Assignment 1

##//Exercise 1

##a) Form three separate vectors from scalars 4,5 and 33; -48, 0 and 45; 7, 3 and 1

a<-c(4, 5, 33)

b<-c(-48, 0, 45)

c<-c(7, 3, 1)

##b) Combine vectors into one 3x3 matrix

m=matrix(c(a,b,c), ncol=3, nrow = 3, byrow = TRUE)

m

##c) Name rows of the matrix as r1, r2 and r3. Name columns of the matrix as A, B and C

rownames(m) <- c("r1","r2","r3")

colnames(m) <- c("A", "B", "C")

m

##Exercise 2

##Generate the following vectors with functions rep() and seq()

##a) [0 0 0 0 0 1 1 1 1 1 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4]

rep(0:4, rep(5,5))

rep(0:4, each=5)

##b) [1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5]

##y <- seq(0:4)

##y

rep(1:5, rep(5,5), length.out=25)

##or

rep(seq(0:4), rep(5,5), length.out=25)

#Exercise 3

#Solar radiation was measured in a greenhouse at eight different times. Observations were 11.1, 10.6, 6.3,

#8.8, 10.7, 11.2, 8.9 and 12.2 units.

#a) Save measurements as variable solar.radiation (set value with a vector containing all

# measurements)

solar.radiation <- c(11.1, 10.6, 6.3, 8.8, 10.7, 11.2, 8.9, 12.2)

#b) Define mean, median and variance of variable solar.radiation (with pre-determined R-functions)

mean(solar.radiation)

median(solar.radiation)

var(solar.radiation)

#c) Add value 10 to every observation and name the new vector as sr10. Define mean, median and

#variance of modified vector. Which of the determined values changed from the previous and how

#much?

sr10 = solar.radiation + 10

sr10

mean(sr10)

median(sr10)

var(sr10)

#the mean and the median changed and increased by 10

#from the previous values while the variance remained the same at 3.525

#d) Multiply the original observations by -2 and name the resulting vector as srm2. What happened to

#mean, median and variance?

srm2=solar.radiation \* -2

srm2

mean(srm2)

median(srm2)

var(srm2)

#the mean became a negative value of -19.95, the median also became

#a negative value of -21.3 while the variance remained positive value but higher at 14.1

#e) Variance is generally determined with following formulas:

#Which of the previous two is used in R-function var()?

#copying the formula didnt work well. showed strange characters.

# I checked the answer online and the second is the correct one i.e b.

#Exercise 4

#Generate 25 random integers from interval [1,100] and place them in row-wise in 5x5 matrice.

rand=ceiling(runif(25, 1, 100))

rand

mat=matrix(rand, ncol=5, nrow=5, byrow=T)

mat

#a) Name the rows as r1-r5 and the columns as c1-c5

rownames(mat) <- c("r1","r2","r3","r4","r5")

colnames(mat) <- c("c1","c2","c3","c4","c5")

mat

#b) Determine the sum of numbers in the matrice

sum(mat)

#c) Substract the minimum of the matrice from every other figure in the Matrix

mat2=mat - min(mat)

mat2

#d) Print the figure that is in column 3 of row 4

mat2[4,3]

FOR 215

Assignment 2 (Give in by Thursday 22.9. 10:00)

Exercise 1

a) Create folder ”assignment2” and set it as your working directory through R-console

b) Download file puudata\_20.xls from Moodle and open it in MS Excel. Column name LPM

stands for diameter, ELAVALARAJA stands for the lower limit of living canopy and

LATVUSLEV stands for width of the canopy.

c) Save the data to your working directory as a text file and name it treedata.txt: save as à

other formats à text tab limited (tab as a column delimiter).

d) Read the file in R and name it properly.

e) Check that the data appears right in R (columns, decimal separators, column names)

f) Find out what are the mean diameter and mean height of the trees in the data.

Exercise 2 (Exercise 1 continues)

a) Create a matrix from treedata columns LPM-, PITUUS- and PUULAJI.

(Hint: choose columns one by one and join them together)

b) Calculate basal area (BA) for every tree and ad the results in the matrix as a new column. In

R, π = pi and BA can be calculated from diameter (d) as (pi\*d^2)/4.

(Hint: add a column which is defined with an existing column)

c) Name the columns for diameter, height, tree species and basal area as: DBH, H, S and BA

d) Finally, save the matrix into your working directory as a properly named text file.

Exercise 3

a) Download file puudata\_300.txt from Moodle, read it into R and name it.

b) Cross tabulate the data by latvuskerros (crown layer) and puulaji (tree species) with function

xtabs()

c) Which is the most common tree species in crown layer 1? How many trees of this species

can be found in the whole data set?

d) Which tree species’ relative portion in second crown layer is the highest (amount in 2.

layer/amount in layer 1 and 2)? Consider only those species that are present in both crown

layers.

Exercise 4

Create the following subsets from file puudata\_300.txt:

a) Trees that are measured from plot (KOEALA) 865

b) Trees that are measured from plots 865 and 490

c) Those spruce trees (PUULAJI=2) that belong to second crown layer and are over 10 meters

tall (the height is given in decimeters)

d) Those trees in the first crown layer, whose diameter is over 150 mm and that are not pines

or spruces (species 1 and 2)

RStudio solution

#clear memory list

rm(list=ls())

#set working directory

setwd("C:/Users/oyeda/Desktop/R\_COURSE/assignment2")

#get working directory

getwd()

?read.table

#read the first table

data<-read.table("puudata\_20.txt", header = TRUE, sep="\t")

summary(data) #summary of the data

#calculate the mean diameter of the trees

meanDiameter <- mean(data$LPM)

meanDiameter

#calulate the mean height of the trees

meanHeight <- mean(data$PITUUS)

meanHeight

#Create a matrix from treedata columns LPM-, PITUUS- and PUULAJI.

#DBH is diameter, h is height and s is the species

?matrix

dbh <- data$LPM

h <- data$PITUUS

s <- data$PUULAJI

mat <- matrix(c(dbh, h, s), nrow=length(dbh))

mat

#Calculate basal area (BA) for every tree and

#add the results in the matrix as a new column. BA is basal area.

data$BA <- with(data, (pi\*h^2)/4)

#joining the BA column to the matrix created earlier

mat <- cbind(mat, data$BA)

mat

#create column names

colnames(mat) <- c("DBH", "H", "S", "BA")

mat

#save the data

?write.table

write.table(mat, file = "treeData20.txt", sep="\t", col.names = TRUE, row.names = FALSE)

######

#load the data "puudata\_300.txt" from the directory.

data300 <- read.table("puudata\_300.txt", sep = "\t", header = TRUE)

?xtabs

#s.tab<-table(data300$LATVKERROS, data300$PUULAJI)

layerSpp <-xtabs(~LATVKERROS+PUULAJI, data300)

#Which is the most common tree species in crown layer 1?

#answer: the most common tree species in the crown layer 1 is species 1

#How many trees of this species

#can be found in the whole data set?

#Answer: there are 91 of these trees in the whole data set.

#d) Which tree species’ relative portion in second crown layer

#is the highest (amount in 2.layer/amount in layer 1 and 2)?

#divide the total in layer 2 by the sum of the total in layer 1 and 2

#Consider only those species that are present in both crown layers.

highSpp2 <- layerSpp[2,] / (layerSpp[2,] + layerSpp[1,])

highSpp2

#answer: species 2 has the highest relative portion of crown layer2 in layers 1 and 2

#crown layers, considering those that have both layers present.

#Exercise 4

#Create the following subsets from file puudata\_300.txt:

#a) Trees that are measured from plot (KOEALA) 865

treesA <- subset(data300, data300$KOEALA==865)

treesA

#b) Trees that are measured from plots 865 and 490

treesB<- subset(data300, data300$KOEALA==865 | data300$KOEALA==490)

treesB

#c) Those spruce trees (PUULAJI=2) that belong to second crown layer

#and are over 10 meters tall (the height is given in decimeters).

treesC <- subset(data300, data300$PUULAJI==2 & data300$LATVKERROS==2 & data300$PITUUS>100)

treesC

#d) Those trees in the first crown layer, whose diameter is over 150 mm and

#that are not pines or spruces (species 1 and 2)

treesD<- subset(data300, data300$LATVKERROS==1 & data300$LPM>150 & data300$PUULAJI!=1

& data300$PUULAJI!=2)

treesD

FOR 215

Assignment 3 (Give in by Monday 26.9. 14:00)

Exercise 1

Search information about function read.table with help-function. What are the data types of the following parameters and how do they affect to the read.table function.

a) header: could be TRUE or FALSE. It is a logical value which shows if the file has variables names in the first line. If set to FALSE, the headings would be omitted. If not included, the it determines the value from the format of the file.

b) dec: this is the character used in the file for decimal points. It is a character string that contains just one single-byte character.

c) nrows: this is an integer data type which specifies the maximum number of rows to read in. Negative and other invalid values are ignored.

d) skip: this is an integer data type and it specifies the number of lines of the data file to skip before beginning to read data.

e) fill: it is a logical data type. If TRUE then in case the rows have unequal length, blank fields are implicitly added. See ‘Details.

Exercise 2

Your research data has been corrupted. The first rows have changed to random letters and some. values are also missing from other rows. However, you try to save what’s left because you desperately need to know the mean height of the measured trees.

a) Modify puudata\_rikki.xls (found in Moodle) data into text file and import it into R with

read.table function. Set the parameters so that the first two rows are left unread (the rows

contain random letters) and only the first 20 rows are imported into R.

b) Define the mean height of the trees with mean-function. Modify parameters if necessary.

Exercise 3

Download file puudata.xls from Moodle and transform it into text file. Define the correlation

between diameter and height with cor function

a) for spruces (PUULAJI = 2) in the first crown layer (LATVKERROS = 1)

b) for birches in the first crown layer (PUULAJI = 3).

c) For which of the previous species the diameter explains greater amount of variation in tree

height?

Exercise 4

Using command function, construct your own function that

a) calculates the product of three inserted parameters

b) prints letter ” R” x times when x is functions parameter

Code solution:

**rm(list=ls())**

**setwd("C:/Users/oyeda/Desktop/R\_COURSE/assignment3")**

**getwd()**

**?read.table**

**read.csv()**

**#Modify puudata\_rikki.xls (found in Moodle) data into text file and**

**#import it into R with read.table function.**

**#Set the parameters so that the first two rows are left unread (the rows**

**#contain random letters) and only the first 20 rows are imported into R.**

**data <- read.table("puudata\_rikki.txt", nrows = 20, header = T, skip = 2, sep = ";")**

**data**

**#calculate the mean height of the trees**

**mean.height<-mean(data$PITUUS)**

**mean.height**

**#read the data "puudata" in my working directory**

**data2<- read.table("puudata.txt", header=T, sep = "\t")**

**data2**

**summary(data2)f**

**#Define the correlation between diameter and height with cor function**

**#a) for spruces (PUULAJI = 2) in the first crown layer (LATVKERROS = 1)**

**spruceLayer1 <- subset(data2, data2$PUULAJI==2 & data2$LATVKERROS==1)**

**#cor.test(spruceLayer1$LPM, spruceLayer1$PITUUS, method = c("pearson", "kendall", "spearman"))**

**#correlation between diameter and the height for birches in the**

**#first crown layer (PUULAJI = 3).**

**cor(spruceLayer1$LPM, spruceLayer1$PITUUS,**

**method = c("pearson", "kendall", "spearman"))**

**#b) for birches in the first crown layer (PUULAJI = 3).**

**birchLayer1<-subset(data2, data2$PUULAJI==3 & data2$LATVKERROS==1)**

**cor(birchLayer1$LPM, birchLayer1$PITUUS,**

**method = "pearson")**

**#c) For which of the previous species the diameter explains**

**#greater amount of variation in tree height**

**#Answer: the diameter of spruce explains greater amount of variation**

**#in the treee height.**

**#Using command function, construct your own function that**

**#a) calculates the product of three inserted parameters**

**prod3<- function(x,y,z){**

**pr<-x\*y\*z**

**return(pr)**

**}**

**prod(74,11,89) #test function**

**#b) function prints letter ” R” x times when x is functions parameter**

**printR<-function(x){**

**rprint<- rep("R", x)**

**return(rprint)**

**}**

**printR(21)**

**Assignment 4**

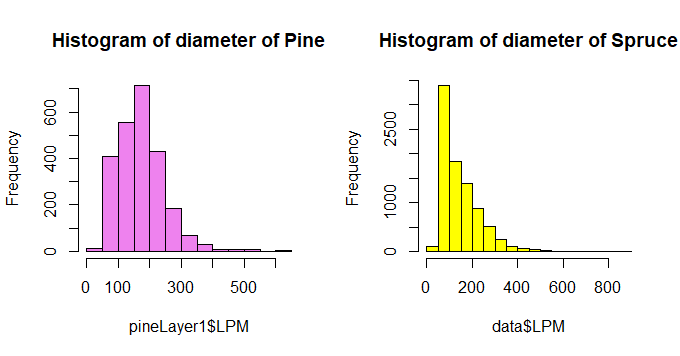
**FOR 215**

**Assignment 4 (Give in by Thursday 29.9. 10:00 am)**

**puudata.xlsx is used in every exercise**

**Exercise 1**

**Form separate histograms for diameters (LPM) of spruces and pines (tree species = PUULAJI, pine= 1 and spruce = 2) in the first crown layer (canopy cover layer = LATVKERROS). Which of the two is visually closer to normal distribution?**



**The pine is visually closer to normal distribution.**

**Exercise 2**

**Plot tree height (PITUUS) as function of tree diameter (LPM) (height on the y-axis and dbh on the**

**x-axis)**

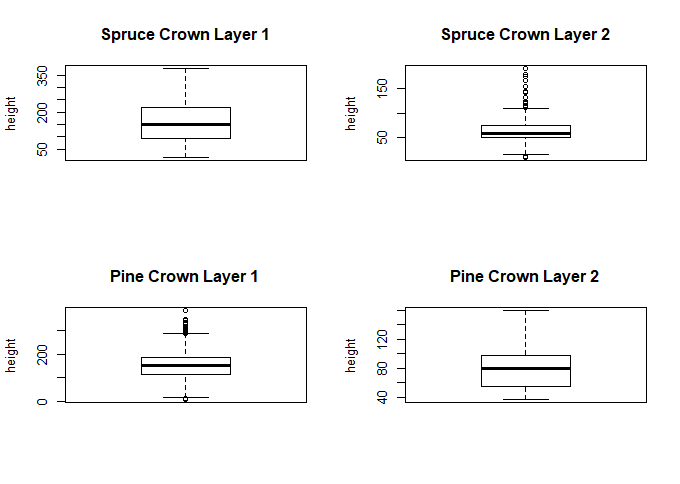
**a) for spruces in the first canopy cover layer**

**b) for pines in the second canopy cover layer**

**Exercise 3**

**Form a box-and-whiskers plot for heights of conifers (species 1 and 2) in first and second crown**

**layer. What can you tell from the difference between height distributions of pines and spruces?**



**The Spruce in the crown layer one appear to be more varying and**

**#less varying in the second crown layer.**

**#Pine, on the other hand, is less varying in the first crown layer but**

**#more varying in the second crown layer.**

**Exercise 4**

**Plot the diameter, height, height of living crown (ELAVALARAJA), and width of the crown**

**(LATVUSLEV) for silver birches (PUULAJI = 3) in the first canopy cover layer with pairs() function.**

**Choose the two variables that seem to have the most linear relationship. Examine the distribution**

**of these two variables by drawing a QQ-plot. What can you tell from the plot?**