Assignment #1

Diego Rios

15 January, 2019

R Markdown

Setting of the working environment

```r
rm(list=ls())
getwd()
```

```
## [1] "/Users/luisdiegoriosreyes/GitHub/QB2019_Rios/2.Worksheets/3.RStudio"
```

```r
setwd("~/GitHub/QB2019_Rios/2.Worksheets/3.RStudio")
```

R as a calculator

```r
#Volume of a cube
l <- 5
l^3
```

```
## [1] 125
```

```r
#Area of a circle with r = 2
r <- 2
pi*r^2
```

```
## [1] 12.56637
```

```r
#Length of the opposite side of a right-triangle
sin(pi/4)*sqrt(2)
```

```
## [1] 1
```

```r
#log (base e) of my favorite number
log(pi)
```

```
## [1] 1.14473
```

Working with Vectors

Basic Features of Vectors

```r
#First part
x <- rnorm(5, mean = 0, sd = 1)
w <- x*14
(x+w)/15
```

```
## [1] -0.3389036  0.4824642 -1.4537640 -0.5273280 -0.9578353
```

```r
k <- 5*x
d <- c(w[1:3],k[1:4])
```

Summary Statistics of Vectors

```r
v <- c(16.4,16,10.1,16.8,20.5,NA,20.2,13.1,24.8,20.2,25,20.5,30.5,31.4,27.1)
max(v,na.rm = T) #max
```

```
## [1] 31.4
```

```r
min(v,na.rm = T) #min
```

```
## [1] 10.1
```

```r
sum(v,na.rm = T) #sum
```

```
## [1] 292.6
```

```r
mean(v,na.rm = T) #mean
```

```
## [1] 20.9
```

```r
median(v,na.rm = T) #median
```

```
## [1] 20.35
```

```r
var(v,na.rm = T) #var
```

```
## [1] 39.44
```

```r
sd(v,na.rm = T)/sqrt(length(v)) # Standard Error of the mean
```

```
## [1] 1.621522
```

Working with Matrices

```r
col1 <- rnorm(5, mean=8,sd=2)
col2 <- rnorm(5,mean=25,sd=10)
tarea1 <- cbind(col1,col2)
```

Answer 1:

The rnorm function generates random numbers according to a normal distribution function. The arguments it has specify the number of observations (n), the mean, and the standard deviation of the desired distribution.

```r
m <- read.table("data/matrix.txt")
m <- t(m)
dim(m)
```
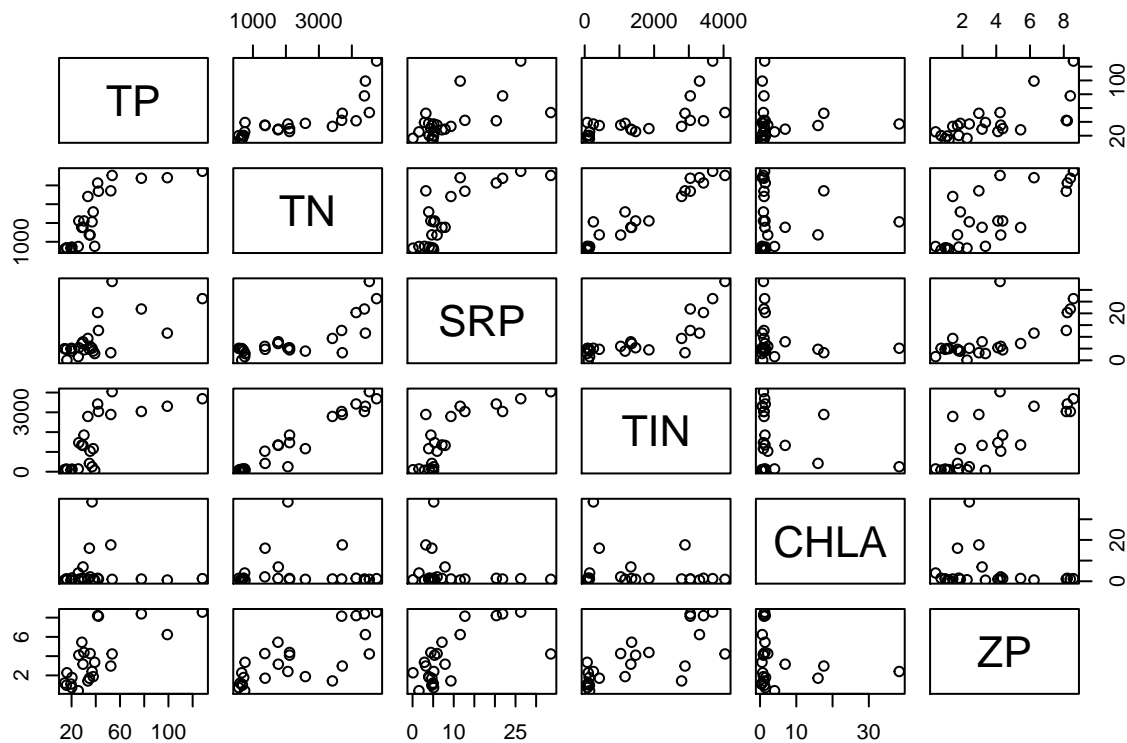
```
## [1]  5 10
```

Answer 2:

5 rows and 10 columns.

```
#Indexing a Matrix
n<-m[-5,-3]
```

Basic data visualization and statistical analysis

```
meso <- read.table("data/zoop_nuts.txt",header = T)
meso.num <- as.matrix(meso[,c(-1,-2)])
pairs(meso.num)
```



```
cor1 <- cor(meso.num)
cor1
```

```
##                TP            TN        SRP         TIN          CHLA
## TP    1.00000000  0.786510407  0.6540957  0.7171143 -0.016659593
## TN    0.78651041  1.000000000  0.7841904  0.9689999 -0.004470263
## SRP   0.65409569  0.784190400  1.0000000  0.8009033 -0.189148017
## TIN   0.71711434  0.968999866  0.8009033  1.0000000 -0.156881463
## CHLA -0.01665959 -0.004470263 -0.1891480 -0.1568815  1.000000000
## ZP    0.69747649  0.756247384  0.6762947  0.7605629 -0.182599904
##               ZP
## TP    0.6974765
## TN    0.7562474
## SRP   0.6762947
## TIN   0.7605629
## CHLA -0.1825999
```

```
## ZP     1.0000000
```

Answer 3:

All variables, except for CHLA, are highly and positively correlated. CHLA is weakly and negatively correlated with the remainder variables.

```r
library(psych)
cor3 <- corr.test(meso.num, method="pearson",adjust = "BH")
print(cor3, digits = 3)
```

```
## Call:corr.test(x = meso.num, method = "pearson", adjust = "BH")
## Correlation matrix
##           TP     TN    SRP    TIN   CHLA     ZP
## TP     1.000  0.787  0.654  0.717 -0.017  0.697
## TN     0.787  1.000  0.784  0.969 -0.004  0.756
## SRP    0.654  0.784  1.000  0.801 -0.189  0.676
## TIN    0.717  0.969  0.801  1.000 -0.157  0.761
## CHLA -0.017 -0.004 -0.189 -0.157  1.000 -0.183
## ZP     0.697  0.756  0.676  0.761 -0.183  1.000
## Sample Size
## [1] 24
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##          TP    TN   SRP   TIN  CHLA    ZP
## TP    0.000 0.000 0.001 0.000 0.983 0.000
## TN    0.000 0.000 0.000 0.000 0.983 0.000
## SRP   0.001 0.000 0.000 0.000 0.491 0.000
## TIN   0.000 0.000 0.000 0.000 0.536 0.000
## CHLA 0.938 0.983 0.376 0.464 0.000 0.491
## ZP    0.000 0.000 0.000 0.000 0.393 0.000
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```

```r
cor4 <- corr.test(meso.num, method="spearman",adjust = "BH")
print(cor4, digits = 3)
```

```
## Call:corr.test(x = meso.num, method = "spearman", adjust = "BH")
## Correlation matrix
##           TP    TN    SRP   TIN   CHLA     ZP
## TP     1.000 0.895  0.539 0.761  0.040  0.741
## TN     0.895 1.000  0.647 0.942  0.021  0.748
## SRP    0.539 0.647  1.000 0.726 -0.064  0.627
## TIN    0.761 0.942  0.726 1.000  0.088  0.738
## CHLA 0.040 0.021 -0.064 0.088  1.000 -0.072
## ZP     0.741 0.748  0.627 0.738 -0.072  1.000
## Sample Size
## [1] 24
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##          TP    TN   SRP   TIN  CHLA    ZP
## TP    0.000 0.000 0.010 0.000 0.914 0.000
## TN    0.000 0.000 0.001 0.000 0.923 0.000
## SRP   0.007 0.001 0.000 0.000 0.884 0.002
## TIN   0.000 0.000 0.000 0.000 0.884 0.000
```

```
## CHLA 0.853 0.923 0.767 0.683 0.000 0.884
## ZP   0.000 0.000 0.001 0.000 0.737 0.000
##
##  To see confidence intervals of the correlations, print with the short=FALSE option
```
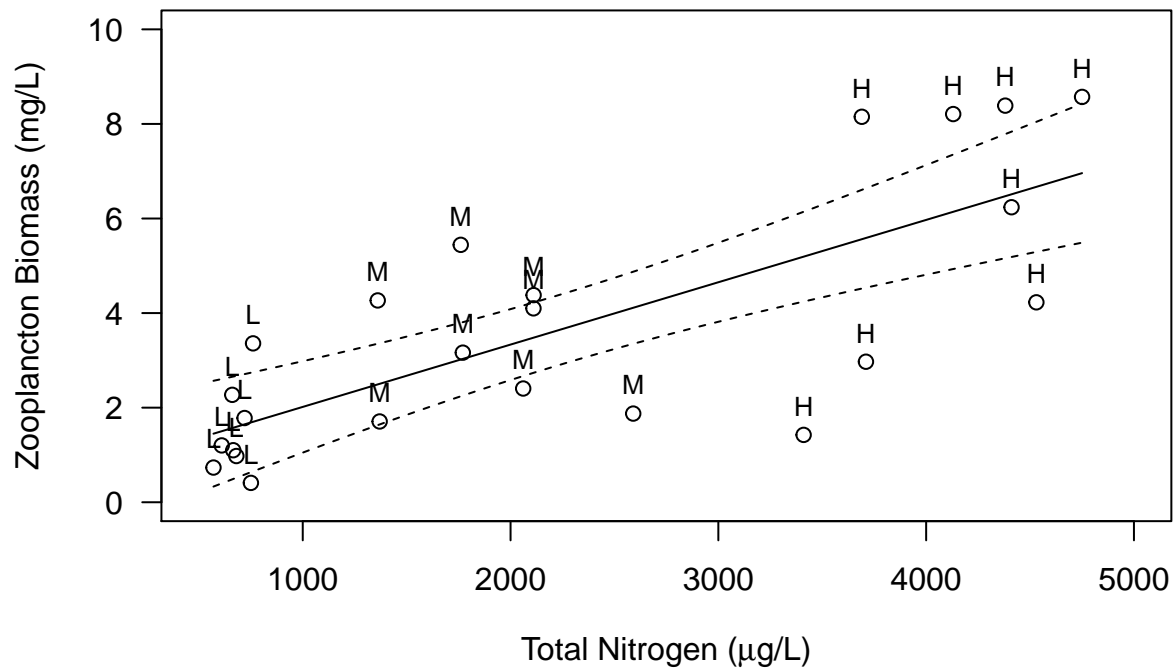
Answer 4

The direction and strength of the correlations don't change with the type of analysis used. One should use non-parametric methods when the data doesn't follow a normal distribution. There doesn't seem to be evidence of false discovery rate given the high significance (p< 0.001) of the correlations. False discovery rate is important because one doesn't want to accept false hypotheses

Linear Regression

```
fitreg <- lm(ZP~TN, data=meso)
summary(fitreg)
```

```
##
## Call:
## lm(formula = ZP ~ TN, data = meso)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7690 -0.8491 -0.0709  1.6238  2.5888
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6977712  0.6496312   1.074    0.294
## TN          0.0013181  0.0002431   5.421 1.91e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.75 on 22 degrees of freedom
## Multiple R-squared:  0.5719, Adjusted R-squared:  0.5525
## F-statistic: 29.39 on 1 and 22 DF,  p-value: 1.911e-05
```

```
plot(meso$TN, meso$ZP, ylim = c(0,10), xlim=c(500,5000),
     xlab= expression(paste("Total Nitrogen (",mu,"g/L)")),
     ylab= "Zooplancton Biomass (mg/L)",las=1)
text(meso$TN, meso$ZP, meso$NUTS, pos=3, cex= 0.8)
newTN <- seq(min(meso$TN), max(meso$TN), 10)
regline <- predict(fitreg, newdata = data.frame(TN=newTN))
lines(newTN, regline)
conf95 <- predict(fitreg, newdata = data.frame(TN=newTN),
                  interval = c("confidence"), level = 0.95, type= "response")
matlines(newTN, conf95[,c("lwr","upr")], type="l", lty = 2, lwd = 1, col = "black")
```

Answer 5

Total nitrogen affects positively Zooplancton Biomass ($p < 0.001$, $R^2 = 0.55$)

ANOVA

```r
NUTS <- factor(meso$NUTS, levels = c('L','M','H'))
zp.means <- tapply(meso$ZP, NUTS, mean)
sem <- function(x){
  sd(na.omit(x))/sqrt(length(na.omit(x)))
}
zp.sem <- tapply(meso$ZP, NUTS, sem)
bp <- barplot(zp.means, ylim = c(0, round(max(meso$ZP), digits = 0)),
              pch = 15, cex = 1.25, las = 1, cex.lab = 1.4, cex.axis = 1.25,
              xlab = "nutrient supply",
              ylab = "zooplankton biomass (mg/L)",
              names.arg = c("low","medium","high"))
arrows(x0 = bp, y0 = zp.means, y1 = zp.means - zp.sem, angle = 90,
       length = 0.1, lwd = 1)
arrows(x0 = bp, y0 = zp.means, y1 = zp.means + zp.sem, angle = 90,
       length = 0.1, lwd = 1)
```
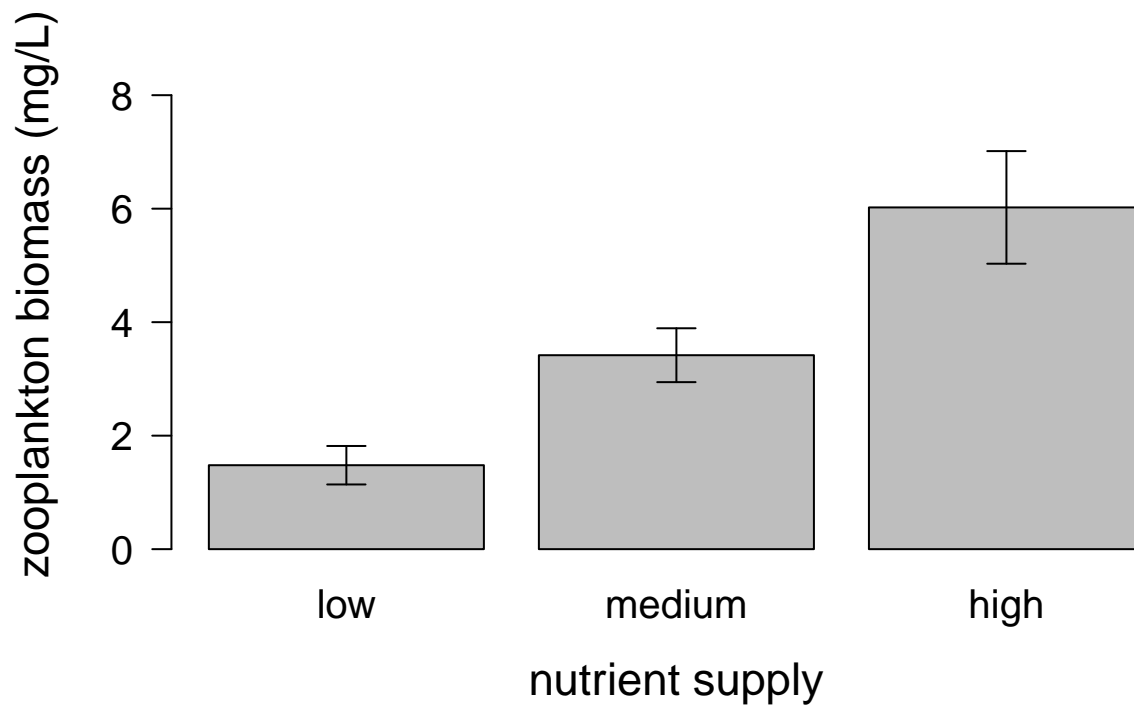
```
fitanova <- aov(ZP ~ NUTS, data = meso)
summary(fitanova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## NUTS         2  83.15   41.58   11.77 0.000372 ***
## Residuals   21  74.16    3.53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Site by Species Matrix

```
meso2 <- read.table("data/zoops.txt",header = T)
meso2 <- meso2[,c(-1,-2)]
site <- as.factor(seq(24))
meso2 <- cbind(site,meso2)
head(meso2)
```

```
##   site  CAL   DIAP  CYCL BOSM   SIMO   CERI NAUP DLUM    CHYD
## 1    1 70.5    0.0  66.1  2.2  417.8  159.8  0.0  0.0   266.9
## 2    2 27.1   19.2 129.6  0.0    0.0   79.4  0.0  0.0   158.7
## 3    3  5.3    8.8  12.7  0.0   73.1  107.5  1.2  0.0  3158.2
## 4    4 79.2   17.9 141.3  3.4    0.0  199.0  0.0  0.0   298.5
## 5    5 31.4    0.0  11.0  0.0  482.0  101.9  0.0  0.0   580.2
## 6    6 22.7  285.1 153.0  0.0  241.5  135.5  1.2  6.6   262.4
```