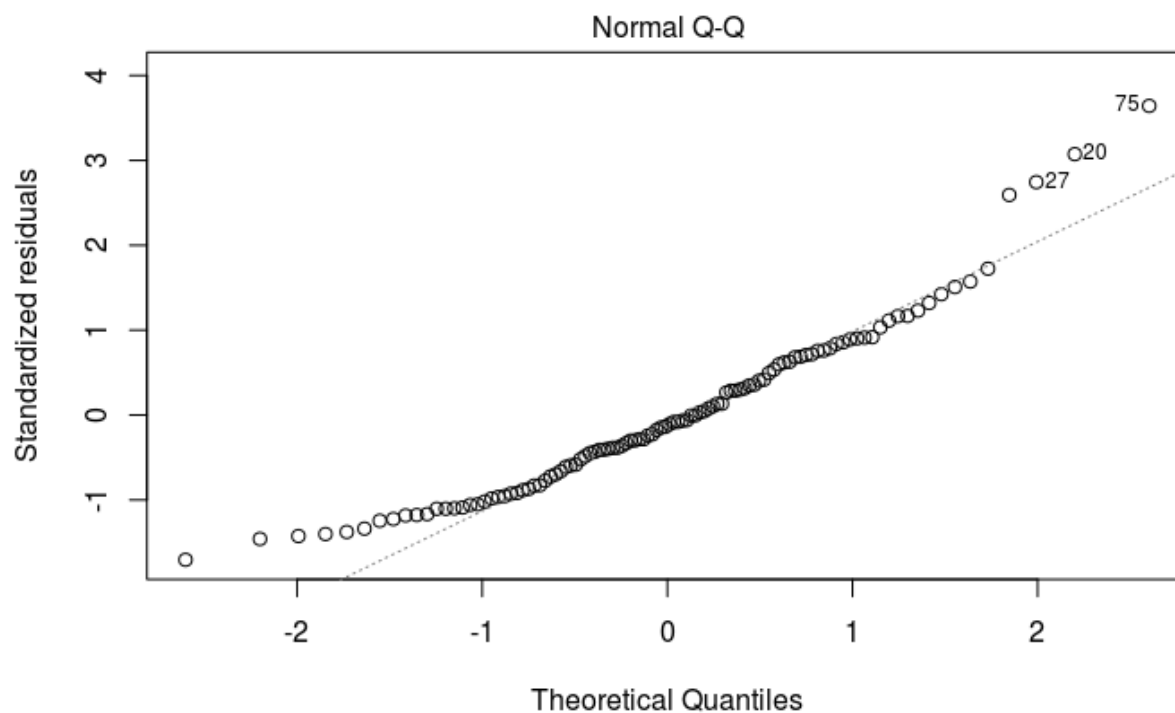


Figura B.26 Gráfico de probabilidade Normal do modelo linear de número de lesões.

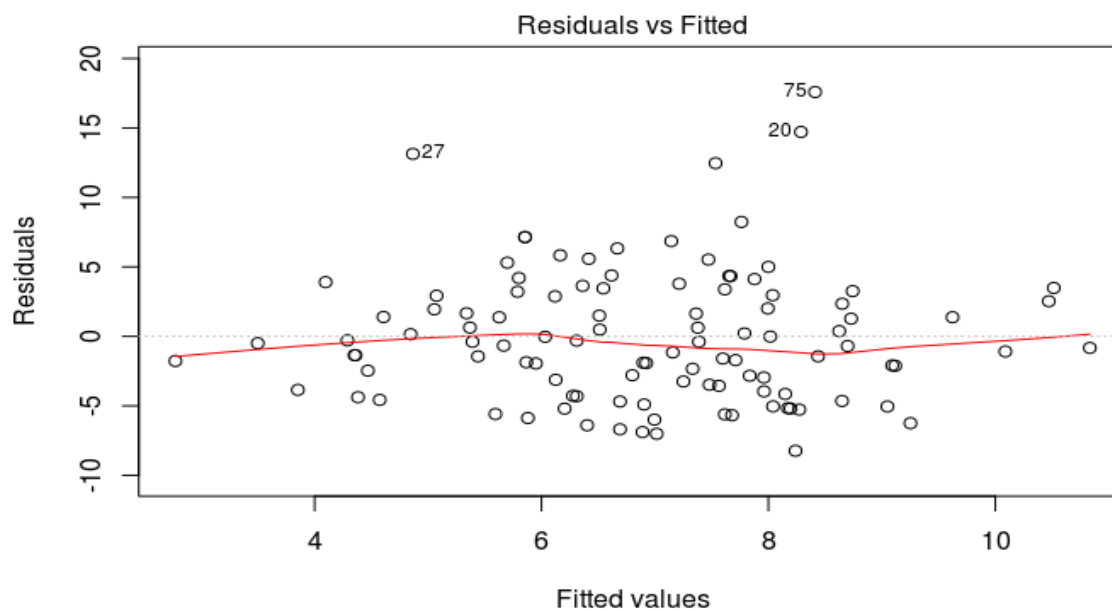


Figura B.27 Gráfico de resíduos pelos valores ajustados do modelo linear de número de lesões.

Gráfico Normal de Probabilidades

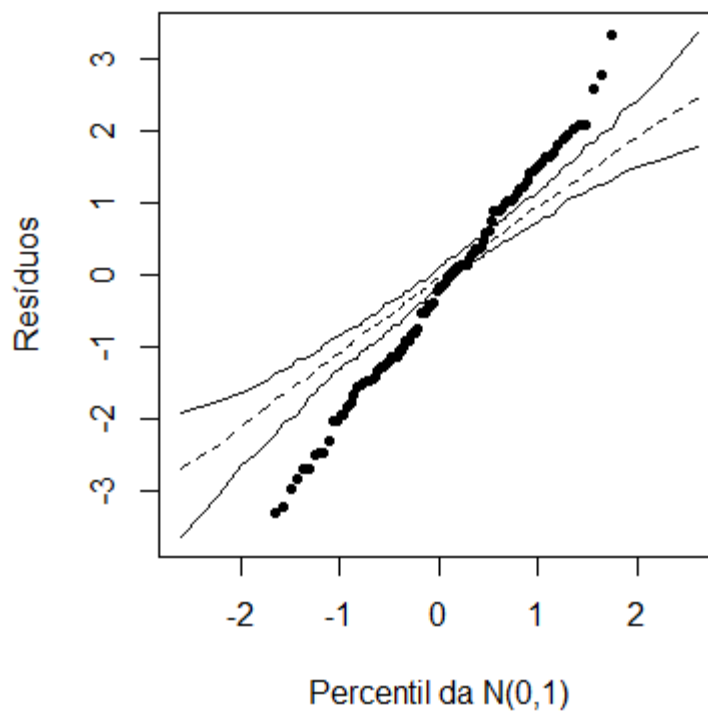


Figura B.28 Gráfico de probabilidade meio normal com envelopes simulados dos resíduos do modelo de Poisson cuja variável resposta é número de lesões.

Gráfico Normal de Probabilidades

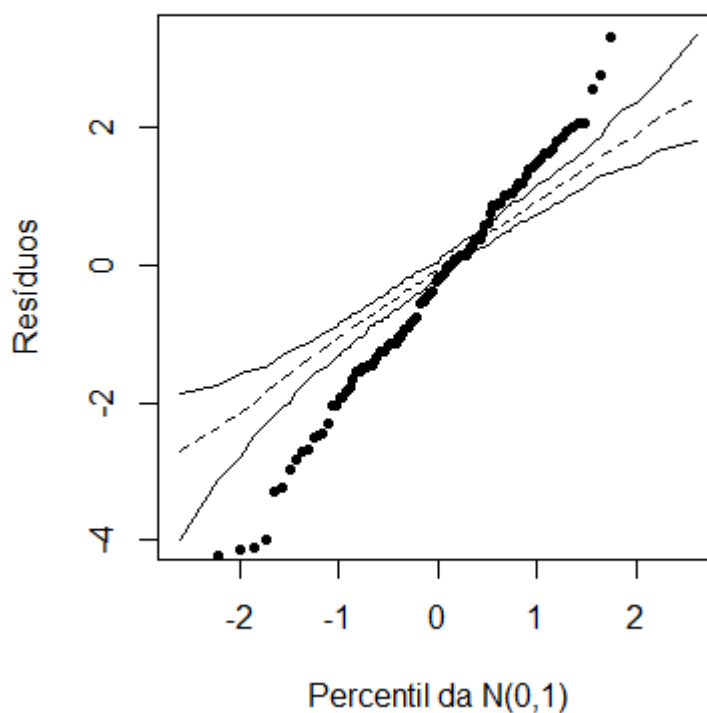


Figura B.29 Gráfico de probabilidade meio normal com envelopes simulados dos resíduos do modelo Binomial Negativa cuja variável resposta é número de lesões.