Estimation of Inland Fish Production of Karnataka for the year 2016-17 using 'R'

April 17, 2017

Contents

Resource and Method:

Inland fish production is estimated by growth and survival of fish seed stocked to different water bodies for cultured fish and area utilization and productivity of different water bodies for natural fish production. District wise resources and fish seed stocking data from field office is collected and compiled in an excel sheet. This work book is used in this 'R' programme for estimation of inland fish production.

R is an open source programming language and software environment for statistical computing and graphics that is supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. Polls, surveys of data miners, and studies of scholarly literature databases show that R's popularity has increased substantially in recent years. R and its libraries implement a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, and others. R is easily extensible through functions and extensions, and the R community is noted for its active contributions in terms of packages.

```r
library(XLConnect)
inland_xl_data<-loadWorkbook("inland_area_stocking_data.xlsx")
#Load excel work book containing resource wise fish seed stocked and Area availabe in each District
```

Read data from Sheet1 to data frame object "resource_data"

```r
resource_data<-readWorksheet(inland_xl_data, sheet = 1 , startRow =1 , startCol =1,
                             endRow =31 , endCol =16, header = TRUE)

str(resource_data)  # structure of resource/stocking data
```

```
## 'data.frame':    30 obs. of  16 variables:
##  $ District    : chr  "Bagalkote" "Bangalore (Urban)" "Bangalore (Rural)" "Belgaum" ...
##  $ stock_R_Y3  : num  5070000 580000 830000 4998000 1611000 ...
##  $ stock_R_Y2  : num  0 363000 0 2308000 3343000 ...
##  $ stock_R_Y1  : num  3380000 93000 0 1630000 3200000 ...
##  $ stock_MJT_Y2: num  1328000 8257000 2386000 83000 5318000 ...
##  $ stock_MJT_Y1: num  1275000 3123200 2237000 3235000 2111000 ...
##  $ stock_MNT_Y1: num  307000 914700 825000 452000 650000 ...
##  $ stock_PVT_Y1: num  119000 127500 25000 706000 5310000 ...
##  $ BW_P        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ MNT_A       : num  27 1905 1982 832 157 ...
##  $ MJT_A       : num  2158 7390 12014 3027 11132 ...
##  $ SR_A        : num  0 0 0 360 280 277 0 885 364 320 ...
##  $ LR_A        : num  0 2322 0 21368 39863 ...
##  $ WL_A        : num  0 91 200 2955 3189 ...
##  $ River_km    : num  206 0 62 271 161 119 227 218 158 109 ...
##  $ Estuary_a   : num  0 0 0 0 0 0 0 0 0 0 ...
```

Field Names used in the source work sheet and thier description is provided below for information and clarity.

```r
desc<-readWorksheet(inland_xl_data, sheet = 1 , startRow = 1 , startCol = 18 ,
                    endRow =17 , endCol =19, header = TRUE)

desc
```

```
##      Field.Name                                 Description
## 1      District                        Name of the District
## 2    stock_R_Y3          Fish stocked to Reservoirs in 2014-15
## 3    stock_R_Y2          Fish stocked to Reservoirs in 2015-16
## 4    stock_R_Y1          Fish stocked to Reservoirs in 2016-17
## 5  stock_MJT_Y2        Fish stocked to Major Tanks  in 2015-16
## 6  stock_MJT_Y1        Fish stocked to Major Tanks  in 2016-17
## 7  stock_MNT_Y1        Fish stocked to Minor Tanks  in 2015-16
## 8  stock_PVT_Y1        Fish stocked to Minor Tanks  in 2016-17
## 9          BW_P          Brackish Water Fish Production in MT
## 10        MNT_A               Minor Tanks Area Available (Ha)
## 11        MJT_A               Major Tanks Area Available (Ha)
## 12         SR_A Small Resrovir (<500 ha) Area Available (Ha)
## 13         LR_A Large Resrovir (>500 ha) Area Available (Ha)
## 14         WL_A                      Water Logged Area (Ha)
## 15     River_km                  River length Available (km)
## 16    Estuary_a                Estuary Area    Available (Ha)
```

Important Assumptions for Estimation

Culture production:

```
assumption<-readWorksheet(inland_xl_data, sheet = 2 , startRow =1 , startCol =1 ,
                          endRow =8 , endCol =5, header = TRUE)
#Read data from Sheet2 to data frame object "assumption"
assumption$cf<- assumption$Growth * assumption$Harvest * assumption$Survival

assumption  # Conversion Factor (cf) for assessing cultured fish production
```

```
##        Resource year Growth Harvest Survival    cf
## 1  Major tanks    Y1    0.7     0.6      0.5 0.210
## 2  Major tanks    Y2    1.7     0.4      0.5 0.340
## 3   Minor tanks    Y1    0.6     1.0      0.6 0.360
## 4 Private Ponds    Y1    1.0     1.0      0.8 0.800
## 5   Reservoirs    Y1    0.7     0.5      0.7 0.245
## 6   Reservoirs    Y2    1.7     0.3      0.6 0.306
## 7   Reservoirs    Y3    2.5     0.2      0.6 0.300
```

Functions for estimation of Culture Fish Production in MT based on above assumtions

Three factors are important for estimation of culture fish production: Survival proportion (S), Harvest proportion (H) and Growth (G) in kg are used to calculate the conversion factor (cf) which is multiplied to fish seed stocked for the respective years (Y1, Y2, Y3)

Function for estimation of 'Reservoir Culture Production'

```
cp_R_f<-function(y1,y2,y3) {

  cf1<- assumption[5,6] # cf is picked from respective row and column of 'asuumption' table
  cf2<- assumption[6,6]
  cf3<- assumption[7,6]

  ((y3 *cf3) + (y2 * cf2) + (y1 * cf1)) / 1000

  }
```

Function for estimation of 'Major Tank Culture Production'

```
cp_MJT_f<-function(y1,y2) {

   cf1<- assumption[1,6]
   # cf is picked from respective row and column of 'asuumption' table
   cf2<- assumption[2,6]

  ((y2 * cf2) + (y1 * cf1))  / 1000

  }
```

Function for estimation of 'Minor Tank Culture Production'

```
cp_MNT_f<-function(y1) {

   cf1<- assumption[3,6]
   # cf is picked from respective row and column of 'asuumption' table

  (y1 * cf1) / 1000

  }
```

Function for estimation of 'Private Ponds Culture Production'

```
cp_PVT_f<-function(y1) {

   cf1<- assumption[4,6]
   # cf is picked from respective row and column of 'asuumption' table

  (y1 *cf1) / 1000

  }
```

Reservoir production for 2016-17

```
R_CP<- cp_R_f(y1 = resource_data$stock_R_Y1, y2 = resource_data$stock_R_Y2,
              y3 = resource_data$stock_R_Y3 )

 # Reserovir culture fish production in MT for each District

sum(R_CP) # Total
```

```
## [1] 24360.65
```

Major Tank Fish Production for 2016-17

```
MJT_CP<- cp_MJT_f(y1 = resource_data$stock_MJT_Y1,
                  y2 = resource_data$stock_MJT_Y2)

# Major Tank culture fish production in MT for each  District

sum(MJT_CP) # Total
```

```
## [1] 81211.62
```

Minor Tank Fish Production for 2016-17

```
MNT_CP<- cp_MNT_f(y1 = resource_data$stock_MNT_Y1)

# Minor Tank culture fish production in MT for each  District
```

```r
sum(MNT_CP) # Total
```

```
## [1] 10024.42
```

Private Ponds Fish Production for 2016-17

```r
PVT_CP<- cp_PVT_f(y1 = resource_data$stock_PVT_Y1)

# Private culture fish production in MT for each  District

sum(PVT_CP) # Total
```

```
## [1] 21939.99
```

Brackish Water cultuteFish Production for 2016-17

```r
BW_CP<- resource_data$BW_P
# Directly obtained from district as in source work sheet


sum(BW_CP) # Total
```

```
## [1] 3396.25
```

A) By Cultre : Total Fish Production in MT

```r
TFP_CP<-sum(BW_CP) + sum(R_CP) +sum(MJT_CP) + sum(MNT_CP) + sum(PVT_CP)

TFP_CP
```

```
## [1] 140932.9
```

Natural Fish Production : Important Assumptions

Fish production potential per Ha for estmation of Natural Fish Production based on area or extant of of resources available. Minor tanks -10% area X 40 Kg. per Ha. Major tanks - 45% area X 45 Kg. per Ha. Reservoirs < 500 Ha. -50% area X 45 Kg. per Ha. Reservoirs > 500 Ha. - 60% area X 45 Kg. per Ha. Waterlogged area - 10 Kg. per Ha. Rivers - 150 Kg. per Km. Esturies - 750 Kg. per Ha.

Conversion Factor (cf) for assessing natural fish production

```r
a_np<-readWorksheet(inland_xl_data, sheet = 3 , startRow =1 , startCol =1 ,
                    endRow =8 , endCol = 3, header = TRUE)

a_np
```

```
##          Resource Area_factor Productivity
## 1 Large_Reservoir        0.60           45
## 2 Small_Reservoir        0.50           45
## 3    Water Logged        1.00           10
## 4      Major Tank        0.45           45
## 5      Minor Tank        0.10           40
## 6           River        1.00          150
## 7         Estuary        1.00          750
```

Function for estimation of Natural Fish Production in MT based on the above assumtions

```
np_f<-function(x, perc_area, productivity) {


  (x * perc_area * productivity) /1000

}
```

Natural fish production in MT for Minor tanks -10% area X 40 Kg. per Ha.

```
MNT_NP<-np_f(resource_data$MNT_A, a_np[5,2], a_np[5,3])

sum(MNT_NP)
```

```
## [1] 209.728
```

Natural fish production in MT for Major tanks - 45% area X 45 Kg. per Ha.

```
MJT_NP<-np_f(resource_data$MJT_A, a_np[4,2], a_np[4,3])

sum(MJT_NP)
```

```
## [1] 4843.699
```

Natural fish production in MT for Reservoirs < 500 Ha. -50% area X 45 Kg. per Ha.

```
SR_NP<-np_f(resource_data$SR_A, a_np[2,2], a_np[2,3])

sum(SR_NP)
```

```
## [1] 182.295
```

Natural fish production in MT for Reservoirs > 500 Ha. - 60% area X 45 Kg. per Ha..

```
LR_NP<-np_f(resource_data$LR_A, a_np[1,2], a_np[1,3])

sum(LR_NP)
```

```
## [1] 5475.789
```

Natural FP Natural fish production in MT for Waterlogged area - 10 Kg. per Ha.

```r
WL_NP<-np_f(resource_data$WL_A, a_np[3,2], a_np[3,3])

sum(WL_NP)
```

```
## [1] 630.91
```

Natural fish production in MT for Rivers - 150 Kg. per Km

```r
River_NP<-np_f(resource_data$River_km, a_np[6,2], a_np[6,3])

sum(River_NP)
```

```
## [1] 877.95
```

Natural fish production in MT for Esturies - 750 Kg. per Ha.

```r
Estuary_NP<-np_f(resource_data$Estuary_a, a_np[7,2], a_np[7,3])

sum(Estuary_NP)
```

```
## [1] 5410.5
```

B) By Natural Production : Total Fish Production

```r
TFP_NP<-sum(MJT_NP) + sum(LR_NP) +sum(MNT_NP) + sum(SR_NP) +
        sum(WL_NP) + sum(River_NP) + sum(Estuary_NP)

TFP_NP
```

```
## [1] 17630.87
```

Total Fish Production of the State for the Year 2016-17

A) Culture Production + B) By Natural Production

```r
TFP<- TFP_CP + TFP_NP

TFP
```

```
## [1] 158563.8
```

District-wise Fish Production in MT and its Value Rs in lakh for 2016 -17

District-wise fish production in MT and its value (Rs in lakh) for 2016-17 is computed and tabluted below;

```
DIST_NP<- MJT_NP + LR_NP + MNT_NP + SR_NP + WL_NP + River_NP + Estuary_NP

DIST_CP<- BW_CP + R_CP + MJT_CP + MNT_CP + PVT_CP

DISTRICT<-resource_data[,1]

Total_FP<- DIST_NP +DIST_CP

Value_Lakh_Rs<-(Total_FP - BW_CP) * 0.85 + (BW_CP * 3.5)


DIST_TFP<-data.frame(DISTRICT, Total_FP, Value_Lakh_Rs)
```

District-wise Fish Production in MT and its Value Rs in lakh for 2016 -17

```
print(DIST_TFP)
```

```
##             DISTRICT  Total_FP Value_Lakh_Rs
## 1          Bagalkote  3348.797     2846.4779
## 2  Bangalore (Urban)  4423.279     3759.7867
## 3  Bangalore (Rural)  2109.529     1793.0994
## 4            Belgaum  4759.949     4045.9564
## 5            Bellary 10388.380     8830.1230
## 6              Bidar  1876.587     1595.0994
## 7            Bijapur  4743.626     4032.0819
## 8       Chamarajnagar  2016.983     1714.4355
## 9         Chikmagalur  5408.599     4597.3094
## 10        Chitradurga  3767.799     3202.6289
## 11   Dakshina Kannada  1203.148     1420.1758
## 12          Davangere  7531.668     6401.9176
## 13            Dharwad  1942.719     1651.3109
## 14              Gadag   757.079      643.5171
## 15           Gulbarga  5023.713     4270.1557
## 16             Hassan  9577.883     8141.2005
## 17             Haveri  2205.835     1874.9598
## 18             Kodagu  3517.451     2989.8333
## 19              Kolar  1608.413     1367.1509
## 20             Koppal  2218.970     1886.1243
## 21             Mandya 12924.390    10985.7317
## 22             Mysore  7331.936     6232.1454
## 23            Raichur 10094.660     8580.4610
## 24            Shimoga 17442.690    14826.2863
## 25             Tumkur  9970.737     8475.1262
## 26              Udupi  3570.670     7871.3195
## 27     Uttara Kannada  7687.760    10300.9087
## 28       Chikballapur  4308.628     3662.3337
## 29         Ramanagara  5019.483     4266.5606
```

```
## 30          Yadagiri  1782.447     1515.0796
```

State Inland Fish Production (in MT) and its Value (Rs in Lakh) for the year 2016-17:

```
Fish_Produnction<- sum(DIST_TFP[,2])
Value<- sum(Value_Lakh_Rs)
Total<-data.frame(Fish_Produnction,Value)
Total
```

```
##   Fish_Produnction    Value
## 1         158563.8 143779.3
```

Write and Save output

New workbook "inland__production.xlsx" is saved in working Directory containing,

Resource Data in Sheet1 Assumptions for Culture Production in Sheet2 Assumptions for Natural Production in Sheet3 District Wise Fish Production nad its value in Sheet4 and State's Total Fish Production and its Value in Sheet5

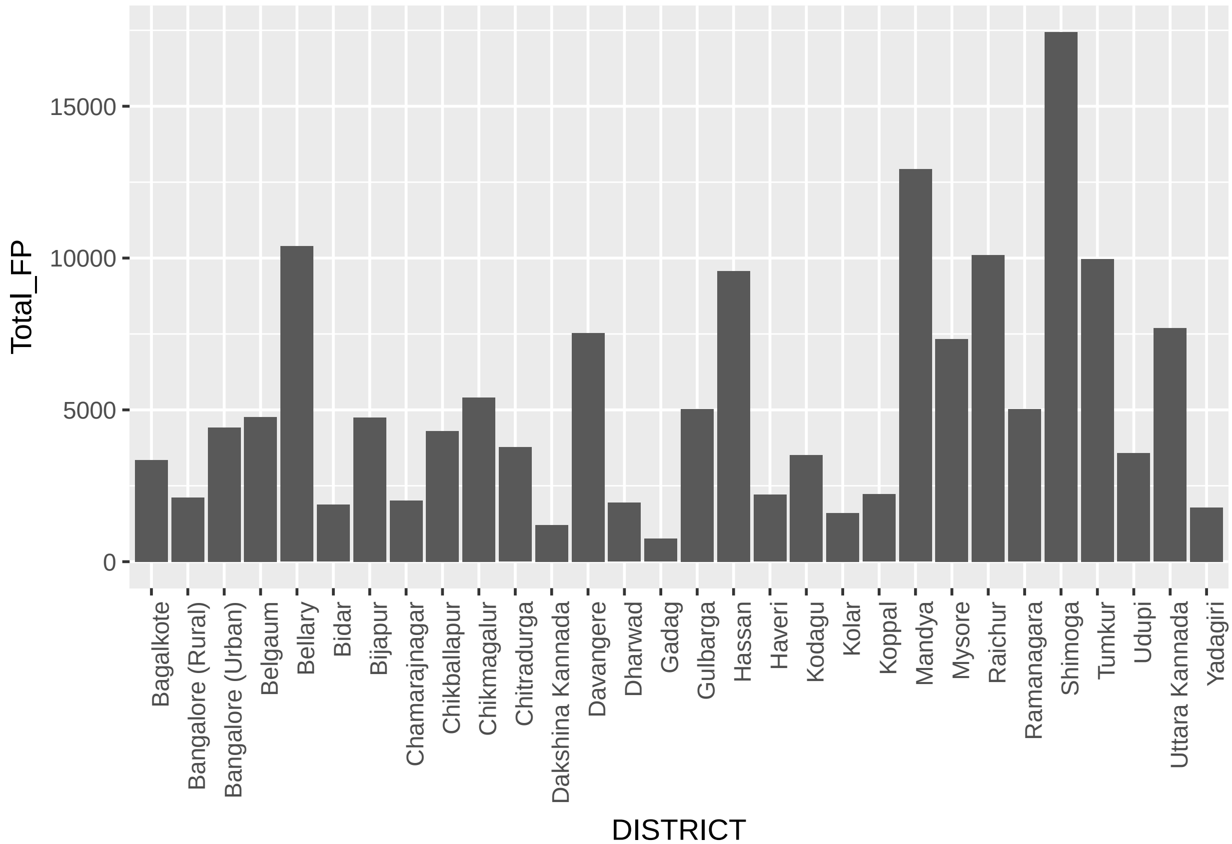
```
writeWorksheet ( inland_xl_data, DIST_TFP, sheet = 5, startRow = 1 , startCol = 1 ,
                 header = TRUE )
writeWorksheet ( inland_xl_data, Total, sheet = 5, startRow = 1 , startCol = 1 ,
                 header = TRUE )
saveWorkbook(inland_xl_data,"inland_production.xlsx")
```

```
library(ggplot2)

ggplot(DIST_TFP, aes(x = DISTRICT, y = Total_FP))+

  geom_bar(stat = "identity") +

  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



Analysis of inland fish production data from 2007-08 to 2015-16 and build a prediction model

Statisitical analysis of Inland fish production data from 2007-08 to 2015-16 is done using R stat tools. For analysis purpose data of each district, stocking in lakh fingerlings (of all category) and resource area in ha (of all category) are taken as data points.

import data in R

```
inland_old_data<-loadWorkbook("inlandcatch.xlsx")
legacy_data<-readWorksheet(inland_old_data, sheet = 2 , startRow =1 , startCol =1 ,
                           endRow =266 , endCol =7, header = TRUE)
```

Structure of data:

```
str(legacy_data)
```

```
## 'data.frame':    265 obs. of  7 variables:
##  $ year_ann     : num  2007 2007 2007 2007 2007 ...
##  $ fin_year     : chr  "2007-08" "2007-08" "2007-08" "2007-08" ...
##  $ District     : chr  "Bagalkote" "Bangalore (Rural)" "Bangalore (Urban)" "Belgaum" ...
```

```
##  $ Qty_MT       : num  1090 3699 3560 4505 9040 ...
##  $ Value_LRs    : num  381 1295 1246 1577 3164 ...
##  $ stocking_LFng: num  20.7 65.3 37.6 67.3 114.2 ...
##  $ Area_ha      : num  2391 24454 11710 28815 54811 ...
```

Summary of the data distribution

```r
legacy_data$District<-factor(legacy_data$District)
legacy_data$fin_year<-factor(legacy_data$fin_year, ordered = TRUE)
summary(legacy_data)
```

```
##     year_ann       fin_year                 District       Qty_MT     
##  Min.   :2007   2010-11:30   Bagalkote         :  9   Min.   :  687  
##  1st Qu.:2009   2011-12:30   Bangalore (Rural):  9   1st Qu.: 2402  
##  Median :2011   2012-13:30   Bangalore (Urban):  9   Median : 4217  
##  Mean   :2011   2013-14:30   Belgaum           :  9   Mean   : 5919  
##  3rd Qu.:2013   2014-15:30   Bellary           :  9   3rd Qu.: 7965  
##  Max.   :2015   2015-16:30   Bidar             :  9   Max.   :43765  
##                 (Other):85   (Other)           :211                  
##    Value_LRs       stocking_LFng       Area_ha     
##  Min.   :  284.8   Min.   :  0.00   Min.   : 1066  
##  1st Qu.: 1269.3   1st Qu.: 29.46   1st Qu.: 6522  
##  Median : 2282.9   Median : 59.52   Median :14263  
##  Mean   : 3422.4   Mean   : 76.40   Mean   :19657  
##  3rd Qu.: 3973.7   3rd Qu.:106.72   3rd Qu.:28815  
##  Max.   :37200.0   Max.   :384.24   Max.   :54813  
## 
```

Aggregate: year_wise fish production

```r
qty_summ<-aggregate(cbind(Qty_MT,Value_LRs,stocking_LFng)~year_ann,
                    data = legacy_data,sum)

print(qty_summ)
```

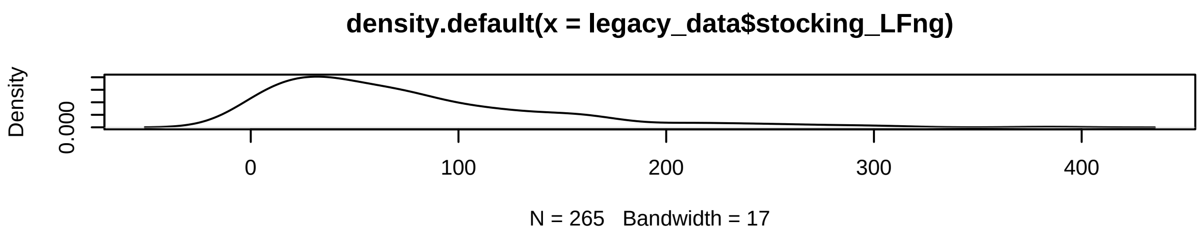
```
##   year_ann   Qty_MT Value_LRs stocking_LFng
## 1     2007 122124.3  42743.52      1599.970
## 2     2008 143717.2  57486.89      1997.440
## 3     2009 159209.7  63683.88      2393.515
## 4     2010 186008.7  77348.29      2704.807
## 5     2011 199054.0  80015.65      2685.390
## 6     2012 168241.4  84604.94      1663.826
## 7     2013 197953.0 159624.00      1519.435
## 8     2014 223420.0 193205.00      3115.826
## 9     2015 168827.0 148217.00      2565.480
```

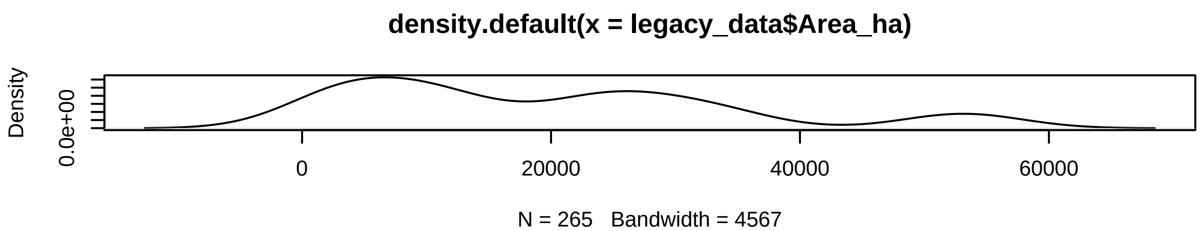
Exploratory data analysis

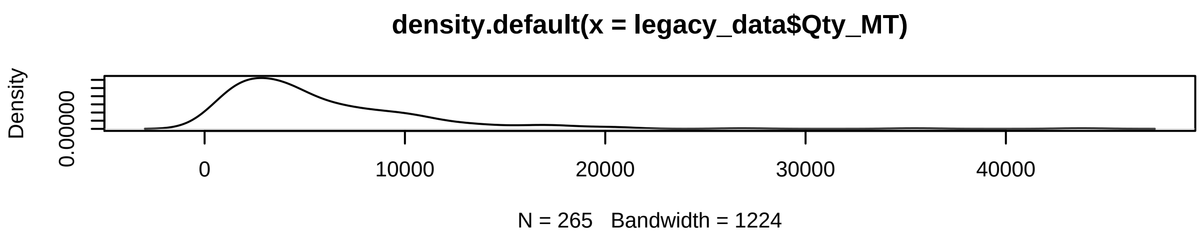
Density plots

Density Plots of dependent variable (Fish Production (Qty_MT)) and independent variables (Stocking and Area) are shwon below

```r
par(mfrow = c(3,1))
plot(density(legacy_data$stocking_LFng))
plot(density(legacy_data$Area_ha))
plot(density(legacy_data$Qty_MT))
```

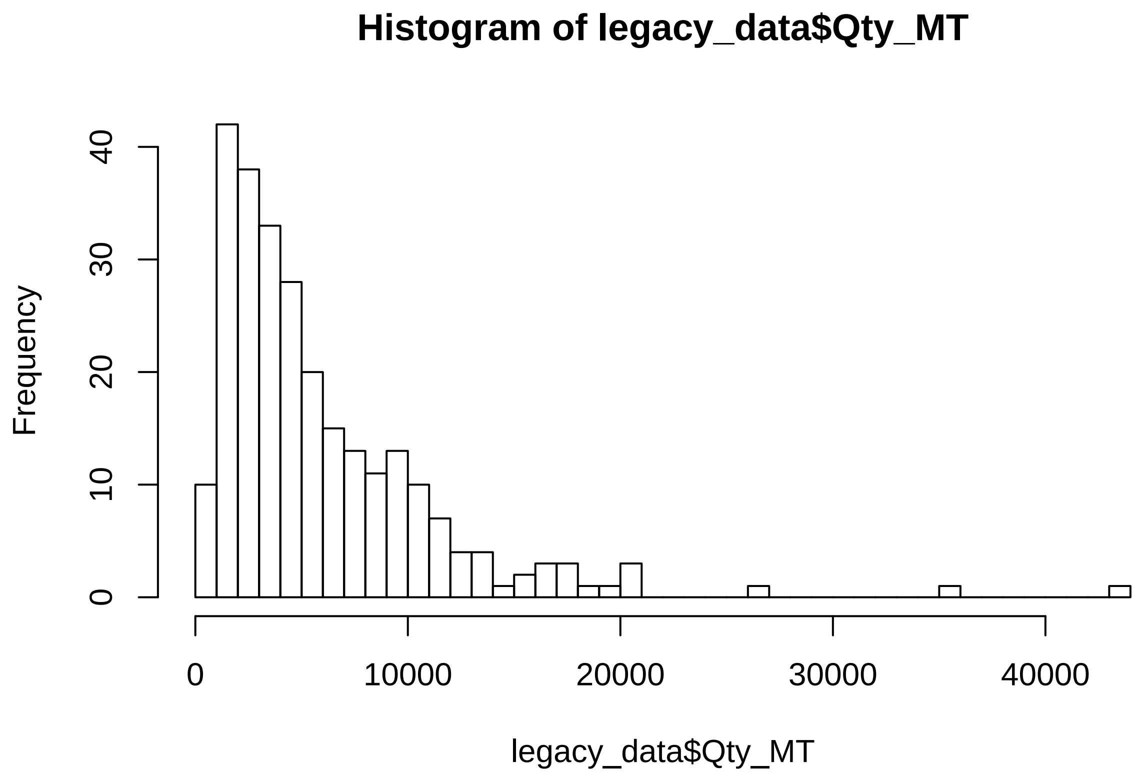




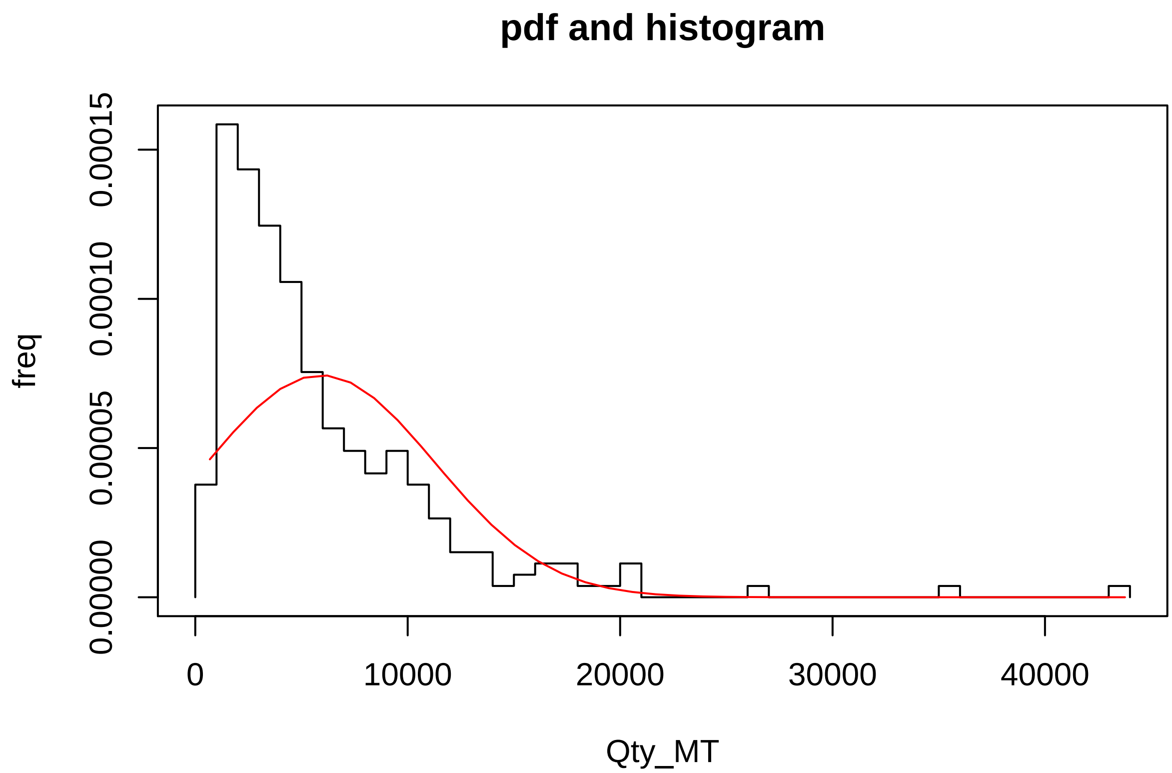


Probabilty Density Function and Histogram plots of dependent variable

```r
h<-hist(legacy_data$Qty_MT,breaks=50)
```




```
Qty_MT<-c(min(h$breaks),h$breaks)
freq<-c(0,h$density,0)
xfit<-seq(min(legacy_data$Qty_MT),max(legacy_data$Qty_MT),length=40)
yfit<-dnorm(xfit,mean=mean(legacy_data$Qty_MT),sd=sd(legacy_data$Qty_MT))
plot(Qty_MT,freq,type="s",ylim=c(0,max(freq,yfit)),
     main= "pdf and histogram")
lines(xfit,yfit, col= "red")
```



Skewness and Krtosis of dependant variable

```
library(fBasics)
skewness(legacy_data$Qty_MT)
```

```
## [1] 2.737128
## attr(,"method")
## [1] "moment"
```

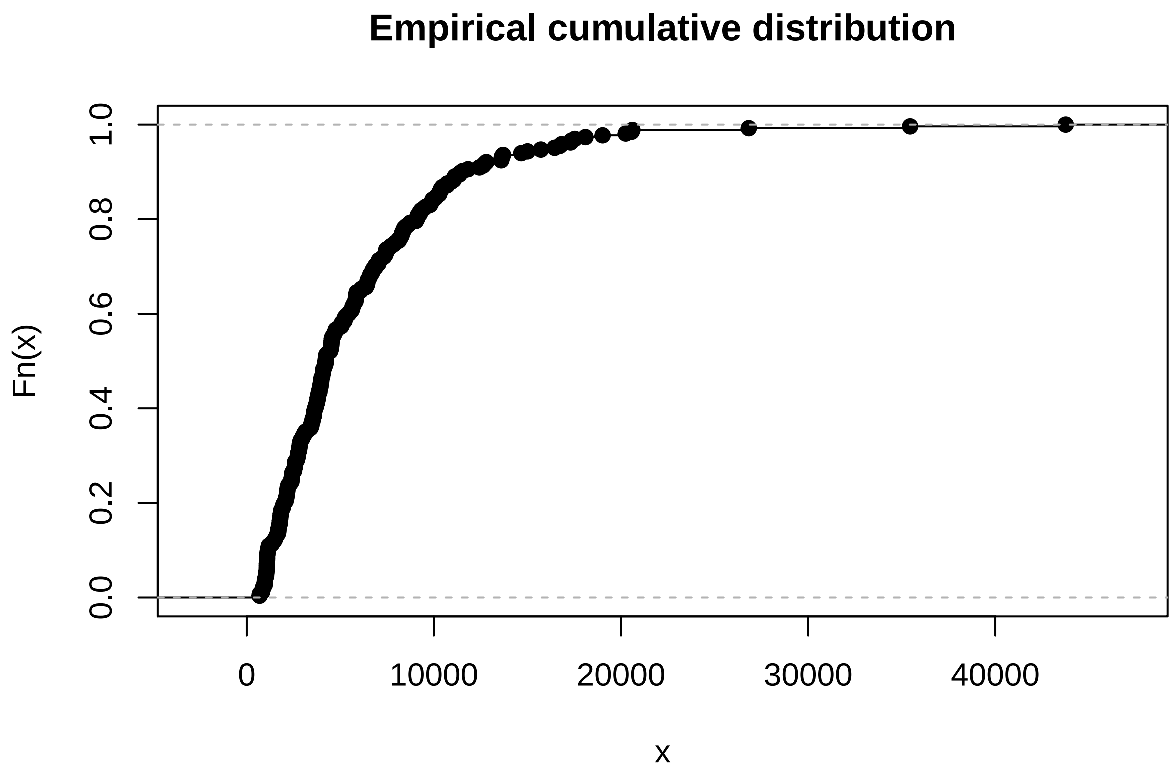
```
kurtosis(legacy_data$Qty_MT)
```

```
## [1] 12.39605
## attr(,"method")
## [1] "excess"
```

Plot clearly shows distribution is skewed to right and has gamma distribution.

Empirical cumulative distribution

```
plot(ecdf(legacy_data$Qty_MT), main = "Empirical cumulative distribution")
```

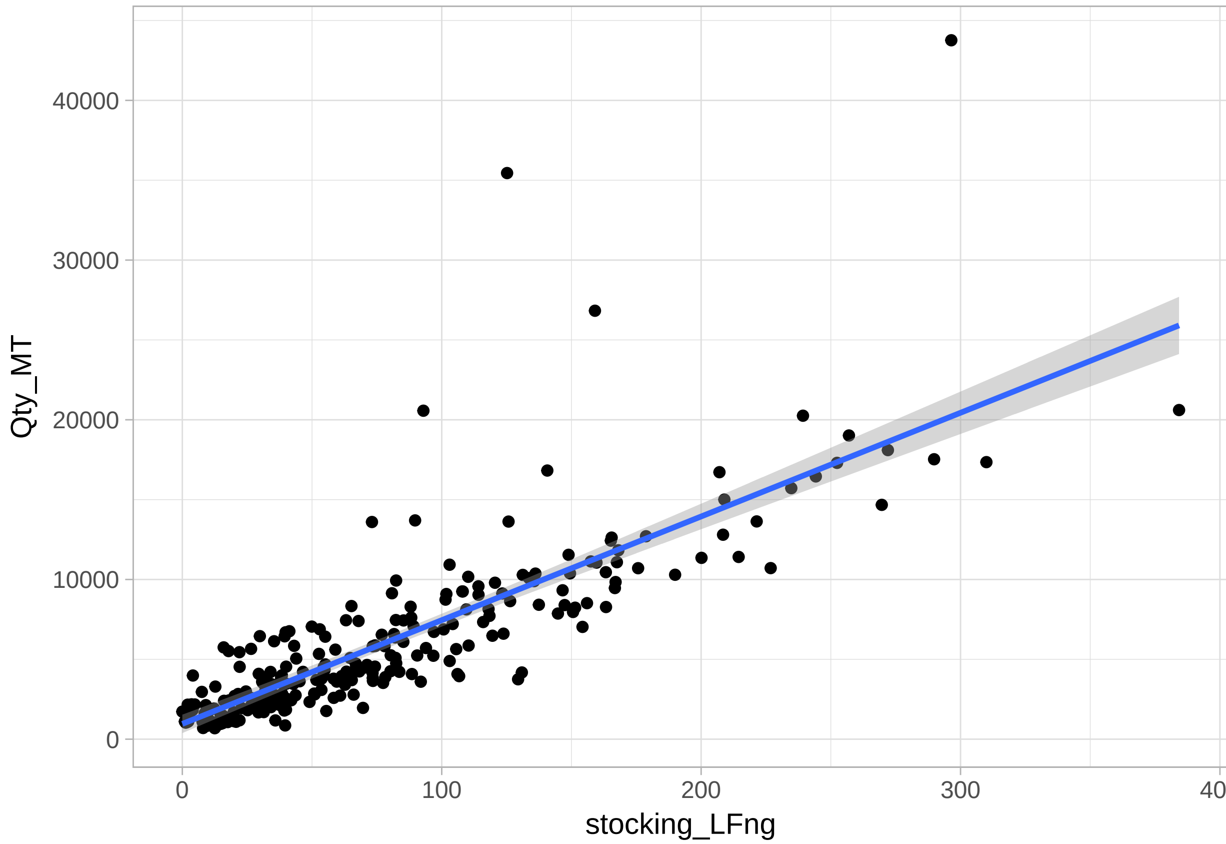



```r
# Empirical cummulitive distribution
summary(ecdf(legacy_data$Qty_MT))
```

```
## Empirical CDF:	  262 unique values with summary
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     687    2419    4236    5961    8086   43760
```

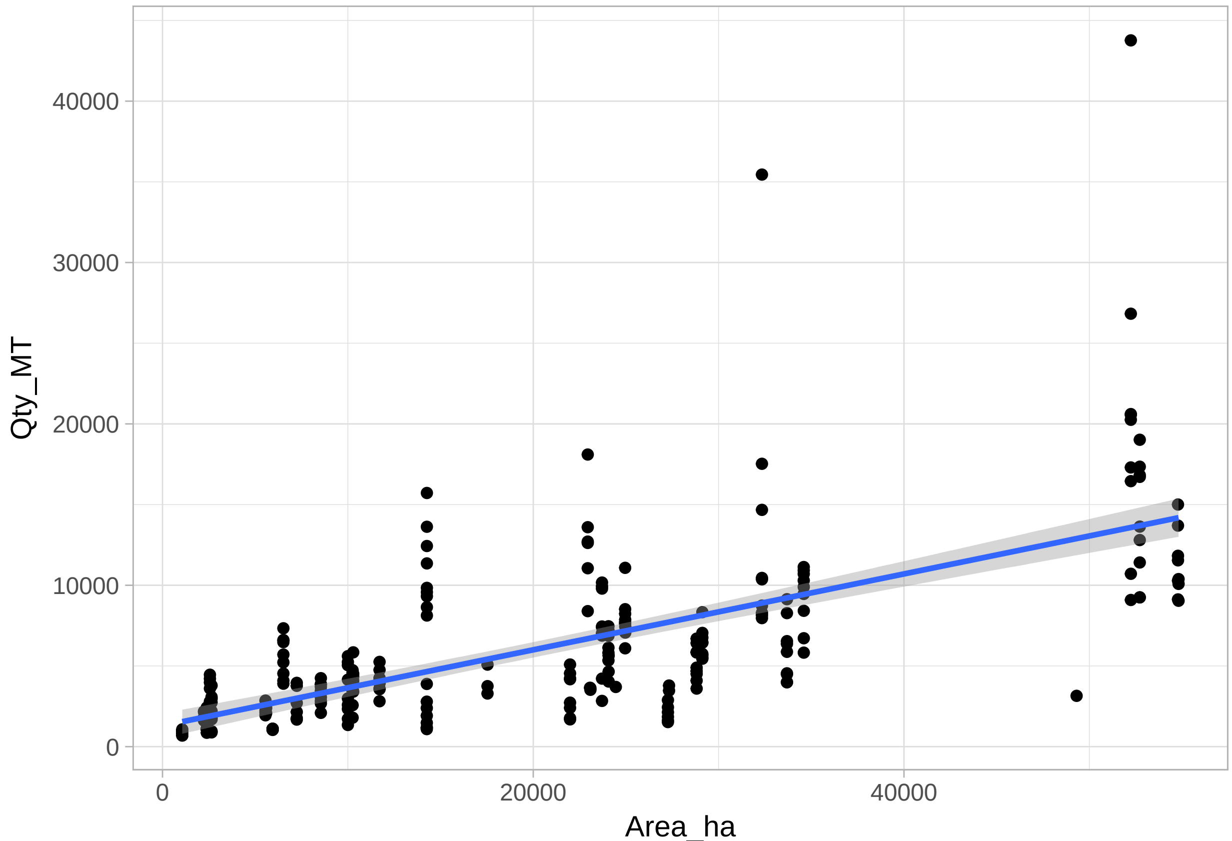
plot showing strong positive relation of Stocking and Production

```r
ggplot(legacy_data, aes(x = stocking_LFng, y = Qty_MT)) +
  geom_point() +
  geom_smooth(method = "lm", se = T) +
  theme_light()
```



plot showing strong positive relation of Stocking and Production

```
ggplot(legacy_data, aes(x = Area_ha, y = Qty_MT)) +
  geom_point() +
  geom_smooth(method = "lm") +
  theme_light()
```



Correlation Coefficient Stocking Vs Production

```
cor.test(legacy_data$stocking_LFng,legacy_data$Qty_MT,method = c("pearson"))
```

```
##
##  Pearson's product-moment correlation
##
## data:  legacy_data$stocking_LFng and legacy_data$Qty_MT
## t = 22.469, df = 263, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7651176 0.8484568
## sample estimates:
##       cor
## 0.8108586
```

Correlation Coefficient Area Vs Production

```
cor.test(legacy_data$Area_ha,legacy_data$Qty_MT,method = c("pearson"))
```

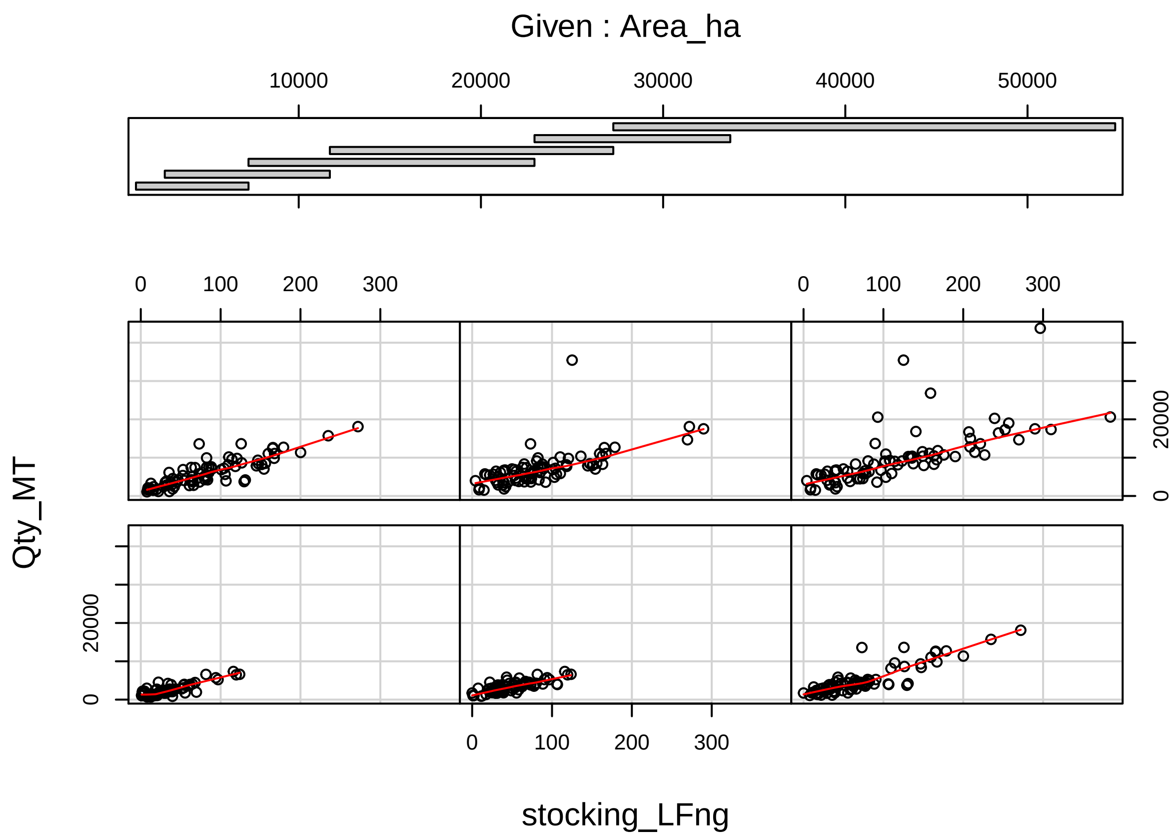
```
##
##  Pearson's product-moment correlation
##
```

```
## data:  legacy_data$Area_ha and legacy_data$Qty_MT
## t = 15.025, df = 263, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6089886 0.7395503
## sample estimates:
##       cor
## 0.6796155
```

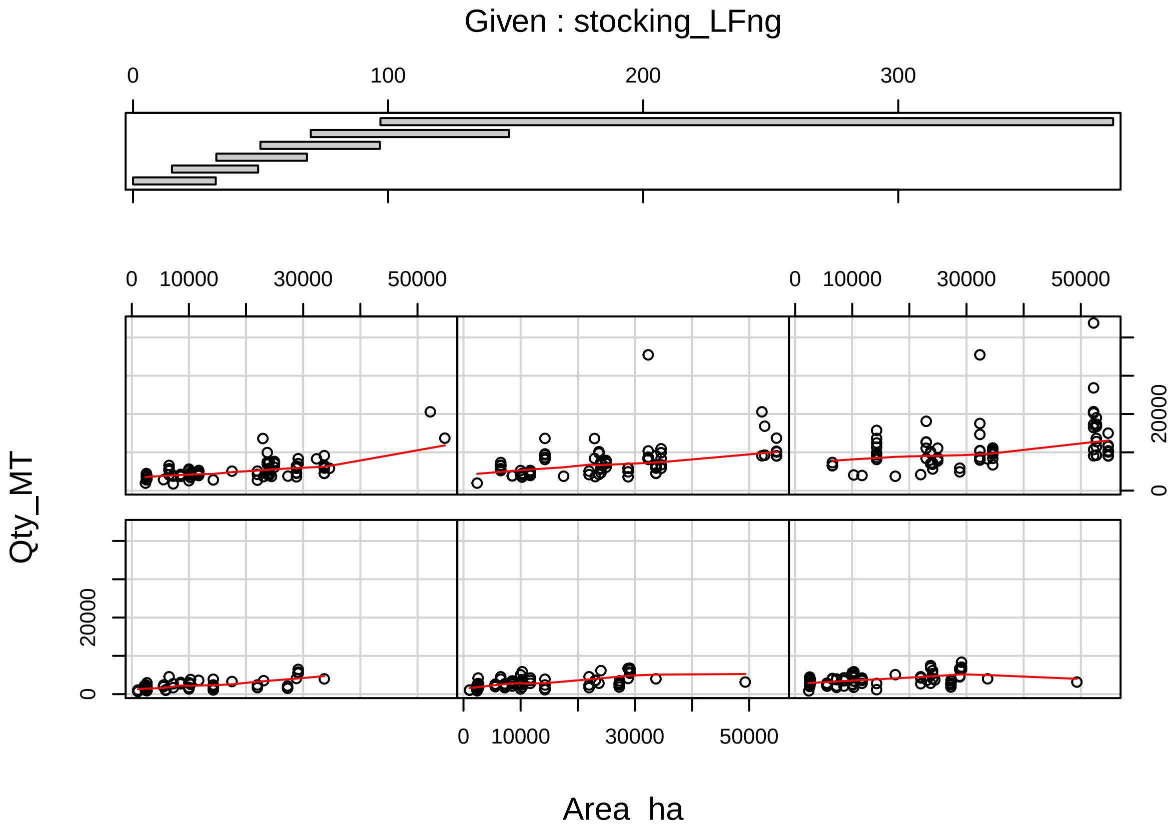
```
    #Pearson's product-moment correlation
```

coplots showing influence of stocking and / or Area on inland fish production

```
par(mfrow = c(2,1))
coplot(Qty_MT~stocking_LFng|Area_ha,panel=panel.smooth,legacy_data)
```




```
coplot(Qty_MT~Area_ha|stocking_LFng,panel=panel.smooth,legacy_data)
```



Fit generalized linear model of Fish Production Vs Stocking * Area Interaction

Model Assumptions

The model fitting is just the first part of the regression analysis since this is all based on certain assumptions. Regression diagnostics are used to evaluate the model assumptions and investigate whether or not there are observations with a large, undue influence on the analysis. Again, the assumptions for linear regression are:

Linearity:

The relationship between X and the mean of Y is linear. Homoscedasticity: The variance of residual is the same for any value of X. Independence: Observations are independent of each other. Normality: For any fixed value of X, Y is normally distributed. Before we go further, let's review some definitions for problematic points.

Outliers:

An outlier is defined as an observation that has a large residual. In other words, the observed value for the point is very different from that predicted by the regression model. ### Leverage points: A leverage point is defined as an observation that has a value of x that is far away from the mean of x. ### Influential observations: An influential observation is defined as an observation that changes the slope of the line. Thus, influential points have a large influence on the fit of the model. One method to find influential points is to compare the fit of the model with and without each observation.

Gamma Distribution [2]

The gamma distribution is continuous and defined for positive real numbers, [0,???). Depending on the values of its parameters, it may be either "ski-slope" shaped or it may be single-peaked, with a more-or-less exaggerated tail on the right. It can be used to represent the density of any variable that is restricted to non-negative values,

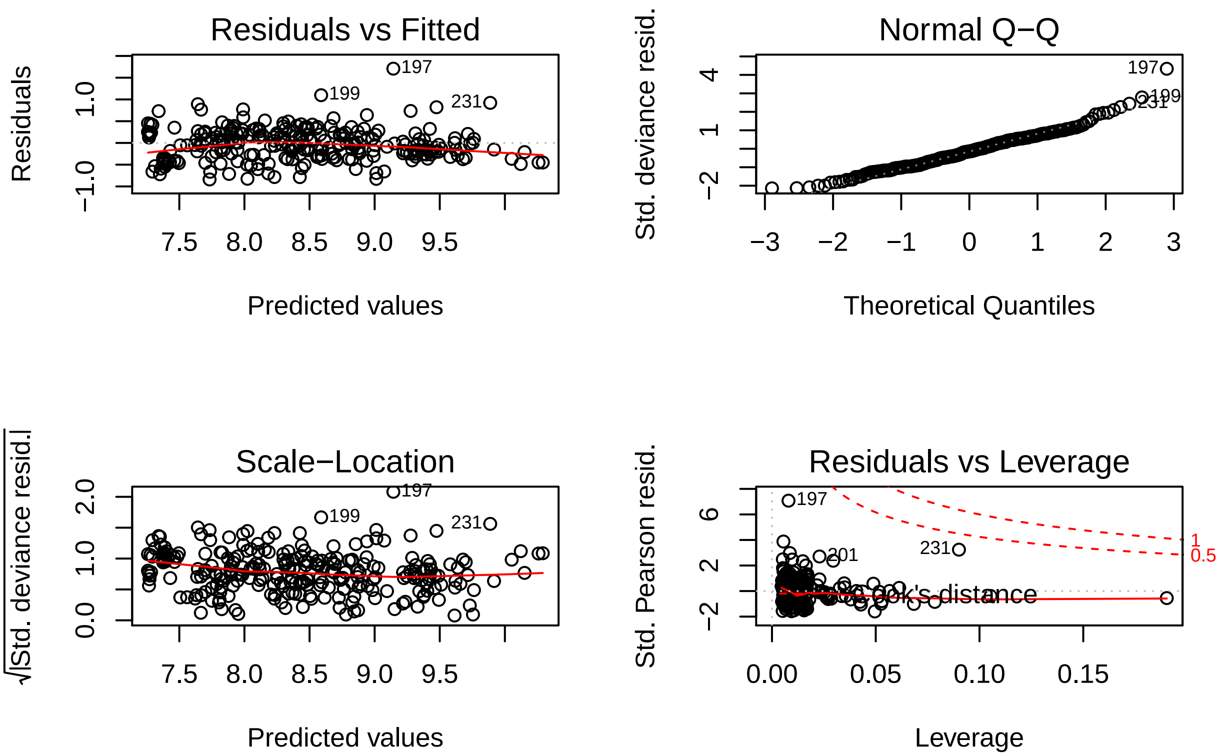
Since Area and Stocking have influemce on inland fish production the data is used to fit glm of Fish Production Vs Stocking * Area Interaction with Gamma family of distribution.

```
Model1<- glm(formula = Qty_MT ~ stocking_LFng * Area_ha, data = legacy_data,
             family = Gamma(link = "log"))
summary(Model1)
```

```
##
## Call:
## glm(formula = Qty_MT ~ stocking_LFng * Area_ha, family = Gamma(link = "log"),
##     data = legacy_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.83969  -0.31536  -0.05488   0.19786   1.70749
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            7.155e+00  5.684e-02 125.866  < 2e-16 ***
## stocking_LFng          1.290e-02  8.168e-04  15.798  < 2e-16 ***
## Area_ha                3.527e-05  2.765e-06  12.757  < 2e-16 ***
## stocking_LFng:Area_ha -1.895e-07  2.182e-08  -8.687 4.16e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.1566796)
##
##     Null deviance: 174.963  on 264  degrees of freedom
## Residual deviance:  35.555  on 261  degrees of freedom
## AIC: 4664.3
##
## Number of Fisher Scoring iterations: 5
```

Diagnostic plots

```
par(mfrow =c(2,2))
plot(Model1)
```



The first plot depicts residuals versus fitted values.

Residuals are measured as follows:

residual = observed y - model-predicted y

The plot of residuals versus predicted values is useful for checking the assumption of linearity and homoscedasticity. If the model does not meet the linear model assumption, we would expect to see residuals that are very large (big positive value or big negative value). To assess the assumption of linearity we want to ensure that the residuals are not too far away from 0 (standardized values less than -2 or greater than 2 are deemed problematic). To assess if the homoscedasticity assumption is met we look to make sure that there is no pattern in the residuals and that they are equally spread around the y = 0 line.

Normal QQ-Plot:

The tests and intervals estimated in summary(Model1) are based on the assumption of normality. The normality assumption is evaluated based on the residuals and can be evaluated using a QQ-plot by comparing the residuals to "ideal" normal observations. Observations lie well along the 45-degree line in the QQ-plot, so we may assume that normality holds here.

Scale Location Plot:

The third plot is a scale-location plot (square rooted standardized residual vs. predicted value). This is useful for checking the assumption of homoscedasticity. In this particular plot we are checking to see if there is a pattern in the residuals.

The assumption of a random sample and independent observations cannot be tested with diagnostic plots. It is an assumption that you can test by examining the study design.

Cook's distance (Residuals Vs Leverage) plot

The fourth plot is of "Cook's distance", which is a measure of the influence of each observation on the regression coefficients. The Cook's distance statistic is a measure, for each observation in turn, of the extent of change in model estimates when that particular observation is omitted. Any observation for which the Cook's distance is close to 1 or more, or that is substantially larger than other Cook's distances (highly influential data points), requires investigation.

Outliers may or may not be influential points. Influential outliers are of the greatest concern. They should never be disregarded. Careful scrutiny of the original data may reveal an error in data entry that can be corrected. If they remain excluded from the final fitted model, they must be noted in the final report or paper.

Model generally fits linear model to a larger extant. However the above plots show few outliers which affect the prediction on slightly higher side. Here are some potentially influential observations in the data

```r
summary(influence.measures(Model1))
```

```
## Potentially influential observations of
##   glm(formula = Qty_MT ~ stocking_LFng * Area_ha, family = Gamma(link = "log"),      data = legacy_da
##
##     dfb.1_ dfb.st_LF dfb.Ar_h dfb.s_LF: dffit   cov.r   cook.d hat
## 19   0.18   0.03     -0.48     0.23     -0.53_*  0.98    0.03   0.05_*
## 25  -0.01   0.03      0.02    -0.05     -0.08    1.06_*  0.00   0.04
## 52   0.01  -0.02     -0.02     0.05      0.07    1.08_*  0.00   0.06_*
## 68   0.03  -0.08      0.00     0.05     -0.08    1.05_*  0.00   0.03
## 80   0.00   0.00      0.00    -0.01     -0.01    1.07_*  0.00   0.05_*
## 81   0.01  -0.02     -0.01     0.03      0.06    1.05_*  0.00   0.04
## 98   0.09  -0.21      0.00     0.14     -0.22    1.06_*  0.01   0.05_*
## 101  0.11  -0.02     -0.04     0.01      0.15    0.95_*  0.01   0.01
## 111  0.02  -0.04     -0.04     0.09      0.13    1.06_*  0.00   0.05_*
## 112 -0.04   0.04      0.08    -0.11     -0.15    1.13_*  0.00   0.11_*
## 128 -0.02   0.04      0.00    -0.03      0.04    1.05_*  0.00   0.03
## 141 -0.09   0.07      0.19    -0.22     -0.31    1.25_*  0.02   0.19_*
## 163  0.21  -0.11     -0.11     0.10      0.23    0.94_*  0.02   0.01
## 176 -0.19   0.03      0.15    -0.06     -0.23    0.95_*  0.01   0.01
## 177 -0.10   0.04      0.00     0.00     -0.16    0.94_*  0.00   0.01
## 180 -0.03  -0.02      0.09    -0.02      0.11    1.05_*  0.00   0.03
## 197 -0.15   0.19      0.22    -0.18      0.43_*  0.72_*  0.10   0.01
## 199 -0.01   0.06      0.12    -0.11      0.23    0.89_*  0.02   0.01
## 210  0.00   0.00      0.00     0.00      0.00    1.06_*  0.00   0.04
## 218  0.13  -0.28     -0.01     0.19     -0.30    1.08_*  0.02   0.08_*
## 229  0.14  -0.30      0.03     0.16     -0.34    1.06_*  0.02   0.07_*
## 231  0.21  -0.23     -0.40     0.59      0.83_*  1.00    0.26   0.09_*
## 232 -0.01   0.03      0.01    -0.05     -0.08    1.05_*  0.00   0.04
## 257  0.07  -0.18      0.08     0.03     -0.30    1.05_*  0.01   0.05_*
## 261  0.00   0.00      0.00     0.00      0.00    1.08_*  0.00   0.06_*
```

Prediction based on model

Stocking of fingerlings and Area available during 2016 -17 is gathered in a dataframe 'data_2016'. Thirty data points representing each disrict is taken in to this dataframe for predction of inland fish production.

```
data_2016<-read.csv("new.txt")
print(data_2016)
```

```
##             District stocking_Lfng  Area_ha
## 1           Bagalkote      50.81000  2158.00
## 2  Bangalore (Urban)      42.58400  9803.00
## 3  Bangalore (Rural)      30.87000 12214.33
## 4             Belgaum      60.23000 27710.00
## 5             Bellary     112.71000 54464.00
## 6               Bidar      76.25000  9564.00
## 7             Bijapur      75.17000 23471.00
## 8        Chamarajnagar      27.50000  7855.00
## 9          Chikmagalur      97.26250 19265.00
## 10         Chitradurga      37.31000 33308.00
## 11    Dakshina Kannada       1.48500  4334.00
## 12           Davangere     114.32000 13468.00
## 13             Dharwad      27.20000  4237.00
## 14               Gadag      15.00000   564.00
## 15            Gulbarga      92.78000  9738.04
## 16              Hassan     124.53500 26923.00
## 17              Haveri      41.01442  4494.00
## 18              Kodagu      55.02500  1991.00
## 19               Kolar      25.50000 20171.67
## 20              Koppal      37.10000  2331.00
## 21              Mandya     148.52000 30688.00
## 22              Mysore     136.08000 21514.00
## 23             Raichur     157.78000 21840.00
## 24             Shimoga     249.06000 46776.00
## 25              Tumkur      99.59000 48476.00
## 26               Udupi       4.39000    16.00
## 27      Uttara Kannada      49.45800 23559.00
## 28        Chikballapur      50.51000 17545.33
## 29          Ramanagara      59.37000  8597.67
## 30            Yadagiri      28.22000  6118.96
```

```
predicted_data_2016<-predict(Model1,
                             data.frame(stocking_LFng=data_2016$stocking_Lfng,
                                        Area_ha = data_2016$Area_ha),
                             type = "response", se.fit = TRUE)

sum(predicted_data_2016$fit)
```

```
## [1] 159868.1
```

As per the fitted model predicted inalnd fish production of the state for the year 2016-17 is "161678.8 MT""

District-wise predicted data

Model has also predicted district-wise fish production for the year 2016-17 based on stocking and Area available in each district.

```
dist_prodn_2016<-data.frame(data_2016$District,predicted_data_2016)
colnames(dist_prodn_2016)<-c("District", "Predicted_fish_production-MT-2016","Std_Error","residual.scale

print(dist_prodn_2016)
```

```
##             District Predicted_fish_production-MT-2016  Std_Error
## 1          Bagalkote                          2605.886  100.57409
## 2  Bangalore (Urban)                          2894.913   84.23132
## 3  Bangalore (Rural)                          2730.563   81.53137
## 4            Belgaum                          5393.086  195.59745
## 5            Bellary                         11692.002  777.73841
## 6              Bidar                          4178.139  143.54405
## 7            Bijapur                          5530.674  164.54449
## 8       Chamarajnagar                         2311.207   77.69722
## 9         Chikmagalur                         6210.808  204.36262
## 10        Chitradurga                         5299.472  279.36887
## 11    Dakshina Kannada                        1518.286   73.76468
## 12          Davangere                         6721.062  310.09848
## 13            Dharwad                         2065.503   78.80657
## 14              Gadag                         1581.883   76.58606
## 15           Gulbarga                         5034.378  206.29047
## 16             Hassan                         8740.912  312.61260
## 17             Haveri                         2458.656   86.51203
## 18             Kodagu                         2735.438  107.43199
## 19              Kolar                         3286.673  114.94969
## 20             Koppal                         2206.365   85.81302
## 21             Mandya                        10826.058  424.80999
## 22             Mysore                         9086.532  403.67160
## 23            Raichur                        11024.090  583.12513
## 24            Shimoga                        18221.609 1449.47182
## 25             Tumkur                        10245.548  605.84416
## 26              Udupi                         1355.214   73.80738
## 27     Uttara Kannada                         4460.114  146.47432
## 28        Chikballapur                        3855.476  104.90269
## 29         Ramanagara                         3385.242  104.69002
## 30           Yadagiri                         2212.309   78.23000
##    residual.scale
## 1       0.3958278
## 2       0.3958278
## 3       0.3958278
## 4       0.3958278
## 5       0.3958278
## 6       0.3958278
## 7       0.3958278
## 8       0.3958278
## 9       0.3958278
```

```
## 10       0.3958278
## 11       0.3958278
## 12       0.3958278
## 13       0.3958278
## 14       0.3958278
## 15       0.3958278
## 16       0.3958278
## 17       0.3958278
## 18       0.3958278
## 19       0.3958278
## 20       0.3958278
## 21       0.3958278
## 22       0.3958278
## 23       0.3958278
## 24       0.3958278
## 25       0.3958278
## 26       0.3958278
## 27       0.3958278
## 28       0.3958278
## 29       0.3958278
## 30       0.3958278
```

Estimation of Inland fish production potential of the state

Using this prediction model an attempt is made to estimate the production potential of inalnd fish production of the state based on the optimum fish seed stockable in difeerent types of water bodies in the state.

Imortant assumptions:

a) Fish seed required at 50% utilization rate to large Reservoirs @ 500 fingerlings per ha
b) Fish seed required at 35% utilization rate to small Reservoirs @ 2000 fingerlings per ha
c) Fish seed required at 0.50% utilization rate to Major Tanks @ 2000 fingerlings per ha
d) Fish seed required at 30% utilization rate to Minor @ 4000 fingerlings per ha
e) Fish seed required at 50% utilization rate to River straech s @ 1000 fingerlings per km
f) Fish seed required at 10% utilization rate to water logged area converted to culture ponds for intensive fish culture @ 10000 fingerlings per ha

Using District-wise resource data stored in 'resource_data' data from let us estimate distict-wise fish seed stockable in these resources based on above assumptions.

```
library(dplyr)
pot_assumptions<-readWorksheet(inland_xl_data, sheet = 4 , startRow =1 ,
                               startCol = 1 , endRow = 7 , endCol = 3,
                               header = TRUE)

pot_cf<- pot_assumptions %>% mutate(cf = rate * utilization / 100000)

pot_cf
```

```
##           resource  rate utilization     cf
```

```
## 1 Large Reserovir   500        0.50 0.0025
## 2 Small Reservoir  2000        0.35 0.0070
## 3     Major Tanks  2000        0.50 0.0100
## 4      Minor Tank  4000        0.30 0.0120
## 5           Ponds 10000        0.10 0.0100
## 6           River  1000        0.50 0.0050
```

Extract resource data

```r
resource<-resource_data[,c(1,10:15)]

colnames(resource)<-c("District","Large Reservoir", "Small Reservoir",
                      "Major Tank", "Minor Tank", "Private Ponds", "River")

print(resource) # Water spread Area in Ha nad River length in km
```

```
##             District Large Reservoir Small Reservoir Major Tank Minor Tank
## 1          Bagalkote           27.00         2158.00        0.0          0
## 2  Bangalore (Urban)         1905.00         7390.00        0.0       2322
## 3  Bangalore (Rural)         1982.13        12014.33        0.0          0
## 4            Belgaum          832.00         3027.00      360.0      21368
## 5            Bellary          157.00        11132.00      280.0      39863
## 6              Bidar          319.00         2564.00      277.0       6409
## 7            Bijapur           12.00         5423.00        0.0      14500
## 8       Chamarajnagar         470.00         6970.00      885.0          0
## 9         Chikmagalur        4635.00         5669.00      364.0      11250
## 10        Chitradurga         270.00        20271.00      320.0       9433
## 11   Dakshina Kannada          87.00            0.00        0.0          0
## 12          Davangere         680.00         9843.00        0.0       3625
## 13            Dharwad        1318.00         1287.00      490.0          0
## 14              Gadag         477.00          564.00        0.0          0
## 15           Gulbarga         146.00         2629.24      355.8       5448
## 16             Hassan        7369.00        17240.00      420.0       8117
## 17             Haveri        2028.00         4494.00        0.0          0
## 18             Kodagu         446.00            0.00      105.0       1886
## 19              Kolar        7076.00        19241.67      462.0          0
## 20             Koppal         148.00         1935.00      396.0          0
## 21             Mandya        1350.00        14283.00      495.0      12924
## 22             Mysore        2935.00         6383.00        0.0       8585
## 23            Raichur         826.00         2544.00      492.0          0
## 24            Shimoga        5113.00         4735.00      544.0      37323
## 25             Tumkur        3954.00        45023.00      396.0       2682
## 26              Udupi          44.00           16.00        0.0          0
## 27     Uttara Kannada        1039.00         2257.00        0.0      17072
## 28        Chikballapur       4280.00        17445.33        0.0          0
## 29         Ramanagara        1637.87         6782.67     1215.0          0
## 30           Yadagiri         869.00         5873.76      245.2          0
##    Private Ponds River
## 1              0   206
## 2             91     0
## 3            200    62
## 4           2955   271
## 5           3189   161
```

```
## 6            314   119
## 7           3548   227
## 8              0   218
## 9           1982   158
## 10          3284   109
## 11          4334   380
## 12             0   115
## 13          2460     0
## 14             0    25
## 15          1305   405
## 16          1146   301
## 17             0     0
## 18             0   211
## 19           468    18
## 20             0    80
## 21          2986   299
## 22          6546   509
## 23         18804   275
## 24          4174   348
## 25           375   290
## 26             0   295
## 27          4230   335
## 28           100   160
## 29           600    20
## 30             0   256
```

Estimation of fish seed stockable in waterbodies

```r
fish_seed_req<- resource %>% mutate(LR = `Large Reservoir`* pot_cf[1,4],
                                    SR = `Small Reservoir` * pot_cf[2,4],
                                    MJR = `Major Tank` * pot_cf[3,4],
                                    MNR = `Minor Tank` * pot_cf[4,4],
                                    PVTR = `Private Ponds` * pot_cf[5,4],
                                    RR = River * pot_cf[6,4], TFSR = LR +
                                      SR + MJR + MNR + PVTR + RR,
                                    TA = `Large Reservoir` +
                                      `Small Reservoir` +
                                      `Major Tank` + `Minor Tank` +
                                      `Private Ponds`)

sum(fish_seed_req$TFSR) # Total Fish seed required in lakh fingerlings
```

```
## [1] 4980.324
```

```r
sum(fish_seed_req$TA) # Total water spread Area available
```

```
## [1] 565627
```

Estimation of district-wise inland fish production potential of the state in MT

Model has also predicted district-wise fish production potential based on stockable quantity and Area available in each district.

```r
predicted_potential<-predict.glm(Model1,data.frame(stocking_LFng=fish_seed_req$TFSR,
                                                    Area_ha = fish_seed_req$TA),
                                 type = "response", se.fit = TRUE)



dist_predn_pot<-data.frame(fish_seed_req$District,
                           fish_seed_req$TFSR,predicted_potential)
colnames(dist_predn_pot)<-c("District", "Fish_seed_Required_LFng","Predicted_fish_production-MT", "Std.
                            "Residual Scale")

print(dist_predn_pot)
```

```
##             District Fish_seed_Required_LFng Predicted_fish_production-MT
## 1          Bagalkote                16.20350                     1692.522
## 2  Bangalore (Urban)                85.26650                     4810.750
## 3  Bangalore (Rural)                91.36564                     5369.368
## 4            Belgaum               314.19000                    36903.222
## 5            Bellary               592.16750                    39824.464
## 6              Bidar               102.15850                     5597.446
## 7            Bijapur               248.60600                    23967.549
## 8        Chamarajnagar              59.90500                     3383.582
## 9         Chikmagalur              210.52050                    17336.533
## 10       Chitradurga               292.35300                    28310.306
## 11   Dakshina Kannada               45.45750                     2588.865
## 12          Davangere              114.67600                     6808.270
## 13            Dharwad               41.80400                     2555.252
## 14              Gadag                5.26550                     1419.619
## 15           Gulbarga              102.77868                     5635.964
## 16             Hassan              253.67150                    21782.393
## 17             Haveri               36.52800                     2467.086
## 18             Kodagu               25.85200                     1923.962
## 19              Kolar              161.77169                    11704.965
## 20             Koppal               18.27500                     1753.205
## 21             Mandya              294.74900                    29683.974
## 22             Mysore              223.04350                    19179.573
## 23            Raichur              214.20800                    17998.154
## 24            Shimoga              542.72350                    42223.727
## 25             Tumkur              366.39000                    24125.062
## 26              Udupi                1.69700                     1310.953
## 27     Uttara Kannada              267.23550                    27578.591
## 28       Chikballapur              134.61731                     8997.269
## 29         Ramanagara               69.82336                     3948.693
## 30           Yadagiri               47.02082                     2822.871
##      Std. Error Residual Scale
## 1      76.39055      0.3958278
## 2     169.74589      0.3958278
## 3     188.04727      0.3958278
## 4    3999.09715      0.3958278
## 5   12999.31898      0.3958278
## 6     253.74237      0.3958278
## 7    2176.63934      0.3958278
## 8     105.97216      0.3958278
## 9    1251.11172      0.3958278
```

```
## 10  2562.06228      0.3958278
## 11    90.74489      0.3958278
## 12   308.56027      0.3958278
## 13    86.04978      0.3958278
## 14    74.00676      0.3958278
## 15   257.24990      0.3958278
## 16  1593.10397      0.3958278
## 17    81.63079      0.3958278
## 18    79.41908      0.3958278
## 19   542.03799      0.3958278
## 20    76.84362      0.3958278
## 21  2771.03756      0.3958278
## 22  1468.91840      0.3958278
## 23  1380.63071      0.3958278
## 24 11527.90202      0.3958278
## 25  3913.46687      0.3958278
## 26    73.15729      0.3958278
## 27  2664.14015      0.3958278
## 28   390.91514      0.3958278
## 29   124.69384      0.3958278
## 30    89.43740      0.3958278
```

```
# District-wise fish production potential based on GLM model
```

Inland fish production potential of the state in MT

```
round(sum(predicted_potential$fit) , -3)
```

```
## [1] 404000
```

references:

[1] GLM with a Gamma-distributed Dependent Variable; Paul E. Johnson October 6, 2014 http://pj.freefaculty.org/guides/stat/Regression-GLM/Gamma/GammaGLM-01.pdf

[2] Distribution Overview: Probability by the Seat of the Pants; Paul Johnson August 30, 2011 http://pj.freefaculty.org/guides/stat/Distributions/DistributionOverview/DistributionReview.pdf

End of the document