Project-2 : CE605

Prateek Kumar Behera
20103082
Geoinformatics
Email: prateekbeh20@iitk.ac.in.com

---

1 Introduction

Analyze discharge data (X) collected independently at four sites. For each site we need to find:

- the magnitude of discharge X_k such that P(X >= X_k) = 0.01 along with 90 percent confidence interval for X_k.
- Test the hypothesis that the mean discharge is equal to 2500 units.
- Perform Goodness of Fit Test for Selected Distribution for 5 percent significance level.

2 Exective Summary

- Plot the histogram of the data sites
- Select the distribution based on the histogram.
- Get the Value of X_k for which the value of P(X >= X_k) = 0.01 or P(X < X_k) < 0.99.
- Get the CI for the estimated parameters for the selected distribution.
- Perform the Hypothesis testing for the mean discharge is equal to 2500 units.
- Perform the Chi-Square Test for goodness-of-fit test at 5 percent significance level.

3 Methodology

- We Start with plotting the histograms of the Data.
- We selected the distribution for the data by fitting the histogram of the data to different distributions.
- After getting the distribution we need to determine the value of x_k such that P(X >= x_k) = 0.01 or P(X <= x_k) = 0.99.
- We can determine the estimated values of the point estimates using Maximum Likelihood estimated
  - Normal Distribution
    - Estimated Mean($\overline{X}$) = 1/n * $\sum X_i$
    - Estimated Std Dev = 1/(n-1) * $\sum (X_i - \overline{X})^2$
  - Exponential Distribution
    - Estimated Lambda = n / $\sum X_i$
- We determine the value of x_k for different distribution
  - For Normal Distribution :- We can use the formula $x_k = 2.33 * \sigma + \mu$ for P(X <= x_k) = 0.99 to get the value of X_k using the standard normal distribution table where z value is 2.33.

- For Exponential Distribution :- by integrating the distribution function of distribution in interval 0 to x_k and the equating it to 0.99, so that we get the value of x_k

- We can obtain the Confidence Interval for different parameters by:-
  - For Normal Distribution:-
    - CI for μ = $\bar{X}$ - $t_{(n-1,\alpha/2)}$ * S/$\sqrt(n)$, $\bar{X}$ + $t_{(n-1,\alpha/2)}$ * S / $\sqrt(n)$
    - CI for σ^2 = (n-1) * S^2 / C_u , (n-1) * S^2 / C_l
    - where C_u = Upper Limit of Chi-Square Distribution, C_l = Lower Limit of Chi-Square Distribution, $\bar{X}$ = Estimated Mean, S = Estimated Std Deviation.
  - For Exponential Distribution:-
    - $\lambda = 1 / \mu$
    - CI for λ= [1/$\bar{X}$ (1 - $Z_{(\alpha/2)}/\sqrt(n)$), 1/$\bar{X}$ (1 + $Z_{(\alpha/2)}/\sqrt(n)$)
    - where λ = Parameter for Exponential Distribution.

- After Obtaining the CI we perform the hypothesis Testing for dof = n - 1 = 99 for which the population variance is not known so we use the T- Distribution to calculate the Rejection Region and calculate the Statistic using the formula:
  - T = ($\bar{X}$ - μ_0)/(S/$\sqrt(n)$)
  - where $\bar{X}$.= Estimated Mean , μ_0 = 2500(given in question).

- Goodness of Fit Test
  - Perform the Chi-Square Test with 5 percent significance level.

-

4 Result

4.1 Data Site 1

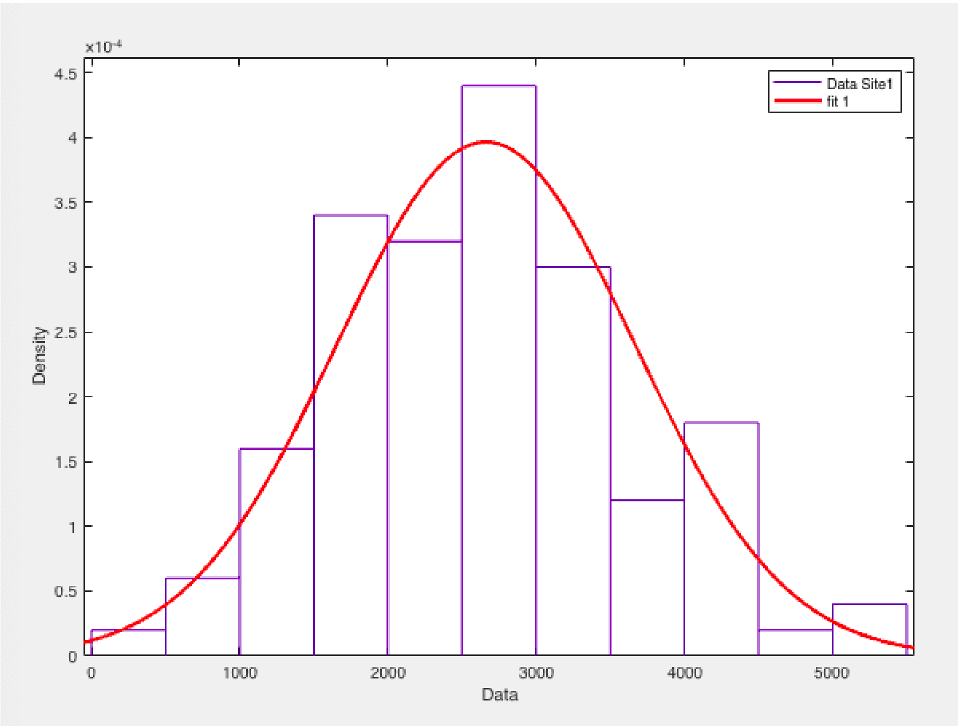

Figure 1: Fitted Data for Data Site 1.(Normal Distibution)

- Selected Distribution which fits the data is Normal Distribution.
- Value of Estimated Mean and Standard Deviation

  - Estimated Mean($\overline{X}$) = 1/n * $\sum X_i$ = 2661.1202
  - Estimated Std Dev = 1/(n-1) * $\sum(X_i - \overline{X})^2$ = 1005.77166728

- Value of X_k
  - 2.33 * 1005.77166728 + 2661.1202 = 5004.56818476.

- 90 percent CI for the estimated Parameters
  - CI for $\mu = \bar{X}$ - $t_{(n-1,\alpha/2)}$ * S/$\sqrt(n)$, $\bar{X} + t_{(n-1,\alpha/2)}$ * S / $\sqrt(n)$ = [2494.1621032315356, 2828.078296768462]
  - CI for σ = (n-1) * S^2 / C_u , (n-1) * S^2 / C_l = [902.3280099776418, 1147.9162358469366]

- Hypothesis Testing that the mean discharge is equal to 2500 units.
  - Significance level = 5 percent
  - Test Statistic = ($\bar{X}$ - μ_0)/(S/$\sqrt(n)$) = 1.60195604272
  - where $\bar{X}$.= Estimated Mean , μ_0 = 2500(given in question).
  - Critical Points for Rejection = [-1.987 , 1.987] for dof = 99 and alpha = 5 percent
  - μ = μ_0 = 2500 H0 cannot be rejected

- Goodness of Fit Test
  - H0:- Assumed Distribution is Normal Distribution
  - HA :- Data does not follow normal Distribution
  - As the Test Statistic is 5.07 which does not lie in the Rejection Region (Critical Value = 11.070) the Null Hypothesis cannot be rejected.
  - Please look at end of this report for Goodness of fit Test Calculations.

4.2 Data Site 2

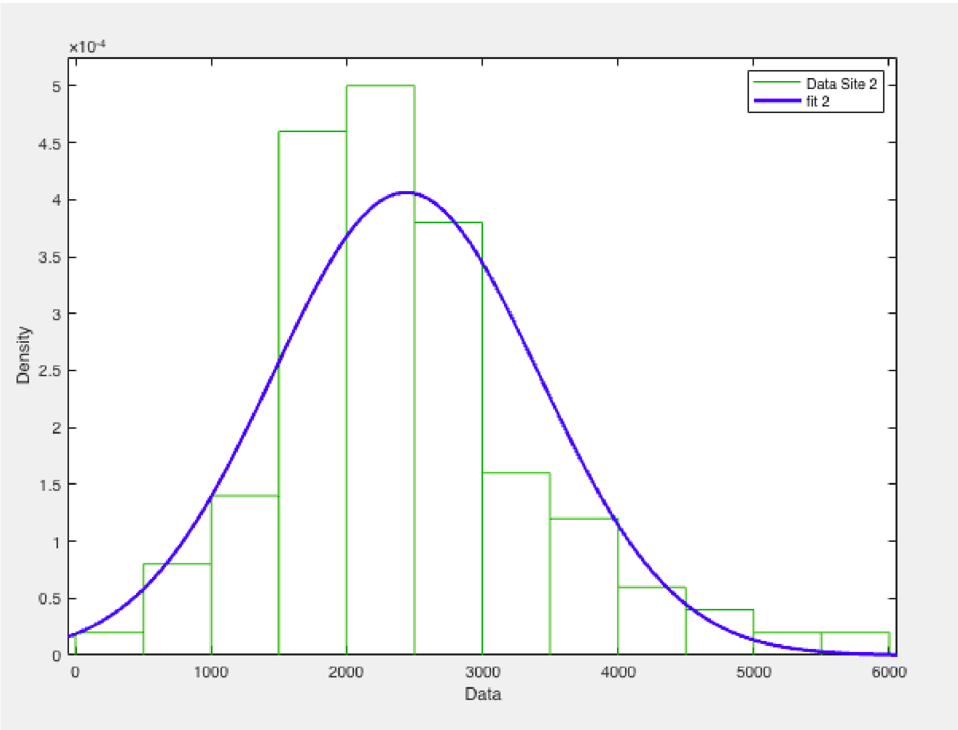

Figure 2: Caption of the sample figure.

- Selected Distribution which fits the data is Normal Distribution.
- Value of Estimated Mean and Standard Deviation
  - Estimated Mean($\overline{X}$) = 1/n * $\sum X_i$ = 2437.4885

  - Estimated Std Dev = 1/(n-1) * $\sum(X_i - \overline{X})^2$ = 981.933137767

- Value of X_k
  - 2.33 * 981.933137767 + 2437.4885 = 4725.392711.

- 90 percent CI for the estimated Parameters
  - CI for $\mu = \bar{X}$ - $t_{(n-1,\alpha/2)}$ * S/$\sqrt(n)$, $\bar{X} + t_{(n-1,\alpha/2)}$ * S / $\sqrt(n)$ = [2274.4875991306562, 2600.4894008693436]
  - CI for σ = (n-1) * S^2 / C_u , (n-1) * S^2 / C_l = [880.9412742046759, 1120.7086340058231]

- Hypothesis Testing that the mean discharge is equal to 2500 units.
  - Significance level = 5 percent
  - Test Statistic = ($\bar{X}$ - μ_0)/(S/$\sqrt(n)$) = -0.636616665592
  - where $\bar{X}$.= Estimated Mean , μ_0 = 2500(given in question).
  - Critical Points for Rejection = [-1.987 , 1.987] for dof = 99 and alpha = 5 percent.
  - $\mu = \mu_0$ = 2500 H0 cannot be rejected.

- Goodness of Fit Test
  - H0:- Assumed Distribution is Normal Distribution
  - HA :- Data does not follow normal Distribution
  - As the Test Statistic is 8.75 which does not lie in the Rejection Region (Critical Value = 11.070) the Null Hypothesis cannot be rejected.
  - Please look at end of this report for Goodness of fit Test Calculations.

4.3 Data Site 3

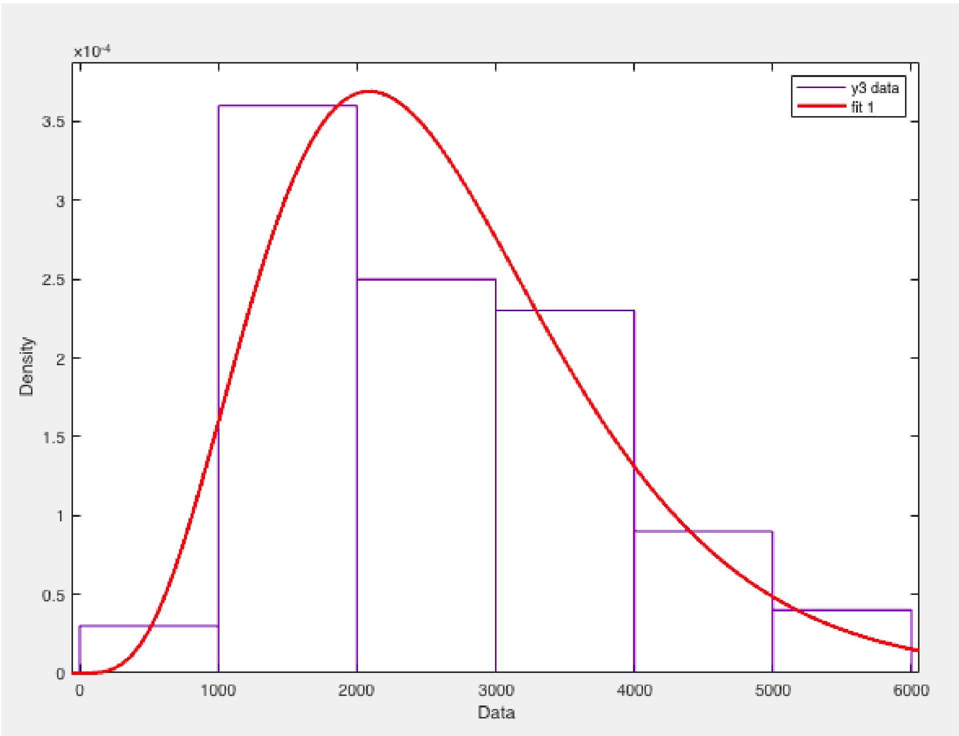

Figure 3: Caption of the sample figure.

- Selected Distribution which fits the data is Chi-Square Distribution.
- Value of Estimated Mean and Standard Deviation
- 90 percent CI for the estimated Parameters
  - CI for Std Deviation= $[(n-1)*S^2/\chi^2(\alpha/2),(n-1)*S^2/\chi^2(\alpha/2)]$ =[2253.9488488326087, 4020.0981716345786].

4.4 Data Site 4

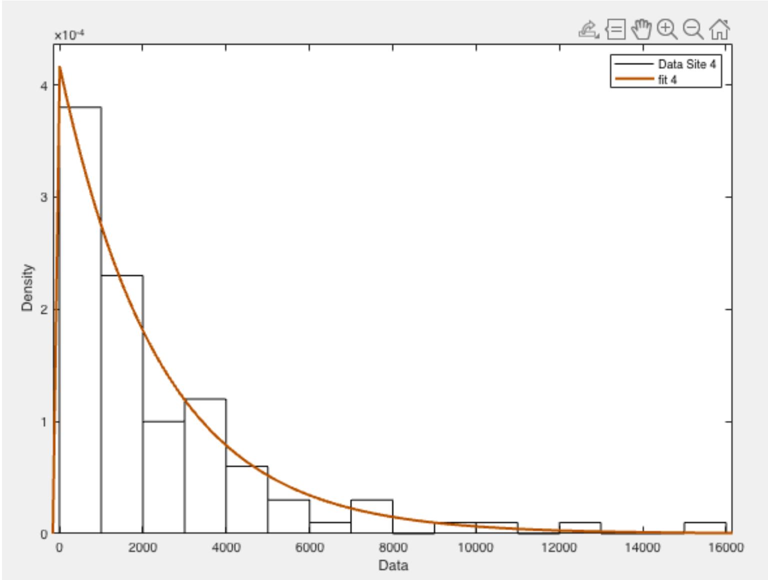

Figure 4: Caption of the sample figure.

- Selected Distribution which fits the data is Exponential Distribution.
- Value of Estimated Mean and Standard Deviation
  - Estimated lambda(λ) = n / $\sum X_i$ = 0.000416769052931
- Value of X_k
  - By integrating $\lambda * exp^{-\lambda * x}$ over 0 to X_k= 0.99 we get the value of X_k = 11050.

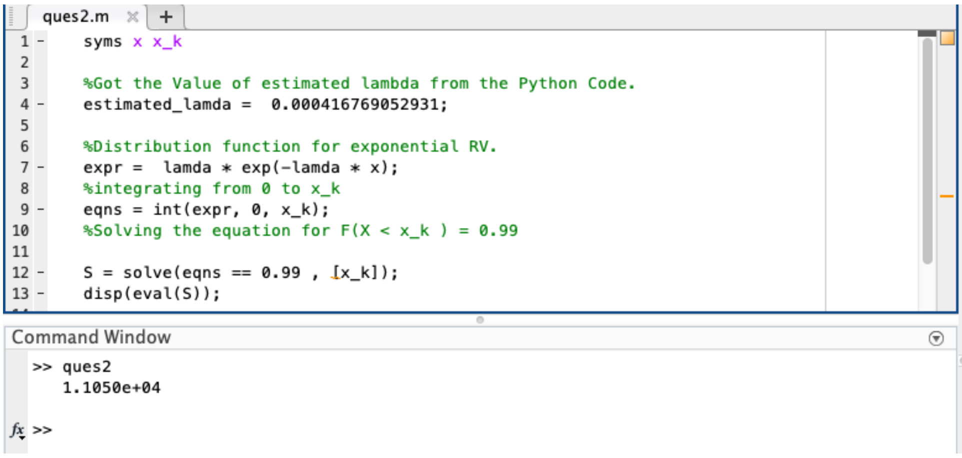

Figure 5: Value of X_k obtained using Matlab Code for Integration.

  –

- 90 percent CI for the estimated Parameters
  - CI for λ= [$1/\bar{X}$ (1 - $Z_{(\alpha/2)}/\sqrt(n)$), $1/\bar{X}$ (1 + $Z_{(\alpha/2)}/\sqrt(n)$) = [0.0003482105, 0.00048532756]
  - where λ = Parameter for Exponential Distribution.
- Hypothesis Testing that the mean discharge is equal to 2500 units.
  - Significance level = 5 percent
  - Test Statistic = ($\bar{X}$ - μ_0)/(S/$\sqrt(n)$) = -0.374701564649
  - where $\bar{X}$.= Estimated Mean , μ_0 = 2500(given in question).

  - Critical Points for Rejection = [-1.987 , 1.987] for dof = 99 and alpha = 5 percent.
  - $\mu = \mu_0 = 2500$ H0 cannot be rejected.

- Goodness of Fit Test
  - H0:- Assumed Distribution is Normal Distribution
  - HA :- Data does not follow normal Distribution
  - As the Test Statistic is 116.172 which lies in the Rejection Region (Critical Value = 7.8154) the Null Hypothesis is rejected.
  - Please look at end of this report for Goodness of fit Test Calculations.

5 Python Code to Evaluate Estimation of Parmeters and their 90 percent Confidence Interval and Perform Hypothesis Testing that that the mean discharge is equal to 2500 units.

```
import math
from scipy.integrate import quad
from matplotlib import pyplot as plt
import numpy as np

f = open("20103082.txt", "r")
f_out = open("20103082_output","w+")


def estimate_parameters(pdf , data):

    n = len(data)
    if pdf == "Normal":
  # by method of Maximum Likelihood
  # calculate estimated paramters
  sum_x = 0
  for i in range(0,n):
      sum_x += data[i]
  estimated_mean = (1.0 * sum_x/n)

  squared_sum = 0
        for i in range(0,n):
      squared_sum += math.pow((data[i] - estimated_mean),2)

  estimated_variance = (1.0*squared_sum/(n-1))
  estimated_stddev = math.sqrt(estimated_variance)
  return [estimated_mean,estimated_stddev]

    if pdf == "Chi-Square":
  #Dof = k - 1 - m
  # k = 2 * sqrt(n) . To get the no. of bins
  # m = 1 for nu(Parameter for Chi-Square.

  k = 2 * math.sqrt(n)
  return k - 1 - 1

    if pdf == "Exponential":
  #By using method of Max. Likelihood
        lamda = 1.0 * n / sum(data)
  return lamda



def confidence_interval(distribution, data, *params):
    #we have to calculate for 90% CI.

    n = len(data)
    if distribution == "Normal":
  #Population variance is not known
  #1. Calculate sample mean
  #2. Calculate S^2

  X_bar = 1.0 * sum(data)/n
  squared_sum = 0
```

```
  for i in range(0,n):
      squared_sum += math.pow((data[i] - X_bar),2)
  S_square = 1.0 * squared_sum/(n-1)

  sample_stddev = math.sqrt(S_square)

  ##Calculate the CI using formula [ X_bar - t(n-1,alpha/2) * S/sqrt(n) , X_bar - t(n-1,alpha
    /2) * S/sqrt(n) ]
  ## Value of t(100,5%) from T-distribution table = 1.66
  low_mean = X_bar - 1.66 * sample_stddev/math.sqrt(n)
  high_mean = X_bar + 1.66 * sample_stddev/math.sqrt(n)

  ## CI for Variance
  ## Population mean is not known
  ## 1.We get the CI using formula [ (n-1) * S^2 / Cu  , (n-1) * S^2 / Cl ]
  ## For the Interval Estimation we use the Chi square dist. table.
  ## Cu = 123 and Cl = 76 for dof = 99 and alpha/2 = 0.05 and 0.95 respectively.

  Cu = 123
  Cl = 76
  low_stddev = math.sqrt((n - 1) * S_square/Cu)
  high_stddev = math.sqrt((n - 1) * S_square/Cl)
  return [[low_mean,high_mean],[low_stddev,high_stddev]]

    if distribution == "Chi-Square":
  ## 2*dof = std. deviation
  ## Estimating the CI for std deviation and from that we can estimate CI for dof.
        ## CI for VARIANCE = [(n-1)*S^2/Chi-Square(alpha/2) ,(n-1)*S^2/Chi-Square(alpha/2) ]
        ## alpha/2 = 0.05

  dof = params[0]
  sample_mean = sum(data)/n
  print "Value of Chi-Square dist for dof = {} for alpha = {} and {} \n is {} {} respectively".
    format(dof,0.05,0.95, 27.587, 8.672)
        squared_sum = 0
        for i in range(0,n):
            squared_sum += math.pow((data[i] - sample_mean),2)
        S_square = 1.0 * squared_sum/(n-1)


        low_variance = 1.0 * (n - 1) * S_square / 27.587
  high_variance = 1.0 * (n - 1) * S_square / 8.672

  return [low_variance, high_variance]

    if distribution == "Exponential":
  ## To Calculate the CI for the Exponential Distribution
  ## As mean = 1/lamda
  ## we use formula [1/X_bar (1 - Z(alpha/2)/sqrt(n)), 1/X_bar (1 + Z(alpha/2)/sqrt(n)) ]
  ## Z(alpha/2) = 1.645  for alpha = 90%
  sample_mean = sum(data)/n

  low_lambda = 1/sample_mean * (1 - 1.645/ math.sqrt(n))
  high_lambda = 1/sample_mean * (1 + 1.645/ math.sqrt(n))
  return [low_lambda,high_lambda]


def hypothesis_testing(data, mean = -1 , variance = 0):

    n = len(data)
    #H0 => mu = mu_0 = 2500
    #HA => mu != mu_0

    mu_0 = 2500
    #(5% significance level)
    significance_level = 5

    if variance == 0:
      #Compute Test Statistic using T-Distribution
      #Population Variance is not known

      sample_mean = sum(data)/n
  squared_sum = 0
      for i in range(0,n):
          squared_sum += math.pow((data[i] - sample_mean),2)
      S_square = 1.0 * squared_sum/(n-1)
```

```
      sample_stddev = math.sqrt(S_square)

  #Test Statistic
  T = 1.0 * (sample_mean - mu_0)/(1.0 * sample_stddev/math.sqrt(n))

  print "Test Statistic for Hypothesis Testing that mean discharge = 2500 is {}".format(T)
  f_out.write("Test Statistic for Hypothesis Testing that mean discharge = 2500 is {}".format(T
    ))
        #For significance level 5 we have to check between
  # range -1.987 to 1.987 for dof = n - 1 = 99

  if T < 1.987 and T > -1.987:
      return ["H0 cannot be rejected",1]
  else:
            return ["H0 can be rejected",0]



data1 = []
data2 = []
data3 = []
data4 = []



header = f.readline()

while True:
    line = f.readline()
    if len(line) != 0:

        val = line.split()
  data1.append(float(val[0]))
  data2.append(float(val[1]))
  data3.append(float(val[2]))
  data4.append(float(val[3]))
    else:
        break



print("###########################################################################")
print("######################## For DATA SITE 1 ###################################")
print("###########################################################################")

f_out.write("###########################################################################")
f_out.write("######################## For DATA SITE 1 ###################################")
f_out.write("###########################################################################")


#Assuming the above distribution to be Normal using Histogram
#1. P(X > x_k) = 0.01 ,so P(X < x_k) = 0.99
#2. P(Z < (x_k - mean)/variance) = 0.99
#3a. According to Std. Normal Distribution table
#3b. (x_k - mean)/std_dev = 2.33

#   Estimate the Paramters.
[est_mean,est_stddev] = estimate_parameters("Normal",data1)

print "est_mean = {}".format(est_mean)
print "est_stddev = {}".format(est_stddev)

f_out.write("est_mean = {}\n".format(est_mean))
f_out.write("est_stddev = {}\n".format(est_stddev))

#Rearranging the eqn (x_k - mean)/stddev = 2.33 from #3b
# we get x_k = 2.33 * std_dev + mean

x_k = 2.33 * est_stddev + est_mean
print "Value of x_k = {}".format(x_k)

f_out.write("Value of x_k = {}\n".format(x_k))

[mean_interval, variance_interval] = confidence_interval("Normal", data1)
```

```
print "mean_interval = {}".format(mean_interval)
print "variance_interval = {}".format(variance_interval)

f_out.write("90 % CI for mean = {}\n".format(mean_interval))
f_out.write("90 % CI for variance = {}\n".format(variance_interval))

H0 = "mu = mu_0 = 2500"
HA = "mu != mu_0"

res = hypothesis_testing(data1)
if res[1] == 1:
    f_out.write("{} {}\n".format(H0,res[0]))
else:
    f_out.write("{} {}\n".format(HA,res[0]))


f_out.write('\n')
print("##########################################################################")
print("######################## For DATA SITE 2 #################################")
print("##########################################################################")
f_out.write("##########################################################################")
f_out.write("######################## For DATA SITE 2 ################################")
f_out.write("##########################################################################")

#   Estimate the Paramters.
[est_mean,est_stddev] = estimate_parameters("Normal",data2)

print "est_mean = {}".format(est_mean)
print "est_stddev = {}".format(est_stddev)

f_out.write("estimated mean = {}\n".format(est_mean))
f_out.write("estimated standard deviation = {}\n".format(est_stddev))

#Rearranging the eqn (x_k - mean)/variance = 2.33 from #3b
# we get x_k = 2.33 * est_stddev + est_mean

x_k = 2.33 * est_stddev + est_mean
print "Value of x_k = {}".format(x_k)

f_out.write("Value of x_k = {}\n".format(x_k))

[mean_interval, variance_interval] = confidence_interval("Normal", data2)
print "mean_interval = {}".format(mean_interval)
print "variance_interval = {}".format(variance_interval)

f_out.write("90 % CI for mean = {}\n".format(mean_interval))
f_out.write("90 % CI for variance = {}\n".format(variance_interval))

res = hypothesis_testing(data2)
if res[1] == 1:
    f_out.write("{} {}\n".format(H0,res[0]))
else:
    f_out.write("{} {}\n".format(HA,res[0]))

f_out.write('\n')
print("##########################################################################")
print("######################## For DATA SITE 3 #################################")
print("##########################################################################")
f_out.write("##########################################################################")
f_out.write("######################## For DATA SITE3 ################################")
f_out.write("##########################################################################")


dof = estimate_parameters("Chi-Square",data3)

print "Degree of freedom = {}".format(dof)
f_out.write("Degree of freedom = {}".format(dof))

variance_interval = confidence_interval("Chi-Square", data3,dof)

stddev_interval = [math.sqrt(variance_interval[0]), math.sqrt(variance_interval[1])]

print "Confidence Interval for Std dev = {} ".format(stddev_interval)
f_out.write("Confidence Interval for Std dev = {} ".format(stddev_interval))
f_out.write('\n')
```

```
print("###########################################################################")
print("######################## For DATA SITE 4 ##################################")
print("###########################################################################")
f_out.write("###########################################################################")
f_out.write("######################## For DATA SITE 4 #################################")
f_out.write("###########################################################################")


#Estimate the Paramters.
lamda = estimate_parameters("Exponential",data4)


print "lambda = {}\n".format(lamda)

f_out.write("lambda = {}\n".format(lamda))
lambda_interval = confidence_interval("Exponential", data4)

print "Confidence Interval for Lambda = {}".format(lambda_interval)

f_out.write("90 % Confidence Interval for Lambda = {}\n".format(lambda_interval))
######################################
######## Calculate x_k ###############
######################################

## Getting the Value of x_k by integrating the distibution
## function in interval 0 to x_k and equating with 0.99
## Got the Value of x_k from Matlab Code(Snippet Pasted in Report)

x_k = 11050
print "x_k value obtained by using matlab code(Snipped Pasted) for integration"
print "Value of x_k = {}".format(x_k)

f_out.write("Value of x_k = {}\n".format(x_k))

res = hypothesis_testing(data4)
if res[1] == 1:
    f_out.write("{} {}\n".format(H0,res[0]))
else:
    f_out.write("{} {}\n".format(HA,res[0]))
```

```
####Output Of Python Code


########################################################################
##
######################## For DATA SITE 1
##################################
########################################################################
##
est_mean = 2661.1202
est_stddev = 1005.77166728
Value of x_k = 5004.56818476
90 % CI for mean = [2494.1621032315356, 2828.078296768462]
90 % CI for variance = [902.3280099776418, 1147.9162358469366]
Test Statistic for Hypothesis Testing that mean discharge = 2500 is
1.60195604272mu = mu_0 = 2500 H0 cannot be rejected

########################################################################
##
######################## For DATA SITE 2
#################################
########################################################################
##
estimated mean = 2437.4885
estimated standard deviation = 981.933137767
Value of x_k = 4725.392711
90 % CI for mean = [2274.4875991306562, 2600.4894008693436]
90 % CI for variance = [880.9412742046759, 1120.7086340058231]
Test Statistic for Hypothesis Testing that mean discharge = 2500 is -
0.636616665592mu = mu_0 = 2500 H0 cannot be rejected

########################################################################
##
######################## For DATA SITE3 ##################################
########################################################################
##
Degree of freedom = 18.0
Confidence Interval for Std dev = [2253.9488488326087,
4020.0981716345786]

########################################################################
##
######################## For DATA SITE 4
#################################
########################################################################
##
lambda = 0.000416769052931
90 % Confidence Interval for Lambda = [0.0003482105437235746,
0.00048532756213776496]
Value of x_k = 11050
Test Statistic for Hypothesis Testing that mean discharge = 2500 is -
0.374701564649mu = mu_0 = 2500 H0 cannot be rejected
```

Goodness of Fit Test

Chi Square Test For Data Site 1

i	interval	observation	Z(lower range)	Z(Higher range)	Z(interval)	Cumulative Area	P	E	Oi^2 /Ei
1	476	0	-2.172580781	-inf	-inf ,-2.17258078068554	0.015	0.015	1.5	0
2	1442	12	-1.212124219	-2.172580781	-2.17258078068554 ,-1.21212421900887	0.1131	0.0981	9.81	14.67889908
3	1925	14	-0.7318959382	-1.212124219	-1.21212421900887 ,-0.731895938170539	0.2327	0.1196	11.96	16.38795987
4	2408	16	-0.2516676573	-0.7318959382	-0.731895938170539 ,-0.251667657332205	0.4013	0.1686	16.86	15.18386714
5	2891	16	0.2285606235	-0.2516676573	-0.251667657332205 ,0.228560623506129	0.5871	0.1858	18.58	13.77825619
6	3374	19	0.7087889043	0.2285606235	0.228560623506129 ,0.708788904344462	0.758	0.1709	17.09	21.12346401
7	3857	10	1.189017185	0.7087889043	0.708788904344462 ,1.1890171851828	0.881	0.123	12.3	8.130081301
8	5306	13	2.629702028	1.189017185	1.1890171851828 ,2.6297020276978	0.988	0.107	10.7	15.79439252
									105.0769201
Formula Used			(X-mu)/sigma			Value from Std.		E = n*P	
In Excel			(B(i) - 2661.1202)/1005.77167	D(i-1)	CONCATENATE(E2 & " ," & D2)	Normal Dist. Table	G(i)-G(i-1)	E = 100 * H(i)	POWER(C(i),2)/I(i)

Dof = 5

T = 105.07 - 100

5.07

T = Test Statistic

Critical Value of Rejection = 11.070

(Obtained from Chi- Square Distribution Table)

est_mean: 2661.1202

est_stddev: 1005.771667

H0:- Assumed Distribution is Normal Distribution

HA :- Data does not follow normal Distribution

As the Test Statistic is 5.07 which does not lie in the Rejection Region (Critical Value = 11.070) the Null Hypothesis cannot be rejected.

Intervals & Frequency Obtained using Matlab Code. For rest of the columns values calculated using formula or the distribution table

Z Higher Range Column is just included to obtain the Z interval Column in which we get the interval of Z.

Chi Square Test For Data Site 2

i	interval	observation	Z(lower range)	Z(Higher range)	Z(interval)	Cumulative Area	P	E	Oi^2 /Ei
1	434.2	0	-2.040147581	-inf	-inf ,-2.04014758112907	0.0207	0.0207	2.07	0
2	956.7	3	-1.508033953	-2.040147581	-2.04014758112907 ,-1.50803395339151	0.0668	0.0461	4.61	1.952277657
3	1479.2	9	-0.9759203257	-1.508033953	-1.50803395339151 ,-0.975920325653947	0.166	0.0992	9.92	8.165322581
4	2001.7	23	-0.4438066979	-0.9759203257	-0.975920325653947 ,-0.443806697916385	0.33	0.164	16.4	32.25609756
5	2524.3	25	0.08840876975	-0.4438066979	-0.443806697916385 ,0.0884087697499322	0.5319	0.2019	20.19	30.95591877
6	3046.8	19	0.6205223975	0.08840876975	0.0884087697499322 ,0.620522397487495	0.7324	0.2005	20.05	18.00498753
7	3569.3	9	1.152636025	0.6205223975	0.620522397487495 ,1.15263602522506	0.8749	0.1425	14.25	5.684210526
8	5659.3	12	3.281090536	1.152636025	1.15263602522506 ,3.28109053617531	0.9995	0.1246	12.46	11.55698234
									108.575797
Formula Used			(X-mu)/sigma			Value from Std.		E = n*P	
In Excel			(B(i) - 2437.4885)/981.9331	D(i-1)	CONCATENATE(E2 & " ," & D2)	Normal Dist. Table	G(i)-G(i-1)	E = 100 * H(i)	POWER(C(i),2)/I(i)

Dof = 5

T = 108.57 - 100

8.57

Critical Value of Rejection = 11.070

(Obtained from Chi- Square Distribution Table)

H0:- Assumed Distribution is Normal Distribution

HA :- Data does not follow normal Distribution

As the Test Statistic is 8.75 which does not lie in the Rejection Region (Critical Value = 11.070) the Null Hypothesis cannot be rejected.

Intervals & Frequency Obtained using Matlab Code. For rest of the columns values calculated using formula or the distribution table

Z Higher Range Column is just included to obtain the Z interval Column in which we get the interval of Z.

est_mean: 2437.4885

est_stddev: 981.9331378

Chi Square Test For Data Site 4

i	interval	observation	Z(lower range)	Z(Higher range)	Z(interval)	Cumulative Area	P	E	Oi^2 /Ei
1	13.2	0	-0.8888806109	-inf	-inf ,-0.888880610914509	0.2177	0.2177	21.77	0
2	1563.5	54	-0.3113801453	-0.8888806109	-0.888880610914509 ,-0.31138014527845	0.3783	0.1606	16.06	181.5691158
3	3113.9	17	0.2661575712	-0.3113801453	-0.31138014527845 ,0.266157571242317	0.6026	0.2243	22.43	12.88452965
4	4664.2	15	0.8436580369	0.2661575712	0.266157571242317 ,0.843658036878376	0.7995	0.1969	19.69	11.42712037
5	6214.5	7	1.421158503	0.8436580369	0.843658036878376 ,1.42115850251443	0.9222	0.1227	12.27	3.993480033
6	15517	7	4.886422053	1.421158503	1.42115850251443 ,4.88642205252375	1	0.0778	7.78	6.298200514
									216.1724464
Formula Used			(X-mu)/sigma			Value from Std.		E = n*P	
In Excel			(B(i) - 2624.1)/1189.8	D(i-1)	CONCATENATE(E2 & " ," & D2)	Normal Dist. Table	G(i)-G(i-1)	E = 100 * H(i)	POWER(C(i),2)/I(i)

Dof = 6 - 1 - 2 = 3

T = 216.172- 100

116.172

Critical Value of Rejection = 7.815

(Obtained from Chi- Square Distribution Table)

H0:- Assumed Distribution is Normal Distribution

HA :- Data does not follow normal Distribution

As the Test Statistic is 116.172 which lies in the Rejection Region (Critical Value = 7.8154) the Null Hypothesis is rejected.

Intervals & Frequency Obtained using Matlab Code. For rest of the columns values calculated using formula or the distribution table

Z Higher Range Column is just included to obtain the Z interval Column in which we get the interval of Z.

est_mean: 2399.4

est_stddev: 2684.5