

MAT02035 - Modelos para dados correlacionados

Equações de Estimação Generalizadas - Exemplos

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, 2019

Equações de estimação generalizadas

- ▶ Modelo de média populacional ou marginal, fornece uma abordagem de regressão para modelos lineares generalizados quando as respostas não são independentes (dados correlacionados/agrupados).
- ▶ O objetivo é fazer inferências sobre a população, levando em consideração a correlação das medidas dentro de indivíduo.
- ▶ Os pacotes gee e geepack são usados para modelos GEE no R.
- ▶ A principal diferença entre gee e geepack é que o geepack contém um método ANOVA que nos permite comparar modelos e realizar testes de Wald.

Equações de estimação generalizadas

- ▶ Sintaxe básica para `geeglm()` do pacote `geepack`; tem uma sintaxe muito parecida com `glm()`

```
geeglm(formula, family = gaussian, data, id,  
       zcor = NULL, constr, std.err = "san.se")
```

- ▶ `formula` Descrição simbólica do modelo a ser ajustado
- ▶ `family` Descrição da distribuição da resposta e função de ligação
- ▶ `data` dataframe opcional
- ▶ `id` vetor que identifica os *clusters* (agrupamentos)
- ▶ `zcor` especifica uma estrutura de correlação definida pelo usuário
- ▶ `constr` estrutura de correlação de trabalho: "independence", "exchangeable", "ar1", "unstructured", "userdefined"
- ▶ `std.err` tipo de erro padrão a ser calculado. O padrão "san.se" é a estimativa robusta (sanduíche); use "jack" para obter uma estimativa da variância aproximada por *jackknife*

Estrutura de correção

- Independence (independência),

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Exchangeable (simetria composta),

$$\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

Estrutura de correção

- ▶ Autoregressive order 1 (autorregressivo de ordem 1),

$$\begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$$

- ▶ Unstructured (não estruturada),

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$$

- ▶ O modelo GEE fornecerá resultados válidos com uma estrutura de correlação mal especificada quando o estimador de variância sanduíche for usado.

Inferência

- ▶ Para um objeto `geeglm` retornado por `geeglm()`, as funções `drop1()`, `confint()` e `step()` não se aplicam; no entanto `anova()` se aplica.
- ▶ A função `esticon()` no pacote `doBy` calcula e testa funções lineares dos parâmetros de regressão para objetos `lm`, `glm` e `geeglm`
- ▶ Sintaxe básica,

```
esticon(obj, cm, beta0, joint.test = FALSE)
```

- ▶ `obj` objeto do modelo
- ▶ `L` matriz especificando funções lineares dos parâmetros de regressão (uma função linear por linha e uma coluna para cada parâmetro)
- ▶ `beta0` vetor de números (H_0 para β)
- ▶ `joint.test` Se `TRUE` um teste de hipóteses de Wald conjunto `Lbeta = beta0` é realizado, *default* é um teste para cada linha, `(Lbeta).i=beta0.i`

Exemplo - GEE

```
# Instala e carrega os pacotes geepack e doBy
install.packages("geepack")
install.packages("doBy")
library(geepack)
library(doBy)

# conjunto de dados ohio do geepack - Efeito da poluição do ar na saúde
# Crianças acompanhadas por quatro anos, com chiado registrado anualmente
data(ohio) # carrega o conjunto de dados
head(ohio)
str(ohio)

# Variável resposta é binária - ajuste um modelo GEE logístico
# tempo (idade; age) como var. contínua
fit.exch <- geeglm(resp ~ age + smoke,
                  family = binomial(link = "logit"),
                  data = ohio, id = id,
                  corstr = "exchangeable", std.err = "san.se")

fit.unstr <- geeglm(resp ~ age + smoke,
                   family = binomial(link = "logit"),
                   data = ohio, id = id,
                   corstr = "unstructured", std.err = "san.se")

summary(fit.exch)
summary(fit.unstr)
```

Exemplo - GEE

```
# tempo (idade; age) como var. categórica
fit <- geeglm(resp ~ factor(age) + smoke,
              family = binomial(link = "logit"),
              data = ohio, id = id,
              constr = "exchangeable", std.err = "san.se")

summary(fit)

# Teste o efeito de smoke usando anova()
fit1 <- geeglm(resp ~ factor(age) + smoke,
               family = binomial(link = "logit"),
               data = ohio, id = id,
               constr = "exchangeable", std.err = "san.se")

fit2 <- geeglm(resp ~ factor(age),
               family = binomial(link = "logit"),
               data = ohio, id = id,
               constr = "exchangeable", std.err = "san.se")

anova(fit1, fit2)

# Teste Wald individual e intervalo de confiança para cada parâmetro
est <- esticon(fit, diag(5))

# Odds ratio and confidence intervals
OR.CI <- exp(cbind(est$estimate, est$lwr, est$upr))
rownames(OR.CI) <- names(coef(fit))
colnames(OR.CI) <- c("OR", "OR 95% LI", "OR 95% LS")
```


Exemplo - GEE

```
# Razão de chance de chiado no peito para uma criança de 9 anos com uma mãe que
# fumou durante o primeiro ano do estudo em comparação com uma criança de 8
# anos com uma mãe que não fumou durante o primeiro ano do estudo.
# Isto é, estimar [smoke+factor(age)0] - [factor(age)-1]
esticon(fit, c(0,-1,1,0,1))
exp(.Last.value$estimate)
# 9 anos de idade com mãe que fumava tem maior risco de chiado no peito
# Teste conjuntamente os efeitos usando esticon()
fit <- geeglm(resp ~ factor(age)*smoke,
              family = binomial(link = "logit"),
              data = ohio, id = id,
              corstr = "exchangeable", std.err = "san.se")

summary(fit)
L = cbind(matrix(0, nrow=3, ncol=5), diag(3))
esticon(fit, L, joint.test=TRUE)
# Também poderia usar anova()
fit1 <- geeglm(resp ~ factor(age)*smoke,
               family = binomial(link = "logit"),
               data = ohio, id = id,
               corstr = "exchangeable", std.err = "san.se")
fit2 <- geeglm(resp ~ factor(age) + smoke,
               family = binomial(link = "logit"),
               data = ohio, id = id,
               corstr = "exchangeable", std.err = "san.se")
anova(fit1, fit2)
```

Avisos

- ▶ **Próxima aula (03/12):** Modelos marginais (GEE) - exemplos.
- ▶ **Para casa:** ler o Capítulo 12 e 13 do livro “**Applied Longitudinal Analysis**”.
 - ▶ Caso ainda não tenha lido, leia também os Caps. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 e 11.
- ▶ **Para casa:** veja o *help* do pacote *geepack* do R.

Bons estudos!

The word "Gee" is written in a large, elegant, cursive script. The letters are filled with a red halftone dot pattern, giving it a textured appearance. The logo is centered on a light-colored, slightly textured rectangular background.

G I R L S ' G E N E R A T I O N