

# Longitudinal Data Analysis

Case study of Trenal.XLS using Linear Mixed Effect Model

Group2:Hugo Blain; Oscar Cabanelas;Wanchang Zhang

2023-03-29

# Contents

<b>Data description</b>	<b>3</b>
Backgrounds of the data . . . . .	3
Data Organization . . . . .	3
Missing Data . . . . .	3
Exploratory Data Analysis for longitudinal data . . . . .	5
<b>Multilevel Data Analysis</b>	<b>5</b>
Two stage model . . . . .	5
Linear Mixed effects model . . . . .	5

## List of Figures

## List of Tables

1	Missing data for each measurement . . . . .	4
---	---	---

# Data description

## Backgrounds of the data

The dataset `Trenal.XLS` contains information on patients who received renal graft(kidney transplant). The patients have been followed for at most 10 years.

People with end-stage kidney disease who receive a kidney transplant generally live longer than people with ESRD who are on dialysis. However, kidney transplant recipients must remain on immunosuppressants (medications to suppress the immune system) for the rest of their life to prevent their body from rejecting the new kidney. The long-term immunosuppression puts them at risk for infections and cancer. Haematocrit level ## Initial Exploratory Analysis ### Import data `Trenal.XLS`

```
library(readxl)
library(knitr)
trenal <- read_excel("Trenal.XLS") # summary(trenal)

trenal = trenal[,-18] #remove a noninformative column const

# Continuous or discrete variables
trenal$id = as.factor(trenal$id)
trenal$j = as.factor(trenal$j)
#trenal$time = as.factor(trenal$time)
trenal$male = as.factor(trenal$male)
trenal$cardio = as.factor(trenal$cardio)
trenal$reject = as.factor(trenal$reject)

# Change the name of respons
colnames(trenal)[19] <- "Hc"
trenal.long = trenal[,13:20] # long table form

# Remove j
trenal.long = trenal.long[,-6]
trenal.long.unique <- trenal.long[match( unique(trenal.long$id), trenal.long$id),]
trenal.long.noNA <- na.omit(trenal.long)

# Wide table form
trenal.wide = as.data.frame(subset(trenal,trenal$j=="1"))[,1:18] # 1160 x 18
```

## Data Organization

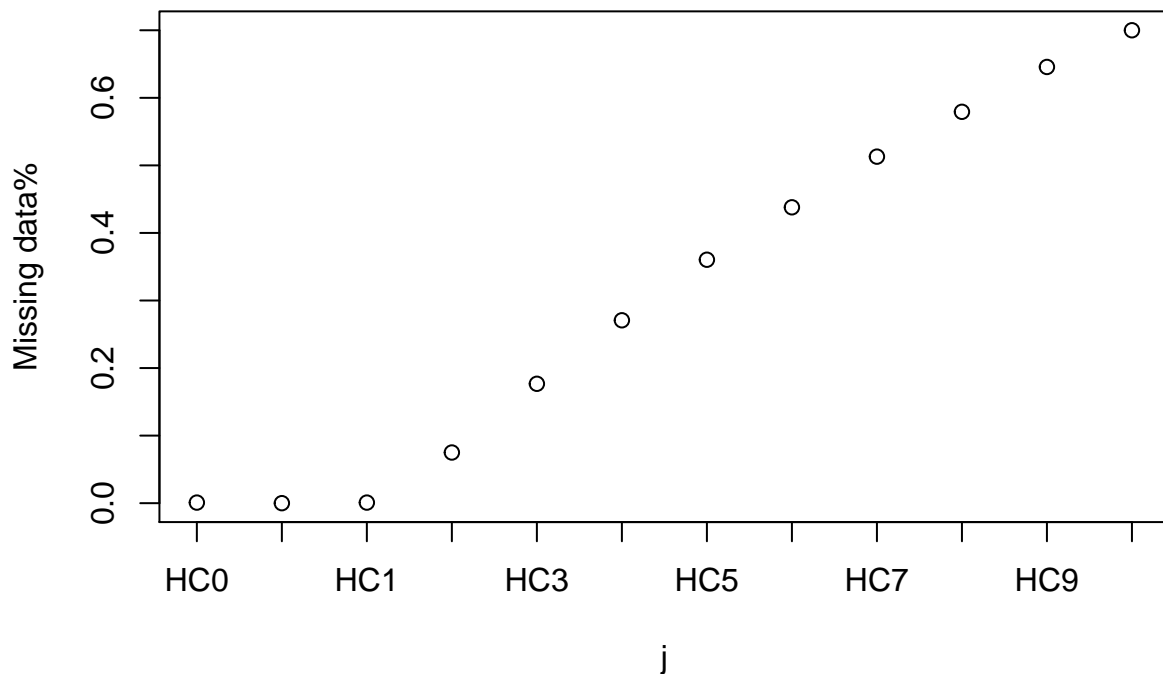
- The input data:
  - id: total 1160 persons
  - age to perform the operation: from 15 to 76 years old, average is 46.43 years old
  - male: we observe 494 females and 666 males
  - cardio: 953 persons has experienced a cardio-vascular problem during the years preceding the transplant, 207 did not.
  - reject: 793 patients shown symptoms of graft rejection during the first three months after the transportation, 367 has not.
- The response variable Hc level: continous from min 14% to max 65%.The Hc level is dependent on the meassured time, individual's age to perform the operation, gender, cardio history and reject history

## Missing Data

Table 1: Missing data for each measurement

	HC0	HC06	HC1	HC2	HC3	HC4	HC5	HC6	HC7	HC8	HC9	HC10
Hc.NA	1.000	0	1.000	87.000	205.000	314.000	418.00	508.000	595.000	672.000	749.000	812.0
Hc.NA.percentage	0.001	0	0.001	0.075	0.177	0.271	0.36	0.438	0.513	0.579	0.646	0.7

```
plot(Hc.NA.percentage,xaxt="n",xlab="j",ylab="Missing data%")
axis(side=1,at=c(1,2,3,4,5,6,7,8,9,10,11,12),labels=colnames(trenal.wide)[1:12])
```



*# Conclusion could be at the first three measurements, there are almost full data  
# More people tends to miss the measurements when time increases*

*##Second We can extract all NA data from the long table to analyse their construction*

```
trenal.long.NA = trenal.long[is.na(trenal.long$Hc),]
```

```
#t = unique(trenal.long.NA$id) # 821 individuals
```

```
trenal.long.NA.unique <- trenal.long.NA[match( unique(trenal.long.NA$id), trenal.long.NA$id),]
summary(trenal.long.NA.unique)
```

```
##      id      age      male      cardio      reject      Hc
##  1      :  1  Min.   :15.00  0:327    0:669    0:588  Min.   : NA
##  2      :  1  1st Qu.:39.00  1:494    1:152    1:233  1st Qu.: NA
##  4      :  1  Median :50.00                      Median : NA
##  7      :  1  Mean   :48.31                      Mean   :NaN
## 14      :  1  3rd Qu.:59.00                      3rd Qu.: NA
## 18      :  1  Max.   :76.00                      Max.   : NA
```

```
## (Other):815   NA's   :1
##      time
## Min.   : 0.000
## 1st Qu.: 3.000
## Median : 5.000
## Mean   : 5.638
## 3rd Qu.: 8.000
## Max.   :10.000
##
```

Exploratory Data Analysis for longitudinal data

Multilevel Data Analysis

Two stage model

Linear Mixed effects model