Analisis\_generico

# Definir directorio de trabajo

Aqui probablemente el codigo te dara problemas, porque queremos definir como directorio de trabajo la carpeta de COPLAS porque los archivos estan ahi. Como este Markdown esta dentro de un Proyecto con su carpeta asociada, R entiende que el directorio es ese (".../Instar/implementaciones/instar\_baza/scripts/git\_Instar\_baza"). Estrictamente hablando deberiamos escribir el codigo de manera que cuando llamemos a un archivo le demos el path apropiado ("...\_validacion\_datos\_validacion\_PLAGAS"). Pero tambien se puede definir el directorio de manera manual (Session > Set Working Directory > Choose Directory), y asi R si acepta que sea una carpeta distinta a donde esta el proyecto. Por mi las dos opciones valen, la que te sea mas facil :)

He definido el WD de manera manual. Definir como directorio la carpeta COPLAS\_PLAGAS # Cargar paquetes

#install.packages("lubridate")  
library(lubridate) # Para las fechas

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

# DATOS INSTAR

## Importar datos INSTAR

* La tabla de datos se construye en excel y se guarda como .csv
* Se abre con procesador de texto, y se comprueba que la fecha está en formato mm-dd-yy
* Se guarda como .txt Las columnas que constituyen la tabla son: fecha, huevo, L1, L2, crisalida, radiacion, tmax, tmin, tmed y exergia.

TOMAMOS EL RODAL GR-023-008 (Sierra de Baza) COMO EJEMPLO

INSTAR<-read.table("instar\_bz008.txt", header=T, sep=",", dec=".")  
INSTAR$fecha <-as.Date(INSTAR$fecha, format="%m-%d-%y")#aunque pongo formato mm/dd/yyyy se pone como yyyy-mm-dd

## Anadir biociclos al data.frame

INSTAR$yday <- yday(INSTAR$fecha) # anadimos el dia del anio con el paquete lubridate  
INSTAR$year <- year(INSTAR$fecha) # anadimos el anio con el paquete lubridate  
INSTAR$biociclo <- NULL  
#write.csv(INSTAR, file = "datos\_instar\_bz008.csv", row.names=FALSE, na="")

Aparece dos chunks para este proceso, uno para Baza y otro para Sierra Nevada, ya que tienen fechas de inicio diferentes.

#Baza  
subset\_1993 <- subset(INSTAR, INSTAR$year==1993)  
for(i in 1:length(subset\_1993$yday)){  
 ifelse(subset\_1993$yday[i]<=79,   
 subset\_1993$biociclo[i] <- 1, subset\_1993$biociclo[i] <- 2)  
}  
  
subset\_1994 <- subset(INSTAR, INSTAR$year==1994)  
for(i in 1:length(subset\_1994$yday)){  
 ifelse(subset\_1994$yday[i]<=79,   
 subset\_1994$biociclo[i] <- 2, subset\_1994$biociclo[i] <- 3)  
}  
  
subset\_1995 <- subset(INSTAR, INSTAR$year==1995)  
for(i in 1:length(subset\_1995$yday)){  
 ifelse(subset\_1995$yday[i]<=79,   
 subset\_1995$biociclo[i] <- 3, subset\_1995$biociclo[i] <- 4)  
}  
  
subset\_1996 <- subset(INSTAR, INSTAR$year==1996)  
for(i in 1:length(subset\_1996$yday)){  
 ifelse(subset\_1996$yday[i]<=79,   
 subset\_1996$biociclo[i] <- 4, subset\_1996$biociclo[i] <- 5)  
}  
  
subset\_1997 <- subset(INSTAR, INSTAR$year==1997)  
for(i in 1:length(subset\_1997$yday)){  
 ifelse(subset\_1997$yday[i]<=79,   
 subset\_1997$biociclo[i] <- 5, subset\_1997$biociclo[i] <- 6)  
}  
  
subset\_1998 <- subset(INSTAR, INSTAR$year==1998)  
for(i in 1:length(subset\_1998$yday)){  
 ifelse(subset\_1998$yday[i]<=79,   
 subset\_1998$biociclo[i] <- 6, subset\_1998$biociclo[i] <- 7)  
}  
  
subset\_1999 <- subset(INSTAR, INSTAR$year==1999)  
for(i in 1:length(subset\_1999$yday)){  
 ifelse(subset\_1999$yday[i]<=79,   
 subset\_1999$biociclo[i] <- 7, subset\_1999$biociclo[i] <- 8)  
}  
  
subset\_2000 <- subset(INSTAR, INSTAR$year==2000)  
for(i in 1:length(subset\_2000$yday)){  
 ifelse(subset\_2000$yday[i]<=79,   
 subset\_2000$biociclo[i] <- 8, subset\_2000$biociclo[i] <- 9)  
}  
  
subset\_2001 <- subset(INSTAR, INSTAR$year==2001)  
for(i in 1:length(subset\_2001$yday)){  
 ifelse(subset\_2001$yday[i]<=79,   
 subset\_2001$biociclo[i] <- 9, subset\_2001$biociclo[i] <- 10)  
}  
  
subset\_2002 <- subset(INSTAR, INSTAR$year==2002)  
for(i in 1:length(subset\_2002$yday)){  
 ifelse(subset\_2002$yday[i]<=79,   
 subset\_2002$biociclo[i] <- 10, subset\_2002$biociclo[i] <- 11)  
}  
  
subset\_2003 <- subset(INSTAR, INSTAR$year==2003)  
for(i in 1:length(subset\_2003$yday)){  
 ifelse(subset\_2003$yday[i]<=79,   
 subset\_2003$biociclo[i] <- 11, subset\_2003$biociclo[i] <- 12)  
}  
  
subset\_2004 <- subset(INSTAR, INSTAR$year==2004)  
for(i in 1:length(subset\_2004$yday)){  
 ifelse(subset\_2004$yday[i]<=79,   
 subset\_2004$biociclo[i] <- 12, subset\_2004$biociclo[i] <- 13)  
}  
subset\_2005 <- subset(INSTAR, INSTAR$year==2005)  
for(i in 1:length(subset\_2005$yday)){  
 ifelse(subset\_2005$yday[i]<=79,   
 subset\_2005$biociclo[i] <- 13, subset\_2005$biociclo[i] <- 14)  
}  
  
subset\_2006 <- subset(INSTAR, INSTAR$year==2006)  
for(i in 1:length(subset\_2006$yday)){  
 ifelse(subset\_2006$yday[i]<=79,   
 subset\_2006$biociclo[i] <- 14, subset\_2006$biociclo[i] <- 15)  
}  
  
subset\_2007 <- subset(INSTAR, INSTAR$year==2007)  
for(i in 1:length(subset\_2007$yday)){  
 ifelse(subset\_2007$yday[i]<=79,   
 subset\_2007$biociclo[i] <- 15, subset\_2007$biociclo[i] <- 16)  
}  
  
subset\_2008 <- subset(INSTAR, INSTAR$year==2008)  
for(i in 1:length(subset\_2008$yday)){  
 ifelse(subset\_2008$yday[i]<=79,   
 subset\_2008$biociclo[i] <- 16, subset\_2008$biociclo[i] <- 17)  
}  
  
subset\_2009 <- subset(INSTAR, INSTAR$year==2009)  
for(i in 1:length(subset\_2009$yday)){  
 ifelse(subset\_2009$yday[i]<=79,   
 subset\_2009$biociclo[i] <- 17, subset\_2009$biociclo[i] <- 18)  
}  
  
subset\_2010 <- subset(INSTAR, INSTAR$year==2010)  
for(i in 1:length(subset\_2010$yday)){  
 ifelse(subset\_2010$yday[i]<=79,   
 subset\_2010$biociclo[i] <- 18, subset\_2010$biociclo[i] <- 19)  
}  
  
subset\_2011 <- subset(INSTAR, INSTAR$year==2011)  
for(i in 1:length(subset\_2011$yday)){  
 ifelse(subset\_2011$yday[i]<=79,   
 subset\_2011$biociclo[i] <- 19, subset\_2011$biociclo[i] <- 20)  
}  
  
subset\_2012 <- subset(INSTAR, INSTAR$year==2012)  
for(i in 1:length(subset\_2012$yday)){  
 ifelse(subset\_2012$yday[i]<=79,   
 subset\_2012$biociclo[i] <- 20, subset\_2012$biociclo[i] <- 21)  
}  
  
subset\_2013 <- subset(INSTAR, INSTAR$year==2013)  
for(i in 1:length(subset\_2013$yday)){  
 ifelse(subset\_2013$yday[i]<=79,   
 subset\_2013$biociclo[i] <- 21, subset\_2013$biociclo[i] <- 22)  
}  
  
subset\_2014 <- subset(INSTAR, INSTAR$year==2014)  
for(i in 1:length(subset\_2014$yday)){  
 ifelse(subset\_2014$yday[i]<=79,   
 subset\_2014$biociclo[i] <- 22, subset\_2014$biociclo[i] <- 23)  
}  
  
INSTAR<- rbind(subset\_1994, subset\_1995, subset\_1996, subset\_1997, subset\_1998, subset\_1999, subset\_2000, subset\_2001, subset\_2002, subset\_2003, subset\_2004, subset\_2005, subset\_2006, subset\_2007, subset\_2008, subset\_2009, subset\_2010, subset\_2011, subset\_2012, subset\_2013, subset\_2014) #Se quitan el primer y ultimo biociclo porque no estan completos  
  
#QUITAR BIOCICLO 2 y 23(incompletos)  
INSTAR<-INSTAR[!(INSTAR$biociclo=="2"),]  
INSTAR<-INSTAR[!(INSTAR$biociclo=="23"),]

#Sierra Nevada  
subset\_2001 <- subset(INSTAR, INSTAR$year==2001)  
for(i in 1:length(subset\_2001$yday)){  
 ifelse(subset\_2001$yday[i]<=79,   
 subset\_2001$biociclo[i] <- 1, subset\_2001$biociclo[i] <- 2)  
}  
  
subset\_2002 <- subset(INSTAR, INSTAR$year==2002)  
for(i in 1:length(subset\_2002$yday)){  
 ifelse(subset\_2002$yday[i]<=79,   
 subset\_2002$biociclo[i] <- 2, subset\_2002$biociclo[i] <- 3)  
}  
  
subset\_2003 <- subset(INSTAR, INSTAR$year==2003)  
for(i in 1:length(subset\_2003$yday)){  
 ifelse(subset\_2003$yday[i]<=79,   
 subset\_2003$biociclo[i] <- 3, subset\_2003$biociclo[i] <- 4)  
}  
  
subset\_2004 <- subset(INSTAR, INSTAR$year==2004)  
for(i in 1:length(subset\_2004$yday)){  
 ifelse(subset\_2004$yday[i]<=79,   
 subset\_2004$biociclo[i] <- 4, subset\_2004$biociclo[i] <- 5)  
}  
subset\_2005 <- subset(INSTAR, INSTAR$year==2005)  
for(i in 1:length(subset\_2005$yday)){  
 ifelse(subset\_2005$yday[i]<=79,   
 subset\_2005$biociclo[i] <- 5, subset\_2005$biociclo[i] <- 6)  
}  
  
subset\_2006 <- subset(INSTAR, INSTAR$year==2006)  
for(i in 1:length(subset\_2006$yday)){  
 ifelse(subset\_2006$yday[i]<=79,   
 subset\_2006$biociclo[i] <- 6, subset\_2006$biociclo[i] <- 7)  
}  
  
subset\_2007 <- subset(INSTAR, INSTAR$year==2007)  
for(i in 1:length(subset\_2007$yday)){  
 ifelse(subset\_2007$yday[i]<=79,   
 subset\_2007$biociclo[i] <- 7, subset\_2007$biociclo[i] <- 8)  
}  
  
subset\_2008 <- subset(INSTAR, INSTAR$year==2008)  
for(i in 1:length(subset\_2008$yday)){  
 ifelse(subset\_2008$yday[i]<=79,   
 subset\_2008$biociclo[i] <- 8, subset\_2008$biociclo[i] <- 9)  
}  
  
subset\_2009 <- subset(INSTAR, INSTAR$year==2009)  
for(i in 1:length(subset\_2009$yday)){  
 ifelse(subset\_2009$yday[i]<=79,   
 subset\_2009$biociclo[i] <- 9, subset\_2009$biociclo[i] <- 10)  
}  
  
subset\_2010 <- subset(INSTAR, INSTAR$year==2010)  
for(i in 1:length(subset\_2010$yday)){  
 ifelse(subset\_2010$yday[i]<=79,   
 subset\_2010$biociclo[i] <- 10, subset\_2010$biociclo[i] <- 11)  
}  
  
subset\_2011 <- subset(INSTAR, INSTAR$year==2011)  
for(i in 1:length(subset\_2011$yday)){  
 ifelse(subset\_2011$yday[i]<=79,   
 subset\_2011$biociclo[i] <- 11, subset\_2011$biociclo[i] <- 12)  
}  
  
subset\_2012 <- subset(INSTAR, INSTAR$year==2012)  
for(i in 1:length(subset\_2012$yday)){  
 ifelse(subset\_2012$yday[i]<=79,   
 subset\_2012$biociclo[i] <- 12, subset\_2012$biociclo[i] <- 13)  
}  
  
subset\_2013 <- subset(INSTAR, INSTAR$year==2013)  
for(i in 1:length(subset\_2013$yday)){  
 ifelse(subset\_2013$yday[i]<=79,   
 subset\_2013$biociclo[i] <- 13, subset\_2013$biociclo[i] <- 14)  
}  
  
subset\_2014 <- subset(INSTAR, INSTAR$year==2014)  
for(i in 1:length(subset\_2014$yday)){  
 ifelse(subset\_2014$yday[i]<=79,   
 subset\_2014$biociclo[i] <- 14, subset\_2014$biociclo[i] <- 15)  
}  
  
INSTAR<- rbind(subset\_2002, subset\_2003, subset\_2004, subset\_2005, subset\_2006, subset\_2007, subset\_2008, subset\_2009, subset\_2010, subset\_2011, subset\_2012, subset\_2013)

## Agregacion INSTAR

Media del vigor y sumas (l1, l2 y l1+l2) por biociclo

sum\_l1<- aggregate(x = INSTAR$L1, by=list(biociclo=INSTAR$biociclo), FUN=sum, na.rm=TRUE)  
names(sum\_l1)[2] <- "sum\_l1"  
  
sum\_l2<- aggregate(x = INSTAR$L2, by=list(biociclo=INSTAR$biociclo), FUN=sum, na.rm=TRUE)  
names(sum\_l2)[2] <- "sum\_l2"  
  
larvas<-cbind(sum\_l1, sum\_l2)  
larvas<-larvas[-c(3)]  
larvas$sum\_l1l2=larvas[,2]+larvas[,3]  
  
sum\_huevo<- aggregate(x = INSTAR$huevo, by=list(biociclo=INSTAR$biociclo), FUN=sum, na.rm=TRUE)  
names(sum\_huevo)[2] <- "sum\_huevo"  
  
sum\_crisalida<- aggregate(x = INSTAR$crisalida, by=list(biociclo=INSTAR$biociclo), FUN=sum, na.rm=TRUE)  
names(sum\_crisalida)[2] <- "sum\_crisalida"  
  
avg\_exergia<-aggregate(INSTAR$exergia, by=list(biociclo=INSTAR$biociclo), FUN=mean, na.rm=TRUE)  
names(avg\_exergia)[2] <- "avg\_exergia"  
  
  
  
agreg\_INSTAR <- cbind(larvas,avg\_exergia, sum\_crisalida, sum\_huevo)  
agreg\_INSTAR\_biociclo <- agreg\_INSTAR[ -c(5, 7, 9) ]  
  
#write.csv(agreg\_INSTAR\_biociclo,"agreg\_INSTAR\_bz008\_biociclo.csv",row.names=FALSE, na="")

## Ln INSTAR (l1, l2 y l1+l2)

\*Update: En la reunión del 9 de septiembre se llegó a la conclusión de que no era necesario transformar las variables.

Se agrega una columna a la tabla 'Agreg\_INSTAR\_xxxxx\_biociclo' con el resultado del Ln de cada columna referente a Larvas (sum\_l1, sum\_l2, sum\_l1l2)

# DATOS COPLAS

## Importar datos COPLAS

COPLAS<-read.csv("zona\_baza.csv", header=TRUE, sep=",")  
names(COPLAS)[1] <- "RODAL"

Llamamos a cada rodal como los tres ultimos numeros de su codigo GRxxxxxx Activar el rodal necesario para el análisis.

#rodal<-subset(COPLAS, RODAL == "GR000006")  
rodal<-subset(COPLAS, RODAL == "GR023008")  
#rodal<-subset(COPLAS, RODAL == "GR023012")  
#rodal<-subset(COPLAS, RODAL == "GR023015")  
#rodal<-subset(COPLAS, RODAL == "GR023017")  
#rodal<-subset(COPLAS, RODAL == "GR023021")  
  
#rodal<-subset(COPLAS, RODAL == "GR000030")  
#rodal<-subset(COPLAS, RODAL == "GR134018")  
#rodal<-subset(COPLAS, RODAL == "GR134020")  
#rodal<-subset(COPLAS, RODAL == "GR134021")  
  
#Eliminar primer y ultimo año porque en Instar no teneos datos de ese año completo.   
rodal<-rodal[-c(1,2, 23),]

## Definir biociclo COPLAS (opcional)

rodal$biociclo <- 3:22 #Sierra de Baza  
#rodal$biociclo <- 1:15 #Sierra Nevada

## Crear una tabla unica en la que tengamos los datos agregados de Instar y los datos de COPLAS

En esta tabla aparecen los datos de Instar agregados, los Ln de sumL1, sumL2 y sumL1+sumL2, el anio y el grado de infestacion de COPLAS.

COPLAS\_INSTAR<-merge.data.frame(agreg\_INSTAR\_biociclo, rodal, by.x = "biociclo")  
COPLAS\_INSTAR<-COPLAS\_INSTAR[-c(8)] #Eliminar la columna con el nombre del rodal

# VALIDACION EXTERNA

library(graphics)  
library(ggplot2)

## glm [avg\_exergia, sum\_l1, sum\_l2, sum\_l1l2 vs COPLAS]

*Se realizara, finalmente, GLM que analicen la relacion entre las variables de Instar (Exergia, sumL1, sumL2 y sumL1L2) y COPLAS. La distribucion es de Poisson.*

Una alternativa a la transformacion de la variable respuesta y a la falta de normalidad es el uso de los modelos linales generalizados. Los GLM son una extension de los modelos lineales que permiten utilizar distribuciones no normales de lso errores (Poisson, como en este caso) y varianzas no constantes.

Los modelos Poisoon se utilizan generalmente para representar datos de conteo. El uso de modelos lineales no sera adecuado ante datos de conteo por las siguientes razones: - El modelo lineal podra predecir valores negativos de la variable respuesta. - La varianza de la variable respuesta aumnetara probablemente a medida que aumenta la media (varianza no constante). - Los errores no estan normalmente distribuidos. - Los ceros son dificiles de manejar en transformaciones de la variable respuesta.

Comprobar dispersion. ?Como se hace? "Overdispersed Poisson occurs when sd >> mean"

### Asunciones y modelo de regresion lineal

#### POISSON

Observar: - Overdispersion

*Duda: en el libro 'Statistics for biology and health' la overdisp. la calcula como residual deviance/df (pg. 224). Antonio pone que se calcula como sumatoria del cuadrado de los residuos de pearson del modelo / n - (n? parametros del modelo +1). Esto sera, para exergia por ejemplo:* *Formula del libro: 3.8162/18 = 0.21* *Formula Antonio: sumar cada residual de pearson al cuadrado calculado antes con 'resid(reg\_exergia, type="pearson")^2' =4.20/18-(1+1)= 0.26* *En los dos casos es menor que 1, se puede decir que no hay overdisp.?* - AIC *Duda: los AICs con Poisson son = Inf.* - Var. explained (explained deviance) *Explained deviance= (100x(null deviance-residual deviance))/null dev.* - Coeficientes

**EXERGIA**

cor(COPLAS\_INSTAR$avg\_exergia, COPLAS\_INSTAR$GRADO.REVISADO)

## [1] -0.3364764

reg\_exergia <- glm(COPLAS\_INSTAR$avg\_exergia ~ COPLAS\_INSTAR$GRADO.REVISADO, family = poisson(link="log"))

## Warning in dpois(y, mu, log = TRUE): non-integer x = 71.839922

## Warning in dpois(y, mu, log = TRUE): non-integer x = 70.619812

## Warning in dpois(y, mu, log = TRUE): non-integer x = 71.138658

## Warning in dpois(y, mu, log = TRUE): non-integer x = 71.067546

## Warning in dpois(y, mu, log = TRUE): non-integer x = 68.993388

## Warning in dpois(y, mu, log = TRUE): non-integer x = 69.658424

## Warning in dpois(y, mu, log = TRUE): non-integer x = 68.267673

## Warning in dpois(y, mu, log = TRUE): non-integer x = 70.897995

## Warning in dpois(y, mu, log = TRUE): non-integer x = 69.348458

## Warning in dpois(y, mu, log = TRUE): non-integer x = 70.633578

## Warning in dpois(y, mu, log = TRUE): non-integer x = 71.929054

## Warning in dpois(y, mu, log = TRUE): non-integer x = 70.086720

summary.glm(reg\_exergia)

##   
## Call:  
## glm(formula = COPLAS\_INSTAR$avg\_exergia ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## family = poisson(link = "log"))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.27835 -0.07260 0.01968 0.09800 0.15847   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 4.256937 0.040152 106.02 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO -0.004702 0.031302 -0.15 0.881   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 0.19967 on 11 degrees of freedom  
## Residual deviance: 0.17703 on 10 degrees of freedom  
## AIC: Inf  
##   
## Number of Fisher Scoring iterations: 3

*"The first two lines tell us which model has been fitted, which is handy if you save the output into a word processor document. Basic numerical information on the residuals is also provided, although in Section 9.8 we present more useful graphical tools that can be used for the model validation process. The estimated intercept and slope are 4.31 and –0.000106, respectively. Keep in mind that distance to the park is expressed in metres. To avoid parameter estimates with lots of zeros, you could (and perhaps should) express it in kilometres, as it will save some ink when presenting the estimated slope on paper. We also get a z-statistic and corresponding p-value for testing the null hypothesis that the slope (and intercept) is equal to 0 and an AIC, which can be used for model selection. The z-statistic is used because we know the variance. In a Gaussian model, the variance is estimated as well, and therefore, a t-statistic is used."* *"The smaller the residual deviance, the better is the model."*

# Mirar overdispersion = residual deviance/df  
3.8162/18

## [1] 0.2120111

## Pearson residuals  
(resid(reg\_exergia, type="pearson"))^2

## 1 2 3 4 5   
## 2.201015e-02 9.867622e-06 1.093094e-02 9.228982e-03 1.170614e-03   
## 6 7 8 9 10   
## 1.238381e-02 7.662320e-02 1.314099e-03 1.188513e-02 1.962077e-03   
## 11 12   
## 2.527036e-02 3.636936e-03

(1.51+0.79+0.58+0.003+0.34+0.091+0.41+0.08+0.001+0.01+0.0022+0.003+0.09+0.049+0.149+0.0057+0.067+0.011+0.0021+0.0015) / (18 - (1+1)) #=0.26

## [1] 0.2622188

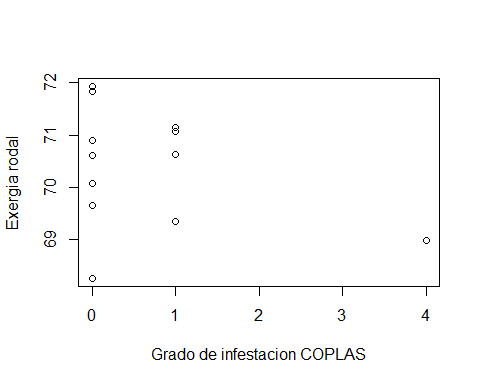
# Mirar explained deviance  
(100\*(3.8163-3.8162))/3.8163

## [1] 0.002620339

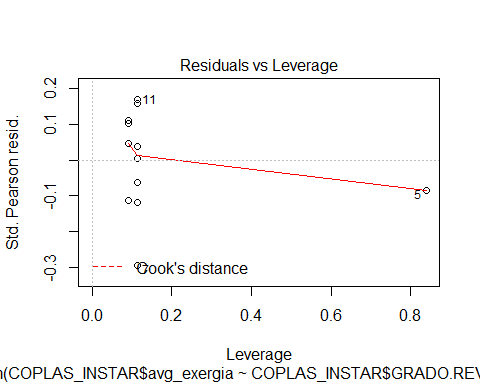
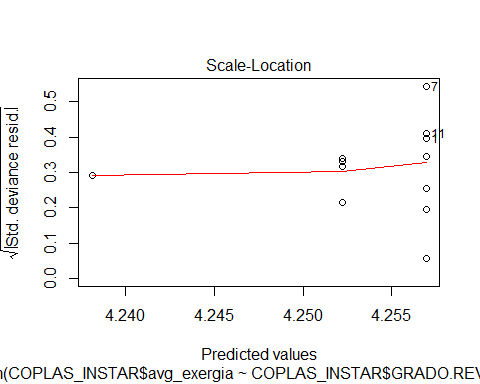
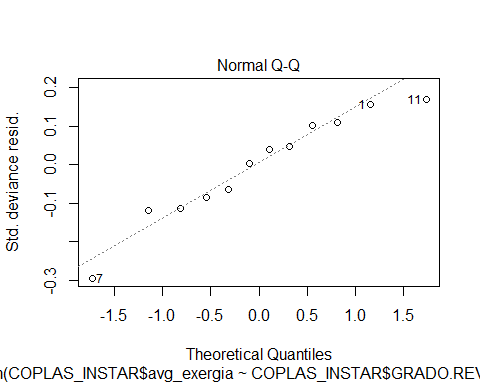
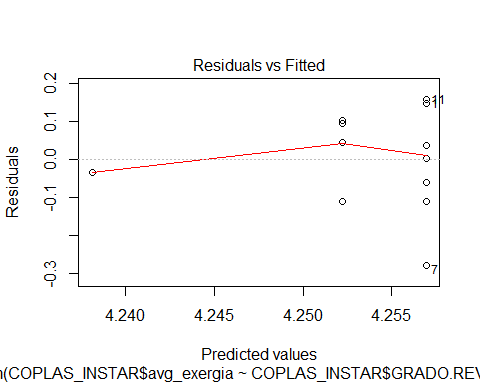
# AIC  
AIC(reg\_exergia)

## [1] Inf

plot(COPLAS\_INSTAR$avg\_exergia ~ COPLAS\_INSTAR$GRADO.REVISADO, xlab="Grado de infestacion COPLAS", ylab = "Exergia rodal")



plot(reg\_exergia)



dev.off()

## null device   
## 1

**L1**

cor(COPLAS\_INSTAR$sum\_l1, COPLAS\_INSTAR$GRADO.REVISADO)

## [1] 0.2814033

reg\_L1 <- glm(COPLAS\_INSTAR$sum\_l1 ~ COPLAS\_INSTAR$GRADO.REVISADO, family = poisson(link="log"))  
summary.glm(reg\_L1)

##   
## Call:  
## glm(formula = COPLAS\_INSTAR$sum\_l1 ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## family = poisson(link = "log"))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2932.3 -927.2 439.2 806.6 1088.3   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.666e+01 8.035e-05 207359 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO 6.353e-02 5.706e-05 1113 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 18039311 on 11 degrees of freedom  
## Residual deviance: 16858780 on 10 degrees of freedom  
## AIC: 16859006  
##   
## Number of Fisher Scoring iterations: 4

#Mirar overdispersion  
49182818/18

## [1] 2732379

Sea cual sea la formula que se use va a salir un valor **MUY ALTO**. Decimos que hay overdisp.?? Hacemos quasipoisson:

reg\_L1\_qpois <- glm(COPLAS\_INSTAR$sum\_l1 ~ COPLAS\_INSTAR$GRADO.REVISADO, family = quasipoisson)  
summary(reg\_L1\_qpois)

##   
## Call:  
## glm(formula = COPLAS\_INSTAR$sum\_l1 ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## family = quasipoisson)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2932.3 -927.2 439.2 806.6 1088.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 16.66233 0.09783 170.314 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO 0.06353 0.06947 0.915 0.382   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for quasipoisson family taken to be 1482333)  
##   
## Null deviance: 18039311 on 11 degrees of freedom  
## Residual deviance: 16858780 on 10 degrees of freedom  
## AIC: NA  
##   
## Number of Fisher Scoring iterations: 4

Da una dispersion de 2178478 ¿¿?? *"The dispersion parameter is estimated as 2178478. This means that all standard errors have been multiplied by 1476 (the square root of 2178478), and as a result, most parameters are no longer significant!"*

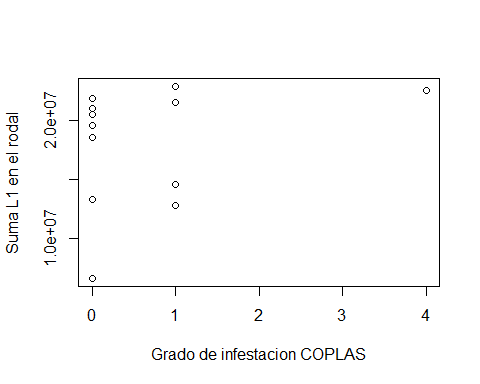
#Mirar explained deviance  
(100\*(53589882-49182818))/53589882#=8.223687

## [1] 8.223687

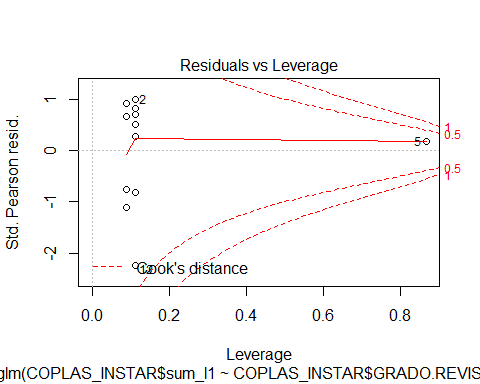
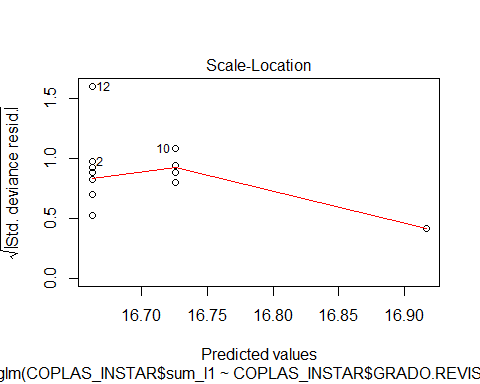
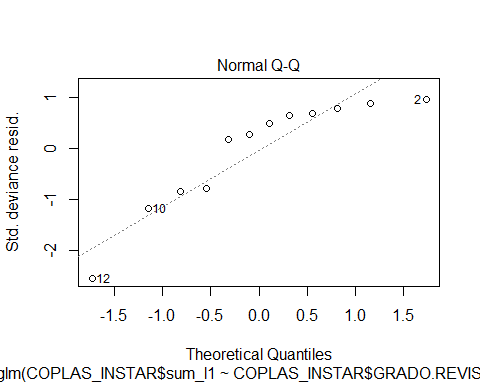
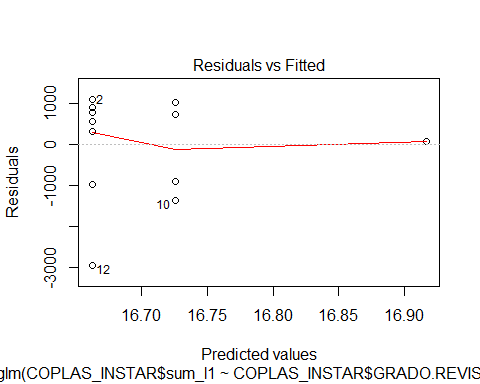
#AIC  
AIC(reg\_L1\_qpois)#NA

## [1] NA

#plots  
plot(COPLAS\_INSTAR$sum\_l1 ~ COPLAS\_INSTAR$GRADO.REVISADO, xlab="Grado de infestacion COPLAS", ylab = "Suma L1 en el rodal")



plot(reg\_L1\_qpois)



dev.off()

## null device   
## 1

**L2**

cor(COPLAS\_INSTAR$sum\_l2, COPLAS\_INSTAR$GRADO.REVISADO)

## [1] 0.320123

reg\_L2 <- glm(COPLAS\_INSTAR$sum\_l2 ~ COPLAS\_INSTAR$GRADO.REVISADO, family = poisson(link="log"))  
summary.glm(reg\_L2)

##   
## Call:  
## glm(formula = COPLAS\_INSTAR$sum\_l2 ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## family = poisson(link = "log"))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1670.5 -1287.8 378.8 733.6 1132.7   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.648e+01 8.810e-05 187000 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO 7.259e-02 6.179e-05 1175 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 14406686 on 11 degrees of freedom  
## Residual deviance: 13101153 on 10 degrees of freedom  
## AIC: 13101377  
##   
## Number of Fisher Scoring iterations: 4

#Mirar overdispersion  
26968257/18

## [1] 1498237

reg\_L2\_qpois <- glm(COPLAS\_INSTAR$sum\_l2 ~ COPLAS\_INSTAR$GRADO.REVISADO, family = quasipoisson)  
summary.glm(reg\_L2\_qpois)

##   
## Call:  
## glm(formula = COPLAS\_INSTAR$sum\_l2 ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## family = quasipoisson)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1670.5 -1287.8 378.8 733.6 1132.7   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 16.47520 0.09756 168.880 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO 0.07259 0.06842 1.061 0.314   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for quasipoisson family taken to be 1226104)  
##   
## Null deviance: 14406686 on 11 degrees of freedom  
## Residual deviance: 13101153 on 10 degrees of freedom  
## AIC: NA  
##   
## Number of Fisher Scoring iterations: 4

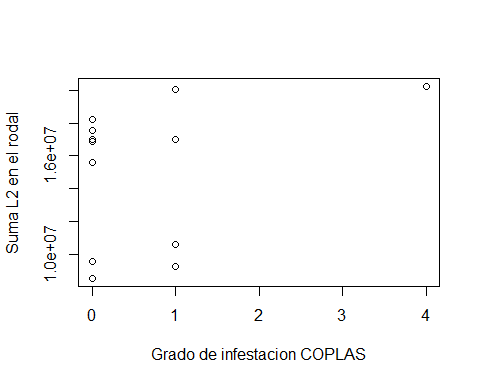
#Dispersion segun quasipoison=1402424  
  
#Mirar explained deviance  
(100\*(26968257-29317185))/26968257#=-8.709973

## [1] -8.709973

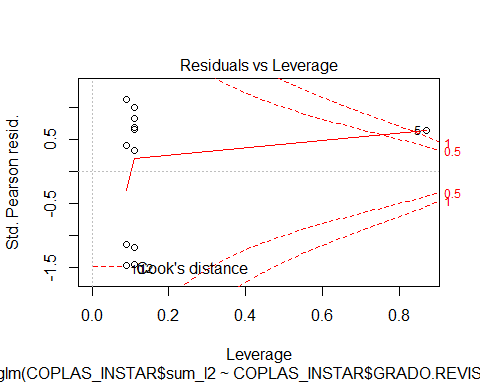
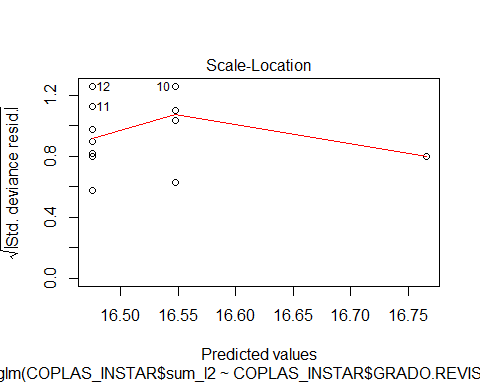
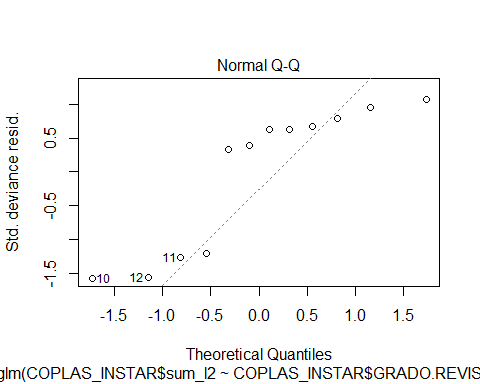
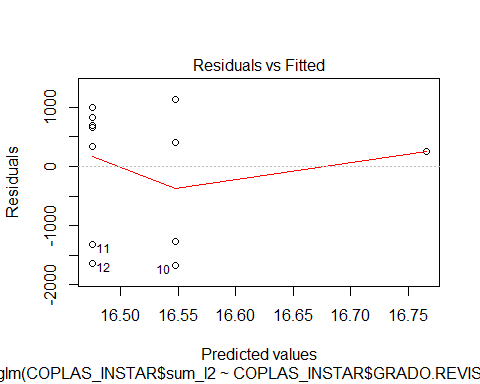
#AIC  
AIC(reg\_L2\_qpois)#NA

## [1] NA

#plots  
plot(COPLAS\_INSTAR$sum\_l2 ~ COPLAS\_INSTAR$GRADO.REVISADO, xlab="Grado de infestacion COPLAS", ylab = "Suma L2 en el rodal")



plot(reg\_L2\_qpois)



dev.off()

## null device   
## 1

**L1+L2**

# cor(COPLAS\_INSTAR$sum\_l1l2, COPLAS\_INSTAR$GRADO.REVISADO)  
reg\_L1L2 <- glm(COPLAS\_INSTAR$sum\_l1l2 ~ COPLAS\_INSTAR$GRADO.REVISADO, family = poisson(link="log"))  
summary.glm(reg\_L1L2)

##   
## Call:  
## glm(formula = COPLAS\_INSTAR$sum\_l1l2 ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## family = poisson(link = "log"))  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3246.6 -1549.6 663.7 1136.1 1523.6   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.727e+01 5.937e-05 290826 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO 6.768e-02 4.192e-05 1614 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for poisson family taken to be 1)  
##   
## Null deviance: 31162746 on 11 degrees of freedom  
## Residual deviance: 28688280 on 10 degrees of freedom  
## AIC: 28688514  
##   
## Number of Fisher Scoring iterations: 4

#Mirar overdispersion  
71354778/18

## [1] 3964154

reg\_L1L2\_qpois <- glm(COPLAS\_INSTAR$sum\_l1l2 ~ COPLAS\_INSTAR$GRADO.REVISADO, family = quasipoisson)  
summary.glm(reg\_L1L2\_qpois)

##   
## Call:  
## glm(formula = COPLAS\_INSTAR$sum\_l1l2 ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## family = quasipoisson)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3246.6 -1549.6 663.7 1136.1 1523.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 17.26628 0.09624 179.412 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO 0.06768 0.06795 0.996 0.343   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for quasipoisson family taken to be 2627617)  
##   
## Null deviance: 31162746 on 11 degrees of freedom  
## Residual deviance: 28688280 on 10 degrees of freedom  
## AIC: NA  
##   
## Number of Fisher Scoring iterations: 4

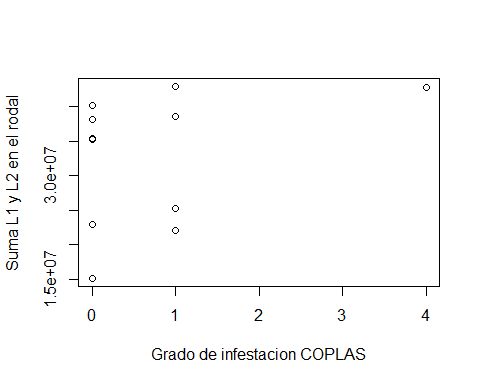
#Dispersion segun quasipoison=3468083  
  
#Mirar explained deviance  
(100\*(71354778-78035661))/71354778#=-8.709973

## [1] -9.362909

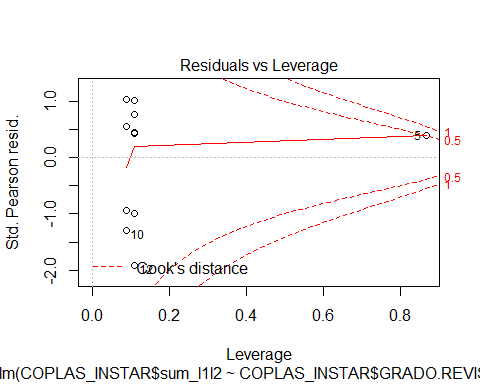
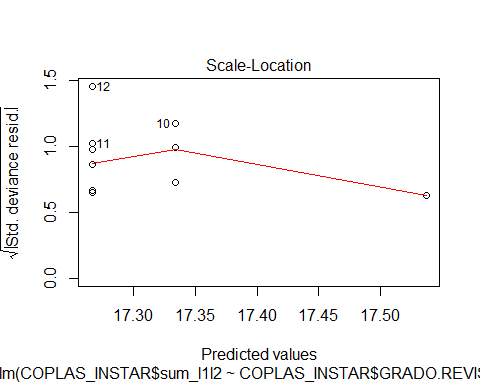
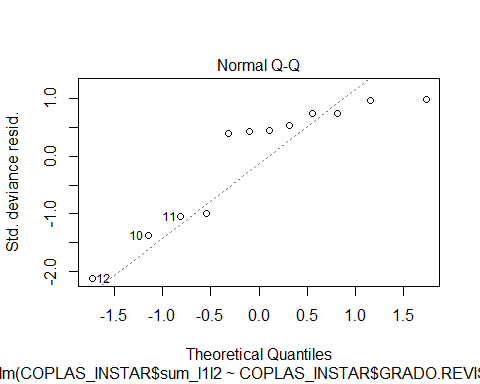
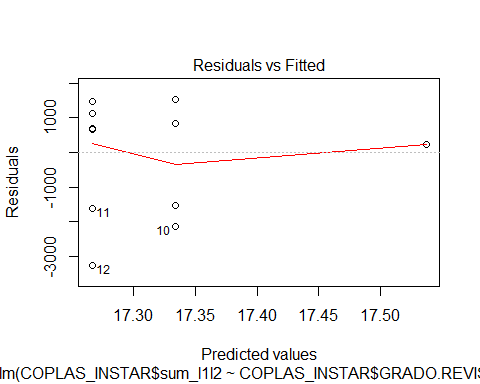
#AIC  
AIC(reg\_L1L2\_qpois)#NA

## [1] NA

#plots  
plot(COPLAS\_INSTAR$sum\_l1l2 ~ COPLAS\_INSTAR$GRADO.REVISADO, xlab="Grado de infestacion COPLAS", ylab = "Suma L1 y L2 en el rodal")



plot(reg\_L1L2\_qpois)



dev.off()

## null device   
## 1

#### NEGATIVE BINOMIAL

**EXERGIA**

library(MASS)

reg\_exergia\_nb<-glm.nb(COPLAS\_INSTAR$avg\_exergia ~ COPLAS\_INSTAR$GRADO.REVISADO, link="log")

## Warning in dpois(y, mu, log = TRUE): non-integer x = 71.839922

## Warning in dpois(y, mu, log = TRUE): non-integer x = 70.619812

## Warning in dpois(y, mu, log = TRUE): non-integer x = 71.138658

## Warning in dpois(y, mu, log = TRUE): non-integer x = 71.067546

## Warning in dpois(y, mu, log = TRUE): non-integer x = 68.993388

## Warning in dpois(y, mu, log = TRUE): non-integer x = 69.658424

## Warning in dpois(y, mu, log = TRUE): non-integer x = 68.267673

## Warning in dpois(y, mu, log = TRUE): non-integer x = 70.897995

## Warning in dpois(y, mu, log = TRUE): non-integer x = 69.348458

## Warning in dpois(y, mu, log = TRUE): non-integer x = 70.633578

## Warning in dpois(y, mu, log = TRUE): non-integer x = 71.929054

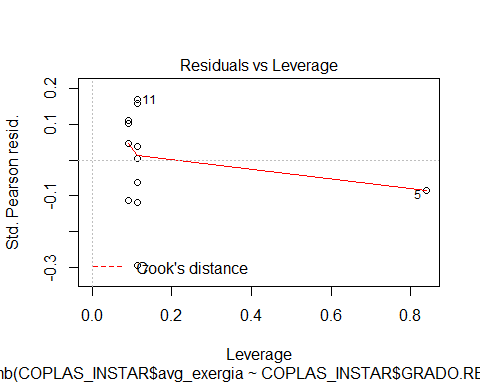
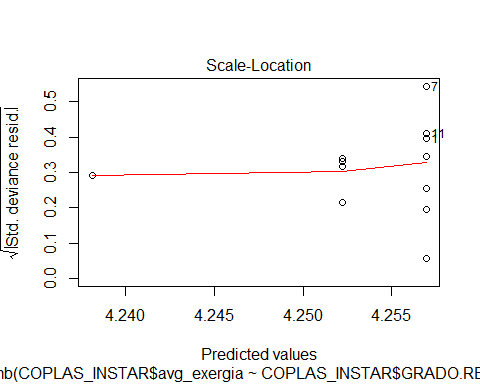
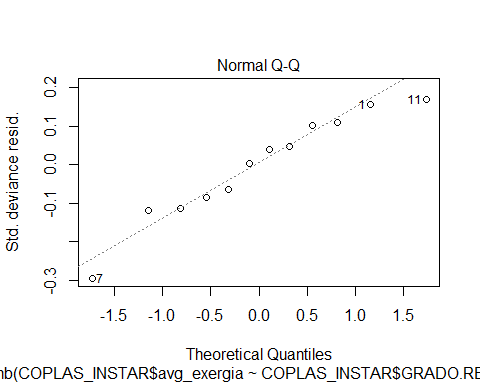
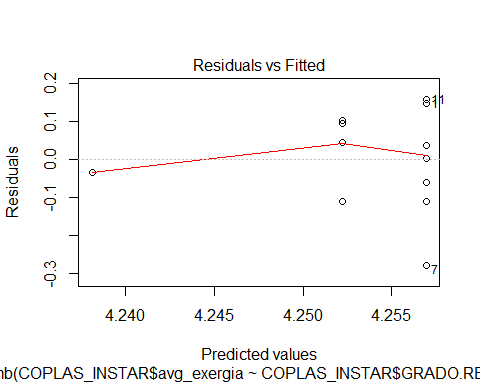
## Warning in dpois(y, mu, log = TRUE): non-integer x = 70.086720

## Warning in theta.ml(Y, mu, sum(w), w, limit = control$maxit, trace =  
## control$trace > : iteration limit reached  
  
## Warning in theta.ml(Y, mu, sum(w), w, limit = control$maxit, trace =  
## control$trace > : iteration limit reached

summary(reg\_exergia\_nb)

##   
## Call:  
## glm.nb(formula = COPLAS\_INSTAR$avg\_exergia ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## link = "log", init.theta = 82358206.66)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.27835 -0.07260 0.01968 0.09800 0.15847   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 4.256937 0.040152 106.02 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO -0.004702 0.031302 -0.15 0.881   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(82358207) family taken to be 1)  
##   
## Null deviance: 0.19967 on 11 degrees of freedom  
## Residual deviance: 0.17703 on 10 degrees of freedom  
## AIC: 79.304  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 82358207   
## Std. Err.: 14740298078   
## Warning while fitting theta: iteration limit reached   
##   
## 2 x log-likelihood: -73.304

plot(reg\_exergia\_nb)



dev.off()

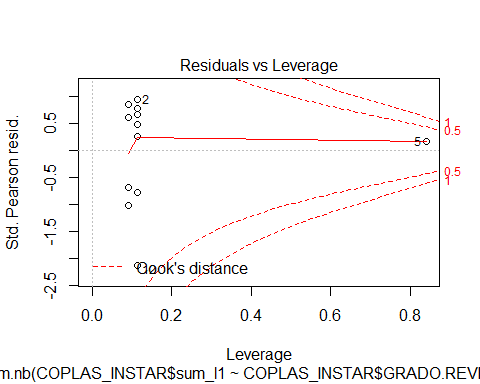
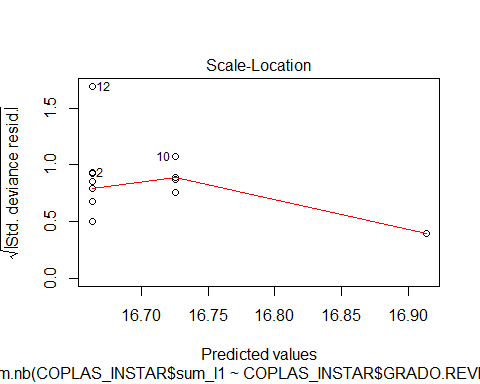
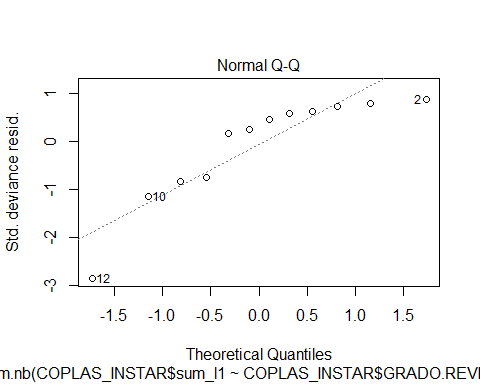
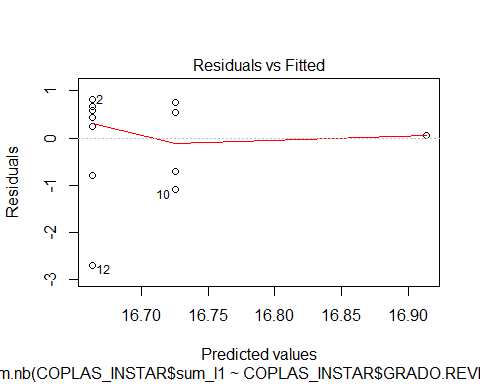
## null device   
## 1

**L1**

reg\_L1\_nb<-glm.nb(COPLAS\_INSTAR$sum\_l1 ~ COPLAS\_INSTAR$GRADO.REVISADO, link="log")  
summary(reg\_L1\_nb)

##   
## Call:  
## glm.nb(formula = COPLAS\_INSTAR$sum\_l1 ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## link = "log", init.theta = 10.56399641)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.6928 -0.7376 0.3353 0.6102 0.8161   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 16.66301 0.10372 160.660 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO 0.06258 0.08034 0.779 0.436   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(10.564) family taken to be 1)  
##   
## Null deviance: 12.838 on 11 degrees of freedom  
## Residual deviance: 12.189 on 10 degrees of freedom  
## AIC: 411.9  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 10.56   
## Std. Err.: 4.25   
##   
## 2 x log-likelihood: -405.903

plot(reg\_L1\_nb)



dev.off()

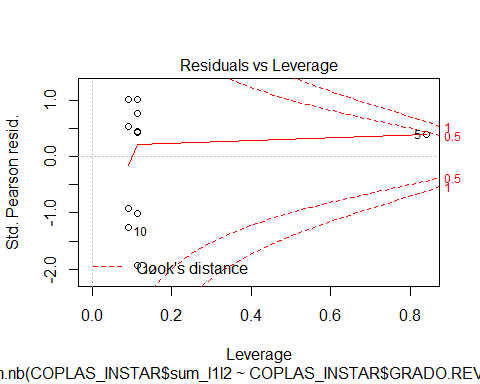
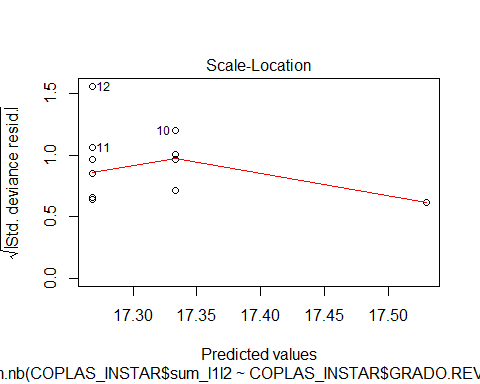
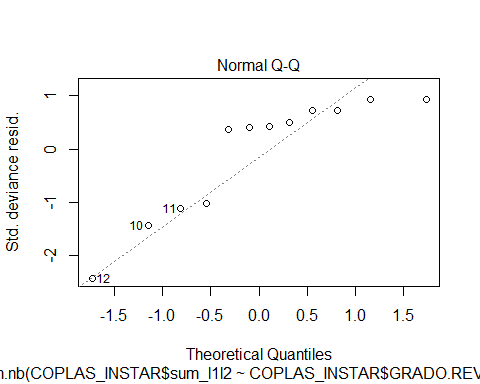
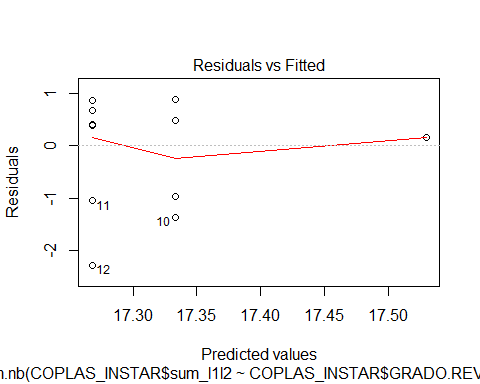
## null device   
## 1

**L1+L2**

reg\_L1L2\_nb<-glm.nb(COPLAS\_INSTAR$sum\_l1l2 ~ COPLAS\_INSTAR$GRADO.REVISADO, link="log")  
summary(reg\_L1L2\_nb)

##   
## Call:  
## glm.nb(formula = COPLAS\_INSTAR$sum\_l1l2 ~ COPLAS\_INSTAR$GRADO.REVISADO,   
## link = "log", init.theta = 12.16580819)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2820 -0.9863 0.3989 0.6780 0.8813   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 17.26783 0.09665 178.669 <2e-16 \*\*\*  
## COPLAS\_INSTAR$GRADO.REVISADO 0.06549 0.07486 0.875 0.382   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(12.1658) family taken to be 1)  
##   
## Null deviance: 13.000 on 11 degrees of freedom  
## Residual deviance: 12.164 on 10 degrees of freedom  
## AIC: 424.88  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 12.17   
## Std. Err.: 4.90   
##   
## 2 x log-likelihood: -418.875

plot(reg\_L1L2\_nb)



dev.off()

## null device   
## 1

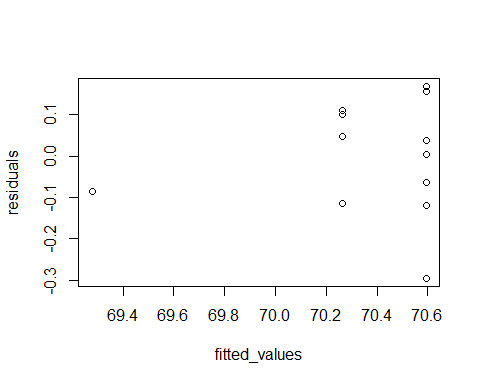
##### ZEROINFLATED

Corregidme si me equivoco, pero este analisis se utiliza cuando la 'response variable', es decir la variable DEPENDIENTE, tiene muchos ceros. En nuestro caso la variable dependiente es Instar, la cual no tiene ceros. La que tiene ceros es COPLAS (variable independiente)

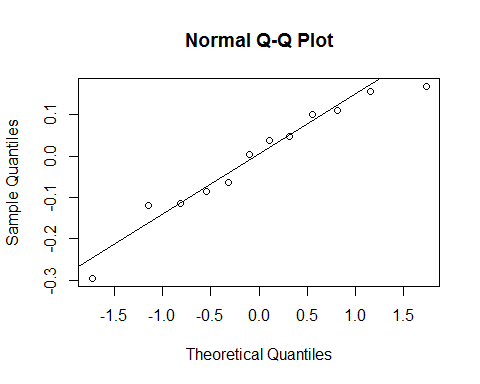
#### MODELOS MIXTOS

*Creo que esto viene al pelo: "This chapter looks at a data set where our first reaction was: 'How in heavens name are we going to analyse these data?'"* Analisis cogido del libro Statistics for biology and health (a partir de p.423)

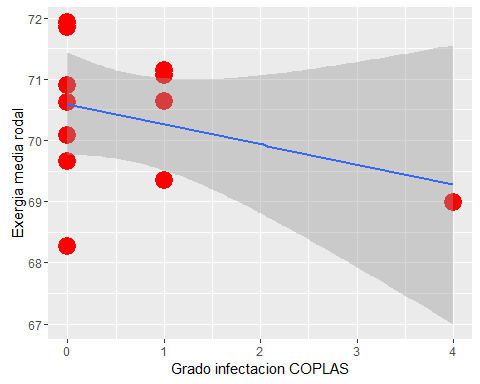
### Residuos estandarizados  
residuals<-rstandard(reg\_exergia)  
### Valores ajustados  
fitted\_values<-fitted(reg\_exergia)  
### Grafica valores ajustados vs residuos (asi evaluamos a ojo la homocedasticidad. Esto es discutible...)  
plot(fitted\_values, residuals)



### Grafico QQ para evaluar la normalidad  
qqnorm(residuals)  
qqline(residuals)



ggplot(COPLAS\_INSTAR, aes(x=GRADO.REVISADO, y=avg\_exergia))+  
geom\_point(colour='red', size=6) +  
geom\_smooth(method=lm)+  
xlab("Grado infectacion COPLAS")+  
ylab("Exergia media rodal")



# VALIDACION INTERNA

## ACP

Autocorrelation plot. Nos ayudara a saber si existen ciclos dentro de toda la serie temporal. ESTE ANALISIS SOLO SE HACE CON LOS DATOS DE BAZA

"*Those plots are showing you the correlation of the series with itself, lagged by x time unitscorrelation of the series with itself, lagged by x time units. So imagine taking your time series of length TT, copying it, and deleting the first observation of copy#1 and the last observation of copy#2. Now you have two series of length T for which you calculate a correlation coefficient. This is the value of of the vertical axis at x=1 in your plots. It represents the correlation of the series lagged by one time unit. You go on and do this for all possible time lags xx and this defines the plot*.

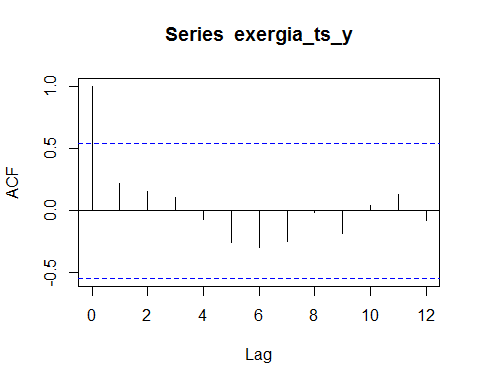
*The answer to your question of what is needed to report a pattern is dependent on what pattern you would like to report. But quantitatively speaking, you have exactly what I just described: the correlation coefficient at different lags of the series. You can extract these numerical values by issuing the command acf(x.ts,100)$acf*.

*In terms of what lag to use, this is again a matter of context. It is often the case that there will be specific lags of interest. Say, for example, you may believe the fish species migrates to and from an area every ~30 days. This may lead you to hypothesize a correlation in the time series at lags of 30. In this case, you would have support for your hypothesis*"

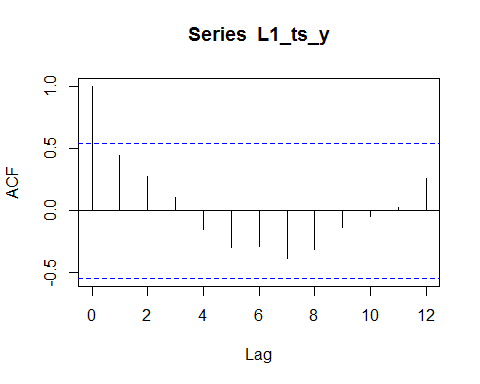
"*The blue lines give the values beyond which the autocorrelations are (statistically) significantly different from zero*"

Hay que definir el LAG. lag.max= Lo hemos definido como el numero de biociclos.

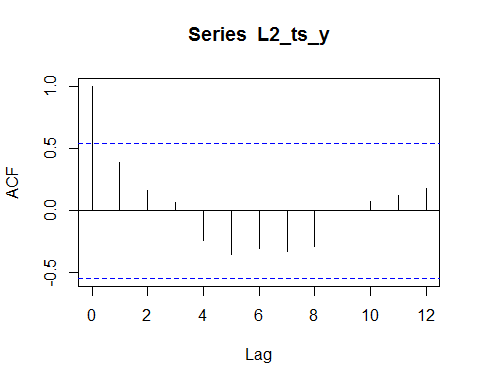
exergia\_ts\_y<-ts(agreg\_INSTAR\_biociclo$avg\_exergia, frequency = 1)  
exergia\_acf\_y<-acf(exergia\_ts\_y, type=c("correlation"), plot=TRUE, lag.max = 19)



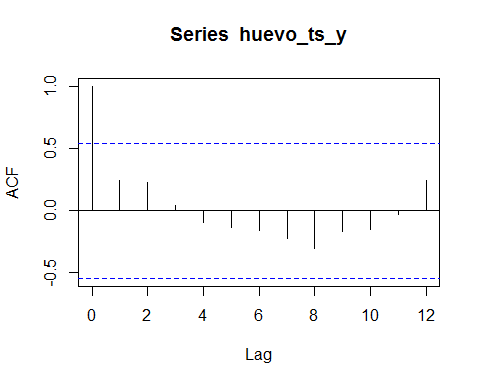
L1\_ts\_y<-ts(agreg\_INSTAR\_biociclo$sum\_l1, frequency = 1)  
L1\_acf\_y<-acf(L1\_ts\_y, type=c("correlation"), plot=TRUE, lag.max = 19)



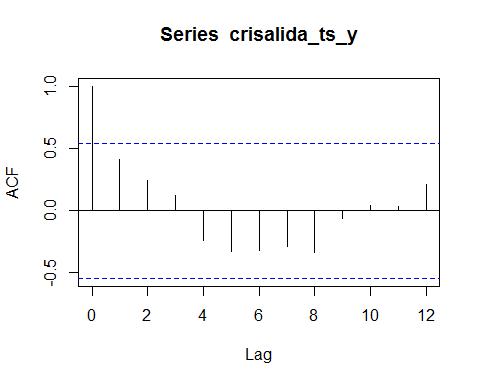
L2\_ts\_y<-ts(agreg\_INSTAR\_biociclo$sum\_l2, frequency = 1)  
L2\_acf\_y<-acf(L2\_ts\_y, type=c("correlation"), plot=TRUE, lag.max = 19)



huevo\_ts\_y<-ts(agreg\_INSTAR\_biociclo$sum\_huevo, frequency = 1)  
huevo\_acf\_y<-acf(huevo\_ts\_y, type=c("correlation"), plot=TRUE, lag.max = 19)



crisalida\_ts\_y<-ts(agreg\_INSTAR\_biociclo$sum\_crisalida, frequency = 1)  
crisalida\_acf\_y<-acf(crisalida\_ts\_y, type=c("correlation"), plot=TRUE, lag.max = 19)



CORRELOGRAMAS El dato de un anio esta significativamente correlacionado positivamente con el anio anterior. Es decir, el dato de la variable X del anio 't' depende del anio 'anterior't-1' de formar que cuanto mas alto el valor de la variable en 't-1', mayor es en 'este't' anio.

Los Lags 3, 4, 5 indican correlacion negativa (no significativo). Quiere decir que lo que pase en 't' esta negativamente correlacionado con't-3', cuanto mayor la variable en 't-3', menos sera en 't'.

Esto puede indicar densodependencia, es decri, regulacion por densidad. Una pena que no sea significativo...

## Analisis de tendencia

Se eligen indicadores de cada variable para calcular su tendencia.

En las variables huevo, L1, L2 y crisalida: - Fecha del maximo. - **Fecha en la que aparece la variable (primer dia en el que la variable es distinta a cero).**

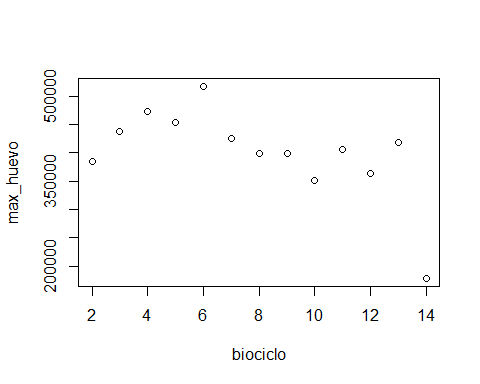
En la variable exergia: - fecha del maximo - fecha del minimo.

\*Nota: se puede calcular el minimo sin contar el 0?

max\_huevo<-aggregate(x = INSTAR$huevo, by=list(biociclo=INSTAR$biociclo), FUN=max, na.rm=TRUE)  
names(max\_huevo)[2] <- "max\_huevo"  
tend\_huevo<-lm(max\_huevo)  
summary(tend\_huevo)

##   
## Call:  
## lm(formula = max\_huevo)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.4485 -1.4620 -0.0386 1.4315 5.5091   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.980e+01 4.761e+00 4.158 0.00159 \*\*  
## max\_huevo -2.944e-05 1.167e-05 -2.523 0.02832 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.237 on 11 degrees of freedom  
## Multiple R-squared: 0.3666, Adjusted R-squared: 0.309   
## F-statistic: 6.366 on 1 and 11 DF, p-value: 0.02832

plot(max\_huevo)



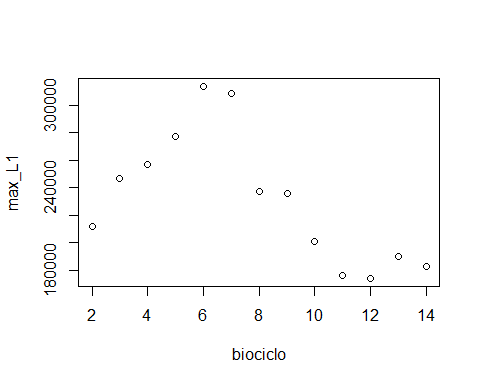
dev.off()

## null device   
## 1

max\_L1<-aggregate(x = INSTAR$L1, by=list(biociclo=INSTAR$biociclo), FUN=max, na.rm=TRUE)  
names(max\_L1)[2] <- "max\_L1"  
tend\_L1<-lm(max\_L1)  
summary(tend\_L1)

##   
## Call:  
## lm(formula = max\_L1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.9969 -0.6963 0.4569 2.1496 3.5312   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.974e+01 4.545e+00 4.343 0.00117 \*\*  
## max\_L1 -5.066e-05 1.924e-05 -2.633 0.02330 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.186 on 11 degrees of freedom  
## Multiple R-squared: 0.3865, Adjusted R-squared: 0.3308   
## F-statistic: 6.931 on 1 and 11 DF, p-value: 0.0233

plot(max\_L1)  
abline(tend\_L1)



dev.off()

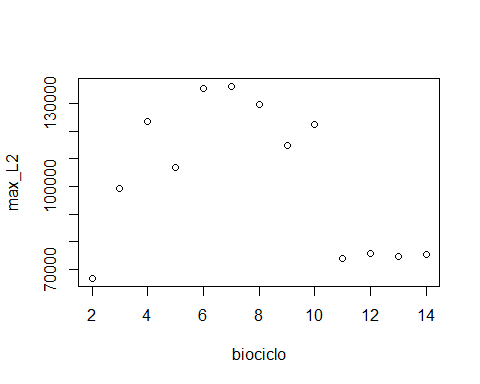
## null device   
## 1

H0=pendiente es cero Segun los coefficients de la regresion no se rechaza H0 ya que p>0.05. Tampoco quiere decir que se acepte.

max\_L2<-aggregate(x = INSTAR$L2, by=list(biociclo=INSTAR$biociclo), FUN=max, na.rm=TRUE)  
names(max\_L2)[2] <- "max\_L2"  
tend\_L2<-lm(max\_L2)  
summary(tend\_L2)

##   
## Call:  
## lm(formula = max\_L2)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.776 -2.783 1.336 2.675 4.644   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.305e+01 4.425e+00 2.950 0.0132 \*  
## max\_L2 -4.923e-05 4.184e-05 -1.177 0.2641   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.833 on 11 degrees of freedom  
## Multiple R-squared: 0.1118, Adjusted R-squared: 0.03107   
## F-statistic: 1.385 on 1 and 11 DF, p-value: 0.2641

plot(max\_L2)  
abline(tend\_L2)



dev.off()

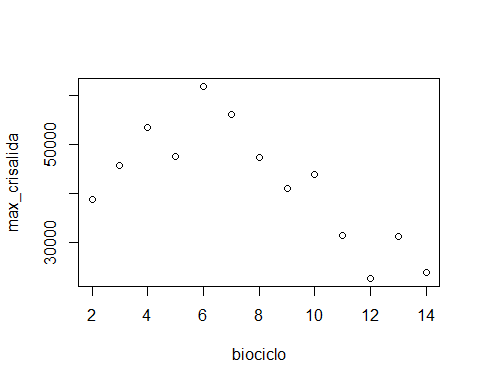
## null device   
## 1

Aqui si existe significativamente una tendencia negativa.Aunque la oendiente es muy muy baja. Podemos relacionarlo con factores abioticos?

max\_crisalida<-aggregate(x = INSTAR$crisalida, by=list(biociclo=INSTAR$biociclo), FUN=max, na.rm=TRUE)  
names(max\_crisalida)[2] <- "max\_crisalida"  
tend\_crisalida<-lm(max\_crisalida)  
summary(tend\_crisalida)

##   
## Call:  
## lm(formula = max\_crisalida)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.6975 -1.4686 0.7943 2.1411 2.6485   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.720e+01 3.087e+00 5.573 0.000167 \*\*\*  
## max\_crisalida -2.199e-04 7.109e-05 -3.094 0.010219 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.974 on 11 degrees of freedom  
## Multiple R-squared: 0.4653, Adjusted R-squared: 0.4166   
## F-statistic: 9.571 on 1 and 11 DF, p-value: 0.01022

plot(max\_crisalida)  
abline(tend\_crisalida)



dev.off()

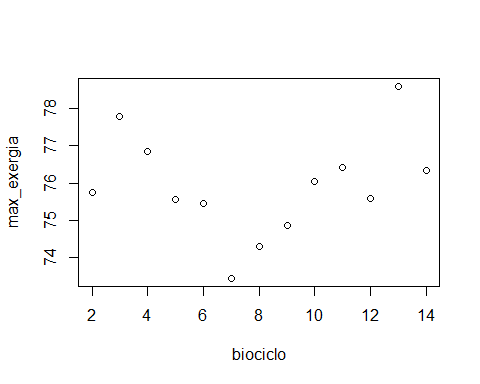
## null device   
## 1

H0=pendiente es cero Segun los coefficients de la regresion no se rechaza H0 ya que p>0.05. Tampoco quiere decir que se acepte.

max\_exergia<-aggregate(x = INSTAR$exergia, by=list(biociclo=INSTAR$biociclo), FUN=max, na.rm=TRUE)  
names(max\_exergia)[2] <- "max\_exergia"  
tend\_exergia\_mx<-lm(max\_exergia)  
summary(tend\_exergia\_mx)

##   
## Call:  
## lm(formula = max\_exergia)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.9423 -2.8705 0.5679 2.8219 5.8552   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -18.4906 64.9856 -0.285 0.781  
## max\_exergia 0.3489 0.8559 0.408 0.691  
##   
## Residual standard error: 4.037 on 11 degrees of freedom  
## Multiple R-squared: 0.01489, Adjusted R-squared: -0.07467   
## F-statistic: 0.1662 on 1 and 11 DF, p-value: 0.6913

plot(max\_exergia)  
abline(tend\_exergia\_mx)



dev.off()

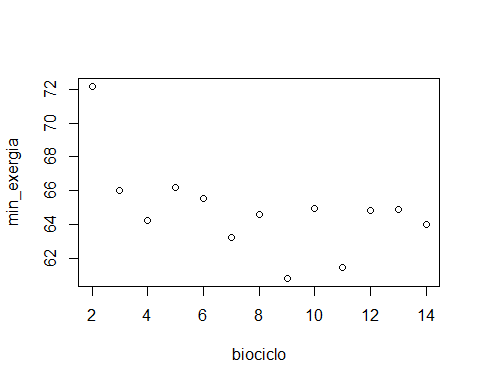
## null device   
## 1

Aqui si existe significativamente una tendencia negativa.Slope=-0.52 Es alta?

min\_exergia<-aggregate(x = INSTAR$exergia, by=list(biociclo=INSTAR$biociclo), FUN=min, na.rm=TRUE)  
names(min\_exergia)[2] <- "min\_exergia"  
tend\_exergia\_mn<-lm(min\_exergia)  
summary(tend\_exergia\_mn)

##   
## Call:  
## lm(formula = min\_exergia)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.4632 -2.1883 -0.1943 2.0786 5.3539   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 59.2172 23.3082 2.541 0.0274 \*  
## min\_exergia -0.7900 0.3592 -2.199 0.0502 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.39 on 11 degrees of freedom  
## Multiple R-squared: 0.3054, Adjusted R-squared: 0.2423   
## F-statistic: 4.836 on 1 and 11 DF, p-value: 0.05016

plot(min\_exergia)  
abline(tend\_exergia\_mn)



dev.off()

## null device   
## 1

Pendiente=-0.85 (alta), pero no significativa.

selectrunning ## Definir ts: huevos, l1, l2, crisalidas, energia, tmax, tmin, tmedia Esto por ahora no sirve.

# plaga\_ts<-ts(INSTAR[2:5], class=c("mts"), frequency=365)#Frecuencia es igual al numero de dias?  
# temp\_ts<-ts(INSTAR[7:9], class=c("mts"), frequency=365)  
# huevos\_ts<-ts(INSTAR$huevo, frequency=365)  
# L1\_ts<-ts(INSTAR$L1, frequency=365)  
# L2\_ts<-ts(INSTAR$L2, frequency=365)  
# crisalidas\_ts<-ts(INSTAR$crisalida, frequency=1)  
# exergia\_ts<-ts(INSTAR$exergia, frequency=1)  
# tmax\_ts<-ts(INSTAR$tmax, frequency=1)  
# tmin\_ts<-ts(INSTAR$tmin, frequency=1)  
# tmed\_ts<-ts(INSTAR$tmed, frequency=1)

# summary(plaga\_ts)  
# cycle (plaga\_ts)  
# class(plaga\_ts)  
# str(plaga\_ts)  
# decompose(plaga\_ts)  
# plot(decompose(plaga\_ts))

NOTA: se supone que hay alguna manera de escribir el Markdown de manera que al verlo desde Github se vean las graficas o los resultados de la ejecucion. Esto estaria muy bien para poder ver los resultados de las regresiones de un vistazo y comparar entre ejecuciones. Seguro que Antonio sabe como se hace, el viernes le preguntamos :D