STATISTICS ASSIGNMENT 2

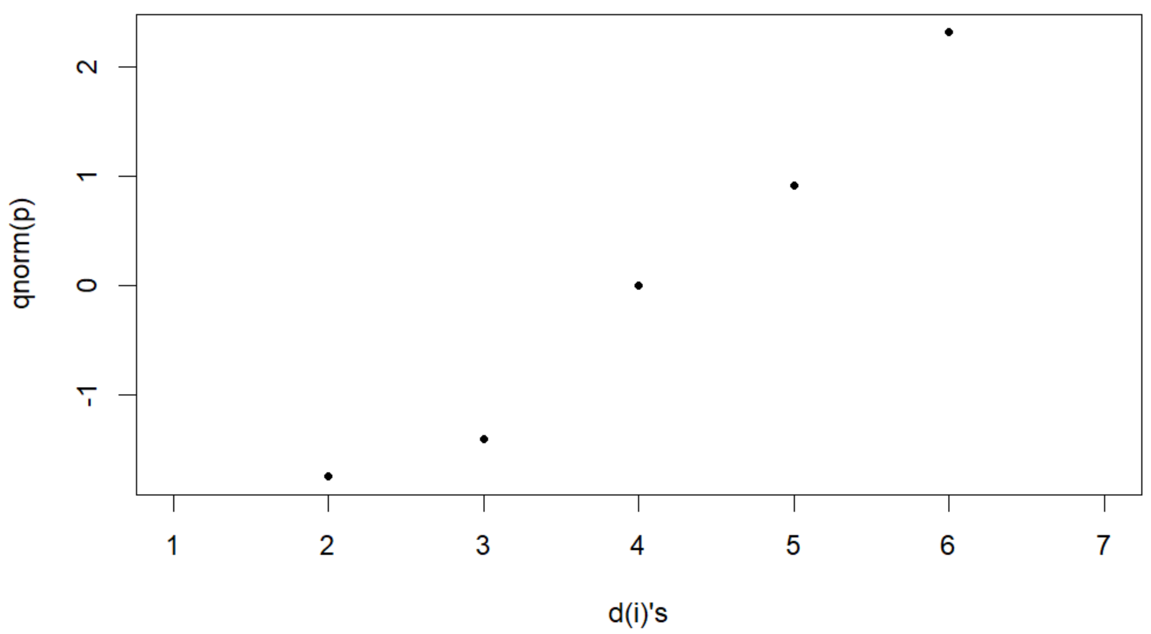
- Aman Singh, BS-2308

Ques 1.

```
#Use the raw data for probit analysis provided in the class webpage, and follow the steps of Finney to produce a linear scatterplot.


library(MASS) # installing MASS library for using functions like qnorm
a = read.table('C:\\Users\\91836\\OneDrive\\Desktop\\Notes\\Stat (AC)\\codes in R\\probit.txt') # reading probit model file
x = as.numeric(a$V3) #extracting the data of daeds and removing non numeric values with NA
y = as.numeric(a$V2) #extracting the size of data and removing non numeric values with NA
x = x[!is.na(x)] #removing NA values
y = y[!is.na(y)] #  removing NA values
plot(qnorm(x/y),pch=20, xlab = "d(i)'s", ylab = 'qnorm(p)') # plotting probit function of dead/size for each doses using qnorm(x) function

#probit model data
#    V1   V2   V3
#1 dose size dead
#2   12  100    0
#3   13  100    4
#4   14  100    8
#5   15  100   50
#6   16  100   82
#7   17  100   99
#8   18  100  100
```



Ques 2.

```r
#Prussian Horse Kick data, Fit a Poisson(λ) distribution to it. The
#Poisson distribution is a discrete distribution with PMF:
#Report the estimated value of λ. Make a table with two columns, one
#column for the fitted Poisson PMFs, and one for the observed relative frequencies.

library(MASS) # installing MASS library
a = read.csv('C:\\Users\\91836\\OneDrive\\Desktop\\Notes\\Stat (AC)\\codes in R\\horsekickdata.csv', head = T) # reading the prussian horse kick data
y = as.numeric(a$y) # extracting the data of death of people from the table and removing non numeric data abd replacing it with NA
y = y[!is.na(y)] # removing NA values
f = fitdistr(y,"Poisson") # estimating paramter for poisson distribution
l = f$estimate[["lambda"]] # estimating value of lambda for poisson distribution
x = 0:max(y) # denifing a vector with the possible values of number of dead people
pmf=dpois(x,l) # finding and creating poisson probability density table
ob_f= table(y)/length(y) # observed frequency density table
table = data.frame('number of deaths'= x,'Observed_Relative_Frequency' = ob_f, 'Fitted_Poisson_PMF' = pmf) # creating table of observed vs estimated probability density of poissson distribution
plot(x, ob_f, type = "b", pch = 16, col = "blue", ylim = c(0, max(c(ob_f, pmf))), xlab = "Number of Deaths", ylab = "Relative Frequency", main = "Observed vs. Poisson PMF") # plotting observed probability density
lines(x, pmf, type = "b", pch = 16, col = "red") # plotting estimated probability density of poissson distribution
legend("topright", legend = c("Observed", "Poisson PMF"), col = c("blue", "red"), lty = c(1, 1), pch = c(16, 16)) # describing legend to dintinguish between the two plots

#estimated value of lambda comes out as 0.7


#table
#  number.of.deaths Observed_Relative_Frequency.y Observed_Relative_Frequency.Freq Fitted_Poisson_PMF
#1                0                             0                      0.514285714        0.496585304
#2                1                             1                      0.325000000        0.347609713
#3                2                             2                      0.114285714        0.121663399
#4                3                             3                      0.039285714        0.028388127
#5                4                             4                      0.007142857        0.004967922
```