MAT02035 - Modelos para dados correlacionados

Estimação e inferência estatística

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

- Até agora, nossa discussão de modelos para dados longitudinais tem sido muito geral, sem menção de métodos para estimar os coeficientes de regressão ou a covariância entre as medidas repetidas.
- **Relembrando:** o modelo de regressão linear geral para o vetor de resposta média

$$\mathrm{E}\left(Y_i|X_i\right) = X_i\beta.$$

- Assumimos que o vetor de respostas, Y_i, tem uma distribuição condicional que é **normal multivariada** com matriz de covariância

$$\mathrm{Cov}\left(Y_i|X_i\right) = \Sigma_i = \Sigma_i(\theta).$$

- Note que θ é um vetor $q \times 1$ de **parâmetros de covariância**.

Introdução

- Com dados balanceados ($n_i = n$), em que uma matriz de covariância **"não estruturada"** é assumida, temos n variâncias e $\frac{n(n-1)}{2}$ covariâncias como elementos do vetor θ (ou seja, $q = \frac{n(n+1)}{2}$).
- Se a covariância é assumida ter um padrão de "simetria composta"[1], então $q = 2$ e θ tem dois elementos.
- Nesta aula, consideramos um método de estimação para os parâmetros desconhecidos, β e θ (ou equivalentemente, Σ_i).

[1] O padrão de simetria composta assume variância constante e mesma covariância para os pares de medidas repetidas.

Estimação: máxima verossimilhança

O método da máxima verossimilhança

- Dado que foram feitas suposições completas sobre a distribuição do vetor de respostas, Y_i ($Y_i \sim N(X_i\beta, \Sigma_i(\theta))$), uma abordagem bem geral de estimação é o **método da máxima verossimilhança (MV)**.

- A ideia fundamental por trás da estimativa de MV é realmente bastante simples e é transmitida por seu nome: use como estimativas de β e θ os valores que são mais prováveis **(ou mais "verossímeis")** para os dados que foram realmente observados.
- As estimativas de máxima verossimilhança de β e θ são aqueles valores de β e θ que maximizam a probabilidade conjunta das variáveis resposta **avaliadas em seus valores observados**.

Estimação: máxima verossimilhança

- A probabilidade das variáveis resposta avaliadas no conjunto fixo de valores observados e consideradas como função de β e $\Sigma_i(\theta)$ é conhecida como **função de verossimilhança**.
  - Assim, a estimativa de β e θ provém **por maximizar** a função de verossimilhança.
- Os valores de β e θ que maximizam a função de verossimilhança são chamados de **estimativas de máxima verossimilhança** de β e θ, e geralmente são indicados por $\hat{\beta}$ e $\hat{\theta}$ (ou $\widehat{\Sigma}_i$, $\Sigma_i(\hat{\theta})$).

Observações independentes

- Suponha que os dados provém de uma série de estudos transversais que são repetidos em n ocasiões diferentes.
- Em cada ocasião, os dados são obtidos em uma amostra de N indivíduos.
  - Aqui é razoável supor que as observações sejam **independentes** umas das outras, uma vez que cada indivíduo é medido em apenas uma ocasião.
- Além disso, para facilitar a exposição, **assumimos que a variância é constante**, digamos σ^2.
- A resposta média está relacionada às covariáveis através do seguinte modelo de regressão linear:

$$\mathrm{E}\left(Y_{ij}|X_{ij}\right) = X_{ij}'\beta.$$

Observações independentes

- Para obter estimativas de máxima verossimilhança de β, devemos encontrar os valores dos parâmetros de regressão que maximizem a **função de densidade de probabilidade normal conjunta**[2] de todas as observações, avaliados nos valores observados da resposta e considerados como uma função de β (e σ^2).
- **Lembrando** que a função de densidade de probabilidade normal (ou gaussiana) univariada para Y_{ij}, dado X_{ij}, pode ser expressa como

$$f(y_{ij}) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2}(y_{ij} - \mu_{ij})^2/\sigma^2\right\}, \; -\infty < y_{ij} < \infty.$$

[2] Aqui, implicitamente, estamos assumindo $e_{ij} \overset{iid}{\sim} N(0, \sigma^2)$.

Observações independentes

- Quando todas as respostas são independentes umas das outras, a função de verossimilhança é simplesmente o produto das funções individuais de densidade de probabilidade normal univariada para Y_{ij} dado X_{ij},

$$L = \prod_{i=1}^{N} \prod_{j=1}^{n} f(y_{ij}).$$

Observações independentes

- É mais comum trabalhar com a função **log-verossimilhança**, que envolverá somas, em vez de produtos, das funções individuais de densidade de probabilidade normal univariada para Y_{ij}.
  - Observe que maximizar a verossimilhança é equivalente a maximizar o logaritmo da verossimilhança (indicado por ℓ).
- Portanto, o objetivo é maximizar

$$\ell = \log\left\{\prod_{i=1}^{N}\prod_{j=1}^{n} f(y_{ij})\right\} = -\frac{K}{2}\log(2\pi\sigma^2) \boxed{-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij} - X'_{ij}\beta)^2/\sigma^2}$$

avaliado nos valores numéricos observados dos dados, em relação aos parâmetros de regressão, β.

- Aqui $K = n \times N$, o **número total de observações**.

Observações independentes

- Observe que β não aparece no primeiro termo da log-verossimilhança.
  - Como resultado, esse termo pode ser ignorado ao maximizar a log-verossimilhança em relação a β.
- Além disso, como o segundo termo tem um sinal negativo, maximizar a log-verossimilhança em relação a β é equivalente a minimizar a seguinte função:

$$\sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij} - X_{ij}'\beta)^2.$$

Observações independentes

Comentários

- Maximizar ou minimizar uma função é um problema matemático comum que pode ser resolvido usando o cálculo.
- Especificamente, a estimativa de máxima verossimilhança de β pode ser obtida igualando a derivada da log-verossimilhança, frequentemente chamada de **função escore**, a zero e encontrando a solução para a equação resultante.
- No entanto, no exemplo considerado aqui, não há necessidade real de recorrer ao cálculo.
- A obtenção da estimativa de máxima verossimilhança de β é equivalente a encontrar a estimativa de **mínimos quadrados ordinários** **(MQO)** de β, ou seja, o valor de β que minimiza a soma dos quadrados dos resíduos.

Observações independentes

- Usando a notação vetorial, a solução dos mínimos quadrados pode ser escrita como

$$\hat{\beta} = \left\{ \sum_{i=1}^{N} \sum_{j=1}^{n} (X_{ij} X'_{ij}) \right\}^{-1} \sum_{i=1}^{N} \sum_{j=1}^{n} (X_{ij} y_{ij}).$$

Observações independentes ($\hat{\beta}$)

- Dado K observações (X_{ij}, y_{ij}), queremos encontrar:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} \sum_{j=1}^{n} (y_{ij} - X'_{ij}\beta)^2.$$

- Lembrando que y_{ij} é um escalar, X'_{ij} é um vetor $1 \times p$ e β um vetor $p \times 1$.
- Assim, queremos encontrar $\hat{\beta}$ para o qual

$$\frac{\partial}{\partial \beta} \sum_{i=1}^{N} \sum_{j=1}^{n} (y_{ij} - X'_{ij}\beta)^2 = 0.$$

Observações independentes ($\hat{\beta}$)

$$
\begin{aligned}
\frac{\partial}{\partial\beta}\sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij}-X_{ij}'\beta)^2 &= \sum_{i=1}^{N}\sum_{j=1}^{n}\frac{\partial}{\partial\beta}(y_{ij}-X_{ij}'\beta)^2\\
&= \sum_{i=1}^{N}\sum_{j=1}^{n} 2\times(y_{ij}-X_{ij}'\beta)\times(-X_{ij}')\\
&= 2\times\left[\sum_{i=1}^{N}\sum_{j=1}^{n}X_{ij}'\beta X_{ij}' - \sum_{i=1}^{N}\sum_{j=1}^{n}y_{ij}X_{ij}'\right]\\
&= 2\times\left[\sum_{i=1}^{N}\sum_{j=1}^{n}\boxed{(X_{ij}X_{ij}')\beta} - \sum_{i=1}^{N}\sum_{j=1}^{n}\boxed{(X_{ij}y_{ij})}\right],
\end{aligned}
$$

pois $(X_{ij}'\beta)X_{ij}' = (\beta' X_{ij})X_{ij}' = X_{ij}(\beta' X_{ij})' = X_{ij}X_{ij}'\beta$ e $y_{ij}X_{ij}' = X_{ij}y_{ij}$.

Observações independentes ($\hat{\beta}$)

$$\frac{\partial}{\partial\beta}\sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij}-X_{ij}'\beta)^2=0$$

$$\Rightarrow \sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}X_{ij}')\hat{\beta}=\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}y_{ij})$$

$$\Rightarrow \left\{\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}X_{ij}')\right\}^{-1}\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}X_{ij}')\hat{\beta}=\left\{\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}X_{ij}')\right\}^{-1}\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}y_{ij})$$

$$\Rightarrow \hat{\beta}=\left\{\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}X_{ij}')\right\}^{-1}\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}y_{ij}),$$

pois $\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}X_{ij}')$ é matriz $p\times p$, e portanto $\left\{\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}X_{ij}')\right\}^{-1}\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij}X_{ij}')=I_p$.

Observações independentes

Comentários

- Essa estimativa de mínimos quadrados é o valor produzido por qualquer *software* estatístico padrão para regressão linear (por exemplo, `PROC GLM` ou `PROC REG` no `SAS`, a função `lm` no `R` e o comando `regress` no `Stata`).
  - **Exercício:** considere o modelo dos dados de nível de chumbo no sangue **(considerando observações independentes)**; compare as estimativas através de uma **implementação sua** de $\hat{\beta}$ com a função `lm`; poste no fórum geral do Moodle.
- Além disso, note que até agora apenas focamos na estimativa de β, ignorando a estimativa de σ^2; a seguir, também consideramos a estimativa da matriz de covariância.

Observações correlacionadas

- Quando há n_i medidas repetidas no mesmo indivíduo, **não se pode assumir que essas medidas repetidas são independentes**.
  - Como resultado, precisamos considerar a função de densidade de probabilidade conjunta para o **vetor de medidas repetidas**.
- Observe, no entanto, que os vetores de medidas repetidas são assumidos como **independentes** uns dos outros.
  - Assim, a função log-verossimilhança, ℓ, pode ser expressa como uma **soma das funções multivariadas individuais** da densidade de probabilidade normal para Y_i dado X_i.

$$\ell = \log\left\{\prod_{i=1}^{N} f(y_i)\right\} = \sum_{i=1}^{N} \log f(y_i),$$

em que

$$f(y_i) = (2\pi)^{-n_i/2}|\Sigma_i|^{-1/2}\exp\left\{-\frac{1}{2}(y_i - \mu_i)'\Sigma_i^{-1}(y_i - \mu_i)\right\}.$$

Observações correlacionadas

- Primeiro **assumimos que** Σ_i (ou θ) **é conhecido** (e, portanto, não precisa ser estimado); depois, relaxaremos essa suposição muito irrealista.
- Dado que assumimos que o vetor de respostas $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})'$ segue uma distribuição condicional que é normal multivariada, devemos maximizar a seguinte função de log-verossimilhança:

$$\ell = -\frac{K}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\log|\Sigma_i| \boxed{-\frac{1}{2}\left\{\sum_{i=1}^{N}(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta)\right\}}$$

em que $K = \sum_{i=1}^{N} n_i$ é o **número total de observações**.

Observações correlacionadas

- Note que β não aparece nos dois primeiros termos na log-verossimilhança.
  - Esses dois termos podem ser ignorados ao maximizar a log-verossimilhança em relação a β.
- Além disso, como o terceiro termo tem um sinal negativo, maximizar a log-verossimilhança em relação a β é equivalente a minimizar

$$\sum_{i=1}^{N}(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta).$$

Observações correlacionadas

- O estimador de β que minimiza essa expressão é conhecido como estimador de **mínimos quadrados generalizados (MQG)** de β e pode ser expresso como

$$\hat{\beta} = \left\{ \sum_{i=1}^{N} (X_i' \Sigma_i^{-1} X_i) \right\}^{-1} \sum_{i=1}^{N} (X_i' \Sigma_i^{-1} y_i).$$

- Veja a função `gls` do pacote `nlme` do `R`.

Observações correlacionadas ($\hat{\beta}$)

$$\begin{aligned}
\frac{\partial}{\partial\beta}\sum_{i=1}^{N}(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta) &= \sum_{i=1}^{N}\frac{\partial}{\partial\beta}\left[(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta)\right]\\
&= \sum_{i=1}^{N}(y_i - X_i\beta)' \times \left[\Sigma_i^{-1} + (\Sigma_i^{-1})'\right] \times (-X_i)\\
&= \sum_{i=1}^{N}\left[y_i' - (X_i\beta)'\right] \times \boxed{(2\Sigma_i^{-1})} \times (-X_i)\\
&= 2 \times \left\{\sum_{i=1}^{N}\left[(X_i\beta)'\Sigma_i^{-1}X_i\right] - \sum_{i=1}^{N}\left(y_i'\Sigma_i^{-1}X_i\right)\right\}\\
&= 2 \times \left[\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\beta - \sum_{i=1}^{N}(X_i'\Sigma_i^{-1}y_i)\right].
\end{aligned}$$

- Note que Σ_i é simétrica, e portanto, Σ_i^{-1} também é simétrica; ou seja, $\boxed{(\Sigma_i^{-1})' = \Sigma_i^{-1}}$.

Observações correlacionadas ($\hat{\beta}$)

$$
\begin{aligned}
&\frac{\partial}{\partial\beta}\sum_{i=1}^{N}(y_i - X_i\beta)'\Sigma_i^{-1}(y_i - X_i\beta) = 0\\
&\Rightarrow \sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\hat{\beta} = \sum_{i=1}^{N}(X_i'\Sigma_i^{-1}y_i)\\
&\Rightarrow \hat{\beta} = \left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}y_i).
\end{aligned}
$$

Propriedades do estimador MQG (Σ_i conhecida)

- A primeira propriedade muito notável é que, para **qualquer escolha** de Σ_i, a estimativa MQG de β é **não enviesada**. Ou seja,

$$\mathrm{E}\,[\hat{\beta}] = \beta.$$

$$\begin{aligned}
\mathrm{E}\,[\hat{\beta}] &= \mathrm{E}\left[\left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}Y_i)\right]\\
&= \left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}\mathrm{E}\,[Y_i])\\
&= \left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\beta\\
&= I_p\beta = \beta.
\end{aligned}$$

Propriedades do estimador MQG (Σ_i conhecida)

- Além disso, **em grandes amostras** (ou assintoticamente), pode se mostrar que a distribuição amostral de $\hat{\beta}$ é uma **distribuição normal multivariada** com **média** β e **covariância**

$$\text{Cov}\,(\hat{\beta}) = \left\{ \sum_{i=1}^{N} (X_i' \Sigma_i^{-1} X_i) \right\}^{-1}.$$

Propriedades do estimador MQG (Σ_i conhecida)

$$
\begin{aligned}
\text{Cov}\,[\hat{\beta}] &= \text{Var}\left[\left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}Y_i)\right] \\
&= \left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}\text{Var}\left[\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}Y_i)\right]\left[\left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}\right]' \\
&= \left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}\sum_{i=1}^{N}\left[X_i'\Sigma_i^{-1}\text{Var}\,(Y_i)(X_i'\Sigma_i^{-1})'\right]\left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1} \\
&= \underbrace{\left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}\sum_{i=1}^{N}\left(X_i'\overbrace{\Sigma_i^{-1}\Sigma_i}^{I_{n_i}}\Sigma_i^{-1}X_i\right)}_{I_p}\left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1} \\
&= \left\{\sum_{i=1}^{N}(X_i'\Sigma_i^{-1}X_i)\right\}^{-1}.
\end{aligned}
$$

Propriedades do estimador MQG (Σ_i conhecida)

- Isso é verdade exatamente quando Y_i tem uma distribuição condicional que é multivariada normal e verdadeiro em **grandes amostras**, mesmo quando a distribuição condicional de Y_i não é normal multivariada.
  - Por "grandes amostras", entendemos que o tamanho da amostra, N, aumenta quando o número de medidas repetidas e parâmetros do modelo permanece fixo.
- Observe também que se Σ_i for assumido como uma matriz diagonal, com variância constante σ^2 ao longo da diagonal, o estimador MQG reduz-se ao estimador MQO considerado mais cedo.
  - **Exercício:** demonstre este resultado.
- Finalmente, embora o estimador MQG de β seja não enviesado para qualquer escolha de Σ_i, pode ser mostrado que o estimador MQG mais eficiente de β (ou seja, o estimador com menor variância ou maior precisão) é aquele que usa o valor verdadeiro de Σ_i.

MQG (Σ_i desconhecida)

- Vamos abordar o caso que geralmente ocorre na prática: não conhecemos Σ_i (ou θ).
  - Neste caso precisamos estimar $\Sigma_i(\theta)$ a partir dos dados disponíveis.
- A estimativa da máxima verossimilhança de θ prossegue da mesma maneira que a estimativa de β, maximizando a log-verossimilhança em relação a θ.
- Especificamente, a estimativa de máxima verossimilhança de θ pode ser obtida igualando a derivada da log-verossimilhança em relação a θ (função escore) a zero e encontrando a solução para a equação resultante.
- Entretanto, em geral, essa **equação é não linear** e não é possível escrever expressões simples, de forma fechada, para o estimador de MV de θ.
  - A estimativa de MV deve ser encontrada resolvendo-se essa equação usando uma técnica **iterativa**.

MQG (Σ_i desconhecida)

- Felizmente, algoritmos computacionais foram desenvolvidos para encontrar a solução.
- Uma vez obtida a estimativa de MV de θ, simplesmente substituímos a estimativa de $\Sigma_i(\theta)$, digamos $\widehat{\Sigma}_i = \Sigma_i(\hat{\theta})$, no estimador de mínimos quadrados generalizados de β para obter a estimativa de MV de β:

$$\hat{\beta} = \left\{\sum_{i=1}^{N} (X_i' \widehat{\Sigma}_i^{-1} X_i)\right\}^{-1} \sum_{i=1}^{N} (X_i' \widehat{\Sigma}_i^{-1} y_i).$$

Propriedades do estimador MQG (Σ_i desconhecida)

- Curiosamente, **em grandes amostras** (ou assintoticamente), o estimador de β resultante que substitui a estimativa de MV de Σ_i tem todas as **mesmas propriedades** de quando Σ_i é realmente conhecido.
- Assim, em termos de propriedades da distribuição amostral de $\hat{\beta}$, não há penalidade por realmente ter que estimar Σ_i a partir dos dados longitudinais em questão.
- No entanto, por mais reconfortante que esse resultado possa parecer, deve-se ter em mente que esta é uma propriedade de grande amostra (ou seja, quando N se aproxima do infinito) de $\hat{\beta}$.
  - Com tamanhos de amostra da magnitude frequentemente encontrados em muitas áreas de aplicação, pode-se esperar que as propriedades da distribuição amostral de $\hat{\beta}$ sejam adversamente influenciadas pela estimativa de um número muito grande de parâmetros de covariância.

Questões de dados ausentes

Questões de dados ausentes

Embora a maioria dos estudos longitudinais seja projetada para coletar dados de cada indivíduo da amostra a cada momento do acompanhamento, **muitos estudos têm algumas observações ausentes**.

Dados ausentes têm três implicações importantes para a análise longitudinal:

1. O **conjunto de dados é** necessariamente **desbalanceado** ao longo do tempo, pois nem todos os indivíduos têm o mesmo número de medições repetidas em um conjunto comum de ocasiões.
    - Como resultado, os métodos de análise precisam ser capazes de lidar com os dados desbalanceados **sem precisar descartar dados de indivíduos com dados ausentes**.

Questões de dados ausentes

2. **Haverá** perda de informações e **redução na precisão** com que mudanças na resposta média ao longo do tempo pode ser estimado.
    - Essa redução na precisão está diretamente relacionada à quantidade de dados ausentes e também será influenciada em certa medida pela maneira como a análise lida com os dados ausentes.
    - Por exemplo, usar apenas os casos completos (ou seja, aqueles indivíduos sem dados ausentes) geralmente **será o método menos eficiente**.
3. A validade de qualquer método de análise exigirá que certas suposições sobre os motivos de qualquer perda, geralmente chamadas de **mecanismo de perda de dados**, sejam sustentáveis.
    - Consequentemente, quando há perda de dados, devemos considerar cuidadosamente os motivos da perda.

Mecanismo de perda de dados

- O mecanismo de perda de dados pode ser pensado como **um modelo** que descreve a probabilidade de uma resposta ser observada ou não (estar ausente) em qualquer ocasião.
- Fazemos uma distinção importante entre mecanismos de dados ausentes que são referidos como **ausentes completamente ao acaso** (*missing completely at random* - MCAR) e **ausentes ao acaso** (*missing at random* - MAR).
- A distinção entre esses dois mecanismos determina a adequação da estimativa de máxima verossimilhança sob o pressuposto de uma distribuição normal multivariada para as respostas, e o MQG sem exigir suposições sobre o formato da distribuição.

MCAR

- Diz-se que os dados são MCAR quando a probabilidade de perda de respostas não está relacionada aos valores específicos que, em princípio, deveriam ter sido obtidos (as respostas ausentes) ou ao conjunto de respostas observadas.
- Ou seja, dados longitudinais são MCAR quando a perda em Y_i é simplesmente o resultado de um mecanismo aleatório que não depende de componentes observados ou não observados de Y_i.

- **A característica essencial do MCAR é que os dados observados podem ser considerados uma amostra aleatória dos dados completos.**
  - Como resultado, os momentos (por exemplo, médias, variâncias e covariâncias) e, de fato, a distribuição dos dados observados **não diferem** dos momentos correspondentes ou da distribuição dos dados completos.

MCAR: consequências para a análise de dados longitudinais

- Qualquer método de análise que produza **inferências válidas na ausência de dados ausentes** também produzirá inferências válidas quando o mecanismo de perda de dados for MCAR e a análise for baseada em todos os dados disponíveis, ou mesmo quando estiver restrito aos "completadores" (ou seja, aqueles sem dados ausentes).
- Dado que estimativas válidas das médias, variâncias e covariâncias podem ser obtidas, o MQG fornece estimativas válidas de β sem exigir nenhuma premissa de distribuição para Y_i.
- O estimador MQG de β é válido, desde que o modelo para a resposta média tenha sido especificado corretamente; não requer nenhuma suposição sobre a distribuição conjunta das respostas longitudinais.

MCAR: consequências para a análise de dados longitudinais

- O estimador de MV de β, pressupondo que as respostas tenham uma distribuição normal multivariada, é também o estimador MQG (com a estimativa MV de $\Sigma_i(\theta)$, por exemplo, $\widehat{\Sigma}_i = \Sigma_i(\hat{\theta})$, substituída).
- Assim, nessa configuração, os estimadores MV e MQG têm exatamente as mesmas propriedades, independentemente da verdadeira distribuição de Y_i.

MAR

- Ao contrário do MCAR, diz-se que os dados são MAR quando a probabilidade de perda de respostas depende do conjunto de respostas observadas, mas não está relacionada aos valores ausentes específicos que, em princípio, deveriam ter sido obtidos.
- Em outras palavras, se os indivíduos são estratificados com base em valores semelhantes para as respostas observadas, a perda é simplesmente o resultado de um mecanismo aleatório que não depende dos valores das respostas não observadas.
- No entanto, como o mecanismo de perda agora depende das respostas observadas, a distribuição de Y_i em cada um dos estratos distintos definidos pelos padrões de perda não é o mesmo que a distribuição de Y_i na população alvo.
- Isso tem consequências importantes para a análise.

MAR: consequências para a análise de dados longitudinais

- Uma destas consequências é que uma **análise restrita aos "completadores" não é válida**.
  - Em outras palavras, os "completadores" são uma amostra tendenciosa da população alvo.
- Além disso, a distribuição dos componentes observados de Y_i, em cada um dos estratos distintos definidos pelos padrões de perda, não coincide com a distribuição dos mesmos componentes de Y_i na população alvo.
  - **Portanto, as médias amostrais, variâncias e covariâncias com base nos "completadores" ou nos dados disponíveis são estimativas tendenciosas dos parâmetros correspondentes na população-alvo**.

MAR: consequências para a análise de dados longitudinais

- Como resultado, o MQG não mais fornece estimativas válidas de β sem fazer suposições corretas sobre a distribuição conjunta das respostas longitudinais.
- Por outro lado, a estimativa de MV de β é válida quando os dados são MAR, desde que a distribuição normal multivariada tenha sido especificada corretamente.
  - **Isso requer a especificação correta não apenas do modelo para a resposta média, mas também do modelo para a covariância entre as respostas.**
- Em certo sentido, a estimativa de MV permite que os valores ausentes sejam validamente "previstos" ou "imputados" usando os dados observados e um modelo correto para a distribuição conjunta das respostas.

Avisos

- **Exercício:** (Re)fazer as provas dos resultados discutidos na aula de hoje.
- **Para casa:** ler o Capítulo 4 do livro "**Applied Longitudinal Analysis**".
  - Caso ainda não tenha lido, leia também os Caps. 1, 2 e 3.
  - Ver também as referências com respeito à derivadas de vetores, matrizes, etc.
- **Próxima aula:** Inferência estatística, máxima verossimilhança restrita.

Bons estudos!