prac1\_acid.R

statisticallyfit

Sat Mar 3 07:33:08 2018

setwd("/datascience/projects/statisticallyfit/github/R/RStatistics/STAT210 Statistical Modelling and Experimental Design/PRACTICALS/Practical 1 - Intro/")  
  
source('/datascience/projects/statisticallyfit/github/R/RStatistics/PLOTTING.R')  
source('/datascience/projects/statisticallyfit/github/R/RStatistics/FORMULAS.R')  
  
  
library(ggplot2)  
  
# setting digits options:  
options(digits=3, show.signif.stars = FALSE)  
  
## Getting data  
acidData <- read.table("acid.txt", header=TRUE)  
acidData

## acid fungus  
## 1 0 33.3  
## 2 0 31.0  
## 3 3 29.8  
## 4 3 27.8  
## 5 6 28.0  
## 6 6 29.0  
## 7 10 25.5  
## 8 10 23.8  
## 9 20 12.5  
## 10 20 15.5  
## 11 30 11.7  
## 12 30 10.0

## Listing names of variables  
m = print(names(acidData)); m

## [1] "acid" "fungus"

## [1] "acid" "fungus"

names(acidData)

## [1] "acid" "fungus"

## Listing first 3 lines  
head(acidData, 3)

## acid fungus  
## 1 0 33.3  
## 2 0 31.0  
## 3 3 29.8

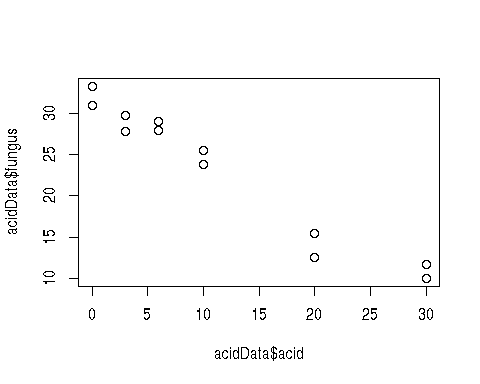
## Fitting Linear Model with Y = fungus, X = acid  
acid.lm <- lm(fungus ~ acid, data=acidData)  
summary(acid.lm)

##   
## Call:  
## lm(formula = fungus ~ acid, data = acidData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.285 -0.907 0.491 1.292 2.414   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 31.7816 0.8327 38.2 3.6e-12  
## acid -0.7498 0.0537 -14.0 6.9e-08  
##   
## Residual standard error: 1.94 on 10 degrees of freedom  
## Multiple R-squared: 0.951, Adjusted R-squared: 0.946   
## F-statistic: 195 on 1 and 10 DF, p-value: 6.89e-08

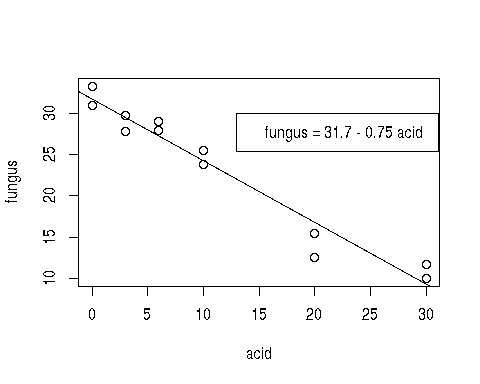
## Anova table  
anova(acid.lm)

## Analysis of Variance Table  
##   
## Response: fungus  
## Df Sum Sq Mean Sq F value Pr(>F)  
## acid 1 733 733 195 6.9e-08  
## Residuals 10 38 4

## Plotting  
# same thing (plotting fungus ~ acid means plotting fungus as function of acid = plot(y ~ x))  
## Adding the least squares line to the plot  
plot(acidData$acid, acidData$fungus)

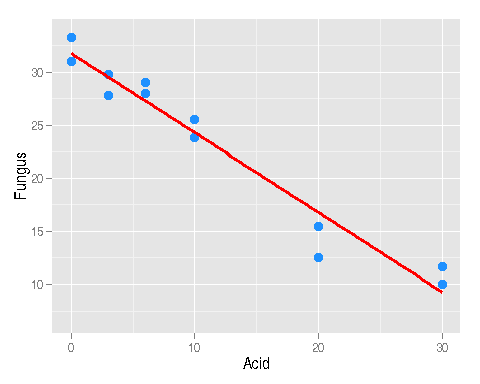


plot(fungus ~ acid, data=acidData)  
abline(acid.lm)  
legend(13, 30, legend="fungus = 31.7 - 0.75 acid ") # coeffs of reg.line



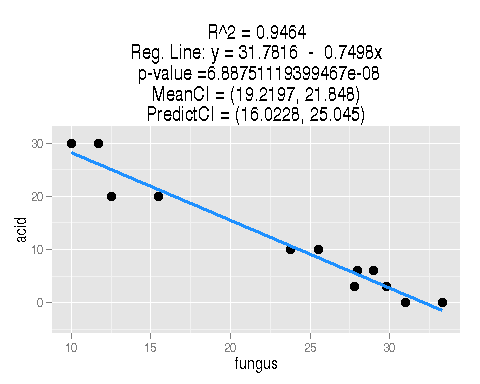
# Plot with ggplot  
g <- ggplot(acidData, aes(x = acid, y = fungus))  
g + geom\_point(shape=19, color="dodgerblue", size=3) +   
 labs(x="Acid", y="Fungus") +  
 stat\_smooth(method="lm", col="red", lwd=1)

## Warning in grid.Call.graphics(L\_polygon, x$x, x$y, index): semi-  
## transparency is not supported on this device: reported only once per page



#geom\_smooth(method="lm", lwd=1, alpha=0.1, fill="red")  
  
# Easier way, all set out  
ggplotRegression(acid.lm, x.value=15)

## Warning in grid.Call.graphics(L\_polygon, x$x, x$y, index): semi-  
## transparency is not supported on this device: reported only once per page



## Prediction intervals  
meanCI(acid.lm, x.value=15, level=0.95)

## [1] 19.2 21.8

predict(acid.lm, new=data.frame(acid=15), interval="confidence", level=0.95)

## fit lwr upr  
## 1 20.5 19.2 21.8

predictCI(acid.lm, x.value=15, level=0.95)

## [1] 16 25

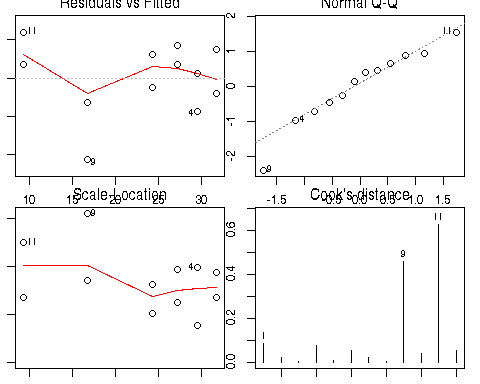
predict(acid.lm, new=data.frame(acid=15), interval="prediction", level=0.95)

## fit lwr upr  
## 1 20.5 16 25

# Confidence intervals for regression parameters  
confint(acid.lm, level=0.95)

## 2.5 % 97.5 %  
## (Intercept) 29.926 33.64  
## acid -0.869 -0.63

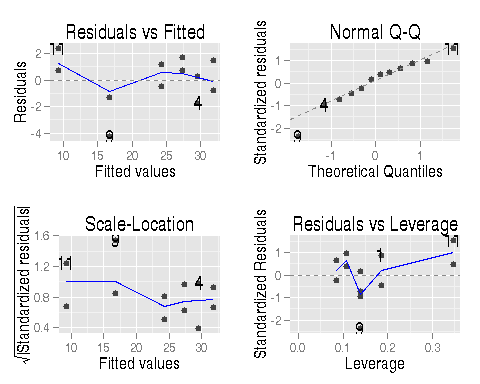
# Diagnostic plots to check model assumptions  
par(mar=c(1,1,1,1)) # to avoid the figure margin error  
par(mfrow=c(2,2))  
plot(acid.lm, which=1:4)



# Diagnostics with ggplot  
library(ggfortify)

## Loading required package: grid  
## Loading required package: scales  
## Loading required package: proto

autoplot(acid.lm) # don't even need par(mfrow) setting



# Shapiro wilk tests normality of residuals  
shapiro.test(acid.lm$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: acid.lm$residuals  
## W = 0.9, p-value = 0.4