**VIT-AP UNIVERSITY, ANDHRA PRADESH**

**CSE2047 – Data Analytics - Lab Sheet : 9**

**Academic year:** 2020-2021  **Branch/ Class:** B.Tech/M.Tech

**Semester:** Fall  **Date:**

**Faculty Name:** Prof. S.Gopikrishnan  **School:** SCOPE

**Student name:** Valiveti Manikanta bhuvanesh **Reg. no.: 19BCD7088**

**LAB 9**

1. **Read the “sample.txt” text file in R. Print the number of characters, number of digits, number of symbols and number of words in the word file.**

df<-read.delim("sample.txt",header = T, sep = " ")

df

nchar(df,allowNA = FALSE, keepNA = NA)

counti=0

countc=0

sapply(df, class)

for (i in names(df)) {

for (j in df[,i]) {

if(is.integer(j)==TRUE){

counti=counti+1 }

if(is.character(j)==TRUE){

countc=countc+1

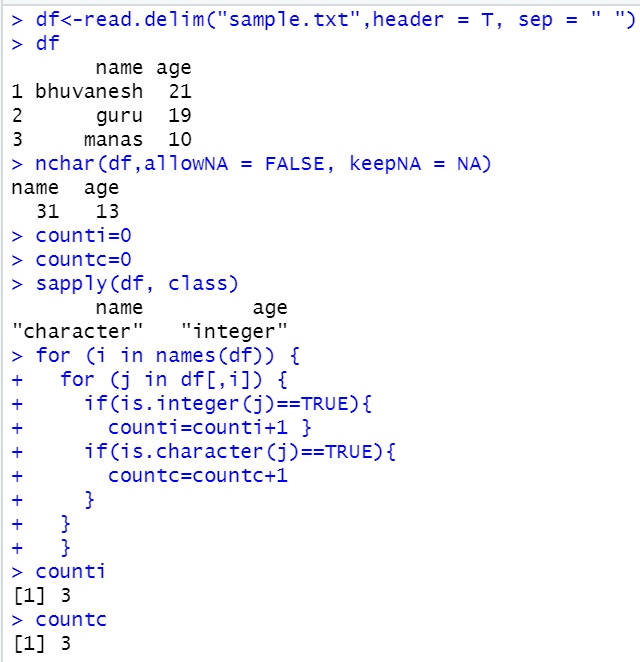
}

}

}

counti

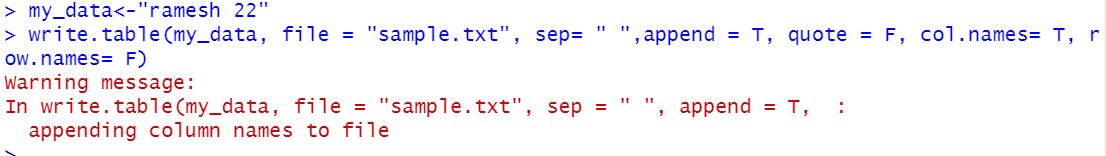
countc



1. **Append a new row to the “sample.txt” text file.**

my\_data<-"ramesh 22"

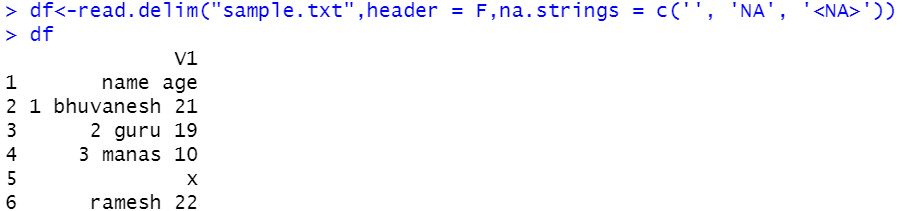
write.table(my\_data, file = "sample.txt", sep= " ",append = T, quote = F, col.names= T, row.names= F)



1. **How to read this text file with missing values?**

df<-read.delim("sample.txt",header = F,na.strings = c('', 'NA', '<NA>'))

df



1. **Read the Iris dataset from csv file and write into a xlsx file in R.**

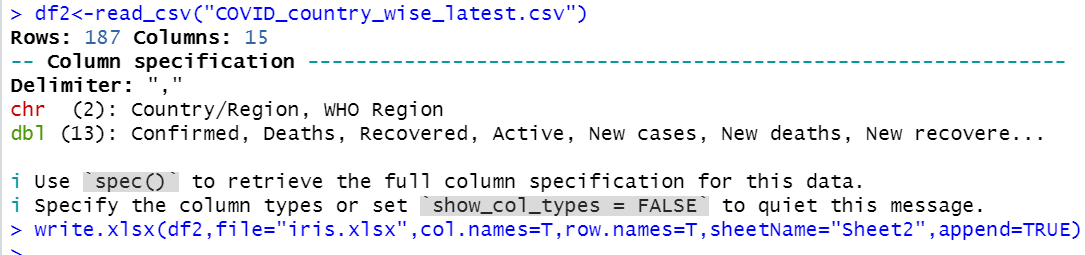
write.xlsx(iris,file="iris.xlsx",col.names=T,row.names=T,sheetName="Sheet1")



1. **Write the covid data set csv file dataset in the second sheet of the xlsx file created for the question2.**

df2<-read\_csv("COVID\_country\_wise\_latest.csv")

write.xlsx(df2,file="iris.xlsx",col.names=T,row.names=T,sheetName="Sheet2",append=TRUE)

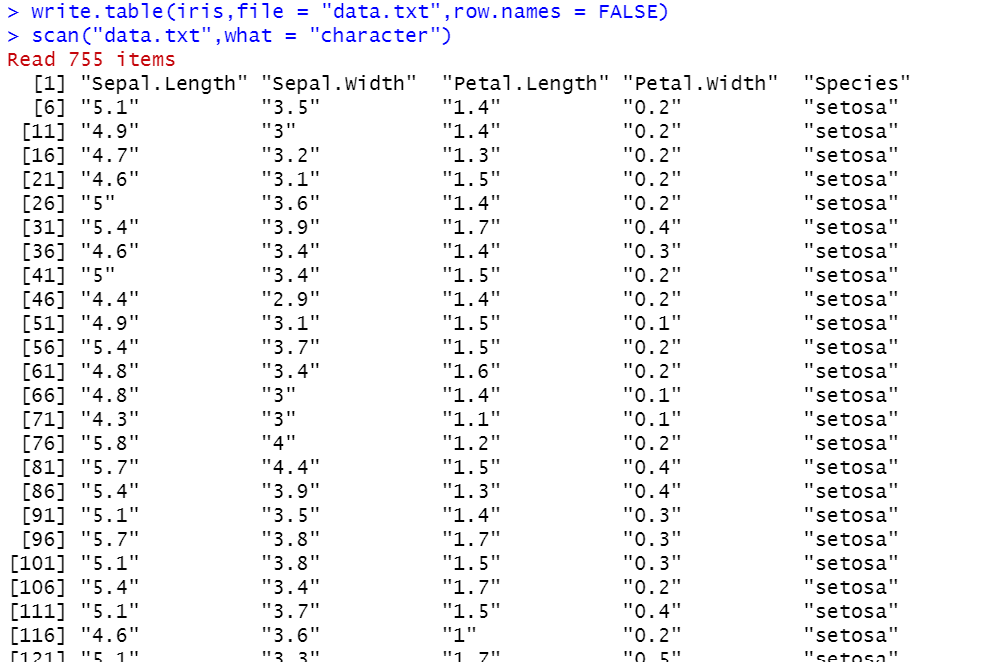


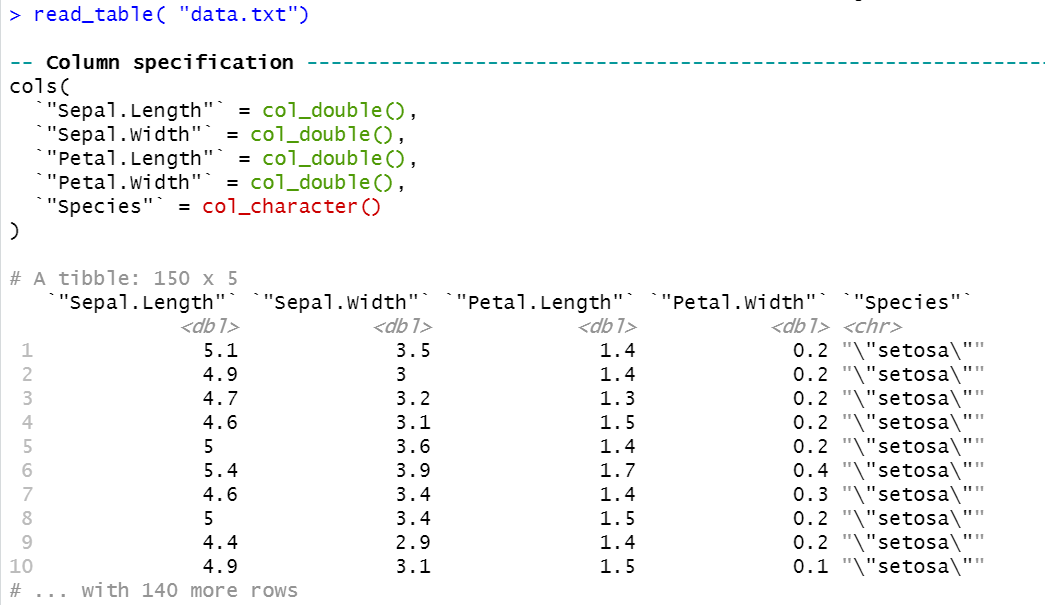
1. **Differentiate scan() and read\_table using Iris data set.**

write.table(iris,file = "data.txt",row.names = FALSE)

scan("data.txt",what = "character")

read\_table( "data.txt")





1. **Use iris dataset and plot the normal distribution on all the numerical columns**

x1 <- iris$Sepal.Length

x2 <- iris$Sepal.Width

x3 <- iris$Petal.Length

x4 <- iris$Petal.Width

* 1. **dnorm()**

v1 <- dnorm(x1, mean = mean(x1), sd = sd(x1))

plot(x1, v1)

v2 <- dnorm(x2, mean = mean(x2), sd = sd(x2))

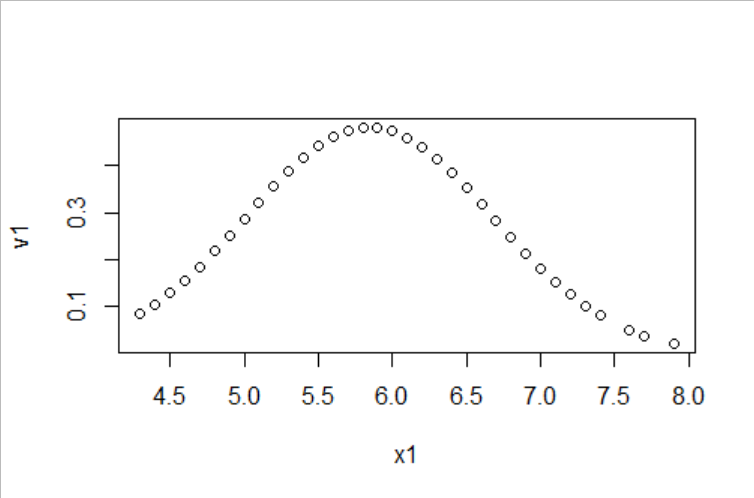
plot(x2, v2)

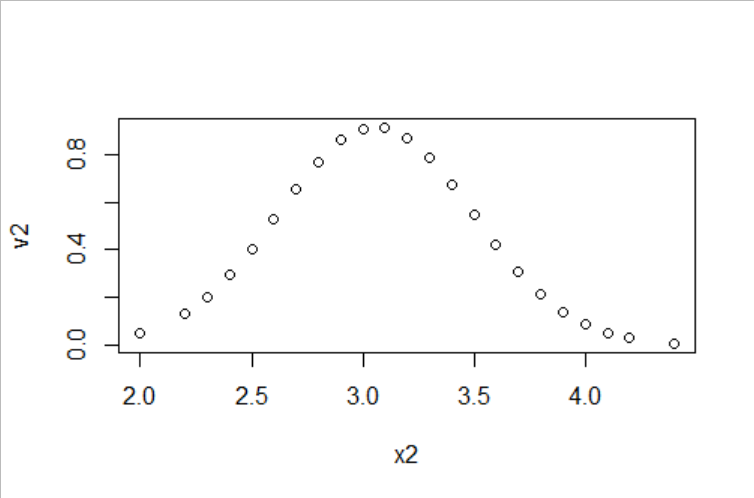
v3 <- dnorm(x3, mean = mean(x3), sd = sd(x3))

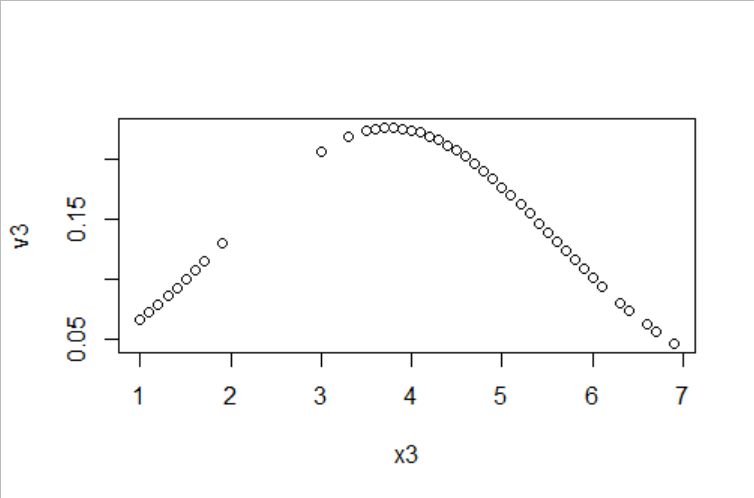
plot(x3, v3)

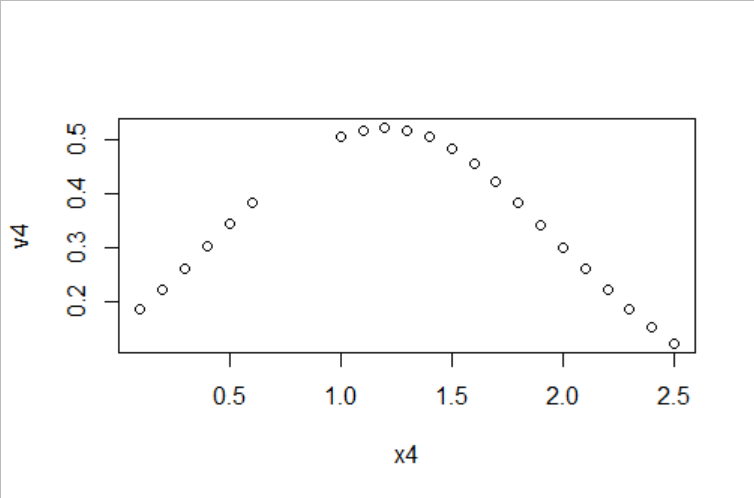
v4 <- dnorm(x4, mean = mean(x4), sd = sd(x4))

plot(x4, v4)









* 1. **pnorm()**

v1 <- pnorm(x1, mean = mean(x1), sd = sd(x1),lower.tail = FALSE)

plot(x1, v1)

v2 <- pnorm(x2, mean = mean(x2), sd = sd(x2),lower.tail = FALSE)

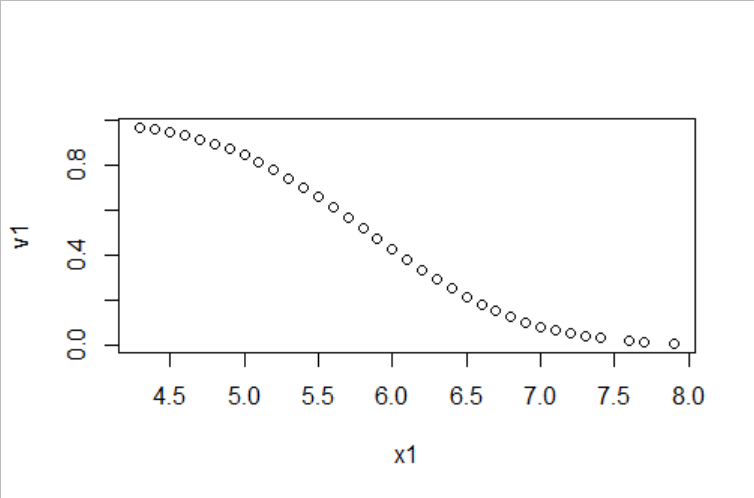
plot(x2, v2)

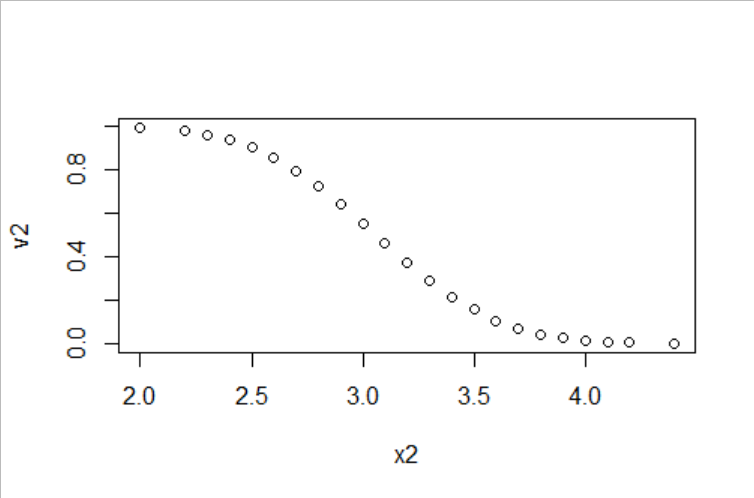
v3 <- pnorm(x3, mean = mean(x3), sd = sd(x3),lower.tail = FALSE)

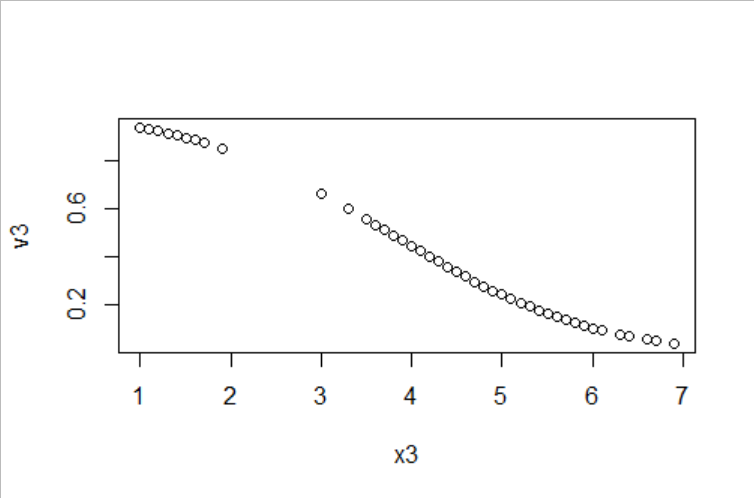
plot(x3, v3)

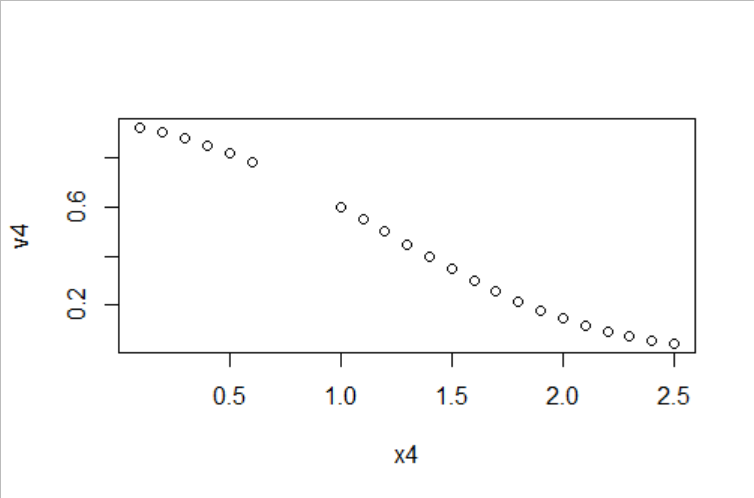
v4 <- pnorm(x4, mean = mean(x4), sd = sd(x4),lower.tail = FALSE)

plot(x4, v4)









* 1. **qnorm()**

x <- seq(0, 1, by = 0.01)

v1 <- qnorm(x, mean = mean(x1), sd = sd(x1))

plot(v1,type = "l")

v2 <- qnorm(x, mean = mean(x2), sd = sd(x2))

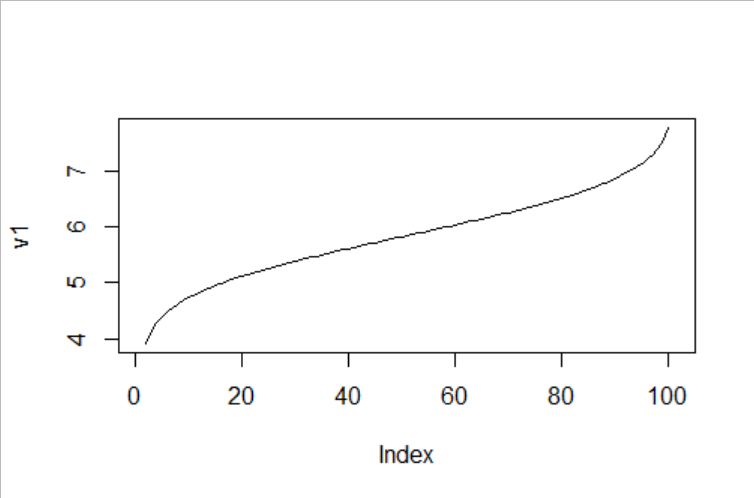
plot(v2,type = "l")

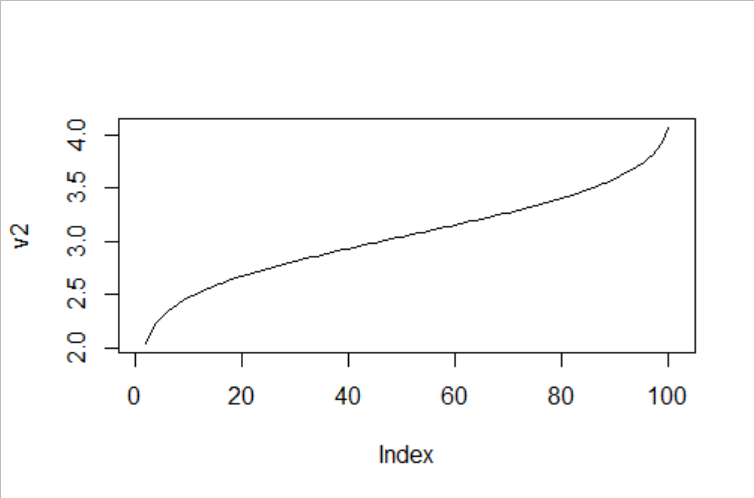
v3 <- qnorm(x, mean = mean(x3), sd = sd(x3))

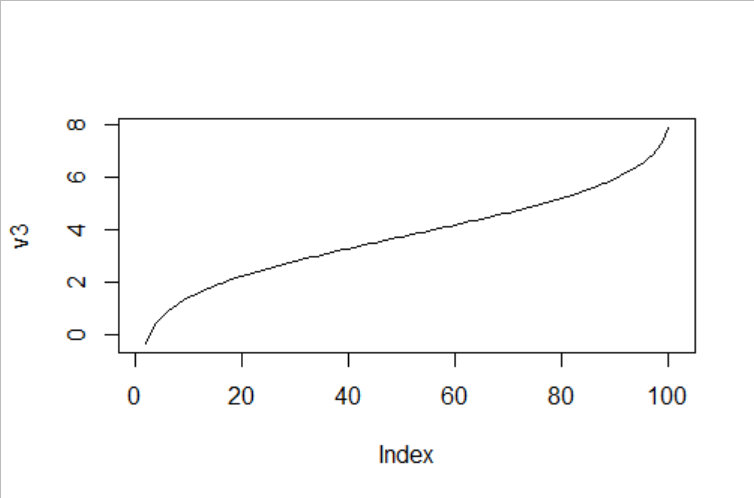
plot(v3,type = "l")

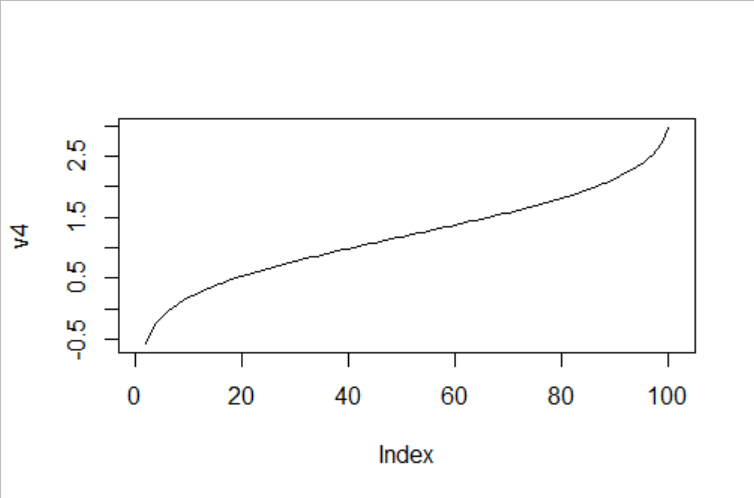
v4 <- qnorm(x, mean = mean(x4), sd = sd(x4))

plot(v4,type = "l")









* 1. **rnorm()**

v1 <- rnorm(x1, mean = mean(x1), sd = sd(x1))

hist(v1)

v2 <- rnorm(x2, mean = mean(x2), sd = sd(x2))

hist(v2)

v3 <- rnorm(x3, mean = mean(x3), sd = sd(x3))

hist(v3)

v4 <- rnorm(x4, mean = mean(x4), sd = sd(x4))

hist(v4)

