

Introduction to **R**

Tomàs Aluja
tomas.aluja@upc.es

September 15, 2008

Install

<http://cran.r-project.org>

A first session:

Clik on **R** icon

```
# a comment
# lets generate 40 observations N(0,1)

x <- rnorm(40)
x <- sort(x) # we need to sort for graphic purposes
later

x # a look at the content of x

# a first summary description

summary(x)
hist(x)

# lets generate the response variable

y <- sin(x*pi)+0.1*rnorm(40)

# lets see the summary of y and its relation with x

summary(y)
hist(y)

plot(x,y)

# creation of a data.frame with x and y

dd <- data.frame(x,y)

# lets si what do we have in the workspace

objects()
```

```

# lets see the class of object dd and its attributes
class(dd)
attributes(dd)

# lets see if a data frame is composed of factors
sapply(dd, is.factor)

# lets verify the min and max of dd
sapply(dd, range)

# linear regression of y function of x
reg1 <- lm(y ~ x, data=dd)
print (reg1)

# lets see the content of object reg1
attributes(reg1)

# lets take the coefficients
reg1$coefficients

# lets see the fit visually
plot(x,y)

# and without closing the plot window
lines(x, reg1$fitted.values)

# lets do a local regression
reg2 <- loess(y ~ x, data=dd)

# and plot the fit in the same display
lines(x,reg2$fitted,col="blue")

# Now, try to do a polinomic regression

# introducing powers of x: reg1 <- lm(y ~ x+x^2,
data=dd) or

# reg1 <- lm(y ~ poly(x,n), data=dd) where n is the
degree of the polnomial

```

```
# What is the degree with a better fit? What is the  
model you will choose to make predictions?
```

```
# Now, we will use this model to predict new  
observations. The problem is that these new  
observations don't follow the model exactly as the  
previous one
```

```
xnew <- rnorm(40, mean=0.3)  
xnew <- sort(xnew)  
ynew <- sin(x*pi*0.9)+0.2*rnorm(40)
```

```
# Use the chosen model to predict this new data
```

```
ypred1 <- predict(your_model,new.data=xnew)
```

```
# Lets see what we have done
```

```
plot(xnew,ynew)  
lines(xnew,ypred1)
```

```
# But we need a numerical measure to assess the  
quality of the fit
```

```
rss <- sum((ynew-ypred1)^2)  
rss
```

```
# Who got the best model?
```

```
# is this measure rss reliable?
```