

MAT02035 - Modelos para dados correlacionados

Revisão de modelos lineares generalizados

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

Introdução

Introdução

- ▶ Quando a variável resposta é categórica (por exemplo, **dados binários** e **de contagem**), **modelos lineares generalizados** (por exemplo, **regressão logística**) podem ser estendidos para lidar com os resultados correlacionados.
- ▶ No entanto, transformações não lineares da resposta média (por exemplo, *logit*) levantam questões adicionais relativas à interpretação dos coeficientes de regressão.
- ▶ Diferentes abordagens para contabilizar a correlação levam a modelos com coeficientes de regressão com interpretações distintas.
- ▶ Neste curso, consideraremos duas extensões principais de modelos lineares generalizados:
 1. modelos marginais;
 2. modelos de efeitos mistos.

Exemplo: Tratamento oral da infecção das unhas dos pés

- ▶ Estudo aleatorizado, duplo-cego, multicêntrico de 294 pacientes comparando 2 tratamentos orais (denotados A e B) para infecção nas unhas dos pés.
- ▶ **Variável desfecho:** variável binária indicando presença de onicólise (separação da placa ungueal do leito ungueal).
- ▶ Pacientes avaliados quanto ao grau de onicólise na linha de base (semana 0) e nas semanas 4, 8, 12, 24, 36 e 48.
- ▶ Interesse na taxa de **declínio da proporção** de pacientes com onicólise ao longo do tempo e nos efeitos do tratamento nessa taxa.

Exemplo: Ensaio clínico de progabida anti-epiléptica

- ▶ Estudo aleatorizado, controlado por placebo, do tratamento de crises epiléticas com progabida.
- ▶ Os pacientes foram aleatorizados para tratamento com progabida ou placebo (em adição à terapia padrão).
- ▶ **Variável desfecho:** Contagem do número de convulsões.
- ▶ Cronograma de medição: medição da linha de base durante 8 semanas antes da aleatorização.
 - ▶ Quatro medições durante intervalos consecutivos de duas semanas.
- ▶ Tamanho da amostra: 28 epiléticos com placebo; 31 epiléticos em progabida.

Introdução

- ▶ **Modelos lineares generalizados** (MLG) são uma classe de modelos de regressão; eles incluem o modelo de regressão linear padrão, mas também muitos outros modelos importantes:
 - ▶ Regressão linear para dados contínuos
 - ▶ Regressão logística para dados binários
 - ▶ Modelos de regressão log-linear / Poisson para dados de contagem
- ▶ Modelos lineares generalizados estendem os métodos de análise de regressão a configurações nas quais a variável resposta pode ser categórica.
- ▶ Nas próximas aulas, consideramos extensões de modelos lineares generalizados para dados longitudinais.
- ▶ Primeiro, revisaremos os modelos logístico e de regressão de Poisson para uma única resposta.

Regressão logística

Regressão logística

- ▶ Até agora, consideramos modelos de regressão linear para uma resposta contínua, Y , da seguinte forma

$$Y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + e.$$

- ▶ A variável resposta Y é assumida como tendo uma distribuição normal com média

$$E(Y) = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

e com variância σ^2 .

Regressão logística

- ▶ Lembre-se de que o intercepto da população (para $X_1 = 1$), β_1 , tem interpretação como o valor médio da resposta quando todas as covariáveis assumem o valor zero.
- ▶ A inclinação da população, digamos β_k , tem interpretação em termos da mudança esperada na resposta média para uma mudança de uma unidade em X_k , uma vez que todas as outras covariáveis permanecem constantes.
- ▶ Em muitos estudos, no entanto, estamos interessados em uma variável resposta dicotômica / binária em vez de contínua.
- ▶ A seguir, consideramos um modelo de regressão para uma resposta binária (ou dicotômica).

Regressão logística

- ▶ Seja Y uma resposta binária, em que
 - ▶ $Y = 1$ representa um “sucesso”;
 - ▶ $Y = 0$ representa uma “falha”.
- ▶ Então a média da variável resposta binária, denominada π , é a proporção de sucessos ou a probabilidade de a resposta assumir o valor 1.
- ▶ Ou seja,

$$\pi = E(Y) = \Pr(Y = 1) = \Pr(\text{“sucesso”}).$$

- ▶ Com uma resposta binária, geralmente estamos interessados em estimar a probabilidade π e relacioná-la a um conjunto de covariáveis.
- ▶ Para fazer isso, podemos usar regressão logística.

Regressão logística

- ▶ Uma estratégia ingênua para modelar uma resposta binária é considerar um modelo de regressão

$$\pi = E(Y) = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p.$$

- ▶ No entanto, em geral, esse modelo não é viável, pois π é uma probabilidade e restringe-se a valores entre 0 e 1.
- ▶ Além disso, a suposição usual de homogeneidade de variância seria violada, uma vez que a variância de uma resposta binária depende da média, ou seja,

$$\text{Var}(Y) = \pi(1 - \pi).$$

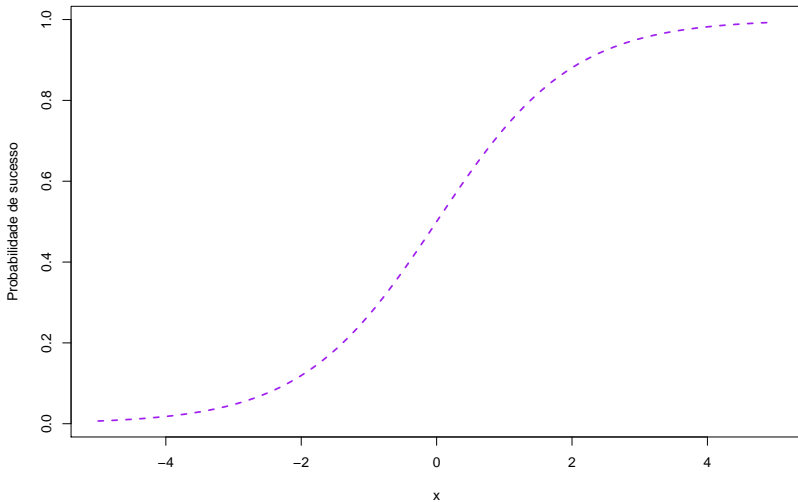
Regressão logística

- ▶ Em vez disso, podemos considerar um modelo de regressão logística em que

$$\log[\pi/(1 - \pi)] = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p.$$

- ▶ Este modelo acomoda a restrição que π está restrita a valores entre 0 e 1.
- ▶ Lembre-se de que $\pi/(1 - \pi)$ é definido como a chance de sucesso.
 - ▶ Portanto, modelar com uma função logística pode ser considerado equivalente a um modelo de regressão linear em que a média da resposta contínua foi substituída pelo logaritmo das chances de sucesso.
- ▶ Observe que a relação entre π e as covariáveis é não linear.

Regressão logística



Regressão logística

- ▶ Partindo do pressuposto de que as respostas binárias são variáveis aleatórias de Bernoulli, podemos usar a estimativa de ML para obter estimativas dos parâmetros de regressão logística.
- ▶ Finalmente, lembre-se a relação entre o “odds” e “probabilidades”.

$$\text{Odds} = \frac{\pi}{1 - \pi};$$

$$\pi = \frac{\text{Odds}}{1 + \text{Odds}}.$$

Regressão logística

- ▶ Dado o modelo de regressão logística

$$\log[\pi/(1 - \pi)] = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

o intercepto populacional, β_1 , tem interpretação como a probabilidade de sucesso logarítmica quando todas as covariáveis assumem o valor zero.

- ▶ A inclinação da população, digamos β_k , tem interpretação em termos da mudança na “log-chance” (log-odds) de sucesso para uma mudança unitária em X_k , uma vez que todas as outras covariáveis permanecem constantes.
- ▶ Quando uma das covariáveis é dicotômica, digamos X_2 , então β_2 tem uma interpretação especial:
 - ▶ $\exp(2)$ é a **razão de chances** de sucesso para os dois níveis possíveis de X_2 (dado que todas as outras covariáveis permanecem constantes).

Regressão logística

Lembre-se de que:

- ▶ π aumenta
 - ▶ odds de sucesso aumenta
 - ▶ log-odds de sucesso aumenta

Da mesma forma, como:

- ▶ π diminui
 - ▶ odds de sucesso diminui
 - ▶ log-odds de sucesso diminui

Regressão logística: exemplo

Desenvolvimento de displasia broncopulmonar (DBP) em uma amostra de 223 crianças com baixo peso ao nascer

- ▶ **Resposta binária:** $Y = 1$ se DBP estiver presente, $Y = 0$ caso contrário.
- ▶ **Covariável:** Peso ao nascer do bebê em gramas.
- ▶ Considere o seguinte modelo de regressão logística

$$\log[\pi/(1 - \pi)] = \beta_1 X_1 + \beta_2 \text{Weight}$$

em que $\pi = E(Y) = \Pr(Y = 1) = \Pr(\text{DBP})$.

Regressão logística: exemplo

- ▶ Para os 223 bebês da amostra, a regressão logística estimada (obtida usando MV) é

$$\log[\hat{\pi}/(1 - \hat{\pi})] = 4.0343 - 0.0042\text{Weight}.$$

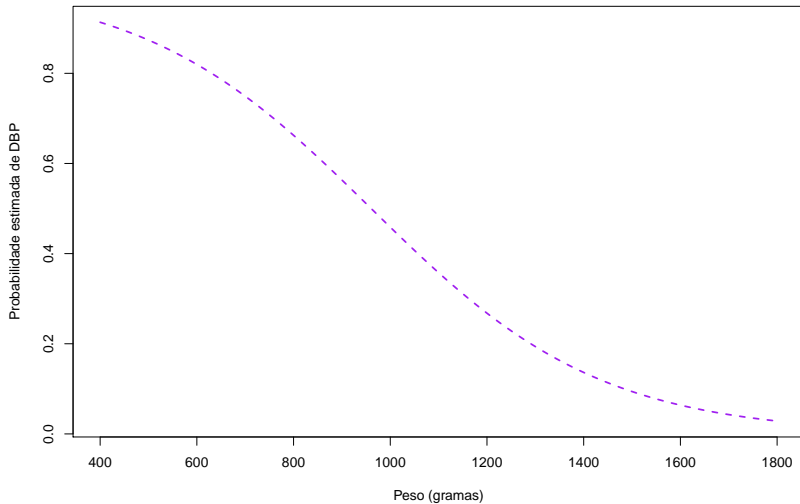
- ▶ A estimativa de MV de β_2 implica que, para cada aumento de 1 grama no peso ao nascer, a log-odds de DBP diminuem 0,0042.
- ▶ Por exemplo, a probabilidade de DBP para um bebê com 1200 gramas são

$$\exp(4.0343 - 1200 \times 0.0042) = \exp(-1.0057) = 0.3658$$

- ▶ Assim, a probabilidade predita de DBP é:

$$0.3658/(1 + 0.3658) = 0.268.$$

Regressão logística: exemplo



Regressão de Poisson

Regressão de Poisson

- ▶ Na regressão de Poisson, a variável resposta é uma contagem (por exemplo, número de casos de uma doença em um determinado período de tempo).
- ▶ A distribuição de Poisson fornece a base da inferência baseada em verossimilhança.
- ▶ Frequentemente, as contagens podem ser expressas como *taxas*.
- ▶ Ou seja, a contagem ou o número absoluto de eventos geralmente não é satisfatório porque qualquer comparação depende quase inteiramente dos tamanhos dos grupos (ou do “tempo em risco”) que gerou as observações.

Regressão de Poisson

- ▶ Como uma proporção ou probabilidade, uma taxa fornece uma base para comparação direta.
- ▶ Em ambos os casos, a regressão de Poisson relaciona as contagens ou taxas esperadas a um conjunto de covariáveis.

Regressão de Poisson

O modelo de regressão de Poisson possui dois componentes:

1. A variável resposta é uma contagem e é assumida como tendo uma distribuição de Poisson. Ou seja, a probabilidade de ocorrer um número específico de eventos, y , é

$$\Pr(y \text{ eventos}) = \frac{e^{-\lambda} \lambda^y}{y!}$$

- Observe que λ é a contagem ou número esperado de eventos e a taxa esperada é dada por λ/t , em que t é uma medida de linha de base relevante (por exemplo, t pode ser o número de pessoas ou o número de pessoas-ano de observação).

Regressão de Poisson

2. $\log(\lambda/t) = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$

Observe que, como $\log(\lambda/t) = \log(\lambda) - \log(t)$, o modelo de regressão de Poisson também pode ser considerado como

$$\log(\lambda) = \log(t) + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

em que o “coeficiente” associado a $\log(t)$ é fixado em 1.

- ▶ Esse termo de ajuste é conhecido como “*offset*”.

Regressão de Poisson

- ▶ Portanto, modelar λ (ou $= \lambda/t$) com uma função logarítmica pode ser considerado equivalente a um modelo de regressão linear em que a média da resposta contínua foi substituída pelo logaritmo da contagem (ou taxa) esperada.
- ▶ Observe que a relação entre λ (ou λ/t) e as covariáveis é não linear.
- ▶ Podemos usar a estimação por MV para obter estimativas dos parâmetros de regressão de Poisson, pressupondo que as respostas sejam variáveis aleatórias de Poisson.

Regressão de Poisson

- ▶ Dado o modelo de regressão de Poisson

$$\log(\lambda/t) = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

o intercepto populacional, β_1 , tem interpretação como a log-taxa esperada quando todas as covariáveis assumem o valor zero.

- ▶ A inclinação da população, digamos β_k , tem interpretação em termos da mudança na log-taxa esperada para uma mudança de unidade única em X_k , uma vez que todas as outras covariáveis permanecem constantes.
- ▶ Quando uma das covariáveis é dicotômica, digamos X_2 , então β_2 tem uma interpretação especial:
 - ▶ $\exp(\beta_2)$ é a razão da taxa (de incidência) para os dois níveis possíveis de X_2 (dado que todas as outras covariáveis permanecem constantes).

Regressão de Poisson: exemplo

Estudo prospectivo de doença cardíaca coronária (CHD)

- ▶ O estudo observou 3154 homens entre 40 e 50 anos em média por 8 anos e registrou incidência de casos de doença coronariana.
- ▶ Os fatores de risco considerados incluem:
 - ▶ **Exposição ao fumo:** 0, 10, 20, 30 cigarros por dia;
 - ▶ **Pressão arterial:** 0 (< 140), 1 (≥ 140);
 - ▶ **Tipo de comportamento:** 0 (tipo B), 1 (tipo A).

Um modelo simples de regressão de Poisson é:

$$\log(\lambda/t) = \log(\text{taxa de CHD}) = \beta_1 + \beta_2 \text{Smoke}$$

ou

$$\log(\lambda) = \log(t) + \beta_1 + \beta_2 \text{Smoke}.$$

Regressão de Poisson: exemplo

Person - Years	Smoking	Blood Pressure	Behavior	CHD
5268.2	0	0	0	20
2542.0	10	0	0	16
1140.7	20	0	0	13
614.6	30	0	0	3
4451.1	0	0	1	41
2243.5	10	0	1	24
1153.6	20	0	1	27
925.0	30	0	1	17
1366.8	0	1	0	8
497.0	10	1	0	9
238.1	20	1	0	3
146.3	30	1	0	7
1251.9	0	1	1	29
640.0	10	1	1	21
374.5	20	1	1	7
338.2	30	1	1	12

Regressão de Poisson: exemplo

- ▶ Neste modelo, a estimativa de MV de β_2 é 0,0318. Ou seja, a taxa de CHD aumenta por um fator de $\exp(0.0318) = 1.032$ para cada cigarro fumado.
- ▶ Alternativamente, a taxa de CHD em fumantes de um maço por dia (20 cigarros) é estimada em $(1.032)^{20} = 1.88$ vezes maior que a taxa de CHD em não fumantes.
- ▶ Podemos incluir os fatores de risco adicionais no seguinte modelo:

$$\log(\lambda/t) = \beta_1 + \beta_2 \text{Smoke} + \beta_3 \text{Type} + \beta_4 \text{BP}$$

Effect	Estimate	Std. Error
Intercept	-5.420	0.130
Smoke	0.027	0.006
Type	0.753	0.136
BP	0.753	0.129

Regressão de Poisson: exemplo

- ▶ A taxa ajustada de CHD (controle da pressão arterial e tipo de comportamento) aumenta em um fator de $\exp(0.027) = 1.028$ para cada cigarro fumado.
- ▶ A taxa ajustada de CHD em fumantes de um maço por dia (20 cigarros) é estimada em $(1.027)^{20} = 1.7$ vezes maior que a taxa de CHD em não fumantes.

Regressão de Poisson

- ▶ Finalmente, observe que quando um modelo de regressão de Poisson é aplicado aos dados consistindo em taxas muito pequenas (digamos, $\lambda/t \ll 0.01$), a taxa é aproximadamente igual à probabilidade correspondente, p , e

$$\log(\text{taxa}) \approx \log(p) \approx \log[p/(1 - p)].$$

- ▶ Portanto, os parâmetros para os modelos de regressão de Poisson e regressão logística são aproximadamente iguais quando o evento em estudo é raro.
- ▶ Nesse caso, os resultados de uma regressão de Poisson e logística não fornecerão resultados discernivelmente diferentes.

Modelos lineares generalizados

Modelos lineares generalizados

- ▶ **Modelos lineares generalizados** (MLG) são uma classe de modelos de regressão; eles incluem o modelo de regressão linear padrão, mas também muitos outros modelos importantes:
 - ▶ Regressão linear para dados contínuos
 - ▶ Regressão logística para dados binários
 - ▶ Modelos de regressão log-linear / Poisson para dados de contagem
- ▶ Modelos lineares generalizados estendem os métodos de análise de regressão a configurações nas quais a variável resposta pode ser categórica.

Notação

- ▶ Assuma N observações independentes de uma única variável resposta Y_i .
- ▶ Associada a cada resposta Y_i , existe um vetor $p \times 1$ de covariáveis, X_{i1}, \dots, X_{ip} .
- ▶ **Objetivo:** o interesse principal está em relacionar a média de Y_i , $\mu_i = E(Y_i | X_{i1}, \dots, X_{ip})$, às covariáveis.

Modelos lineares generalizados

Em modelos lineares generalizados:

1. Assume-se que distribuição da variável resposta, Y_i , pertence a família de distribuições conhecida como **família exponencial**, por exemplo, as distribuições:
 - ▶ normal;
 - ▶ Bernoulli;
 - ▶ binomial;
 - ▶ Poisson, entre outras.
2. Um **componente sistemático** que especifica os efeitos das covariáveis na média da distribuição de Y_i

$$\eta_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} = \sum_{k=1}^p \beta_k X_{ik}.$$

3. A transformação da média da resposta, μ_i , tem uma relação linear com as covariáveis por meio de uma **função de ligação** apropriada:

Modelos lineares generalizados

$$g(\mu_i) = \eta_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip},$$

em que a função de ligação $g(\cdot)$ é uma função conhecida, por exemplo, $\log(\mu_i)$.

A família exponencial

- ▶ Todas as distribuições que pertencem a família exponencial podem ser expressas como

$$f(y_i; \theta_i, \phi) = \exp[\{y_i \theta_i - a(\theta_i)\} / \phi + b(y_i, \phi)],$$

em que $a(\cdot)$ e $b(\cdot)$ são funções específicas que distinguem um membro da família de outro.

- ▶ A família exponencial expressa desta forma tem θ_i como um parâmetro de locação (“canônico”) e ϕ como um parâmetro de escala (ou dispersão).
- ▶ **Exercício:** Identifique as distribuições normal, Bernoulli e Poisson como membros da família exponencial.

Média e variância das distribuições na família exponencial

- ▶ Distribuições na família exponencial compartilham algumas propriedades estatísticas comuns.
 - ▶ Por exemplo, $E(Y_i) = \mu_i = \frac{\partial a(\theta_i)}{\partial \theta}$ e $\text{Var}(Y_i) = \phi \frac{\partial^2 a(\theta_i)}{\partial \theta^2}$.
- ▶ Assim, a variância de Y_i pode ser expressa em termos de

$$\text{Var}(Y_i) = \phi v(\mu_i),$$

em que o parâmetro de escala $\phi > 0$.

- ▶ A **função de variância**, $v(\mu_i)$, descreve como a variância da resposta está funcionalmente relacionada μ_i , a média de Y_i .

A função de ligação

- ▶ A função de ligação aplica uma transformação à média e, em seguida, vincula as covariáveis à média transformada,

$$g(\mu_i) = \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip}$$

em que a função de ligação $g(\cdot)$ é uma função conhecida, por exemplo, $\log(\mu_i)$.

- ▶ Isso implica que é a resposta média transformada que muda linearmente com as mudanças nos valores das covariáveis.

Modelos lineares generalizados

Funções de ligação canônicas e variância para as distribuições normais, Bernoulli e Poisson.

Distribuição	Função de variância	Ligação canônica
Normal	$v(\mu) = 1$	Identidade: $\mu = \eta$
Bernoulli	$v(\mu) = \mu(1 - \mu)$	Logit: $\log \left[\frac{\mu}{1-\mu} \right] = \eta$
Poisson	$v(\mu) = \mu$	Log: $\log(\mu) = \eta$

em que $\eta = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$.

Extensões de modelos lineares generalizados para dados longitudinais

Extensões de MLG para dados longitudinais

- ▶ Quando a variável de resposta é categórica (por exemplo, dados binários e de contagem), modelos lineares generalizados (por exemplo, regressão logística) podem ser estendidos para lidar com as respostas correlacionadas.
- ▶ No entanto, transformações não lineares da resposta média (por exemplo, logit) levantam questões adicionais relativas à interpretação dos coeficientes de regressão.
- ▶ Como veremos, modelos diferentes para dados longitudinais discretos têm objetivos de inferência um tanto diferentes.

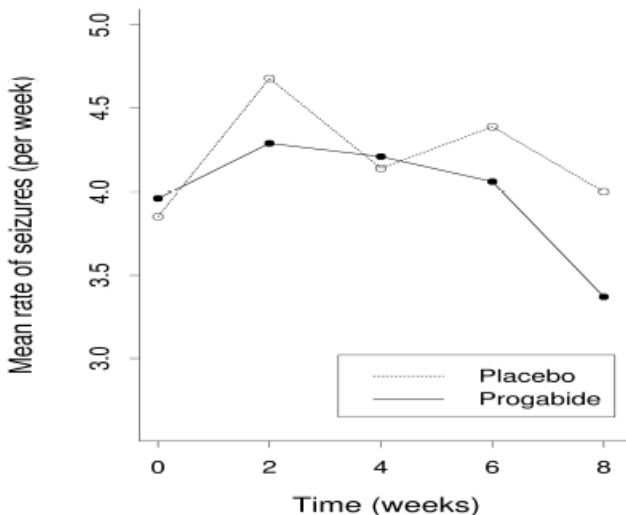
Exemplo: Tratamento oral da infecção das unhas dos pés

- ▶ Estudo aleatorizado, duplo-cego, grupo paralelo, multicêntrico de 294 pacientes comparando 2 tratamentos orais (denotados A e B) para infecção nas unhas dos pés.
- ▶ **Variável resposta:** variável binária indicando presença de onicólise (separação da placa ungueal do leito ungueal).
- ▶ Pacientes avaliados quanto ao grau de onicólise (separação da placa ungueal do leito ungueal) na linha de base (semana 0) e nas semanas 4, 8, 12, 24, 36 e 48.
- ▶ Interesse está na taxa de declínio da proporção de pacientes com onicólise ao longo do tempo e os efeitos do tratamento nessa taxa.

Exemplo: Ensaio clínico de progabida anti-epiléptica

- ▶ Estudo aleatorizado, controlado por placebo, do tratamento de crises epilépticas com progabida.
- ▶ Os pacientes foram aleatorizados para tratamento com progabida ou placebo, além da terapia padrão.
- ▶ **Variável resposta:** Contagem do número de convulsões
- ▶ Cronograma de medição: medição da linha de base durante 8 semanas antes da randomização. Quatro medições durante intervalos consecutivos de duas semanas.
- ▶ Tamanho da amostra: 28 epilépticos com placebo; 31 epilépticos em progabide

Exemplo: Ensaio clínico de progabida anti-epiléptica



MLG para dados longitudinais

- ▶ Em seguida, focamos em várias abordagens distintas para analisar respostas longitudinais.
- ▶ Essas abordagens podem ser consideradas extensões de modelos lineares generalizados para dados correlacionados.
- ▶ A ênfase principal será em dados de resposta discretos, por exemplo, dados de contagem ou respostas binárias.
- ▶ **Nota:** nos modelos lineares (efeitos mistos) para respostas contínuas, a interpretação dos coeficientes de regressão é independente da correlação entre as respostas.

MLG para dados longitudinais

- ▶ Com dados discretos de resposta, esse não é mais o caso.
- ▶ Com modelos não lineares para dados discretos, diferentes abordagens para contabilizar a correlação levam a modelos com coeficientes de regressão com interpretações distintas.
 - ▶ Voltaremos a esta questão importante no decorrer do curso.
- ▶ No restante desta aula, examinaremos brevemente três extensões principais de modelos lineares generalizados.

MLG para dados longitudinais

- ▶ Suponha que $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in})$ é um vetor de respostas correlacionadas do i -ésimo indivíduo.
- ▶ Para analisar esses dados correlacionados, precisamos especificar ou pelo menos fazer suposições sobre a distribuição multivariada ou conjunta,

$$f(Y_{i1}, Y_{i2}, \dots, Y_{in}).$$

- ▶ A maneira pela qual a distribuição multivariada é especificada produz três abordagens analíticas um tanto diferentes:
 1. Modelos marginais
 2. Modelos de efeitos mistos
 3. Modelos de transição

Modelos marginais

- ▶ Uma abordagem é especificar a distribuição marginal em cada momento:

$$f(Y_{ij}) \quad \text{para } j = 1, 2, \dots, n.$$

juntamente com algumas suposições sobre a estrutura de covariância das observações.

- ▶ A premissa básica dos **modelos marginais** é fazer inferências sobre as **médias populacionais**.
- ▶ O termo “**marginal**” é usado aqui para enfatizar que a resposta média modelada é condicional apenas para covariáveis e não para outras respostas (ou efeitos aleatórios).

Ilustração

- ▶ Considere o **estudo tratamento oral da infecção das unhas dos pés**.
- ▶ Estudo aleatorizado, duplo-cego, grupo paralelo, multicêntrico, de 294 pacientes comparando 2 tratamentos orais (denotados A e B) para infecção das unhas dos pés.
- ▶ **Variável resposta:** variável binária que indica presença de onicólise (separação da placa ungueal do leito ungueal).
- ▶ Pacientes avaliados quanto ao grau de onicólise (separação da placa ungueal do leito ungueal) na linha de base (semana 0) e nas semanas 4, 8, 12, 24, 36 e 48.
- ▶ O interesse encontra-se na taxa de declínio da proporção de pacientes com onicólise ao longo do tempo e nos efeitos do tratamento nessa taxa.

Ilustração

- Suponha que a probabilidade marginal de onicólise siga um modelo logístico,

$$\text{logit} \{ \Pr(Y_{ij} = 1) \} = \beta_1 + \beta_2 \text{Mês}_{ij} + \beta_3 \text{Trt}_i + \beta_4 (\text{Trt}_i \times \text{Mês}_{ij}),$$

em que $\text{Trt} = 1$ se o grupo de tratamento B e 0, caso contrário.

- Este é um exemplo de um modelo marginal.
- Note, no entanto, que a estrutura de covariância ainda precisa ser especificada.

Modelos de efeitos mistos

- ▶ Outra possibilidade é supor que um subconjunto dos parâmetros de regressão no modelo linear generalizado varie de indivíduo para indivíduo.
- ▶ Especificamente, poderíamos assumir que os dados de um único indivíduo são observações independentes com uma distribuição pertencente à família exponencial, mas que os coeficientes de regressão podem variar de indivíduo para indivíduo.
- ▶ Ou seja, **condicional** aos efeitos aleatórios, supõe-se que as respostas para um único indivíduo sejam observações independentes de uma distribuição pertencente à família exponencial.

Ilustração

- ▶ Considere o **estudo tratamento oral da infecção das unhas dos pés**.
- ▶ Suponha, por exemplo, que a probabilidade de onicólise para os participantes do estudo seja descrita por um modelo logístico, mas que o risco para um indivíduo dependa de seu “nível de resposta aleatória” **latente** (talvez determinado ambiental e geneticamente).
- ▶ Podemos considerar um modelo em que

$$\text{logit} \{ \Pr(Y_{ij} = 1) \} = \beta_1 + \beta_2 \text{Mês}_{ij} + \beta_3 \text{Trt}_i + \beta_4 (\text{Trt}_i \times \text{Mês}_{ij}) + b_i.$$

- ▶ Observe que esse modelo também requer especificação da distribuição de efeitos aleatórios, $F(b_i)$.
- ▶ Este é um exemplo de um modelo linear generalizado de efeitos mistos.

Modelos de transição (Markov)

- ▶ Finalmente, outra abordagem é expressar a distribuição conjunta como uma série de distribuições condicionais,

$$f(Y_{i1}, Y_{i2}, \dots, Y_{in}) = f(Y_{i1}) \times f(Y_{i2}|Y_{i1}) \times \dots \times f(Y_{in}|Y_{i1}, \dots, Y_{i,n-1}).$$

- ▶ Isso é conhecido como **modelo de transição** (ou modelo para as transições) porque representa a distribuição de probabilidade em cada ponto do tempo como condicional ao passado.
- ▶ Isso fornece uma representação completa da distribuição conjunta.

Ilustração

- ▶ Considere o **estudo tratamento oral da infecção das unhas dos pés**.
- ▶ Poderíamos escrever o modelo de probabilidade como

$$\begin{aligned} f(Y_{i1}, Y_{i2}, \dots, Y_{in} | X_i) &= f(Y_{i1} | X_i) \times f(Y_{i2} | Y_{i1}, X_i) \times f(Y_{i3} | Y_{i1}, Y_{i2}, X_i) \\ &\quad \times \dots \times f(Y_{i7} | Y_{i1}, Y_{i2}, \dots, Y_{i6}, X_i). \end{aligned}$$

- ▶ Ou seja, a probabilidade de onicólise no tempo 2 é modelada condicionalmente à presença/ausência de onicólise no tempo 1 e assim por diante.

Ilustração

- ▶ Por exemplo, um modelo logístico de “**primeira ordem**”, permitindo dependência apenas da resposta anterior, é fornecido por

$$\text{logit} \{ \Pr(Y_{ij} = 1 | Y_{i,j-1}) \} = \beta_1 + \beta_2 \text{Mês}_{ij} + \beta_3 \text{Trt}_i + \beta_4 (\text{Trt}_i \times \text{Mês}_{ij}) + \beta_5 Y_{i,j-1}.$$

Em resumo

- ▶ Discutimos as principais características dos modelos lineares generalizados
- ▶ Descrevemos brevemente três extensões principais de modelos lineares generalizados para dados longitudinais:
 1. Modelos marginais
 2. Modelos de efeitos mistos
 3. Modelos transitórios
- ▶ No restante do curso, focaremos em (i) Modelos Marginais e (ii) Modelos de efeitos mistos.

Em resumo

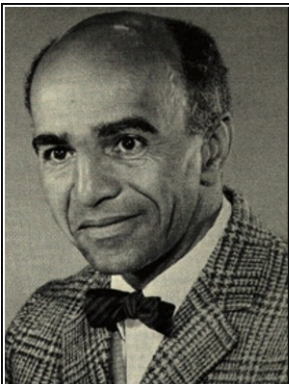
- ▶ Em geral, os modelos de transição são um pouco menos úteis para modelar efeitos de covariáveis.
- ▶ Especificamente, inferências de um modelo de transição podem ser potencialmente enganosas se um tratamento ou exposição alterar o risco ao longo do período de acompanhamento.
- ▶ Nesse caso, o risco condicional, dado o histórico anterior do resultado, é alterado de maneira menos nitidamente.

Avisos

Avisos

- ▶ **Próxima aula:** Modelos marginais (GEE).
- ▶ **Para casa:** ler o Capítulo 11 do livro “**Applied Longitudinal Analysis**” (em particular a Seção 11.7).
 - ▶ Caso ainda não tenha lido, leia também os Caps. 1, 2, 3, 4, 5, 6, 7, 8, 9 e 10.
 - ▶ Veja o *help* da função `glm` do R; rode os exemplos apresentados no *help* da função.
- ▶ **Não deixe de ver:** Another mixed effects model visualization

Bons estudos!



Basically, I'm not interested in doing research and I never have been... I'm interested in understanding, which is quite a different thing. And often to understand something you have to work it out yourself because no one else has done it.

— *David Blackwell* —

AZ QUOTES