MAT02035 - Modelos para dados correlacionados

Estimação e inferência estatística (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, 2022

Inferência estatística

Inferência estatística

- Para construir **intervalos de confiança** e **testes de hipóteses** sobre β, podemos fazer uso direto das estimativas de MV $\hat{\beta}$ e da sua matriz de covariância estimada

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

Inferência estatística

- Para um único componente de β, digamos β_k, um método natural de construção de **limites de confiança** de 95% é tomar o $\hat{\beta}_k$ mais ou menos $1,96$ vezes o erro padrão de $\hat{\beta}_k$.
  - Diferentes limites de confiança podem ser obtidos escolhendo os quantis apropriados da distribuição normal padrão.
- O erro padrão de $\hat{\beta}_k$ é simplesmente a raiz quadrada do elemento da diagonal principal de $\widehat{\text{Cov}}(\hat{\beta})$ correspondente a $\hat{\beta}_k$,

$$\sqrt{\widehat{\text{Var}}(\hat{\beta}_k)}.$$

Inferência estatística

- De maneira similar, um **teste da hipótese** nula, $H_0 : \beta_k = 0$ versus $H_A : \beta_k \neq 0$, pode ser baseado na seguinte estatística de Wald:

$$Z = \frac{\hat{\beta}_k}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_k)}},$$

em que $\widehat{\text{Var}}(\hat{\beta}_k)$ denota o elemento da diagonal principal de $\widehat{\text{Cov}}(\hat{\beta})$ correspondente a $\hat{\beta}_k$.

- Esta estatística de teste pode ser comparada com uma distribuição normal padrão.

Inferência estatística

- De maneira mais geral, pode ser de interesse construir intervalos de confiança e testes de hipóteses com respeito a certas **combinações lineares** dos componentes de β.
- Seja L um vetor ou uma matriz de pesos **conhecidos** e suponha que é de interesse testar $H_0 : L\beta = 0$.
- A combinação linear dos componentes de β, $L\beta$, representa um **contraste** de interesse científico.

Exemplo

- Suponha que $\beta = (\beta_1, \beta_2, \beta_3)'$ e $L = (0, 0, 1)$, então $H_0 : L\beta = 0$ é equivalente a $H_0 : \beta_3 = 0$.
  - Agora, se considerarmos $L = (0, 1, -1)$, então $H_0 : L\beta = 0$ é equivalente a $H_0 : \beta_2 - \beta_3 = 0$ ou $H_0 : \beta_2 = \beta_3$.

Inferência estatística

- Uma estimativa natural de $L\beta$ é $L\hat{\beta}$, e pode ser mostrado que $L\hat{\beta}$ tem distribuição normal multivariada com média $L\beta$ e matriz de covariância $L\text{Cov}\,(\hat{\beta})L'$.
- Nos dois exemplos anteriores, L é um único vetor linha 1×3, $L = (0, 0, 1)$ ou $L = (0, 1, -1)$.
- Se L é um único vetor linha, então $L\text{Cov}\,(\hat{\beta})L'$ é um escalar e a sua raiz quadrada fornece uma estimativa do erro padrão para $L\hat{\beta}$.
- Assim um intervalo de confiança de aproximadamente 95% para $L\beta$ é dado por

$$L\hat{\beta} \pm 1,96\sqrt{L\widehat{\text{Cov}}\,(\hat{\beta})L'}.$$

Inferência estatística

- De forma similar, para testar $H_0 : L\beta = 0$ versus $H_A : L\beta \neq 0$, podemos usar a estatística de Wald

$$Z = \frac{L\hat{\beta}}{\sqrt{L\widehat{\text{Cov}}(\hat{\beta})L'}},$$

e comparar esta estatística de teste com a distribuição normal padrão.

- Um teste idêntico para $H_0 : L\beta = 0$ versus $H_A : L\beta \neq 0$ usa a estatística

$$W^2 = (L\hat{\beta})\{L\widehat{\text{Cov}}(\hat{\beta})L'\}^{-1}(L\hat{\beta}),$$

e comparar W^2 com a distribuição qui-quadrado com 1 grau de liberdade[1].

[1] Lembre que se $Z \sim N(0, 1)$, então $Z^2 \sim \chi^2_{(1)}$.

Inferência estatística

- Esta última observação nos ajuda a motivar como o teste de Wald prontamente generaliza quando L tem mais que uma linha, permitindo o **teste simultâneo** de uma **hipótese multivariada**.
- Por exemplo, suponha que $\beta = (\beta_1, \beta_2, \beta_3)'$ e é de interesse testar a igualdade dos três parâmetros de regressão.
  - A hipótese nula pode ser expressa como $H_0 : \beta_1 = \beta_2 = \beta_3$. Fazendo

$$L = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

Inferência estatística

- $H_0 : \beta_1 = \beta_2 = \beta_3$ pode ser expressa como $H_0 : L\beta = 0$, pois se

$$
\begin{aligned}
L\beta &= \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \\
&= \begin{pmatrix} \beta_1 - \beta_2 \\ \beta_1 - \beta_3 \end{pmatrix} = 0,
\end{aligned}
$$

e então

$$
\begin{pmatrix} \beta_1 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix}.
$$

Inferência estatística

- Em geral, suponha que L tem r linhas (representando r contrastes de interesse científico), então um teste simultâneo de $H_0 : L\beta = 0$ versus $H_A : L\beta \neq 0$ é dado por

$$W^2 = (L\hat{\beta})'\{L\widehat{\text{Cov}}(\hat{\beta})L'\}^{-1}(L\hat{\beta}),$$

que tem uma distribuição qui-quadrado com r graus de liberdade ($\chi^2_{(r)}$).

- Este é o **teste de Wald multivariado**.

Inferência estatística

- Uma alternativa para o teste de Wald é o **teste da razão de verossimilhanças**.

TRV

O teste da razão de verossimilhanças de $H_0 : L\beta = 0$ versus $H_A : L\beta \neq 0$ é obtido comparando as log-verossimilhanças maximizadas para dois modelos:

- um modelo incorpora a restrição que $L\beta = 0$ (por exemplo, $\beta_3 = 0$ ou $\beta_2 = \beta_3$); este será chamado de **modelo reduzido**;
- e outro modelo sem restrição; este será chamado de **modelo completo**.
- **Observação:** note que estes dois modelos são **encaixados**, no sentido que o modelo "reduzido" é um caso especial do modelo "completo".

Inferência estatística

- Uma estatística de teste é obtida por

$$G^2 = 2(\hat{\ell}_{\text{comp}} - \hat{\ell}_{\text{red}}),$$

e comparamos esta estatística com uma **distribuição qui-quadrado** com graus de liberdade igual a diferença entre o número de parâmetros nos modelos completo e reduzido[2].

Observação

O teste da razão de verossimilhanças só pode ser usado quando o número de observações para os modelos completo e reduzido for o mesmo; **atenção com dados ausentes nas covariáveis**.

[2] Em geral, a estatística da razão de verossimilhanças é definida como $\lambda = \hat{L}_{\text{red}}/\hat{L}_{\text{comp}}$. Quando G^2 é "grande", rejeita-se H_0.

Inferência estatística

- O uso da verossimilhança também pode fornecer limites de confiança para β ou $L\beta$.
- Para um único componente de β, digamos β_k, podemos definir uma função **log-verossimilhança perfilada**, $\ell_p(\beta_k)$, obtida maximizando a log-verossimilhança sobre os parâmetros restantes, mantendo β_k em algum valor fixo.

Inferência estatística

- Um intervalo de confiança baseado na verossimilhança para β_k é obtido considerando-se valores de β_k que são razoavelmente consistentes com os dados.
  - Especificamente, um intervalo de confiança aproximado de 95% baseado na verossimilhança é dado pelo conjunto de todos os valores de β_k que satisfazem

$$2 \times \{\ell_p(\hat{\beta}_k) - \ell_p(\beta_k)\} \leq 3,84,$$

em que o valor crítico $3,84$ é obtido da distribuição qui-quadrado com 1 grau de liberdade.

- De maneira mais geral, os intervalos de confiança para $L\beta$ podem ser obtido invertendo o teste correspondente de $H_0 : L\beta = 0$ de maneira semelhante.

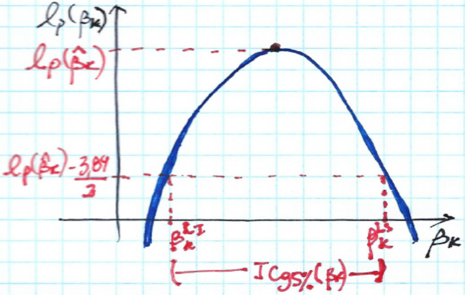
Inferência estatística

IC baseado na verossimilhança perfilada

$$2 \times \{ \ell_p(\hat{\beta}_k) - \ell_p(\beta_k) \} \le 3,84$$

$$\Rightarrow \ell_p(\hat{\beta}_k) - \ell_p(\beta_k) \le \frac{3,84}{2}$$

$$\Rightarrow \boxed{\ell_p(\beta_k) \ge \ell_p(\hat{\beta}_k) - \frac{3,84}{2}}$$



Comentários

- Embora a construção de testes de razão de verossimilhanças e intervalos de confiança com base em verossimilhança seja mais enredada (por exemplo, exigindo um ajuste adicional do modelo sob a hipótese nula) do que os correspondentes testes e intervalos de confiança com base na estatística de Wald, os testes e intervalos de confiança baseados em verossimilhança geralmente têm **propriedades superiores**.
- Este é especialmente o caso quando a **variável resposta é discreta**.
    - Por exemplo, na regressão logística com dados binários, os testes de razão de verossimilhanças têm melhores propriedades que os testes de Wald correspondentes.
- Portanto, em caso de dúvida, recomenda-se o uso de testes e intervalos de confiança baseados em verossimilhança.

Comentários

- Notamos que testes de razão de verossimilhanças também podem ser usados para hipóteses sobre os **parâmetros de covariância**.
- No entanto, existem alguns problemas em potencial com o uso padrão do teste da razão de verossimilhanças para comparar modelos encaixados para a covariância; retornaremos a este tópico quando discutirmos a modelagem da estrutura de covariância.
- Em geral, não é recomendado testar hipóteses sobre os parâmetros de covariância usando testes de Wald.
  - Em particular, a distribuição amostral da estatística do teste Wald para um parâmetro de variância não possui uma distribuição normal aproximada quando o tamanho da amostra é relativamente pequeno e a variância populacional é próxima de zero.
  - Como a variância tem um limite inferior igual a zero, são necessárias amostras muito grandes para justificar a aproximação normal para a distribuição amostral da estatística do teste de Wald quando a variância está próxima de zero.

Estimação por máxima verossimilhança restrita (REML)

REML

- **Relembrando:** a estimativa de MV de β e θ (ou Σ_i) é obtida maximizando a seguinte função de log-verossimilhança

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

Embora os estimadores de máxima verossimilhança (EMV) tenham as propriedades usuais de grandes amostras (ou assintóticas), o EMV de Σ_i **tem viés bem conhecido em pequenas amostras** (por exemplo, **os elementos da diagonal de Σ_i são subestimados**).

REML

- Para ver o problema, considere a regressão linear com erros independentes.
- Se as K observações são independentes (n estudos transversais com N indivíduos cada), maximizamos

$$-\frac{K}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij} - X_{ij}'\beta)^2/\sigma^2.$$

- Isso fornece o estimador de mínimos quadrados usual de β, mas o EMV de σ^2 é

$$\hat{\sigma}^2 = \sum_{i=1}^{N}\sum_{j=1}^{n}(Y_{ij} - X_{ij}'\hat{\beta})^2/K.$$

REML

- Pode ser mostrado que

$$\mathrm{E}\left(\hat{\sigma}^2\right) = \left(\frac{K - p}{K}\right)\sigma^2,$$

em que p é a dimensão de β.

- Como resultado, o EMV de σ^2 será enviesado em pequenas amostras e subestima σ^2 (note que $(K - p)/K < 1$).

REML

- Com efeito, o viés aparece porque o EMV de σ^2 não leva em consideração que β também é estimado pelos dados.
  - Ou seja, no estimador de σ^2 substituímos β por $\hat{\beta}$.
- Não deveria ser muito surpreendente que problemas semelhantes apareçam na estimativa de Σ_i ou (θ).
- Um estimador não enviesado é dado por usar $K - p$ como denominador em vez de K.

REML

- A teoria da estimação por máxima verossimilhança restrita[3] **(REML)**[4] foi desenvolvida para resolver esse problema.
- A principal ideia por trás da REML é eliminar os parâmetros β da verossimilhança de modo que seja definida apenas em termos de Σ_i.
- Uma maneira possível de obter a verossimilhança restrita é considerar transformações dos dados em um conjunto de combinações lineares de observações que têm uma distribuição da qual não depende de β.
- Por exemplo, os resíduos após estimar β por **mínimos quadrados ordinários** podem ser usados[5].
- A verossimilhança desses resíduos dependerá apenas de Σ_i, e não de β.

---

[3] Também referida na literatura como estimação por máxima verossimilhança **residual**.

[4] Do inglês *restricted maximum likelihood estimation*.

[5] Veja uma justificativa desta modificação na log-verossimilhança em SINGER *et. al*, *Análise de dados longitudinais*, 2018, p. 47–48).

REML

- Assim, ao invés de maximizar

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

a REML maximiza a seguinte log-verossimilhança ligeiramente modificada

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

$$\boxed{-\frac{1}{2}\log\left|\sum_{i=1}^{N}X_i'\Sigma_i^{-1}X_i\right|}.$$

- Quando a verossimilhança restrita é maximizada, obtemos uma estimativa menos enviesada de $\Sigma_i(\theta)$.
- Ou seja, o termo determinante extra efetivamente faz uma correção ou ajustes que são análogos à correção do denominador em $\hat{\sigma}^2$.

REML

- Quando a estimativa REML é usada, obtemos as estimativas MQG de β,

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

em que $\widehat{\Sigma}_i = \Sigma_i(\hat{\theta})$ é a estimativa REML de Σ_i.

Comentário

- Recomenda-se o uso da REML para estimação de Σ_i (com β estimado usando o estimador MQG que substitui a estimativa REML, $\widehat{\Sigma}_i$, por Σ_i).
- A log-verossimilhança REML também deve ser usado para comparar modelos encaixados **para a covariância**.
- No entanto, a construção de testes de razão de verossimilhanças comparando modelos encaixados **para a média deve sempre ser baseada na log-verossimilhança MV**, e não no REML.

Avisos

- **Exercício:** utilize os dados do estudo dos níveis de chumbo no sangue (TLC).
  - Proponha um modelo de regressão linear para a média com base nas questões de pesquisa.
  - Encontre as estimativas para os coeficientes de regressão do modelo proposto (use a função `gls` do pacote `nlme` do `R`; explore diferentes especificações da covariância - veja a documentação de `corClasses`).
  - **Com base nas estimativas**, faça a exposição das suas conclusões.
  - **Compartilhe** a sua solução no fórum geral do Moodle.
- **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:** Modelando a média através da análise de perfis de respostas.

Bons estudos!