# A short course in Longitudinal Data Analysis
# ESRC Research Methods and Short Course

Lab 2 - June 2011

### Exploring Longitudinal Data

The exploration of longitudinal data can be separated into the exploration of the mean and correlation structure, respectively. We will cover tabular and graphical summaries of the data.

## 1 Mean response profiles

We start by examining the mean response profiles within each treatment group. When follow-up times are common to all subjects, it is easier to conver the data to the `balanced` format befor doing so, as follows.

```
> mental.bal<-to.balanced(mental,id.col=1,time.col=2,Y.col=3,other.col=4:7)
> dim(mental.bal)
[1] 150  11
```

Subsets of a longitudinal data-set can be extracted using the function `subset(object,subset)`. For example, to extract the subset of schizophrenia data for individuals under treatment 1,

```
> mental.bal1 <- subset(mental.bal, mental.bal$treat == 1)
> dim(mental.bal1)
[1] 50 11
> mental.bal1[1:6, ]
  id Y.t0 Y.t1 Y.t2 Y.t4 Y.t6 Y.t8 treat n.obs surv.time cens.ind
2   2   44   NA   NA   NA   NA   NA     1     1     0.740        0
8   8   61   81   NA   NA   NA   NA     1     2     1.035        1
9   9   59   59   NA   NA   NA   NA     1     2     1.039        0
10 10   59   72   NA   NA   NA   NA     1     2     1.116        1
11 11   58   55   NA   NA   NA   NA     1     2     1.193        1
12 12   46   46   NA   NA   NA   NA     1     2     1.238        0
```

**Exercise.** Create subsets of data for patients under treatments 2 and under treatment 3.

To calculate and plot the mean response profile for subjects receiving treatment 1, with pointwise approximate 95% confidence limits proceed as follows.

```
> mean.treat1 <- apply(mental.bal1[,2:7],2,mean,na.rm=T)
> sd.treat1<-sqrt(apply(mental.bal1[,2:7],2,var,na.rm=T))
> n.treat1<-apply(!is.na(mental.bal1[,2:7]),2,sum)
> se.treat1<-sd.treat1/sqrt(n.treat1)
> lower<-mean.treat1-2*se.treat1
> upper<-mean.treat1+2*se.treat1
> range(c(lower,upper))
[1] 46.0026 60.9556
> week<-c(0,1,2,4,6,8)
> plot(week,mean.treat1,type="l",lwd=3,ylim=c(40,65))
> lines(week,lower,type="l",lwd=2,lty=2)
> lines(week,upper,type="l",lwd=2,lty=2)
```

Notice that pointwise standard errors are valid because observations from different subjects are independent.

**Exercise.** Produce a suitable graphic that gives the same information for all three treatments.

**Exercise.** How many subjects drop-out of the study at each time point under each of the three treatments?

A plot displaying all of the individual longitudinal profiles can be very cluttered, whereas displaying only means and standard errors as each follow-up time hides the pattern of within-patient variation over time. A good compromise is to show means, plus and minus two standard errors, together with a sample of individual profiles plottes as less prominent lines, as in the following.

```
> plot(week,mean.treat1,type="l",lwd=2,ylim=c(20,90))
> lines(week,lower,type="l",lwd=2,lty=2)
> lines(week,upper,type="l",lwd=2,lty=2)
n1<-dim(mental.bal1)[1]
select<-sample(1:n1,10)
for (i in 1:10) {
   lines(week, mental.bal1[select[i],2:7],col="gray")
   }
```

**Exercise.** What does this plot tell you about the correlaiton structure of the `mental` data-set?

## 2 Exploring correlation structure

For balanced designs the correlation structure can be explored by constructing a matrix of correlations for all pairs of time points.

```
> round(var(mental.bal1[,2:7],na.rm=T,use="pairwise.complete.obs"),3)
        Y.t0    Y.t1    Y.t2    Y.t4    Y.t6    Y.t8
Y.t0 109.149  93.570  72.713  60.283  85.648  63.638
Y.t1  93.570 163.915 117.509  83.083  94.294  99.904
Y.t2  72.713 117.509 190.137  95.924  81.909  75.758
Y.t4  60.283  83.083  95.924 138.248  84.937 104.225
Y.t6  85.648  94.294  81.909  84.937 175.257 146.912
Y.t8  63.638  99.904  75.758 104.225 146.912 172.129
> round(cor(mental.bal1[,2:7],use="pairwise.complete.obs"),3)
> round(cor(mental.bal1[,2:7],use="pairwise.complete.obs"),3)
      Y.t0  Y.t1  Y.t2  Y.t4  Y.t6  Y.t8
Y.t0 1.000 0.701 0.489 0.478 0.586 0.433
Y.t1 0.701 1.000 0.766 0.634 0.596 0.596
Y.t2 0.489 0.766 1.000 0.759 0.534 0.438
Y.t4 0.478 0.634 0.759 1.000 0.546 0.650
Y.t6 0.586 0.596 0.534 0.546 1.000 0.901
Y.t8 0.433 0.596 0.438 0.650 0.901 1.000
```

Notice how the variances tend to increase slightly over time, whilst the correlations tend to become weaker with increasing timew-separation.

A suitable graphic display is a matrix of scatterplots, that is scatterplots of responses across all pairs of time points. Notice that the number of points in the scatterplot decreases with time, owing to dropout during the study.

```
> pairs(mental.bal1[,2:7], pch = 19, cex = 0.5)
```

For unbalanced data, the *variogram* is a more convenient way to explore correlation structure. Here, we apply the variogram to a balanced data-set to allow you to compare the result with earlier results for the same data.

```
mental1<-mental[mental$treat==1,]
vargm <- variogram(mental1$id,mental1$time,mental1$Y)
names(vargrm)
[1] "svar"   "sigma2"
vargrm$sigma2
[1] 154.2389
plot(vargrm,ylim=c(0,160))
```

To exit from `R`, type `q()`, but **remember to save your workspace** for the later practical classes.

PJD, 25.05.2011

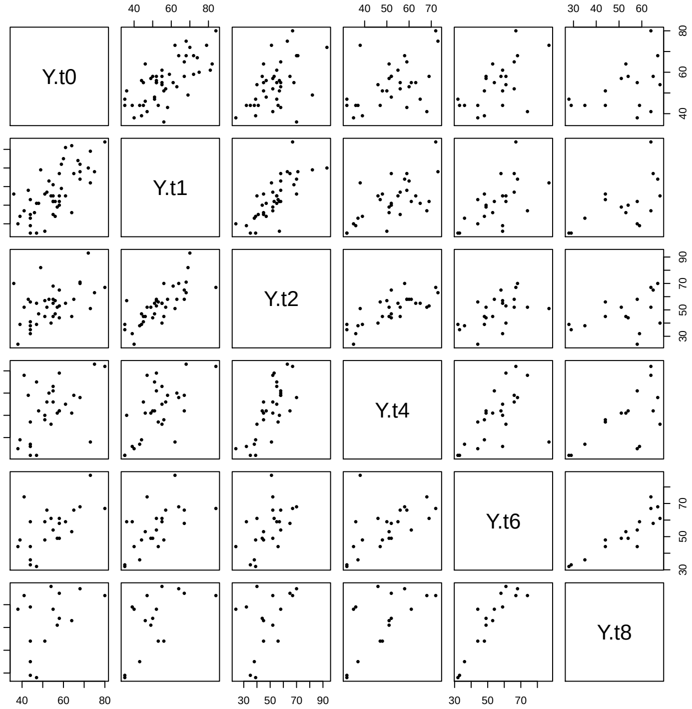

Figure 1: matrix of scatterplots

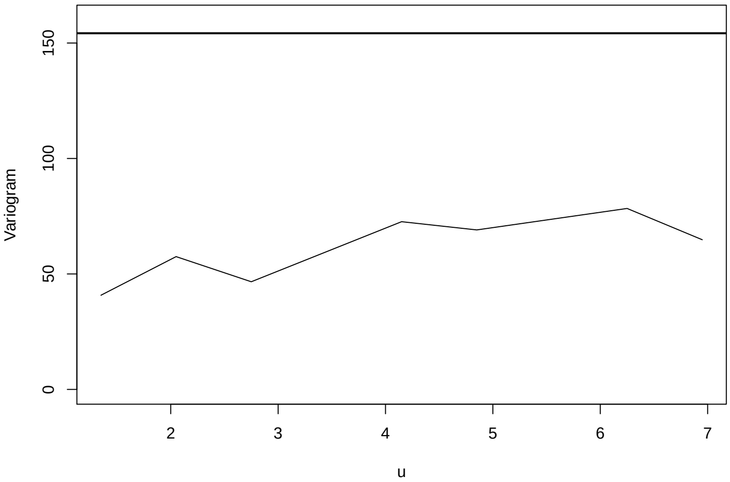

Figure 2: Sample variogram for subjects in the `mental` data-set allocated to treatment 1