

**UNIVERSIDADE FEDERAL DO PARANÁ
CURSO DE BACHARELADO EM ESTATÍSTICA**

FABIO HENRIQUE SCHROEDER – GRR20124691
BRUNO HENRIQUE ABREU – GRR20159983
ANDRE LUIZ GRION – GRR20159284

Relatório: CURVA DE CRESCIMENTO USANDO MODELO MISTO:
UMA APLICAÇÃO NA PROGRESSÃO DA DOENÇA DE MACHADO-
JOSEPH

Curitiba/PR
2018

RELATÓRIO: CURVA DE CRESCIMENTO USANDO MODELO MISTO: UMA
APLICAÇÃO NA PROGRESSÃO DA DOENÇA DE MACHADO-JOSEPH

Resumo apresentada para a disciplina Análise
de dados longitudinais, no curso de Estatística,
da Universidade Federal do Paraná – UFPR.

Prof. José Luiz Padilha Silva

Curitiba/PR
2018

INTRODUÇÃO

O artigo em questão foi escrito objetivando ajustar curvas de crescimento evolução para descrever a progressão da DMJ (doença de Machado-Joseph), quantificada pelo escore NESSCA, utilizando o modelo de coeficientes aleatórios. Essa doença também é conhecida como ataxia spinocerebellar tipo 3 (SCA3), é uma desordem neurodegenerativa autossômica dominante caracterizada pela ataxia cerebelar progressiva. É uma doença hereditária, progressiva, de manifestação tardia (geralmente na idade adulta) e devida a uma mutação em um gene localizado no cromossoma 14q32.1. A enfermidade em questão tem como quadro clínico a perda da coordenação motora, onde inicialmente é afetado a capacidade de andar e posteriormente a fala, também podendo haver alterações oculares, por fim em estágios avançados da doença o portador não se move, e possuem grande dificuldade em falar. O primeiro caso a ser diagnosticado com tal enfermidade ocorreu recentemente, nos anos de 1970.

Para o estudo em questão foram utilizados uma coorte de pacientes do Hospital de Clínicas de Porto Alegre (HCPA) portadores de DMJ, que foi acompanhada por um período de 10 anos. A coorte em questão teve o objetivo de estudar a história natural da doença, através de avaliações clínicas realizadas com esses pacientes. Como indicador de comprometimento clínico do paciente foi utilizado o escore de exame neurológico para ataxia spinocerebellar (Neurological Examination Score for Spinocerebellar Ataxia - NESSCA), que é um escore baseado em uma avaliação quantitativa do exame neurológico padronizado e seu foco está nas principais características, em geral, da SCA e, em particular, da DMJ. A escala NESSCA produz um escore que varia de 0 a 40, onde 40 é comprometimento total do paciente, essa escala é produzida através de 18 itens.

Os dados em questão apresentam estrutura longitudinal não-regulares com estrutura desbalanceada, ou seja, há sujeitos que possuem informações faltantes ao longo do período acompanhado, entretanto para alguns sujeitos foi realizada uma imputação clínica do tipo recordatório, essa imputação ocorreu para as observações faltantes no intervalo entre uma avaliação e outra quando o escore de NESSCA não houve variação ou ainda quando não houve uma grande variação, sendo assim utilizado os sintomas retratados pelo paciente para estimar o momento que houve alteração no escore.

Em relação ao comprimento da mutação no gene ATXN3 que causa a DMJ, a mensuração desta sequência repetitiva aumentada definiu a variável comprimento da mutação. Com relação ao início da doença, considerado como o aparecimento dos primeiros sintomas, ressalta-se que este foi datado pelo relato do próprio sujeito ou familiar próximo. Sujeitos sem a DMJ têm entre 10 a 40 repetições nesta sequência, já portadores da DMJ têm entre 62 a 86.

ANÁLISE

Inicialmente foi realizado uma análise exploratória dos dados para verificar o comportamento dos sujeitos em determinadas variáveis, objetivando verificar medidas centrais e de dispersão.

Posteriormente foi realizado o ajuste ao modelo de coeficientes aleatórios, que tem como objetivo descrever aritmeticamente a relação entre as observações e o tempo. No modelo em questão são estimadas as curvas de crescimento para cada sujeito, e também a curva de crescimento dos grupos, que consideram os parâmetros das curvas de crescimentos individuais, e, ainda, a curva de crescimento média. No artigo em questão foi considerada linear a relação entre o escore de NESSCA e o tempo, que segundo os autores acontecem nas aplicações mais comuns a essa modelo. Os autores ainda indicam a importância da flexibilidade desses modelos no trabalho realizado, viabilizando o uso de dados desbalanceados. O modelo de coeficientes aleatórios é um caso especial do modelo misto, pois os coeficientes da regressão das curvas individuais, serão considerados efeitos aleatórios e os coeficientes de regressão das curvas médias, efeitos fixos. No modelo em questão os efeitos aleatórios permitem que além, das covariáveis, características intrínsecas de cada indivíduo também sejam consideradas, sendo a variabilidade dos efeitos aleatórios denominada de variabilidade entre sujeitos, pois mensura a evolução distinta de cada enfermo. Já a variabilidade do erro aleatório é denominada de variabilidade intra-sujeitos, pois mede o quanto as observações de um mesmo sujeito se desviam da reta média.

No que tange o objetivo do estudo, foi utilizado o NESSCA como variável resposta e idade no início da doença, comprimento da mutação como variáveis explicativas, sendo essas duas ora incluídas separadamente, ora incluídas juntamente na formulação do modelo. Resultando assim em 3 modelos distintos.

- (1)
- $$E[\text{nessca}_{ij}] = (\alpha_0 + \alpha_1 \text{idinic}_i) + (\beta_0 + \beta_1 \text{idinic}_i) \text{tempo}_{ij}$$
- (2)
- $$E[\text{nessca}_{ij}] = (\alpha_0 + \alpha_1 \text{cagexp}_i) + (\beta_0 + \beta_1 \text{cagexp}_i) \text{tempo}_{ij}$$
- (3)
- $$E[\text{nessca}_{ij}] = (\alpha_0 + \alpha_1 \text{idinicd}_i + \alpha_2 \text{cagexpd}_i) + (\beta_0 + \beta_1 \text{idinic}_i + \beta_2 \text{cagexp}_i) \text{tempo}_{ij}$$

Onde no modelo (1) apresenta a covariável idade no início da doença de forma contínua, (2) em sua formulação foi utilizado a covariável comprimento da mutação contínua e por fim (3) apresenta as covariáveis idade no início da doença e comprimento da mutação dicotomizados.

RESULTADOS

Os resultados encontradas na análise exploratória não levam em conta o fator tempo de evolução da enfermidade, logo os resultados apenas, indicam que há uma grande amplitude na idade no início da doença quando considerado a amostra total, quanto às análises, ainda exploratória, provenientes da amostra dicotomizadas é verificado que para a dicotomização do atributo comprimento da mutação, os sujeitos com comprimento da mutação inferior a 74, possuem, em média, idade no início da doença maior e escore NESSCA menor, em relação aos sujeitos com comprimento da mutação maior ou igual a 74; Já para o atributo idade no início da doença, os sujeitos com idade no início da doença inferior a 34 anos apresentam, em média, escore NESSCA maior em relação aos sujeitos com idade no início da doença maior ou igual a 34 anos. Logo a análise exploratória dá apenas uma sugestão de que uma menor idade no início da doença e um maior comprimento da mutação estejam relacionados com um maior grau de comprometimento na progressão da DMJ.

No artigo é apresentado um gráfico de curvas individuais para NESSCA, onde é possível verificar uma grande variabilidade no comportamento das curvas, indicando que há uma considerável variação na manifestação da doença ao longo dos anos, além de apresentar um comportamento crescente.

Com o uso do modelo longitudinal pode ser testado de maneira confiável as diferenças encontradas. Com a utilização do modelo (1), é possível afirmar que quanto maior a idade inicial, menor é o acréscimo na inclinação da reta, portanto, de forma mais lenta é a progressão da doença. Através da matriz de variância-covariância dos efeitos aleatórios estimados foi verificado que a variação nos interceptos é 35,75, enquanto que 0,47 representa a variação nas inclinações. Segundo os autores, isso indica que a variabilidade entre os sujeitos é maior no que diz respeito ao escore da NESSCA no início da doença do que na velocidade de evolução da doença. Já a covariância entre os interceptos e as inclinações é -2,88, indicando relação negativa entre eles. Além disso, a variância residual (variabilidade intrasujeitos) é igual a 0,95.

No modelo (2), considerando como covariável o comprimento da mutação de maneira contínua, é possível estimar que, com o acréscimo de uma unidade no comprimento da mutação, ocorre o aumento de 0,16 na inclinação da reta média. Ou seja, quanto maior o comprimento da mutação mais rápida acontece a progressão da doença, como no modelo (1) é observado uma relação negativa entre os interceptos e inclinações das retas individuais, além de ser verificado através da variância residual que há mais variabilidade entre sujeitos que intrasujeitos. Com o auxílio do gráfico das retas ajustadas, é verificado que sujeitos com maior comprimento da mutação possuem um valor baixo no escore NESSCA no começo da doença, porém com um desenvolvimento acelerado. Em contrapartida indivíduos com comprimento da mutação baixo no começo apresentam escores mais altos e o desenvolvimento da doença se dá de forma mais lenta. Sugerindo que o comprimento da mutação, pode ser considerado fator de risco para a doença.

No modelo (3), que contém de maneira dicotomizadas a idade do indivíduo no início da doença e o comprimento da mutação, é possível afirmar que considerando uma mesma categoria de idade inicial, que os sujeitos com comprimento da mutação maior ou igual a 74 possuem inclinação 0,59 unidades

superior, em relação aos sujeitos de comprimento da mutação inferior a 74. Portanto, as informações sugerem que os sujeitos com comprimento da mutação inferior a 74 possuem progressão da doença de forma mais lenta em relação a outra classe. Para esse modelo em questão também é observada um valor alto da variância do efeito aleatório em relação à variância residual o que indica que há mais variabilidade entre sujeitos que dentro de sujeitos. E com base no gráfico das retas ajustadas, é verificado que sujeitos que com comprimento de mutação igual ou maior de 74, independentemente da idade no início da doença, iniciam com escores mais baixos, mas possuem avanço mais acentuado da doença ao longo dos anos.

Em relação a idade inicial, foi identificado que enfermos com idade no início da doença inferior a 34 anos, possuem um escore mais alto no início da doença e apresentam desenvolvimento da doença de forma mais acelerada em relação aos sujeitos com idade no início da doença maior ou igual a 34 anos, independente da categoria de comprimento da mutação.

Os autores ainda ressaltam que os resultados são passíveis ao viés de recordação.

ANÁLISE CRÍTICA DO TRABALHO

A introdução do trabalho foi toda focada no problema de análise de dados longitudinais, dando ênfase a solução via modelos com efeitos aleatórios. Os autores explicitam o objetivo de estudar a evolução da doença na introdução mas não fazem qualquer menção às covariáveis estudadas. A explicação delas só veio na seção Material e Métodos, não deixando claro pro leitor, na introdução, o estudo que seria realizado em função da idade no início da doença e o comprimento da mutação que acabaram sendo a maior parte da discussão do artigo.

Também faltou descrever o número de observações, ainda que média, que foi realizada em cada indivíduo. Na verdade essa é uma falha bastante considerável, principalmente pelo uso de imputação. Portanto, os autores deveriam ter, não só descrito mais adequadamente o número de observações efetivamente realizadas, como explicitar o número de observações imputadas. Ainda sobre a imputação, pelo que foi descrito, ela ainda pode ter sido a fonte de linearidade e de menor variabilidade observada entre indivíduos.

Os autores também não explicaram o porquê de considerar as variáveis dicotomizadas quando avaliadas conjuntamente. Nossa suspeita é de uma possível facilidade de interpretação para este caso.

Para análise dos dados, foram utilizados três *softwares*. O SPSS para análise descritiva, o SAS para os modelos mistos e *Microsoft Excel* para os gráficos. Esse tipo de atitude ou demonstra uma falta de capacidade dos *softwares* de realizar todo tipo de análise ou, pior, uma falta de conhecimento dos autores de conduzir em um mesmo *software* todas análises necessárias. Uma explicação mais plausível também seria a repartição do trabalho entre os autores, onde cada um teria ficado responsável por uma análise e a conduziram nos *softwares* os quais estariam mais familiarizados.

Nenhum comentário foi feito sobre dados faltantes na variável comprimento da mutação. Isso só é observado pelos valores demonstrados nas tabelas.

Na descrição do gráfico que representa o modelo 1 fala da diminuição do intercepto pela idade do início da doença quando na verdade esse efeito não foi significativo pelo teste de Wald, como demonstrado em uma tabela. O mesmo ocorre na

interpretação do modelo 2, onde foi dito que o aumento do comprimento da mutação leva a diminuição do valor do intercepto mesmo este não apresentando valor significativo. Neste último caso, mesmo citando a não significância estatística, os autores ainda afirmam que o maior comprimento da mutação diminui o intercepto da reta média, mesmo comentando que este resultado contraria a prática médica. Os autores insistem na explicação do problema relatando as dificuldades relacionadas com o viés recordatório e mesmo de aspectos mais técnicos de diagnóstico de doença. Lembrando, tudo isso para explicar um efeito que na verdade não foi significativo e cuja a explicação mais sensata do ponto de vista estatístico seria de que não existe efeito de comprimento da mutação no momento do diagnóstico da doença (intercepto), quando ao escore NESSCA. Os autores ainda insistem nesta discussão ao comentar o modelo 3 em que o intercepto relacionado ao comprimento da mutação novamente não foi significativo. Ainda com base no modelo 3, os autores perdem tempo explicando os efeitos da variável idade do início da doença que, também, não foram significativas.

Os autores até citam o fato de na avaliação conjunta da idade com o comprimento da mutação, o efeito da inclinação da idade deixar de ser significativa enquanto o efeito da inclinação do comprimento da mutação se mantém significativa mas não fala nada do confundimento e o quanto estas duas variáveis estão relacionadas para a explicação deste fato.

No final, os autores sugerem a possibilidade de uso de "modelos mistos não lineares" de forma superficial e colabora na confusão causada com o uso do termo "linear" como sinônimo reta na ilustração da relação entre as variáveis e o uso mais relacionado a consideração da combinação dos parâmetros a serem estimados.

Não poderíamos deixar de relatar também os pontos positivos do trabalho, como a boa qualidade na análise descritiva e exploratória das covariáveis, o detalhamento dos modelos, pelo menos da parte sistemática e a inclusão de códigos utilizados (do SAS). Principalmente sobre a descrição dos modelos e sobre os códigos expostos, acabam sendo muito úteis para um dos objetivos dos autores, explicitado somente no final do artigo, que é a demonstração do uso do método dos efeitos aleatórios e adequação de dados longitudinais, especialmente de dados desbalanceados, para que possam ser conhecidos e utilizados por pesquisadores de outras áreas.

Referência

HAUSER, L. et al. *CURVA DE CRESCIMENTO USANDO MODELO MISTO: UMA APLICAÇÃO NA PROGRESSÃO DA DOENÇA DE MACHADO-JOSEPH*. Rev HCPA., 2009;29.