

Introducción a los modelos mixtos (SIA 3011003)

Profesor Juan Carlos Salazar-Uribe
jcsalaza@unal.edu.co



ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de barras interactivo en 3D. (Ustedes deben ejecutar este código)

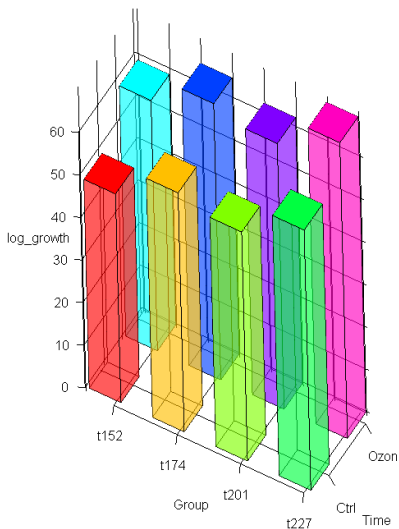
```
library(lattice)
library(ggplot2)
library(gridExtra)
library(Rcpp)
library(MASS)
library(barplot3d)
library(questionr)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
data$ID<-data$i..ID
data$Time<-factor(data$time1)

b<-wtd.table(x = data$Group, y=data$Time, weights = data$log_growth)
barplot3d(rows=2,cols=4,z=b,scalexy=10,alpha=0.4,theta=30,phi=50,
  topcolors=rainbow(8),sidecolors=rainbow(8),ylabels =c("Ctrl","Ozon"),
  xlabels = c("t152","t174","t201","t227"),
  xsub="Group",ysub="Time",zsub="log_growth",gap=0.5)
#rgl.close()
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de barras interactivo en 3D.



Boxplot un diagrama de caja o boxplot es un método estadístico visual para mostrar la localización y dispersión de uno o más conjuntos de datos numéricos. Se adapta relativamente fácil con datos categóricos.

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot. Por ejemplo, un boxplot global (ignorando la naturaleza longitudinal de los datos) de los logaritmos de crecimiento de acuerdo al grupo se puede obtener con:

```
#Global boxplot ignoring longitudinality
library(ggplot2)
data<-read.csv(file="SPRUCE1.csv",header=T,
               sep=',',dec='.')
#data$ID<-data$ï..ID
ggplot(data, aes(x=Group, y=log_growth,col=Group,group=Group)) +
  geom_boxplot(aes(x=Group, y=log_growth))
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot. Un boxplot global de los logaritmos de crecimiento de acuerdo al grupo:

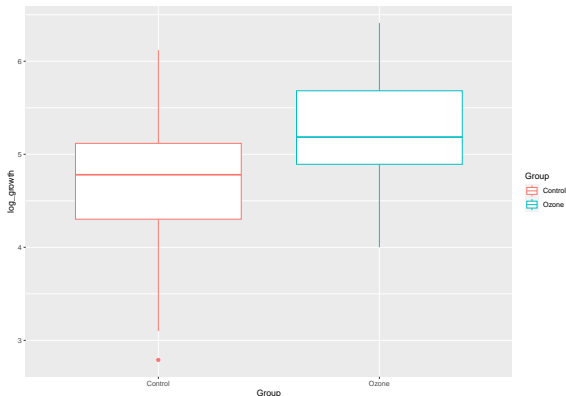


Figura 2: Gráfico de boxplot. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot. Una versión un poco más informativa se puede obtener con:

```
ggplot(data=data, aes(x=Group, y=log_growth,col=Group,group=Group,  
                      fill=Group)) +  
  geom_boxplot(position=position_dodge())+  
  scale_fill_manual(values=c("blue","purple"))+  
  labs(fill = "Group",y="Log_Growth",x="Group")+  
  stat_summary( aes(x=Group, y=log_growth,col=Group,group=Group,  
                   fill=Group),fun=mean, geom="point", shape=20,  
               size=5, color="cyan", fill="cyan")
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot. Una versión un poco más informativa:

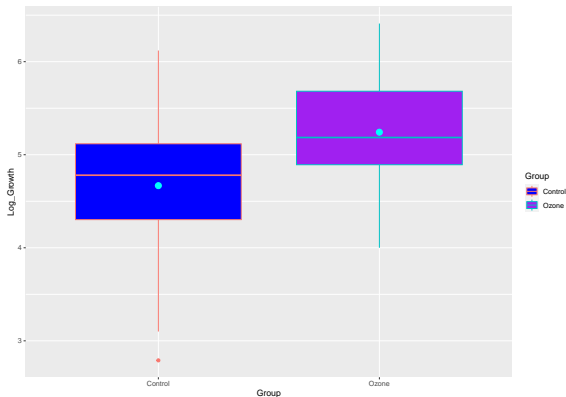


Figura 3: Gráfico de boxplot con promedios. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot. Respecto a la variable tiempo:

```
ggplot(data=data, aes(x=factor(time1),y=log_growth,  
                      group=factor(time1),fill=factor(time1))) +  
  geom_boxplot(position=position_dodge())+  
  scale_fill_manual(values=c("blue","purple","red","grey"))+  
  stat_summary( fun=mean, geom="point",  
               shape=20,  
               size=5, color="cyan", fill="cyan")+  
  labs(fill="Time",x="Time")
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot. Respecto a la variable tiempo:

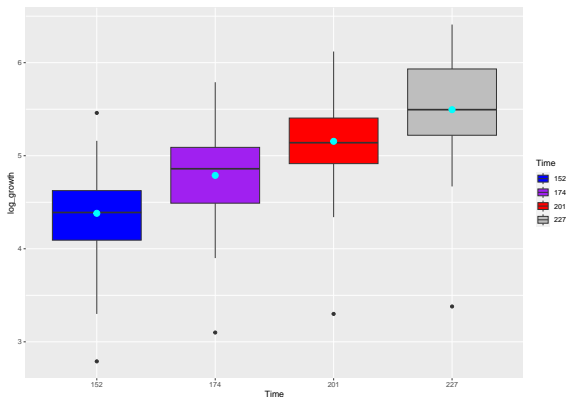


Figura 4: Gráfico de boxplot con promedios. Abetos

Gráfico de violín. Un diagrama de violín permite comparar la distribución de varios grupos mostrando sus densidades. Es similar a un boxplot en esencia.

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de violín. Ejemplo de un violin plot de acuerdo a la variable categórica "Group". Abetos:

```
library(questionr)
library(ggplot2)
library(gridExtra)
library(ggpubr)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$"..ID"
data$Time<-as.factor(data$time1)
ggplot(data, aes(x=Group, y=log_growth, fill=Group)) +
  # fill=name allow to automatically dedicate a color for each group
  geom_violin()+
  stat_summary(aes(x=Group, y=log_growth,
                    fill=Group),fun=mean, geom="point",
               shape=20,
               size=5, color="yellow", fill="cyan",position=position_dodge(width = .75))
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de violín. Ejemplo de un violin plot de acuerdo a la variable categórica "Group". Abetos:

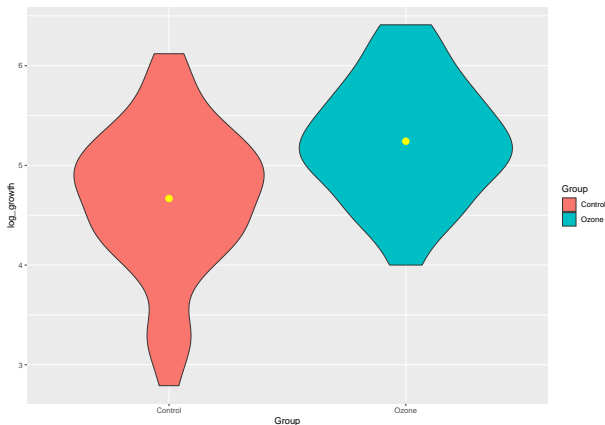


Figura 5: Gráfico de violín de ambos grupos. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de violín. Ejemplo de un violin plot de acuerdo a la variable categórica "Time". Abetos:

```
library(questionr)
library(ggplot2)
library(gridExtra)
library(ggpubr)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$"..ID"
data$Time<-as.factor(data$time1)
ggplot(data, aes(x=Time, y=log_growth, fill=Time)) +
  # fill=name allow to automatically dedicate a color for each group
  geom_violin()+
  stat_summary(aes(x=Time, y=log_growth,
                    fill=Group),fun=mean, geom="point",
               shape=20,
               size=4, color="yellow", fill="cyan",position=position_dodge(width = .75))
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de violín. Ejemplo de un violin plot de acuerdo a la variable categórica "Time". Abetos:

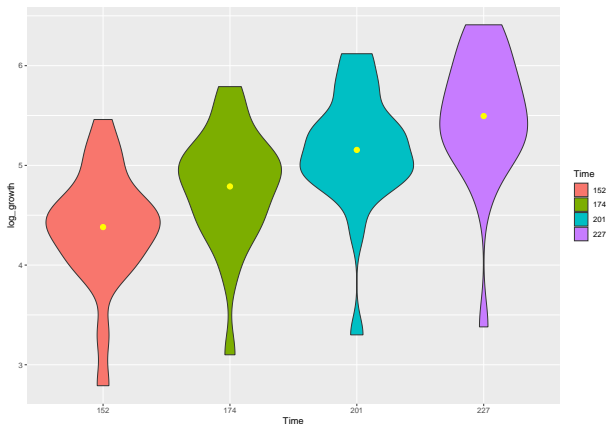


Figura 6: Gráfico de violín de acuerdo a Time. Abetos

Gráfico de rejilla (Trellis plot). Un gráfico de rejilla es una serie o cuadrícula de pequeños gráficos o tablas similares, lo que permite compararlos fácilmente. Se puede adaptar de acuerdo a los niveles de una variable categórica. En realidad, los trellis plots son gráficos que muestran una variable o la relación entre variables, condicionada a una o más variables.

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de rejilla (Trellis plot). Ejemplo de un trellis plot de acuerdo a la variable categórica “Group”. Abetos:

```
# lattice library: xyplot function
library(lattice)
library(ggplot2)
library(gridExtra)
data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$"..ID
Control<-data[data$Group=="Control",]
Ozone<-data[data$Group=="Ozone",]

C<-xyplot(log_growth ~ time1 | as.factor(ID), data=Control, as.table=T)
O<-xyplot(log_growth ~ time1 | as.factor(ID), data=Ozone, as.table=T)

grid.arrange(C,O)
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de rejilla (Trellis plot). Ejemplo de un trellis plot de acuerdo a la variable categórica “Group”. Abetos:

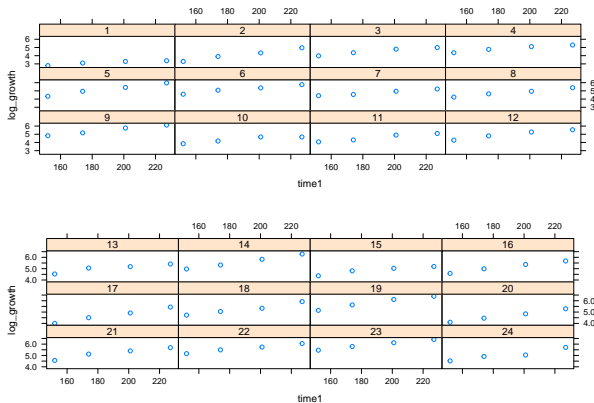


Figura 7: Gráfico de rejilla de ambos grupos. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de rejilla (Trellis plot). Ejemplo de un trellis plot de acuerdo a la variable categórica “Group”. Abetos:

```
C1<-suppressWarnings(xyplot(log_growth ~ time1 | as.factor(ID), data=Control,  
  prepanel = function(x, y) prepanel.loess(x, y, family="gaussian"),  
  xlab = "Time", ylab = "Log(Growth)",  
  panel = function(x, y) {  
    panel.xyplot(x, y)  
    panel.loess(x,y, family="gaussian")  
    panel.text(x=x,y=y,labels=y,pos=1,cex=0.5)},  
  ylim=c(0, 8), as.table=T))  
  
O1<-suppressWarnings(xyplot(log_growth ~ time1 | as.factor(ID), data=Ozone,  
  prepanel = function(x, y) prepanel.loess(x, y, family="gaussian"),  
  xlab = "Time", ylab = "Log(Growth)",  
  panel = function(x, y) {  
    panel.xyplot(x, y)  
    panel.loess(x,y, family="gaussian")  
    panel.text(x=x,y=y,labels=y,pos=1,cex=0.5)},  
  ylim=c(0, 8), as.table=T))  
  
grid.arrange(C1,O1)
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de rejilla (Trellis plot). Ejemplo de un trellis plot de acuerdo a la variable categórica "Group". Abetos:

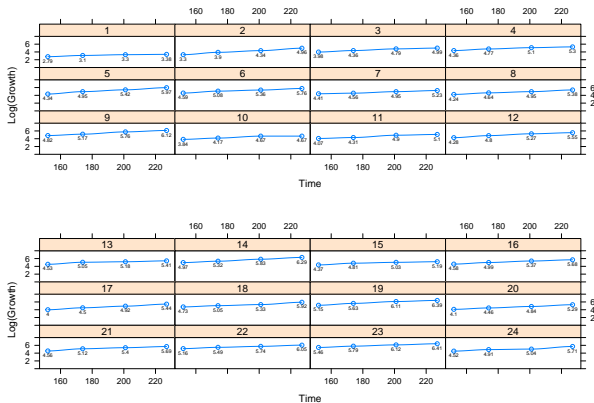


Figura 8: Gráfico de rejilla con líneas de ambos grupos. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de rejilla (Trellis plot). Ejemplo de un trellis plot con boxplots, de acuerdo a la variable categórica “Group”. Abetos:

```
library(lattice)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$i..ID

bwplot(log_growth ~ Group | as.factor(time1), data=data)
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de rejilla (Trellis plot). Ejemplo de un trellis plot con boxplots, de acuerdo a la variable categórica “Group”. Abetos:

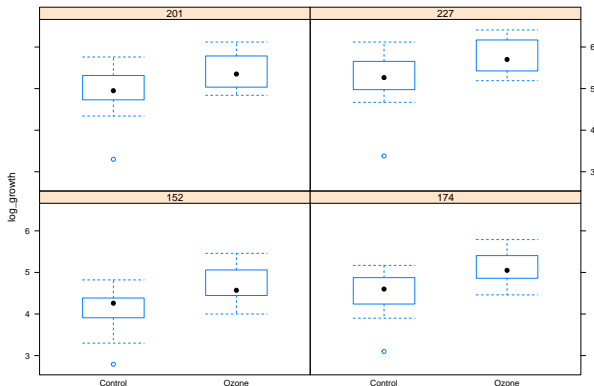


Figura 9: Trellis plot+Boxplots de acuerdo al grupo y tiempo. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión del gráfico anterior:

```
library(ggplot2)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$i..ID

ggplot(data, aes(x=factor(time1), y=log_growth, fill=Group)) +
  geom_boxplot(alpha=0.5)
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión del gráfico anterior:

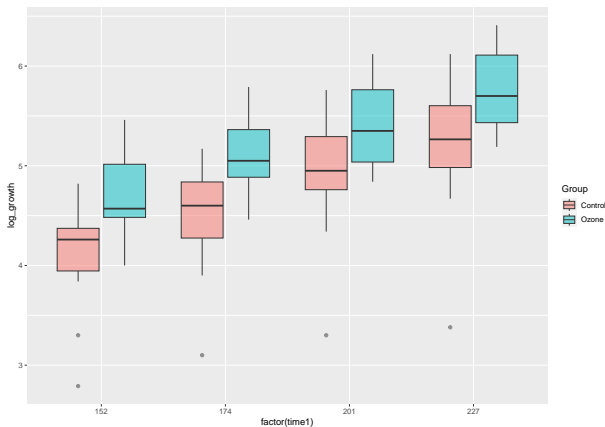


Figura 10: Boxplots de acuerdo al grupo y tiempo. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión del gráfico anterior con los promedios señalados:

```
library(ggplot2)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$i..ID

df<-data

ggplot(df)+
  geom_boxplot(aes(x=factor(time1),y= log_growth, fill=Group),color="purple")+
  stat_summary(aes(x=factor(time1), y=log_growth,
                    fill=Group,group=Group),fun=mean, geom="point",
               shape=20,
               size=4, color="yellow", fill="cyan",position=position_dodge(width = .75))+
  labs(fill = "Group",y="Log-Growth",x="Time")+
  labs(title = "Boxplot of log-growth by Group",
       subtitle = "",
       caption = "Fuente: Elaboración propia.")
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión del gráfico anterior con los promedios señalados:



Figura 11: Boxplots de acuerdo al tiempo y grupo. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión del gráfico anterior de acuerdo al Group:

```
library(ggplot2)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$i..ID

ggplot(data, aes(x=Group, y=log_growth, fill=factor(time1))) +
  geom_boxplot(alpha=0.5)
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión del gráfico anterior de acuerdo al Group:

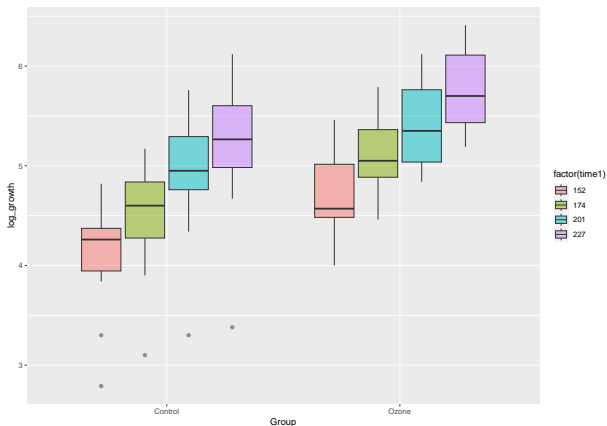


Figura 12: Boxplots de acuerdo al tiempo y grupo. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión del gráfico anterior con los promedios señalados:

```
library(ggplot2)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$i..ID

df<-data

ggplot(df)+
  geom_boxplot(aes(x=Group,y= log_growth, fill=factor(time1)),color="purple")+
  stat_summary(aes(x=Group, y=log_growth, fill=factor(time1),group=factor(time1)),fun=mean,
               geom="point", shape=20,
               size=4, color="yellow", fill="cyan",position=position_dodge(width = .75))+
  labs(fill = "Time",y="Log-Growth",x="Group")+
  labs(title = "Boxplot of log-growth by Time",
       subtitle = "",
       caption = "Fuente: Elaboración propia.")
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión del gráfico anterior con los promedios señalados:

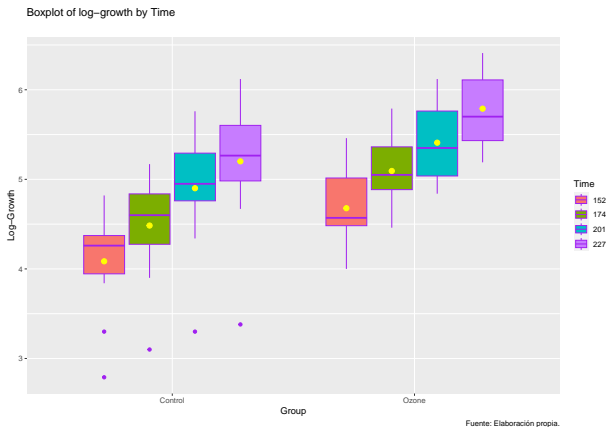


Figura 13: Boxplots de acuerdo al tiempo y grupo. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión de boxplot:

```
library(ggplot2)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$`.ID

Control<-data[data$Group=="Control",]
Ozone<-data[data$Group=="Ozone",]

ggplot(data=Control, aes(x=time1, y=log_growth)) +
  geom_boxplot(aes(x=time1, y=log_growth, group=factor(time1), fill="Control"),
               col="red", alpha=0.4, width=3.8)+
  geom_boxplot(data=Ozone, aes(x=time1, y=log_growth, group=factor(time1), fill="Ozone"),
               col="blue", alpha=0.3, width=3.8)+
  stat_summary(data=Control, fun=mean, geom="point", size=3, col="red", shape=19)+
  stat_summary(data=Ozone, fun=mean, geom="point", size=3, col="blue", shape=19)+
  labs(fill = "Group", y="Log_Growth", x="Time")
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Boxplot Otra versión de boxplot:

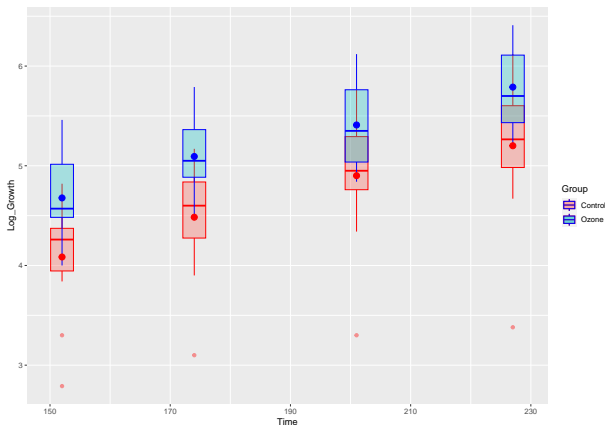


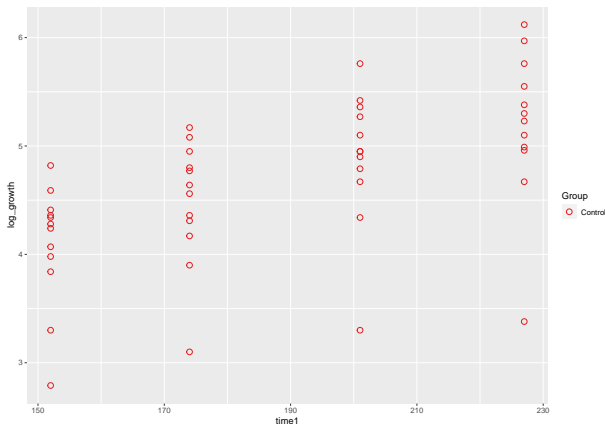
Figura 14: Boxplots de acuerdo al tiempo y grupo. Abetos

Gráfico de dispersión, nube de puntos o scatter plot. Un diagrama de dispersión (también conocido como gráfico de dispersión) usa puntos para representar valores para **dos variables numéricas** diferentes. La posición de cada punto en el eje horizontal y vertical indica valores para un punto de datos individual.

Gráfico de dispersión, nube de puntos o scatter plot. Es en realidad, un tipo de gráfico estadístico que utiliza coordenadas cartesianas para mostrar valores de, típicamente, dos variables numéricas de un conjunto de datos. Si los puntos están codificados, se puede mostrar una variable adicional. En este gráfico el tiempo se ubica en el eje horizontal y los valores de la variable que se estudia en el eje vertical y los valores asociados a cada sujeto se representan como puntos en el plano.

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de dispersión, nube de puntos o scatter plot Ejemplo de un scatter plot para el grupo de control de los árboles de abeto:



ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de dispersión, nube de puntos o scatter plot Ejemplo de un scatter plot para un sujeto del grupo de control de los árboles de abeto:

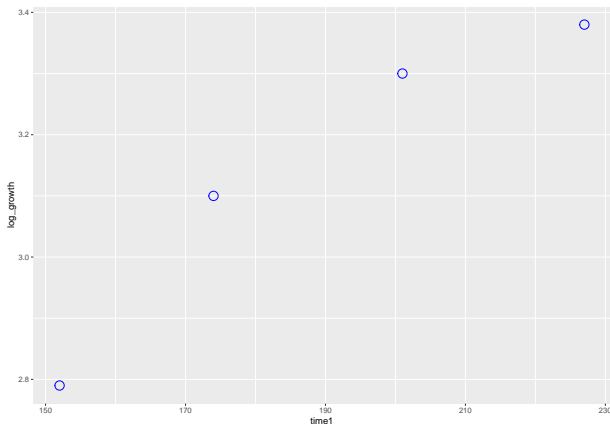


Figura 16: Scatter plot para el sujeto 1 del grupo de Control. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Matriz de gráficos de nube de puntos Si se tiene al menos dos variables continuas, en R, se puede visualizar rápidamente, la forma de la asociación, si esta existe o si es muy evidente al graficarla.

```
library(readxl)
library(ggplot2)
library(GGally)

data_wide<-read.csv(file="Sitka_Spruce_ancho.csv",header=T,sep=',',dec='.')
ggpairs(data_wide,columns=c(3,4,5,6),
        lower=list(continuous=wrap("smooth", colour="blue")),
        diag=list(continuous=wrap("densityDiag",colour="tomato")))
```

Matriz de gráficos de nube de puntos

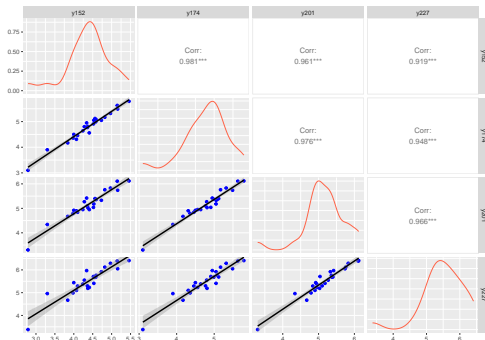


Figura 17: Scatter plot matrix. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Matriz de gráficos de nube de puntos respecto a una categórica

```
library(readxl)
library(ggplot2)
library(GGally)

data_wide<-read.csv(file="Sitka_Spruce_ancho.csv",header=T,
                    sep=',',dec='.')
ggpairs(data_wide,columns=c(3,4,5,6),
mapping=ggplot2::aes(colour = Group,alpha=0.2))
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Matriz de gráficos de nube de puntos respecto a una categórica

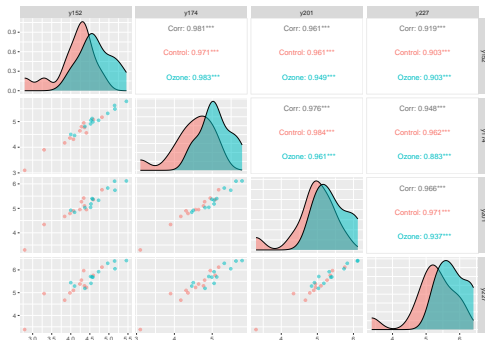


Figura 18: Scatter plot matrix. Abetos

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Gráfico de trayectoria o de perfil. Este gráfico es simplemente un gráfico de dispersión con el tiempo en el eje horizontal y los valores de la variable que se estudia en el eje vertical y los puntos asociados a cada sujeto se unen con una línea recta.

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Ejemplo de un gráfico de perfiles individuales

```
#Ejemplos clase 1 INTROMLM
library(lattice)
library(ggplot2)
library(gridExtra)
library(Rcpp)
library(MASS)
library(lattice)
library(nlme)
library(lme4)
library(lmerTest)
library(MCMCglmm)
library(MCMCpack)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$i..ID
data_control<-data[data$Group=="Control",]

#Gráfico de perfiles

ggplot(data_control, aes(x = time1, y = log_growth,group=ID)) +
  geom_line(aes(color = Group, group = ID),size=1.4,alpha=0.6)+
  scale_colour_brewer(palette="Set1")+
  geom_point(size=2.8,aes(color = Group, group = ID, shape=Group))+
  scale_shape_discrete(solid=FALSE)
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Ejemplo de un gráfico de perfiles individuales

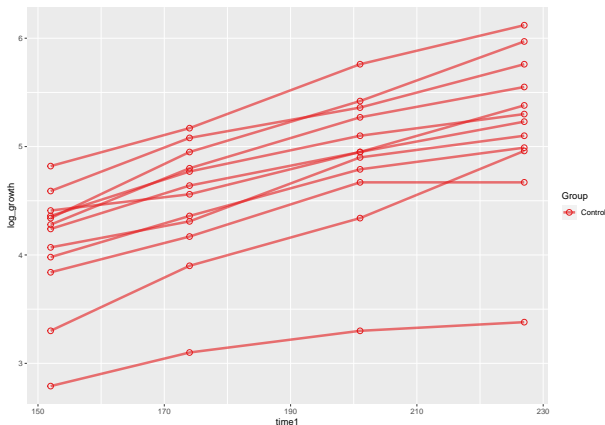


Figura 19: Perfiles individuales grupo Control. Abetos

Ejemplo de un gráfico de perfiles de acuerdo a la variable categórica "Group":

```
#Ejemplos clase 1 INTROMLM
library(lattice)
library(ggplot2)
library(gridExtra)
library(Rcpp)
library(MASS)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$í..ID

#Gráfico de perfiles

ggplot(data, aes(x = time1, y = log_growth,group=ID)) +
  geom_line(aes(color = Group, group = ID),size=1.4,alpha=0.6)+
  scale_colour_brewer(palette="Set1")+
  geom_point(size=2.8,aes(color = Group, group = ID, shape=Group))+
  scale_shape_discrete(solid=FALSE)
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Ejemplo de un gráfico de perfiles de acuerdo a la variable categórica “Group”:

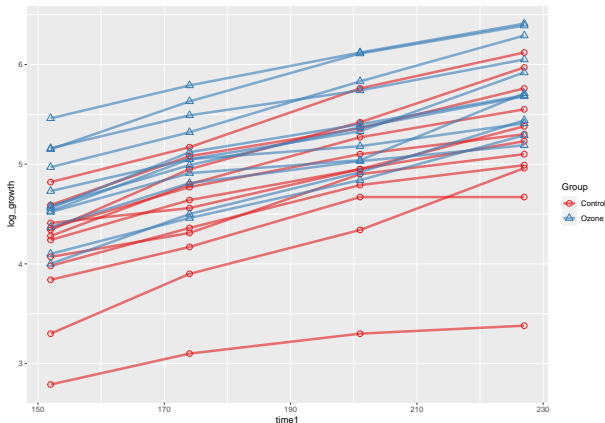


Figura 20: Perfiles individuales de ambos grupos. Abetos

Gráfico de perfil promedio. Un gráfico de perfil promedio es un gráfico de líneas en la que cada punto indica la media marginal estimada de una variable dependiente (ajustada por cualquier covariable) en un nivel de un factor. Los niveles de un segundo factor se pueden usar para hacer líneas separadas. Para los análisis multivariados, se recomienda crear gráficos de perfil para cada variable dependiente.

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Ejemplo de un gráfico de perfiles promedio de acuerdo a la variable categórica “Group”:

```
#Ejemplos clase 3 INTROMLM
library(lattice)
library(ggplot2)
library(gridExtra)
library(Rcpp)
library(MASS)

data<-read.csv(file="SPRUCE1.csv",header=T,sep=',',dec='.')
#data$ID<-data$i..ID

#Gráfico de perfiles promedio

ggplot(data, aes(x = time1, y = log_growth,group=Group)) +
  stat_summary(aes(group=Group,col=Group),
               geom = "line", fun = mean, size = 1.2)+
  xlab("Tiempo")
```

ANÁLISIS EXPLORATORIO DE DATOS LONGITUDINALES

Ejemplo de un gráfico de perfiles promedio de acuerdo a la variable categórica “Group”:

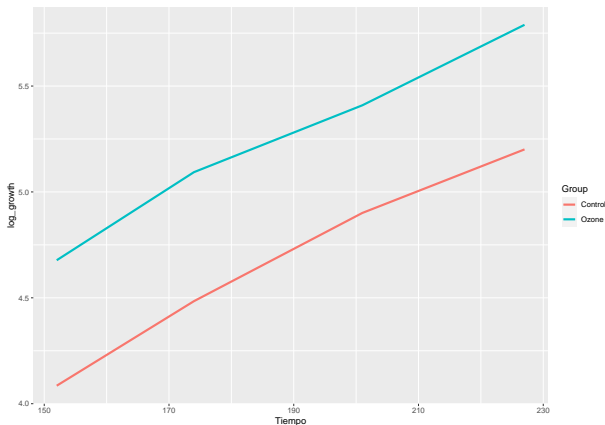


Figura 21: Perfiles promedio de ambos grupos. Abetos