Continuous longitudinal data analysis (Part I) - Classical models

Juan R Gonzalez

Contents

1 Introduction

Objectives

- Learn how to perform logitudinal data analysis in the presence of continuous data using classical methods
- Peform data analyses where the scientific question is to determine factors associated with a continuous outcome that has been repeteadly measured over time.

2 Data

Data can be in width or long format

```
data.width <- read.table("../data/hypothetical_ancho.txt", header=TRUE)
head(data.width)
##   id time1 time2 time3 time4 group
## 1  1    31    29    15    26     A
## 2  2    24    28    20    32     A
## 3  3    14    20    28    30     A
## 4  4    38    34    30    34     B
## 5  5    25    29    25    29     B
## 6  6    30    28    16    34     B
```

```
data.long <- read.table("../data/hypothetical_largo.txt", header=TRUE)
head(data.long, n=12)
##    id time score group
## 1   1    1    31     A
## 2   1    2    29     A
## 3   1    3    15     A
## 4   1    4    26     A
## 5   2    1    24     A
## 6   2    2    28     A
## 7   2    3    20     A
## 8   2    4    32     A
## 9   3    1    14     A
## 10  3    2    20     A
## 11  3    3    28     A
## 12  3    4    30     A
```

Data are normally collected in width format since they can be easily collected in a database. You can use the function `make.rm` to transform data into long format as following

```
source("../R/make.rm.R")
data.long2 <-make.rm(constant=c("id", "group"),
                repeated=c("time1", "time2", "time3", "time4"),
                data=data.width)
head(data.long2, n=12)
##    id group repdat contrasts
## 1   1     A     31        T1
## 2   2     A     24        T1
## 3   3     A     14        T1
## 4   4     B     38        T1
## 5   5     B     25        T1
## 6   6     B     30        T1
## 7   1     A     29        T2
## 8   2     A     28        T2
## 9   3     A     20        T2
## 10  4     B     34        T2
## 11  5     B     29        T2
## 12  6     B     28        T2
```

Let's check everything has been properly created

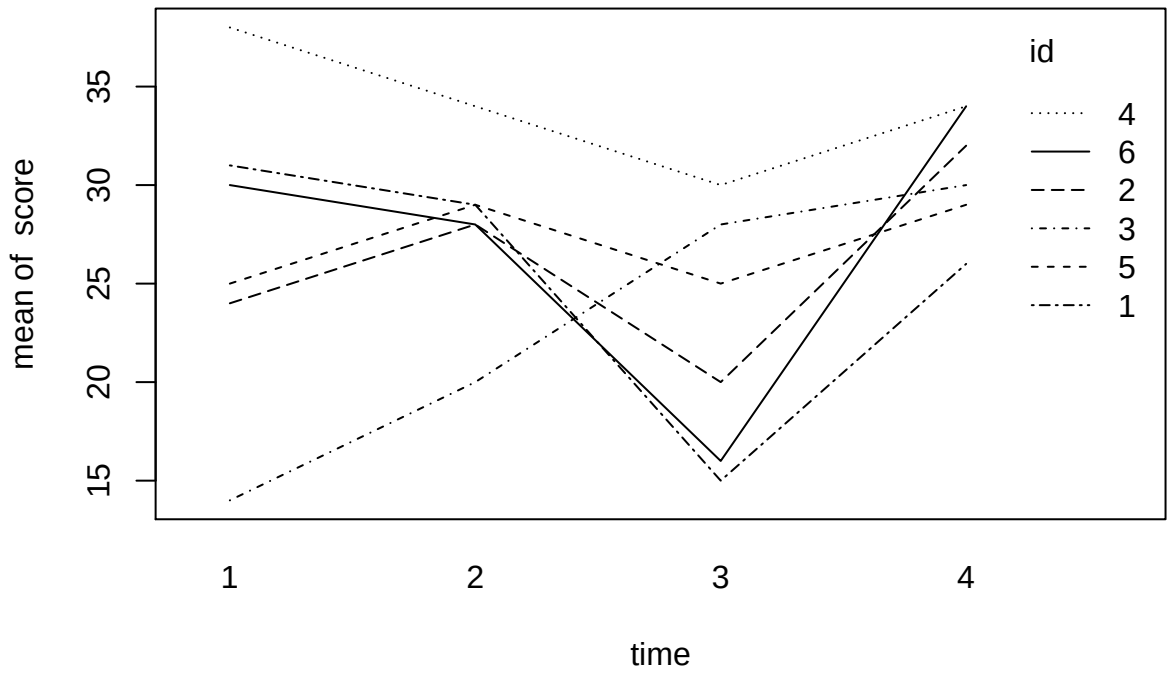
```
data.long[data.long$id==1,]
##   id time score group
## 1  1    1    31     A
## 2  1    2    29     A
## 3  1    3    15     A
## 4  1    4    26     A
data.long2[data.long2$id==1,]
##    id group repdat contrasts
## 1   1     A     31        T1
## 7   1     A     29        T2
## 13  1     A     15        T3
## 19  1     A     26        T4
```

Notice that the variable time has been replaced by the variable `contrasts`.

3 Data visualization

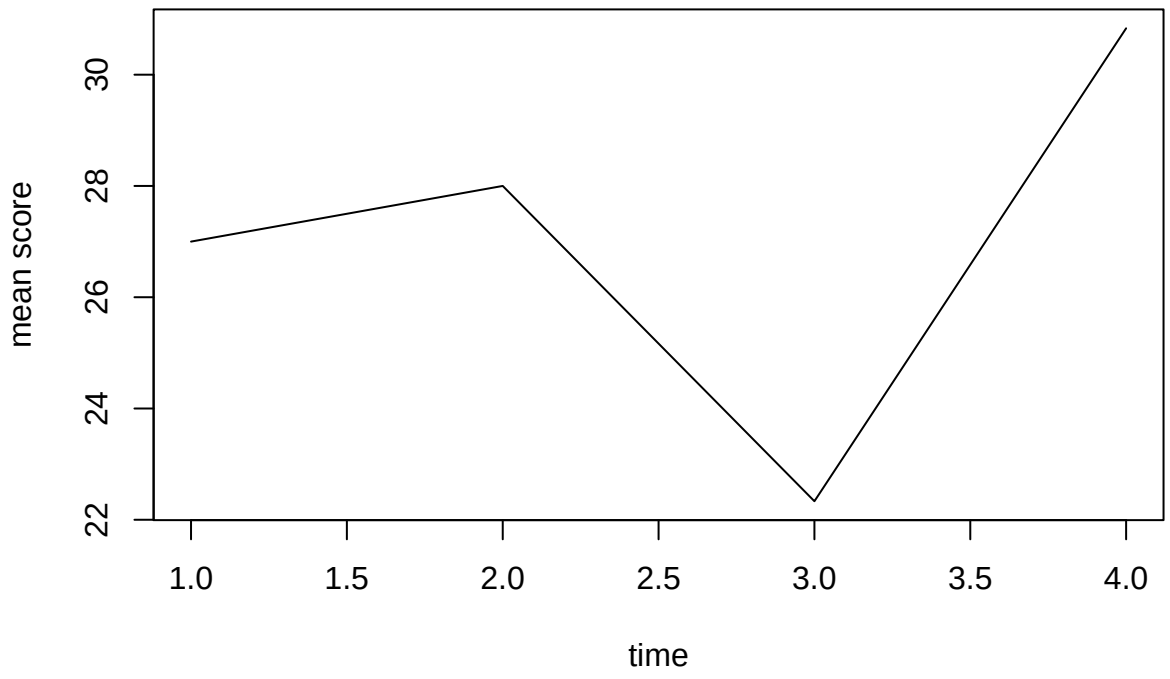
Individual profiles can be obtained by typing:

```
with(data.long, interaction.plot(time, id, score))
```



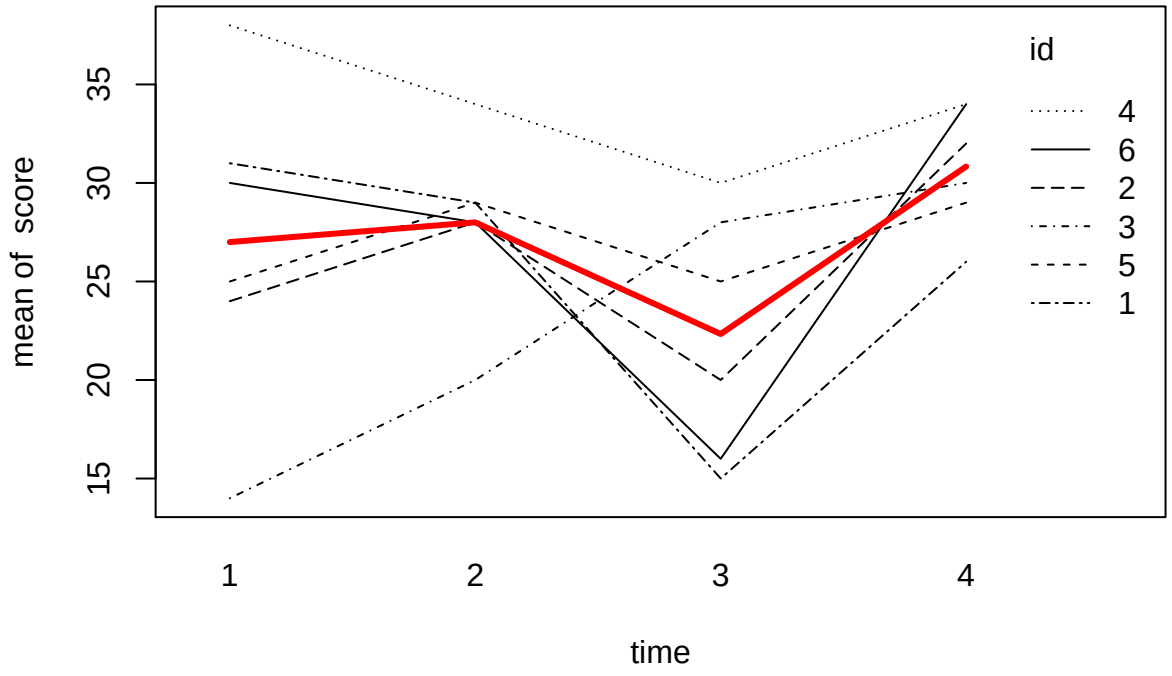
Overall profile is represented by

```
datos.agg <- aggregate(score ~ time, data = data.long, mean)
with(datos.agg, plot(time, score, ylab="mean score", type="l"))
```



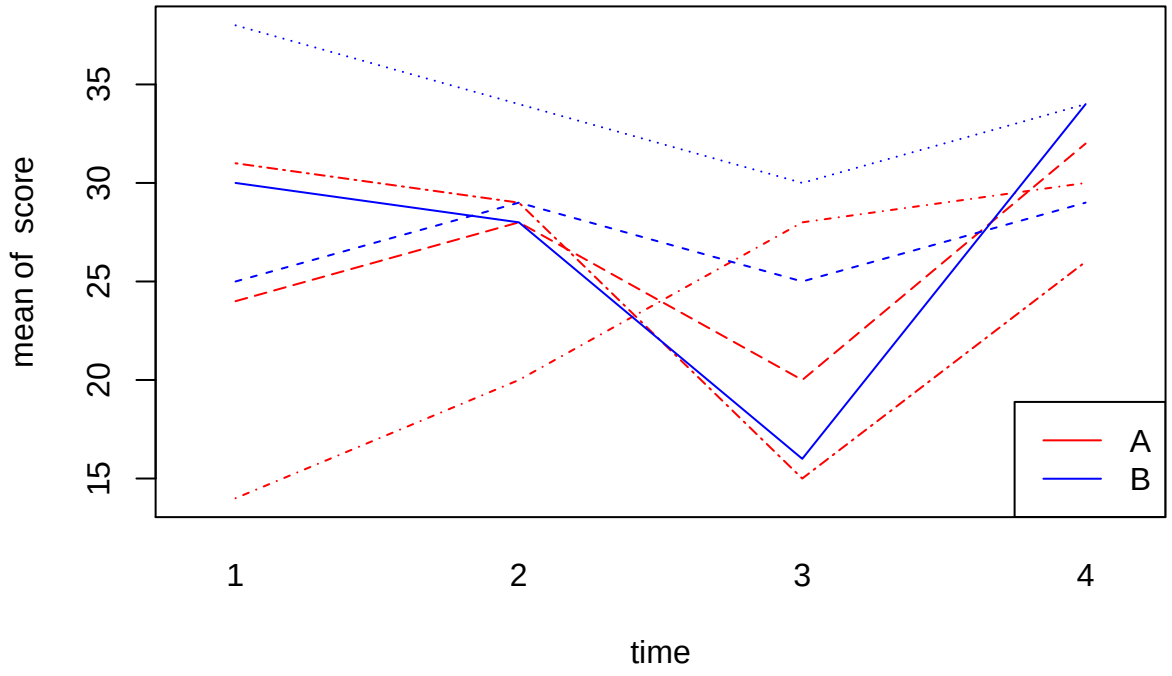
Both figures can be represented in the same plot by:

```
with(data.long, interaction.plot(time, id, score))
with(datos.agg, lines(time, score, ylab="mean score",
                      col="red", lwd=3))
```



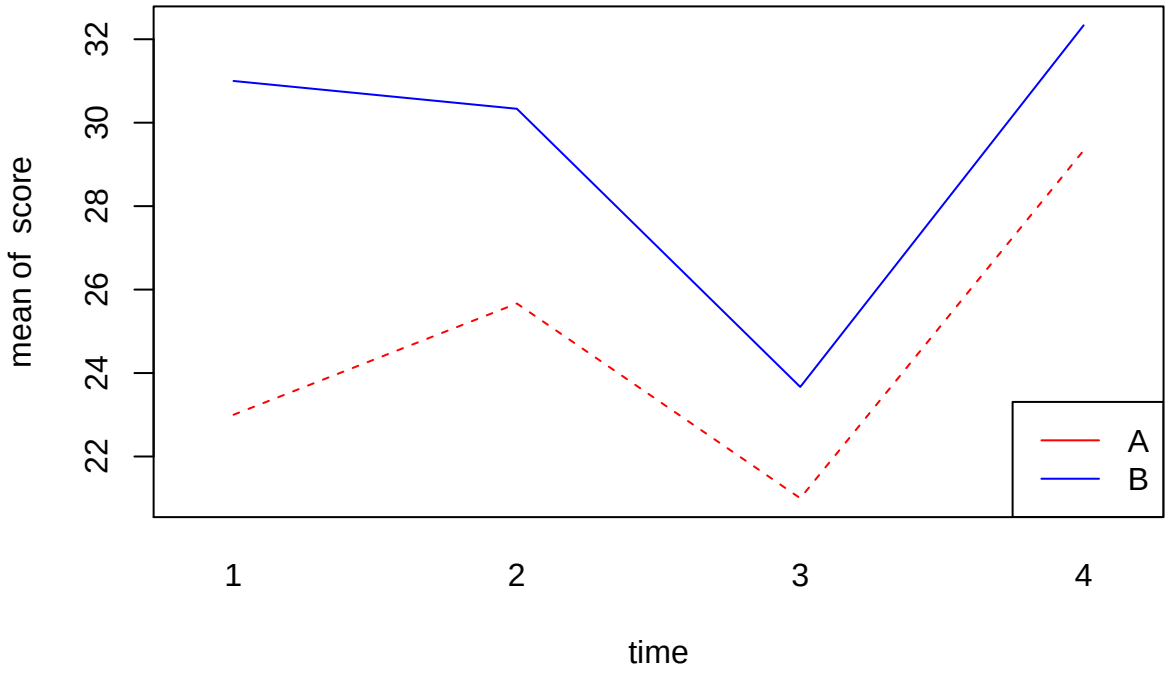
If groups are present, profiles can be drawn by each group

```
with(data.long, interaction.plot(time, id, score,
       col=c(rep("red",3), rep("blue",3)), legend=FALSE))
legend("bottomright", lty=1, c("A","B"), col=c("red", "blue"))
```



or the overall profile in each group can be visualize by

```
datos.agg <- aggregate(score ~ time + group, data = data.long, mean)
with(datos.agg, interaction.plot(time, group, score,
       col=c("red", "blue"), legend=FALSE))
legend("bottomright", lty=1, c("A","B"), col=c("red", "blue"))
```



4 Data Analysis

Each type of analysis requires to have data in width or long format

4.1 Strategy 1: End-point analysis

```
mod <- aov(time4 ~ time1 + group, data.width)
summary(mod)
##             Df Sum Sq Mean Sq F value Pr(>F)
## time1        1   3.95   3.951   0.335  0.603
## group        1   9.55   9.549   0.811  0.434
## Residuals    3  35.33  11.778
```

4.2 Strategy 2: ANOVA repeated measurements

The naive version is obtained by

```
mod <- aov(score ~ as.factor(time)*group, data.long)
summary(mod)
##                       Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(time)        3  224.8   74.93   2.291 0.1173
## group                  1  126.0  126.04   3.854 0.0673 .
```

```
## as.factor(time):group  3   26.8    8.93   0.273 0.8439
## Residuals             16  523.3   32.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The proper analysis has to be performed by

```
mod <- aov(score ~ as.factor(time)*group + Error(id), data.long)
summary(mod)
##
## Error: id
##       Df Sum Sq Mean Sq
## group  1  29.57   29.57
##
## Error: Within
##                        Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(time)         3  224.8   74.93   2.567 0.0932 .
## group                   1  182.0  182.03   6.237 0.0246 *
## as.factor(time):group   3   26.8    8.93   0.306 0.8207
## Residuals              15  437.8   29.18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4.3 Strategy 3: MANOVA

```
mod <- manova(cbind(time1, time2, time3, time4) ~ group, data.width)
summary(mod)
##           Df  Pillai approx F num Df den Df Pr(>F)
## group      1 0.49226  0.24238      4      1 0.8879
## Residuals  4
```

Other tests can be applied

```
summary(mod, test = "Wilks")
##           Df   Wilks approx F num Df den Df Pr(>F)
## group      1 0.50774  0.24238      4      1 0.8879
## Residuals  4
summary(mod, test = "Hotelling-Lawley")
##           Df Hotelling-Lawley approx F num Df den Df Pr(>F)
## group      1          0.96951  0.24238      4      1 0.8879
## Residuals  4
```

5 Recommended lectures

In the GitHub folder there is a paper further describing classical and modern methods to analyze continuous longitudinal data (file Modelos_clasicos_y_modernos_datos_longitunidales_continuos.pdf).

6 Exercise (to deliver)

Exercise 1:

Para estudiar las diferencias entre dos procedimientos diferentes de recuperación de pacientes de un infarto, se consideraron dos grupos experimentales en sendos hospitales, de 8 pacientes cada uno. La variable respuesta es el índice de Bartel, que varía entre 0 y 100, y que constituye una medida de la habilidad funcional con la que se valoran diferentes capacidades, de forma que valores más altos se corresponden con una mejor situación del paciente. De cada uno de los 16 pacientes se dispone de su respuesta cada semana a lo largo de 5 semanas consecutivas. Los datos se pueden encontrar en el archivo *recuperainfarto.txt*.

1. ¿Cuál de los dos procedimientos presenta una mejor recuperación de los pacientes? ¿Es esta diferencia estadísticamente significativa? [NOTA: contesta a estas preguntas planteando la pregunta estadística que creas conveniente]

Exercise 2:

En un estudio sobre la agudeza visual se dispone de la respuesta de siete individuos. La respuesta en cada ojo es el retraso en milisegundos entre la emisión de una luz y la respuesta en a la misma por el cortex. Cada ojo se somete a cuatro mediciones correspondientes a cuatro instantes consecutivos. Se tienen mediciones en el ojo izquierdo y derecho. Los datos se pueden encontrar en el archivo *agudezavisual.txt*

1. Crea una nueva base de datos agregando la información para cada una de la medida repetida (ojo) [NOTA: toma la media - usa la función `aggregate`]
2. ¿Existen diferencias entre la medición final y la basal?
3. ¿Existe un efecto temporal en la respuesta

---

7 References

-

8 Session information

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 14393)
##
## locale:
## [1] LC_COLLATE=Spanish_Spain.1252  LC_CTYPE=Spanish_Spain.1252
## [3] LC_MONETARY=Spanish_Spain.1252 LC_NUMERIC=C
## [5] LC_TIME=Spanish_Spain.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] knitr_1.15.1    BiocStyle_2.2.1
##
## loaded via a namespace (and not attached):
##  [1] backports_1.0.5 magrittr_1.5    rprojroot_1.2   tools_3.3.2
##  [5] htmltools_0.3.5 yaml_2.1.14     Rcpp_0.12.9     stringi_1.1.2
```

```
##  [9] rmarkdown_1.3   stringr_1.2.0   digest_0.6.11   evaluate_0.10
```