



Universitat d'Alacant
Universidad de Alicante

Departamento de Ecología

ECOLOGY

Degree of Biology

Sciences University College

GRAPHICAL REPRESENTATION AND STATISTICAL ANALYSIS OF DATA

Carlos Sanz Lázaro



Introduction

The study of changes in biological communities in relation to a process or hypothesis under study, involves interpreting changes in the populations of a given community in relation to the physical-chemical characteristics of the environment. The graphic representation of these variables according to the changes of the physico-chemical variables, environmental gradients or treatments of the considered factors is a fundamental step to define and communicate the results of the study, commonly supported by appropriate statistical analyses in accordance with the type of the data.

Aim

Using the data taken in the field trip to Maimó, we want to evaluate the effect of the variations of environmental factors between the sunny and shady areas of a hill.

Below, it is the script for the R program as a possibility to make the necessary graphic representations and the corresponding statistical analyses.



Script of the plots and the analyses needed

R version 3.0.1 (2013-05-16) -- "Good Sport"

```
#Save the spreadsheet matrix as text (delimited by tabs) (* .txt)
```

```
#Load the matrix
```

```
dat=read.table(file.choose(), header=T)
```

```
#File: "Data_EN.txt"
```

```
dat
```

```
#We install GAD package
```

```
#Packages> Install packages> select a mirror and look for the GAD  
package
```



GRAPHS

Scatter plot with regression

If we want to do the regression of the diameter of the canopy vs P50 only for the sunny side, we will have to save the txt file only with that specific data. If we use sunny and shadow side data, we will get the regression of both sides together

```
plot(Diam_canopy ~ P50, data=dat)
```

```
# We name the axis
```

```
plot(Diam_canopy ~ P50, data=dat, ylab=" Canopy diameter (cm)", xlab=" P50 (cm)")
```

```
# We do the regression
```

```
fit1<-lm(Diam_canopy ~ P50, data=dat)
```

```
summary(fit1)
```

```
# We add the regression model to the graph
```

```
abline(fit1)
```



```
# We add the limits of the axis and the interval of the ticks on each axis
```

```
plot(Diam_canopy ~ P50, data=dat, ylab=" Canopy diameter (cm)", xlab=" P50 (cm)", xaxt = "n", yaxt = "n", xlim=c(50,90),ylim=c(2500,3500), type="p" )
```

```
axis(side = 1, at = seq(50,90, 5))
```

```
axis(side = 2, at = seq(2500,3500, 200))
```

```
abline(fit1)
```



Bar chart with standard error

#We calculate the means and standard error (SE) of the cover of the species in each part of the side of the hill.

```
means<-tapply(dat$Cov, dat$Side_hill, mean)
```

```
se<-tapply(dat$ Cov, dat$ Side_hill, sd)/sqrt(tapply(dat$ Cov, dat$ Side_hill, length))
```

#We do the bar chart

```
barx<-barplot(means, ylim=c(0,max(means+3*se)), col="grey",  
ylab="Cover (%)")
```

In the axis “y”, we define the limits (ylim) between 0 and the data range, adding to the mean three times the SE, so the axis is not too close to the SE. This value can be modified according to the range of our data.

#We add the SE

```
arrows(barx, means, barx, means+se, angle=90, code=3, length=0.2)
```



ANALYSIS OF THE VARIANCE

```
# We load GAD package
```

```
library(GAD)
```

```
# We must bear in mind that to use GAD our data must be balanced (all levels / treatments of a factor must have the same number of replicates)
```

```
# We define name of our factors and if they are fixed or random (ours will always be fixed). The defined factor cannot be called exactly the same as the original variable.
```

```
Side <- as.fixed(dat$Side_hill)
```

```
#We define our model:
```

```
mod <- lm(Cov ~ Side, data=dat)
```

```
#We test ANOVA assumptions:
```

```
#1)Homogeneity of variances
```

```
C.test(mod)
```

```
#2)Normality of data
```

```
shapiro.test(dat$Cov)
```



If the data do not comply with the assumptions of the ANOVA, we transform the data (transformation sinus arc is adequate because data are percentages) and recheck the assumptions. If the data still does not fulfill the assumptions, the ANOVA will be performed using a significance level (α ; value to which the p value obtained in ANOVA will be compared to test the null hypothesis) of 0.01, instead of the value of α most frequently used (0.05).

#We do ANOVA

gad(mod)

If there are significant differences in the main effect, we do the post-hoc test to find between which levels differences occur

snk.test(mod, term = "Side")



Class Interval Histograms

As in the regression, if we want to do the histogram only for the sunny side of the hill data, we will have to save the txt file with only these data.

```
hist (dat$Shrub_height, breaks=5, ylim=c(0,5), col="grey", xlab="Height  
(cm)", ylab="Frequency")
```

