

Testes de Hipóteses em Modelos Multivariados de Covariância Linear Generalizada

Estudante: Lineu Alberto Cavazani de Freitas - Avaliador: Daniel Furtado Ferreira*
Departamento de Estatística-Universidade Federal de Lavras-Cx. Postal 37 - Lavras-MG -Brasil
(Dated: 31 de agosto de 2021)

I. VISÃO GERAL

A dissertação (projeto de qualificação) tem como foco principal é o desenvolvimento de testes de hipóteses para os McGLMs, por se tratar de uma classe de modelos flexível e com alto poder de aplicação a problemas práticos.

A proposta é adaptar o teste Wald para a realização de testes de hipóteses gerais sobre parâmetros de McGLMs. Além disso tem-se como objetivos à implementação funções para efetuar tais testes, bem como funções para efetuar ANOVAs e MANOVAs.

As propriedades e comportamento dos testes propostos são verificados com base em estudos de simulação e o potencial de aplicação das metodologias discutidas será apresentado com base na aplicação a conjuntos de dados reais.

Agradeço aos orientadores Prof. Dr. Wagner Hugo Bonat e Prof. Dr. Marco Antonio Zanata Alves, bem como ao estudante Lineu Alberto Cavazani de Freitas a possibilidade de participar e colaborar um pouco com o trabalho nesta banca.

II. QUESTIONAMENTOS SOBRE A INTRODUÇÃO

O texto como um todo (não só da introdução) está bem escrito e algumas correções foram anotadas no arquivo pdf que será enviado ao estudante ou ao seu orientador. Seguem as principais questões sobre a introdução:

- a) (Página 9) Os modelos multivariados de covariância linear generalizada (McGLM) foram propostos por Bonat e Jørgensen (2016). O que tem de diferente em sua proposta em relação ao trabalho deles?
- b) Pelo que entendi do trabalho, e que provavelmente você me respondeu na pergunta do item (a) anterior, a diferença é a proposição de novos testes baseados no teste de Wald? Pode me explicar a ideia do teste que foi apresentada na página 11?

III. QUESTIONAMENTOS SOBRE O REFERENCIAL

Seguem as principais questões sobre o referencial:

- a) (página 14) Você escreveu: “Podemos imaginar que, substituindo esta matriz identidade no modelo (2.1) por uma matriz qualquer que reflita a relação entre os indivíduos da amostra teremos uma extensão do Modelo Linear Generalizado para observações dependentes. É justamente essa a ideia dos modelos de covariância linear generalizada, o cGLM.” Que é o autor desta proposta? Bonat e Jørgensen (2016)? Não teria que citar? Isso acontece na próxima seção da página 14 e 15. Me parece que é a proposta de seu orientador com Bent, logo deve ser citado. Acho que se esqueceu disso, pois seu trabalho estende o artigo dos dois.
- b) Na página 14, modelo 2.3 o que significa a constante D (Componentes) da fatoração apresentada?
- c) Nos modelos multivariados de covariância linear generalizada (McGLMs) da página 15 você pode considerar R modelos lineares diferentes, R funções de ligações diferentes e R vetores de médias e funções de variâncias diferentes. Isso não torna o modelo superparametrizado não?
- d) O que não ficou claro para mim é o surgimento da matriz Σ_b ($R \times R$) de covariância entre observações. De repente some-se com a $\Omega_r(\tau)$ e surge a Σ_b . Na verdade quando temos uma estrutura de vetores aleatórios multivariados correlacionados, o modelo de variância da matriz Y em (2.4) é o natural. Então sair da estrutura 2.3 modelando cada componente univariado para o 2.4, modelando o caso multivariado está correto, mas não há conexão entre os dois casos, em meu ponto de vista.
- e) Você apresenta uma revisão sobre ANOVA e MANOVA. No caso de modelos lineares generalizados qual é o nome apropriado para a “ANOVA correspondente”?

IV. QUESTIONAMENTOS SOBRE A PROPOSTA

Seguem as principais questões sobre a proposta:

- a) Uma de suas justificativas no início da seção 3 é que os McGLMs não são ajustados com base na maximização da função de verossimilhança para se usar o método de Wald. Isso é verdade, mas a literatura em peso retrata as estimativas como estimativas de máxima verossimilhança. Acho que não pode haver confusão do processo numérico de estimação, como o uso de quadrados mínimos iterativos ponderados ou o

* <http://www.des.ufla.br/~danielff>; danielff@ufla.br

método de Newton Raphson com o método estatístico de estimação. Assim, **acho que o uso do método de Wald é muito mais apropriado do que um mera adaptação como você menciona**, pelo fato de se ter as estimativas dos parâmetros e de sua matriz de covariâncias. Obs. Esta observação é feita anteriormente no texto, mas não marquei onde está.

- b) Na página 25, esta primeira parte da seção em questão, está parecendo uma introdução em que se faz justificativas, detalha objetivos e lança-se alguns questionamentos. Tem de colocar mais em um formato temporal do que vai ser feito do que questionamentos, metas e hipóteses lançadas.
- c) Na página 26, você usa h como número total de parâmetros e em sua revisão usou Q em um lugar e p em outro. Estou mencionando este exemplo aqui, para sugerir que haja uma padronização destes símbolos ao longo de todo o texto.
- d) Na página 26, equação 3.2, podemos usar uma justificativa muito forte. Se assumirmos que temos estimadores de máxima verossimilhança dos parâmetros, então 3.2 é uma FQ e assim, tem distribuição assintótica qui-quadrado com s graus de liberdade.
- e) Na página 29, está escrito que podemos ter diferentes preditores lineares para diferentes variáveis. Esta era uma pergunta que estava preparando para te fazer ao final da arguição. Então obviamente é possível. Estou colocando desta forma aqui, para chamar atenção que estas explicações deveriam estar no referencial e não na seção métodos. Tem muitas observações neste contexto marcada no texto pdf que te enviei (ou vou enviar) via seu orientador.
- f) Na página 30 seção 3.2, conforme já argui anteriormente, deveriam chamar análise de deviance e análise de deviance multivariada (ANOD e MANOD ou algo assim).
- g) Na página 30, você menciona os tipos I, II e III de SQ (no caso deviances). No caso de regressão, modelos de posto coluna completo, em minha opinião só tem sentido falar em tipo I e II. A menos que você irá ajustar modelos com fatores qualitativos com vários níveis. Irá?
- h) Nas páginas 32 e 33 você apresenta os casos tipo II e III com um modelo com x_1x_2 nele. Filosoficamente você pode considerar este produto uma interação, que

na verdade é o componente linear dela, mas em minha opinião a melhor forma de ajustar é um efeito para todos os outros, ou seja, ajustando o modelo completo e zerando o componente de interesse via hipótese nula (seria a tipo III). Nunca pensei em ignorar um efeito multiplicativo de duas variáveis em um modelo. Estou surpreso e gostaria que opinasse sobre isso.

- i) Ainda sobre a MANOVA e ANOVA qual é o teste aplicado para cada FV? Na seção seguinte você menciona que usou o pacote *car* para isso.
- j) Faltou você planejar como seria sua simulação nesta seção; Como vai avaliar o desempenho do teste, o que vai medir, o que vai considerar em termos de configurações.

V. QUESTIONAMENTOS SOBRE OS RESULTADOS

Seguem as principais questões sobre os resultados:

- a) Na seção 4.2 você menciona a ideia de implementar testes de comparações múltiplas. Daí pela primeira vez fala em variáveis categóricas (fatores), saindo do modelo clássico de regressão de posto coluna completo. Louvável. Mas se isso for realmente feito, isso deve ser planejado desde a introdução, objetivos, revisão. Na seção método, deverá ser descrito como será feito, etc. Pessoalmente, acho muito coisa para ser feita em uma dissertação de mestrado.
- b) Somente no primeiro parágrafo da página 37 é mencionado de forma ainda superficial a validação do teste. O local é incorreto (deveria ser seção 3) e teria que dar mais detalhes.

VI. CONCLUSÕES

Minha recomendação é de aprovar a qualificação. O texto precisa ser modificado, dando detalhes na metodologia e não na seção de resultados, separar da seção de métodos partes que são da introdução e partes que são referenciais teóricos. Em resumo, há de se fazer uma reestruturação no texto. Minha avaliação que o trabalho em seu intento e conteúdo é muito bom, ambicioso e repletos de mérito. O texto deve ser reestruturado para uma leitura mais limpa e clara.