eg1: respiratory data set

quote from the following source: "The data are from a clinical trial of patients with respiratory illness, where 111 patients from two different clinics were randomized to receive either placebo or an active treatment. Patients were examined at baseline and at four visits during treatment. At each examination, respiratory status (categorized as 1 = good, 0 = poor) was determined."

Source : http://staff.pubhealth.ku.dk/~pd/mixed-jan.2006/R-mixed-geeglm-Lecture.pdf

install.packages(“HSAUR”)

install.packages(“gee”)

library(HSAUR)

library(gee)

data(respiratory)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Center | id | treat | Sex | age | baseline | visit1 | visit2 | visit3 | visit4 |
| 1 | 1 | P | M | 46 | 0 | 0 | 0 | 0 | 0 |
| 1 | 2 | P | M | 28 | 0 | 0 | 0 | 0 | 0 |
| 1 | 3 | A | M | 23 | 1 | 1 | 1 | 1 | 1 |
| 1 | 4 | P | M | 44 | 1 | 1 | 1 | 1 | 0 |
| 1 | 5 | P | F | 13 | 1 | 1 | 1 | 1 | 1 |
| 1 | 6 | A | M | 34 | 0 | 0 | 0 | 0 | 0 |
| 1 | 7 | P | M | 43 | 0 | 1 | 0 | 1 | 1 |
| 1 | 8 | A | M | 28 | 0 | 0 | 0 | 0 | 0 |

Respiratory data for eight subjects. Measurements on

the same individual are alike.

eg2: OME data set

積液性中耳炎（**Otitis media** with effusion）

Otitis 耳炎

[OME function | R Documentation](https://www.rdocumentation.org/packages/MASS/versions/7.3-53/topics/OME)

install.packages(“MASS”)

library(MASS)

data(OME)

library(gee)

fm <- gee(cbind(Correct, Trials-Correct) ~ Loud + Age + OME, id = ID, data = OME, family = binomial, corstr = "exchangeable")

otitis media with effusion 中耳積水

otitis: inflammation of the ear

OME

**Format**

The OME data frame has 1129 rows and 7 columns:

ID

Subject ID (1 to 99, with some IDs missing). A few subjects were measured at different ages.

OME

"low" or "high" or "N/A" (at ages other than 30 and 60 months).

Age

Age of the subject (months).

Loud

Loudness of stimulus, in decibels.

Noise

Whether the signal in the stimulus was "coherent" or "incoherent".

Correct

Number of correct responses from trials.

Trials

Number of trials performed.

Naive independent model:

Disadvantage:

Generalized estimating equations (GEE)

eg1. Consider iid random variables y1,...,yn with common mean c?

How to estimate c? Least squares estimator:

Estimating equation:

eg2: y1,...,yn are iid random variables with density function

f(y,c), where c is an unknown parameter.

How to estimate c? maximum likelihood estimator:

Estimating equation

From eg1 and eg2, we observe that the expectation of the estimating equations are usually 0.

Therefore, a natural way to construct an estimating equation is

Specifications for GEE

Data: There are n subjects (n responses), and for each subject i there are ni repeated observations.

(a)Systematic part: Like generalized linear model, link the expectation of the each response variable to the linear predictor via the link function.

(b)Random part: specify how the variance of the response variable is associated with its mean.

(c) correlation matrix:

eg.

Generalized Estimating equation (Liang and Zeeger 1986):

Source: https://cran.r-project.org/web/packages/HSAUR2/HSAUR2.pdf

Format

A data frame with 555 observations on the following 7 variables.

centre the study center, a factor with levels 1 and 2.

treatment the treatment arm, a factor with levels placebo and treatment.

gender a factor with levels female and male.

age the age of the patient.

status the respiratory status (response variable), a factor with levels poor and good.

month the month, each patient was examined at months 0, 1, 2, 3 and 4.

subject the patient ID, a factor with levels 1 to 111.

Details

In each of two centres, eligible patients were randomly assigned to active treatment or placebo.

During the treatment, the respiratory status (categorised poor or good) was determined at each of

four, monthly visits. The trial recruited 111 participants (54 in the active group, 57 in the placebo

group) and there were no missing data for either the responses or the covariates. The question of

interest is to assess whether the treatment is effective and to estimate its effect.

Note that the data are in long form, i.e, repeated measurments are stored as additional rows in the

data frame.

Source

C. S. Davis (1991), Semi-parametric and non-parametric methods for the analysis of repeated measurements

with applications to clinical trials. Statistics in Medicine, 10, 1959–1980.

The baseline status, i.e., the status for month == 0, will enter the models as

an explanatory variable and thus we have to rearrange the data.frame respiratory

in order to create a new variable baseline:

install.packages(“HSAUR”)

install.packages(“gee”)

library(HSAUR)

library(gee)

data(respiratory)

#library(HSAUR)

#library(gee)

#data(respiratory)

#attach(respiratory)

data("respiratory", package = "HSAUR2")

resp <- subset(respiratory, month > "0")

resp$baseline <- rep(subset(respiratory, month == "0")$status,rep(4, 111))

resp$nstat <- as.numeric(resp$status == "good")

resp$month <- resp$month[, drop = TRUE]

#resp\_gee2 <- gee(nstat ~ centre + treatment + gender + baseline+age, data = resp, family = "binomial", id = subject,

#corstr = "exchangeable", scale.fix = TRUE, scale.value = 1)

resp\_gee2 <- gee(nstat ~ centre + treatment + baseline+age, data = resp, family = "binomial", id = subject,

corstr = "exchangeable", scale.fix = TRUE, scale.value = 1)

Please use other correlation structures such as “AR1” and “unstructured” and do the analysis again.

Please find an R package that can deal with generalize mixed effects model (response is binary and id as random intercept) to analyze the data again.

[Mixed Effects Logistic Regression | R Data Analysis Examples (ucla.edu)](https://stats.oarc.ucla.edu/r/dae/mixed-effects-logistic-regression/)

**require**(ggplot2)

**require**(GGally)

**require**(reshape2)

**require**(lme4)

**require**(compiler)

**require**(parallel)

**require**(boot)

**require**(lattice)

hdp <- **read.csv**("https://stats.idre.ucla.edu/stat/data/hdp.csv")

hdp <- **within**(hdp, {

Married <- **factor**(Married, levels = 0:1, labels = **c**("no", "yes"))

DID <- **factor**(DID)

HID <- **factor**(HID)

CancerStage <- factor(CancerStage)

})

m <- **glmer**(remission ~ IL6 + CRP + CancerStage + LengthofStay + Experience +

(1 | DID), data = hdp, family = binomial, control = **glmerControl**(optimizer = "bobyqa"),

nAGQ = 10)

*# print the mod results without correlations among fixed effects*

**print**(m, corr = FALSE)

Source: <http://www.unc.edu/courses/2010spring/ecol/562/001/docs/lectures/lecture14.htm>

<http://www.stats4stem.org/r-respiratory-data.html>

<http://staff.pubhealth.ku.dk/~pd/mixed-jan.2006/R-mixed-geeglm-Lecture.pdf>

From the source: https://cran.r-project.org/web/packages/MASS/MASS.pdf

" The experiment was to study otitis media with effusion (OME), a very common childhood condition

where the middle ear space, which is normally air-filled, becomes congested by a fluid. There is

a concomitant fluctuating, conductive hearing loss which can result in various language, cognitive

and social deficits. The term ‘binaural hearing’ is used to describe the listening conditions in which

the brain is processing information from both ears at the same time. The brain computes differences

in the intensity and/or timing of signals arriving at each ear which contributes to sound localisation

and also to our ability to hear in background noise.

Some years ago, it was found that children of 7–8 years with a history of significant OME had

significantly worse binaural hearing than children without such a history, despite having equivalent

sensitivity. The question remained as to whether it was the timing, the duration, or the degree of

severity of the otitis media episodes during critical periods, which affected later binaural hearing. In

an attempt to begin to answer this question, 95 children were monitored for the presence of effusion

every month since birth. On the basis of OME experience in their first two years, the test population

was split into one group of high OME prevalence and one of low prevalence."

#variables: "centre" "treatment" "sex" #"age" "status" "month" "subject"

respiratory\_new=respiratory[respiratory$month!=0,]

attach(respiratory\_new)

geefit\_res<-gee(status~age+factor(treatment)+factor(sex)

+factor(centre),id=subject,family=binomial, corstr="exchangeable",scale.fix=T,data=respiratory)

+ rblval + factor(year)\*factor(period) + rblval\*factor(period) + size + I(size^2), id=strtno, family=binomial, corstr="exchangeable", scale.fix=T, data=sm)

geefit.ex<-gee(nesting2~factor(year) + factor(period) + factor(deply) + rblval + factor(year)\*factor(period) + rblval\*factor(period) + size + I(size^2), id=strtno, family=binomial, corstr="exchangeable", scale.fix=T, data=sm)

install.packages("HSAUR2")

library(HSAUR2)

[epilepsy function | R Documentation](https://www.rdocumentation.org/packages/HSAUR2/versions/1.1-18/topics/epilepsy)

data("epilepsy", package = "HSAUR2")

itp <- interaction(epilepsy$treatment, epilepsy$period)

tapply(epilepsy$seizure.rate, itp, mean)

per <- rep(log(2),nrow(epilepsy))

epilepsy$period <- as.numeric(epilepsy$period)

names(epilepsy)[names(epilepsy) == "treatment"] <- "trt"

fm <- seizure.rate ~ base + age + trt + offset(per)

epilepsy\_glm <- glm(fm, data = epilepsy, family = "poisson")

epilepsy\_gee1 <- gee(fm, data = epilepsy, family = "poisson", id=subject, corstr = "exchangeable", scale.fix = TRUE)

# epilepsy\_gee1 <- gee(fm, data = epilepsy, family = "poisson", id = subject, corstr = "independence", scale.fix = TRUE, + scale.value = 1)

# epilepsy\_gee2 <- gee(fm, data = epilepsy, family = "poisson", id = subject, corstr = "exchangeable", scale.fix = TRUE, + scale.value = 1)

# epilepsy\_gee1 <- gee(fm, data = epilepsy, family = "poisson", id = subject, corstr = "independence", scale.fix = TRUE, scale.value = 1)

# epilepsy\_gee2 <- gee(fm, data = epilepsy, family = "poisson", id = subject, corstr = "exchangeable", scale.fix = TRUE, scale.value = 1)

Source code:

[A Handbook of Statistical Analyses Using R (r-project.org)](https://cran.r-project.org/web/packages/HSAUR2/vignettes/Ch_analysing_longitudinal_dataII.pdf)

tapply(epilepsy$seizure.rate, itp, var)

install.packages("sos")

library("sos")

findFn("{generalized estimating equation}")

Model:

HW:

Source:(1) [Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence](http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195152968.001.0001/acprof-9780195152968)

(2) http://www.ats.ucla.edu/stat/r/examples/alda/ch3.htm

or https://stats.idre.ucla.edu/r/examples/alda/r-applied-longitudinal-data-analysis-ch-3/

quote from [Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence](http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195152968.001.0001/acprof-9780195152968)

"Each record contains four variables: (1) *ID*; (2) *AGE*, the child’s age (in years) at each assessment (1.0, 1.5, or 2.0);

(3) *COG*, the child’s cognitive performance score at that age; and (4) *PROGRAM*, a dichotomy that describes whether the child participated in the early intervention program."

This is an longitudinal study. We would like to find out whether the early intervention has significant effect on

cognition level. Please use GEE to analysis this dataset. Write down your model and use children’s age as an

adjusted variable. (response: cognition value Yij, intervention is the target covariate and age is the adjusted

variable)

**early.int <- read.table("J:/統計諮詢2020fall/earlyint\_pp.txt")**

require(stats)

id<-early.int[,3]

age<-early.int[ , 4]

cog<-early.int[ , 5]

program<-early.int[ ,6]

library(lattice)

time<-age-1

xyplot(cog~age | id, data=early.int,

panel = function(x, y){

panel.xyplot(x, y)

panel.lmline(x, y)

}, ylim=c(50, 150), as.table=T)

require(nlme)