

MAT02035 - Modelos para dados correlacionados

Visão geral de modelos lineares para dados longitudinais
(continuação)

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2021

Modelando a média

Introdução

- ▶ Distinguem-se duas abordagens principais:
 1. a análise de perfis de resposta;
 2. curvas paramétricas ou semi-paramétricas.
- ▶ A análise dos dados longitudinais enfoca as mudanças na resposta média ao longo do tempo e na relação dessas mudanças com as covariáveis.
- ▶ O fato de as medidas obtidas no mesmo indivíduo não serem independentes, mas estarem correlacionadas positivamente é uma consideração importante em sua análise, mas para a maioria dos estudos longitudinais a correlação geralmente **não é de interesse científico** em si.
- ▶ Antes de discutir abordagens para modelar a resposta média ao longo do tempo, é importante esclarecer a distinção entre **parâmetros substantivos** e **de incômodos (de perturbação)** no contexto de um estudo longitudinal.

Parâmetros substantivos e de incômodo para dados longitudinais

- ▶ Nos modelos de regressão para dados longitudinais, os parâmetros de regressão β relacionam as mudanças na resposta média ao longo do tempo às covariáveis e são geralmente considerados de **interesse primário** ou intrínseco.
 - ▶ Podem ser definidos para resumir aspectos importantes das **questões de pesquisa**.
 - ▶ Nos referimos a esses parâmetros como **parâmetros substantivos**.
- ▶ Por outro lado, em muitas aplicações, parâmetros que resumem aspectos da covariância ou correlação entre as medidas repetidas são considerados de **interesse secundário**.
 - ▶ Os parâmetros associados a esses aspectos secundários dos dados costumam ser chamados de parâmetros **de incômodo**.
 - ▶ Para a análise de dados longitudinais, os parâmetros de correlação ou covariância são frequentemente considerados parâmetros de incômodo, uma vez que não há interesse intrínseco neles.

Parâmetros substantivos e de incômodo para dados longitudinais

SECUNDÁRIO \neq NEGLIGENCIÁVEL

- ▶ Em algumas configurações em que dados correlacionados surgem, pode haver uma reversão completa de funções.
- ▶ Em estudos com famílias, o objetivo é determinar se a presença de doença em um membro da família aumenta o risco de doença para os parentes.
 - ▶ As **correlações** entre irmãos e entre pais e filhos são de **interesse principal** (e, neste caso, os parâmetros de regressão β são de incômodo) porque suas magnitudes relativas podem ser usadas para fornecer evidências indiretas de risco genético para a doença devido ao compartilhamento do mesmo conjunto de genes.

Parâmetros substantivos e de incômodo para dados longitudinais

SECUNDÁRIO \neq NEGLIGENCIÁVEL

- ▶ Um exemplo adicional surge quando os pesquisadores estão interessados em avaliar a **heterogeneidade de um efeito de tratamento** em uma população.
 - ▶ Nesse cenário, a **variância do efeito** do tratamento é de **interesse primário**.

Modelando a resposta média ao longo do tempo

Análise de perfis

- ▶ Permite descrever padrões arbitrários na resposta média ao longo do tempo.
 - ▶ Nenhuma tendência de tempo específica é assumida. Em vez disso, os **tempos de medição** são considerados como **níveis de um fator discreto**.
 - ▶ Só é aplicável quando todos os indivíduos são medidos no mesmo conjunto de ocasiões e o número de ocasiões geralmente é pequeno ($n_i = n$ e n pequeno). (Note, neste caso, p está relacionado com n)

Curva paramétrica

- ▶ Assume uma tendência linear ou quadrática, por exemplo, para a resposta média ao longo do tempo.
 - ▶ Pode reduzir drasticamente o número de parâmetros do modelo.
 - ▶ Descrevem a **resposta média** como uma **função explícita do tempo**.
 - ▶ Não há necessidade que todos os indivíduos no estudo tenham o mesmo conjunto de tempos de medição, nem o mesmo número de medidas repetidas.

Modelando a covariância

Modelando a covariância

- ▶ A contabilização da correlação entre medidas repetidas completa a especificação de qualquer modelo de regressão para dados longitudinais e geralmente aumenta a eficiência ou a precisão com a qual os parâmetros de regressão podem ser estimados.
- ▶ Quando um modelo apropriado para a covariância é adotado, erros padrão corretos são obtidos e inferências válidas sobre os parâmetros de regressão podem ser feitas.

Modelando a covariância

- ▶ Além disso, quando há dados ausentes, a modelagem correta da covariância é frequentemente um requisito para obter estimativas válidas para os parâmetros de regressão.
- ▶ Distinguem-se três abordagens principais:
 1. covariância não estruturada;
 2. modelos de padrões de covariância;
 3. estruturas de covariância de (induzidas por) efeitos aleatórios.

Covariância não estruturada

- ▶ Permite qualquer padrão arbitrário de covariância entre as medidas repetidas.
 - ▶ Isso resulta no que normalmente é chamado de covariância “não estruturada”.
 - ▶ Assim, quando existem n medidas repetidas, as n variâncias em cada ocasião e as $n \times (n - 1)/2$ covariâncias (ou correlações) aos pares são estimadas.
- ▶ Historicamente, a matriz de covariância não estruturada tem sido o modelo de escolha para a covariância na análise de perfis de resposta (mas também pode ser usada na análise de curvas paramétricas).

Covariância não estruturada

Existem duas **desvantagens** em potencial com essa abordagem.

1. O número de parâmetros de covariância pode ser bastante grande.
 - ▶ Se houver n ocasiões de medição, a matriz de covariância $n \times n$ terá $n \times (n + 1)/2$ parâmetros únicos.
 - ▶ Assim, em um estudo longitudinal com 10 ocasiões de medição, uma covariância não estruturada possui 55 parâmetros (10 variâncias e 45 covariâncias).
 - ▶ Quando o número de parâmetros de covariância a ser estimado é grande em relação ao tamanho da amostra, é provável que as estimativas sejam **instáveis**.
2. É aplicável apenas quando todos os indivíduos são medidos no mesmo conjunto de ocasiões.

Modelos de padrões de covariância

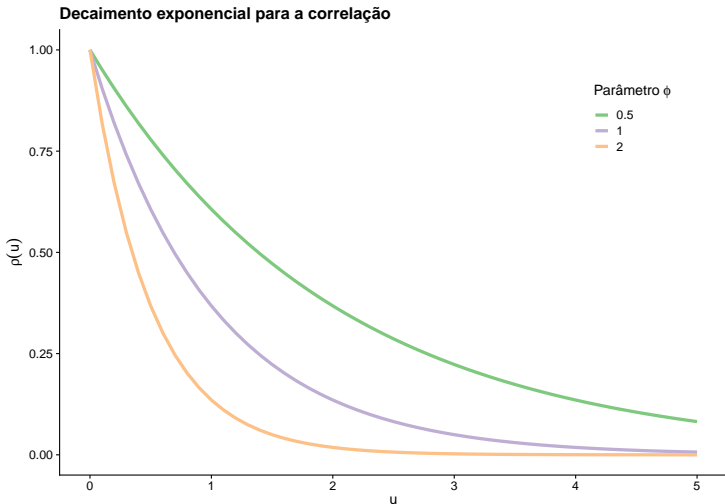
- ▶ Esta abordagem toma emprestado ideias da literatura da análise de séries temporais.
- ▶ Espera-se que medidas repetidas realizadas mais próximas no tempo sejam mais altamente correlacionadas do que medidas repetidas mais distantes no tempo.
 - ▶ Um caso especial é o **modelo de decaimento exponencial**

$$\rho(u) = \exp\{-\phi u\}, \phi > 0,$$

em que u representa a **separação entre os tempos de medida**.

- ▶ Isso implica que a correlação decai à medida que a separação entre os tempos de medida aumenta ($\rho(u) \rightarrow 0$ conforme $u \rightarrow \infty$).

Modelos de padrões de covariância



Modelos de padrões de covariância

- ▶ Outras funções (explícitas) da separação do tempo também podem representar a correlação entre medidas repetidas (por exemplo, $\rho(u) = \exp\{-\phi u^2\}$, $\phi > 0$).
- ▶ Esses modelos podem ser usados com observações desigualmente espaçadas.
- ▶ **Parcimônia:** modelos paramétricos podem descrever adequadamente a estrutura de covariância entre as medidas repetidas com apenas alguns (poucos) parâmetros.

Estruturas de covariância de efeitos aleatórios

- ▶ Uma estratégia alternativa e um tanto indireta para impor estrutura à covariância é através da introdução de **efeitos aleatórios**.
 - ▶ Uma das primeiras abordagens para analisar dados de medidas repetidas.
- ▶ No chamado modelo **ANOVA univariada de medidas repetidas**, a correlação entre medidas repetidas é explicada pela inclusão de um único efeito aleatório específico individual.
 - ▶ Esse efeito pode ser pensado como um intercepto variando aleatoriamente, representando uma agregação de todos os fatores não observados ou não mensurados que tornam alguns indivíduos “altos respondedores” e outros “baixos respondedores”.
- ▶ A consequência de adicionar um único efeito aleatório específico do indivíduo a todas as medidas em qualquer indivíduo é que as medidas repetidas resultantes serão **correlacionadas positivamente**.
 - ▶ Assim, a inclusão de efeitos aleatórios impõe estrutura à covariância.

Abordagens históricas

Abordagens históricas

- ▶ De uma perspectiva histórica, três métodos para a análise de dados de medidas repetidas podem ser destacados:
 - (1) análise de variância univariada de medidas repetidas (ANOVA)
 - (2) análise multivariada de variância de medidas repetidas (MANOVA)
 - (3) métodos baseados em medidas resumo.
- ▶ Todas essas três abordagens tiveram graus variados de popularidade e algumas ainda são amplamente utilizadas em diferentes áreas de aplicação.

Abordagens históricas

- ▶ Muitas dessas abordagens são desnecessariamente restritivas em suas suposições e objetivos analíticos.
 - ▶ Por exemplo, ANOVA e MANOVA se concentram na comparação de grupos em termos de sua tendência de resposta média ao longo do tempo, mas fornecem poucas informações sobre como os indivíduos mudam ao longo do tempo.
 - ▶ Além disso, a ANOVA e a MANOVA têm inúmeras características que limitam sua utilidade para a análise de dados longitudinais.

Avisos

- ▶ **Para casa:** ler o Capítulo 3 do livro “**Applied Longitudinal Analysis**”.
 - ▶ Caso ainda não tenha lido, leia também os Caps. 1 e 2.
- ▶ **Próxima aula:** Estimação e inferência estatística.

Bons estudos!

