**DISTRIBUTIONAL ANALYSIS REPORT**

Name-Rajrishi Sarkar

Roll-203110061

# Data description

Metallic Glass Descriptors:

Relevant papers:

There is no paper published on this data set at the present time. The data was assembled primarily by Vanessa Nilsen under the guidance of Prof. Dane Morgan at UW Madison (ddmorgan@wisc.edu). A previous study of reduced glass transition temperature as a GFA descriptor can be found in reference [5]: https://www.sciencedirect.com/science/article/pii/S0022309300000648

X features:

The metallic glass dataset gives two columns with information about the material Composition. The first is the overall composition, and the second is the highest Composition element. The columns from four to the end are the MAGPIE features that have been generated from the material composition column and give values such as properties averaged over the material composition as well as features that are only for the majority element in each alloy [3]. The majority element features are labelled as "site1".

Y property:

The reduced glass transition temperature (Trg) has historically been used as a rough predictor for Glass Forming Ability (GFA). By making a model to predict Trg for an arbitrary alloy, it could be possible to use these values to estimate GFA directly, or as input for another model to then predict GFA.

**Lets analyze the distribution of each of the features one by one**

# Glass forming temperature (target)

# 

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: 0.223 max: 0.688

median: 0.584

mean: 0.577388

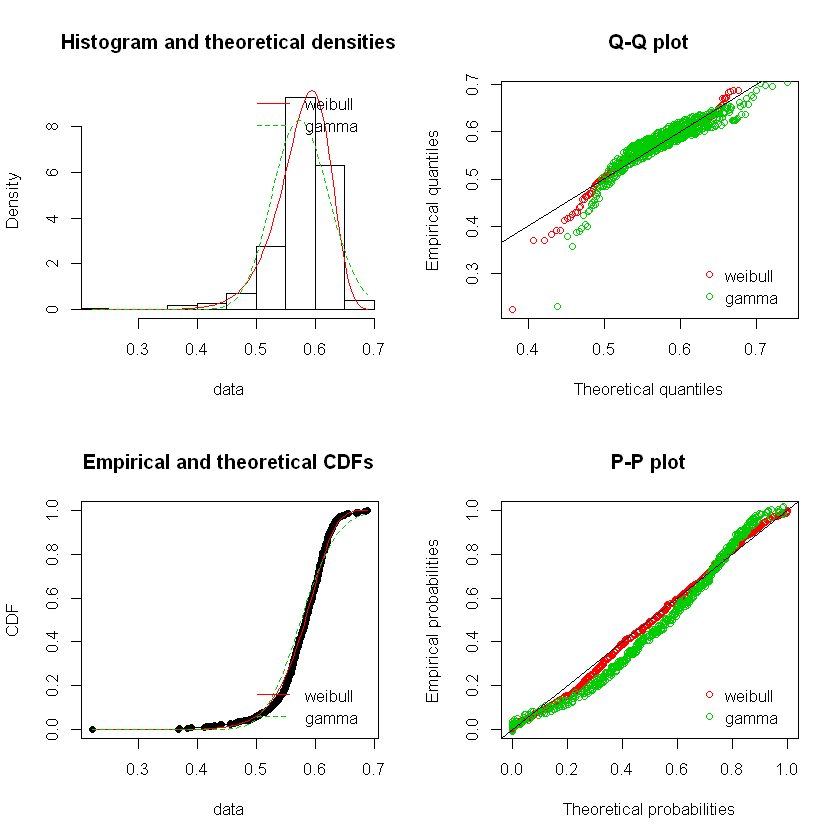
estimated sd: 0.04834414

estimated skewness: -1.759822

estimated kurtosis: 10.12794

**the observation lies below the lognormal curve, we will try to fit the distributions to the data.**

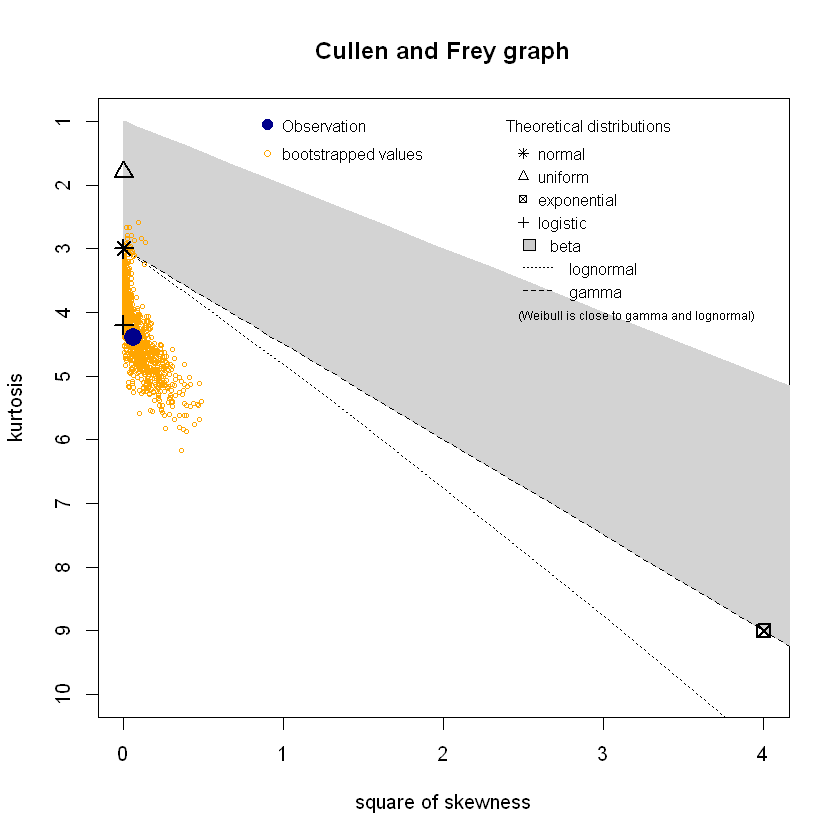
**Fitting distributions:**



both the distributions seem to fit pretty good comparing the AIC values of both the distributions. WEIBULL DIST(AIC)=-2004.26; GAMMA DIS(AIC)=-1794.17 due to low AIC gamma dist is a better fit for the target feature.

# 2) Density composition average

# ****visualizing the skew kurtosis plot****



**summary statistics**

min: -2.296437 max: 4.418691

median: 0.1412969

mean: 1.709401e-12

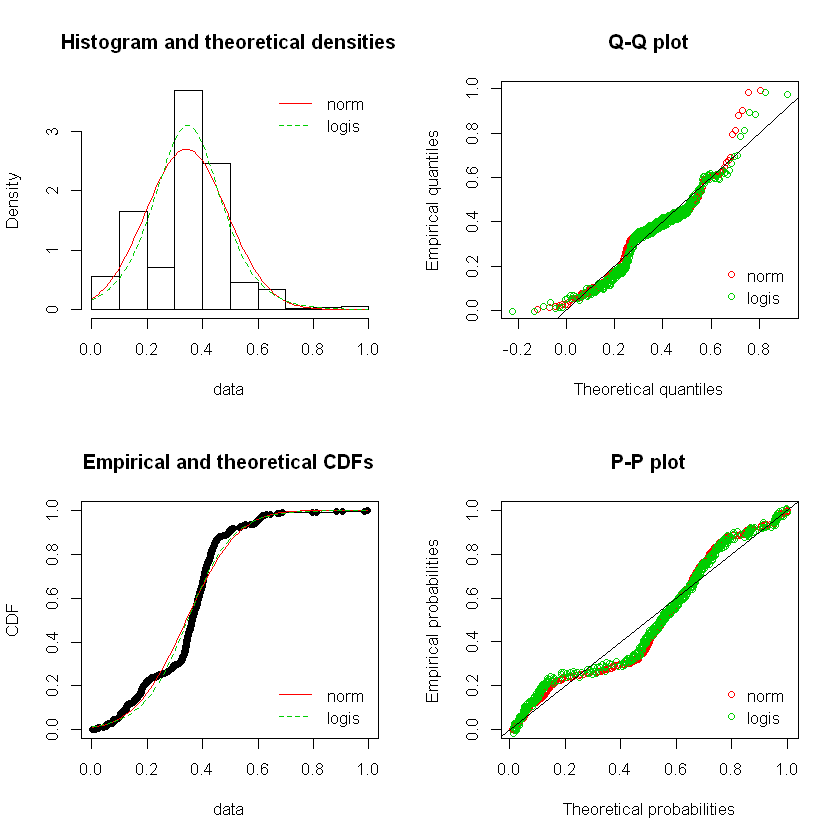
estimated sd: 1.000856

estimated skewness: 0.2426468

estimated kurtosis: 4.387157

the observation lies near the logistic distribution point and some bootstrap samples also lies near normal distribution

**fitting distributions:**



Normal(AIC)=575.87 ; Logistic(AIC)=-596.3

Logistic is a better fit

# IsBoron composition average

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -0.5856 max: 6.479041

median: -0.5856

mean: 3.589747e-11

estimated sd: 1.000856

estimated skewness: 3.714128

estimated kurtosis: 22.47165

the observation clearly falls under the beta distribution

**fitting distributions**

# 

# judging from the aic values BETA DISTRIBUTION is suitable for this particular distribution

# IsDBlock composition average

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -2.091415 max: 1.029882

median: 0.3744096

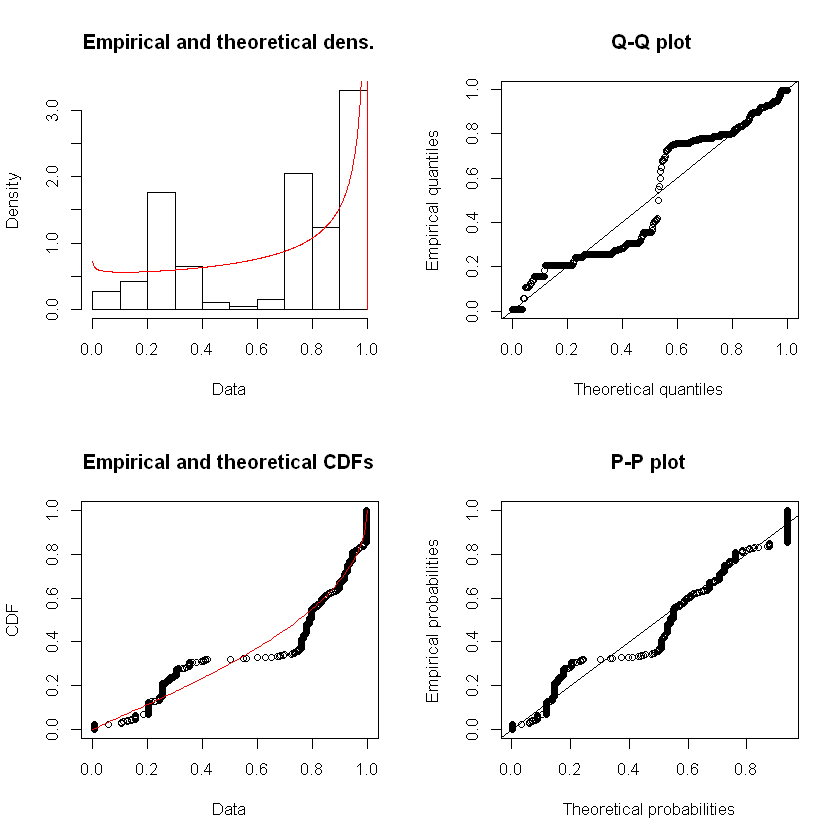
mean: 1.36752e-11

estimated sd: 1.000856

estimated skewness: -0.6624578

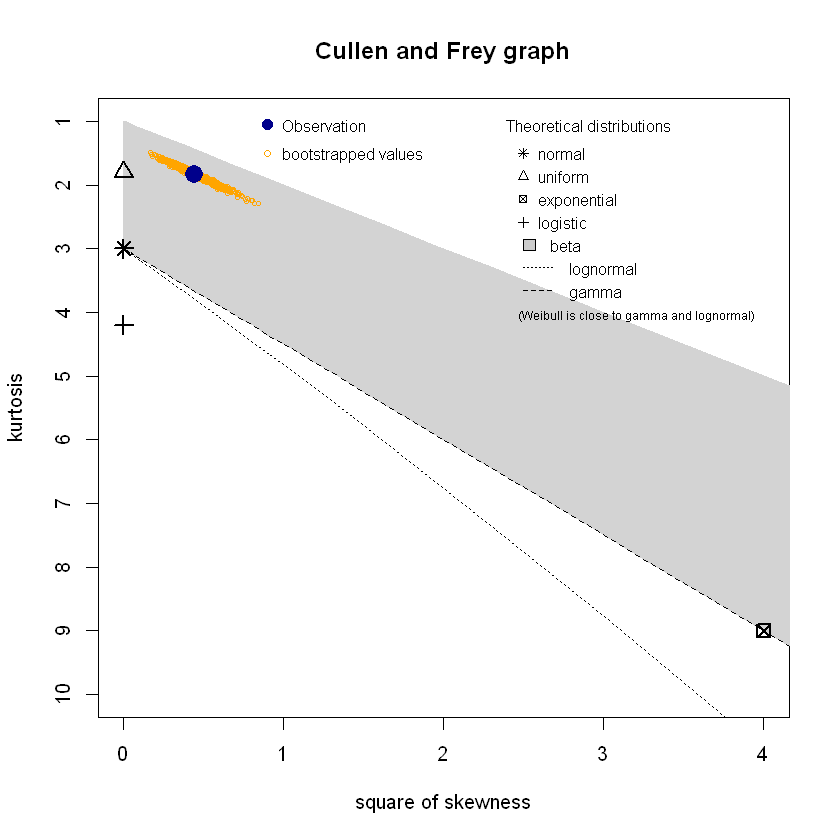
estimated kurtosis: 1.828959

**fitting beta distribution**



# 5) IsTransitionMetal composition average

# ****visualizing the skew kurtosis plot****



**summary statistics**

min: -2.091415 max: 1.029882

median: 0.3744096

mean: 1.36752e-11

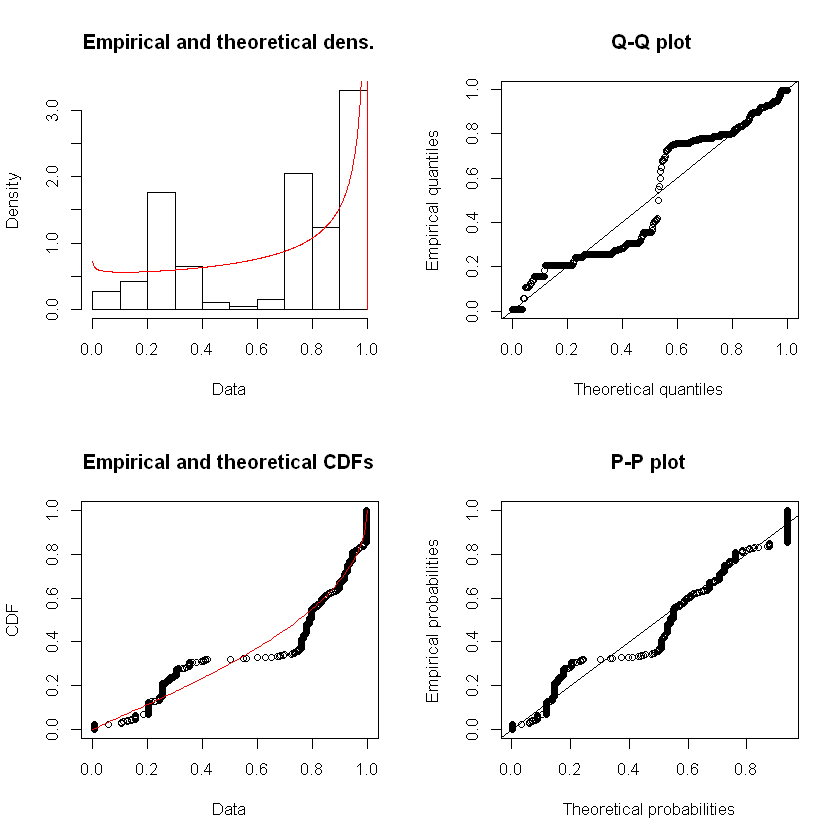
estimated sd: 1.000856

estimated skewness: -0.6624578

estimated kurtosis: 1.828959

the observation lies in the beta region

**fitting beta distribution**



# 6) NValance composition average

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.922857 max: 6.27228

median: 0.05548036

mean: 2.564103e-11

estimated sd: 1.000856

estimated skewness: 1.637832

estimated kurtosis: 9.921015

**fitting to lognormal distribution**

# 

# 7) HeatVaporization max value

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.922857 max: 6.27228

median: 0.05548036

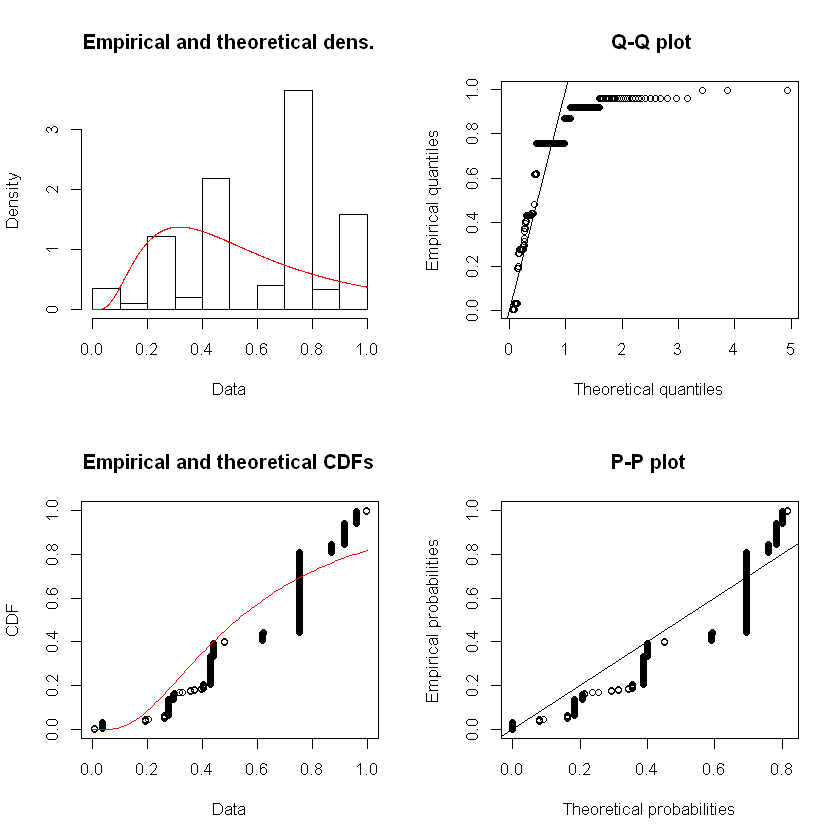
mean: 2.564103e-11

estimated sd: 1.000856

estimated skewness: 1.637832

estimated kurtosis: 9.921015

**fitting the lognormal distribution**



# 8) BoilingT difference

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -2.118943 max: 3.247621

median: 0.1006106

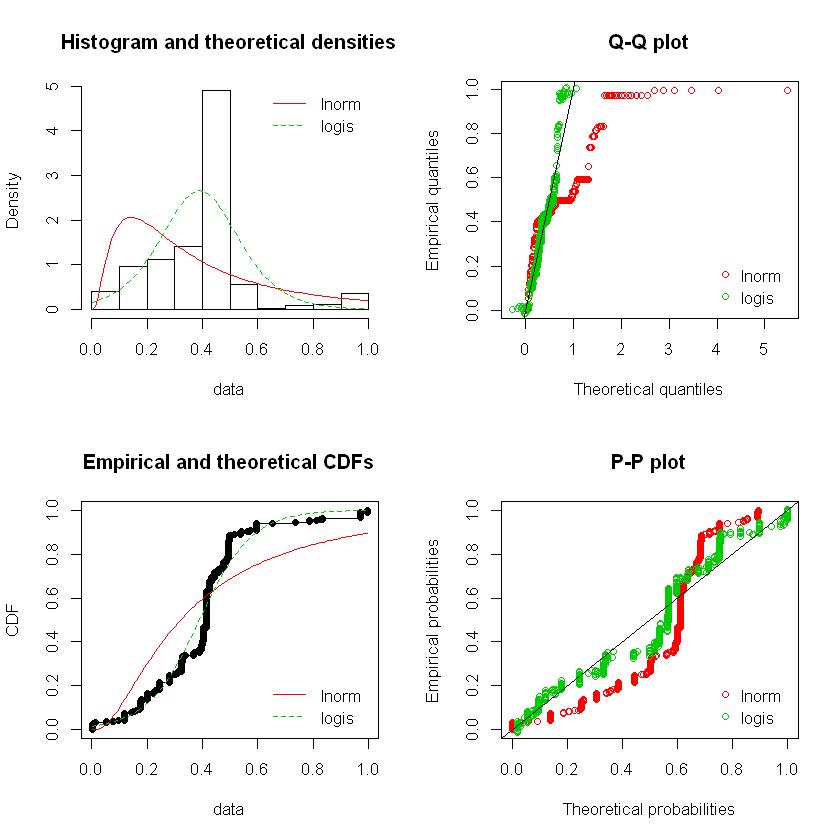
mean: -4.444444e-11

estimated sd: 1.000856

estimated skewness: 0.7919673

estimated kurtosis: 5.295467

**fitting to distributions**



# judging from the observations the LOGISTIC DIST seems to fit better

# 9) HeatVaporization difference

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.844594 max: 3.084747

median: 0.1760586

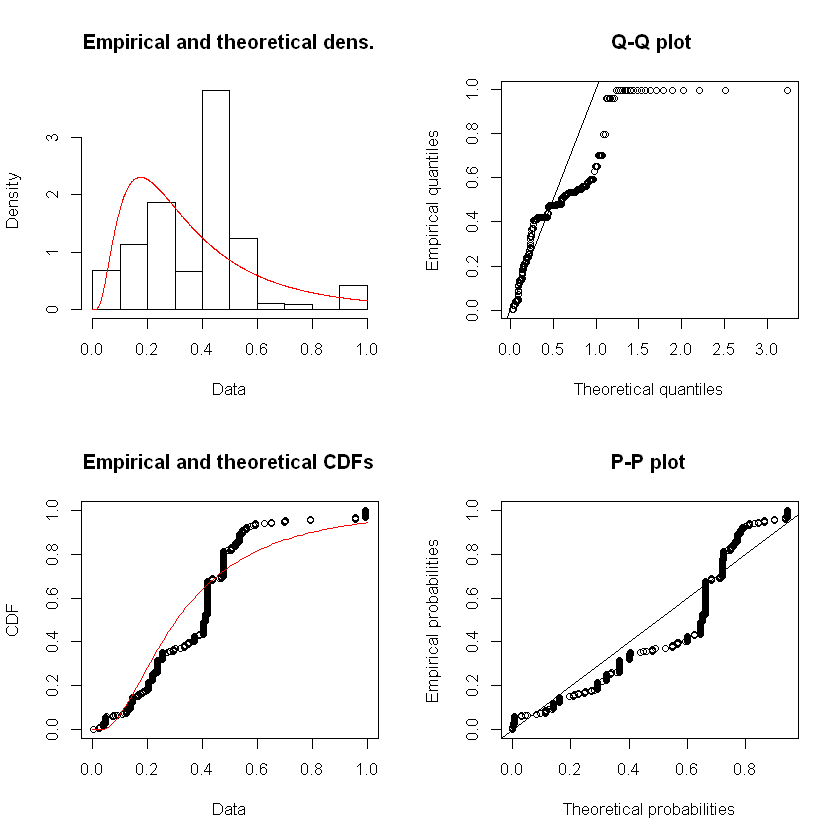
mean: -4.273503e-11

estimated sd: 1.000856

estimated skewness: 0.8393281

estimated kurtosis: 4.674056

**fitting to lognormal distribution**



# 10) MeltingT\_difference

# 

**summary statistics**

min: -1.435136 max: 3.822408

median: -0.2355999

mean: -8.888888e-11

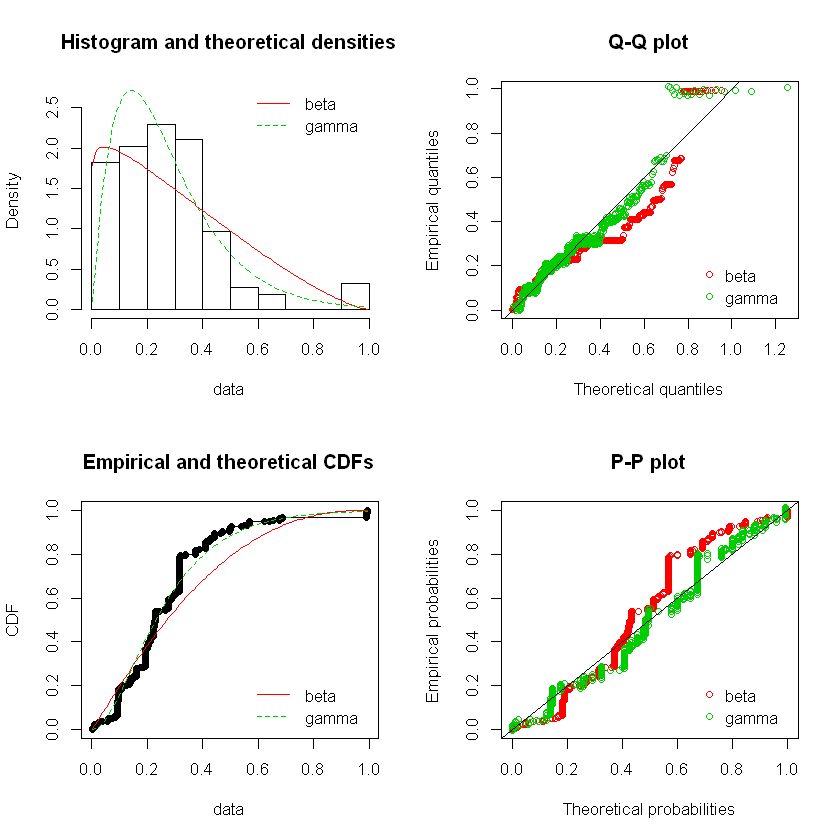
estimated sd: 1.000856

estimated skewness: 1.862502

estimated kurtosis: 7.748327

observation lies in the beta region, bootstrapped values are hitting the gamma distribution line

**fitting the distributions**



GAMMA DIST(AIC=-497) is a better fit

# 11) NdValence difference

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.654388 max: 1.005911

median: 0.4738515

mean: 5.128207e-11

estimated sd: 1.000856

estimated skewness: -0.6532023

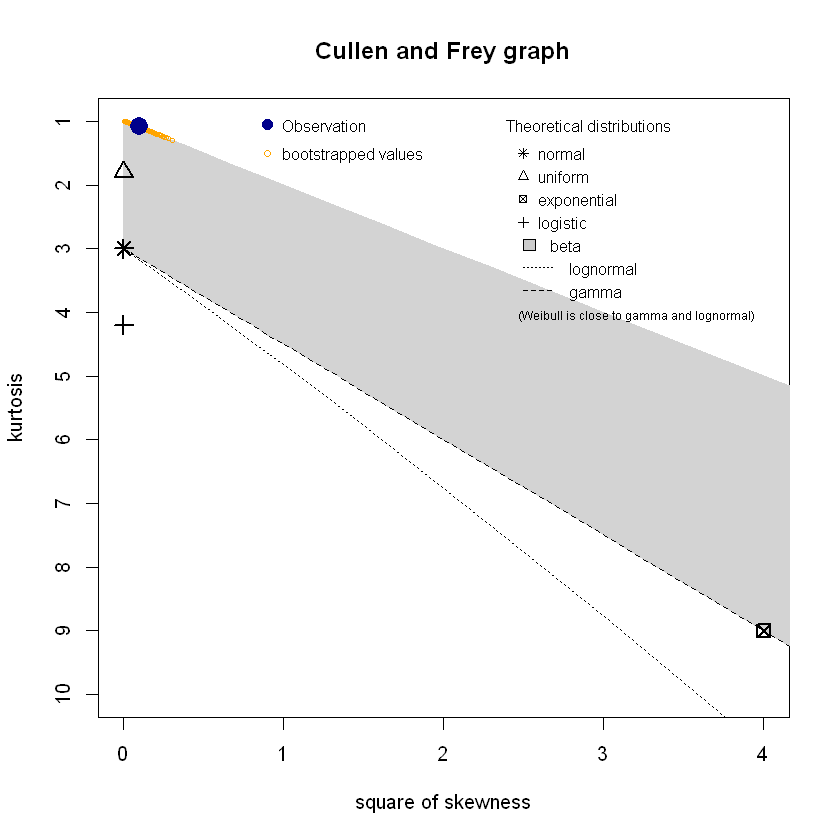
estimated kurtosis: 1.898246

# ****fitting the beta distribution****

# 

# 12) NsUnfilled difference

# ****visualizing the skew kurtosis plot****



**summary statistics**

min: -1.165706 max: 0.8578491

median: 0.8578491

mean: -2.051282e-10

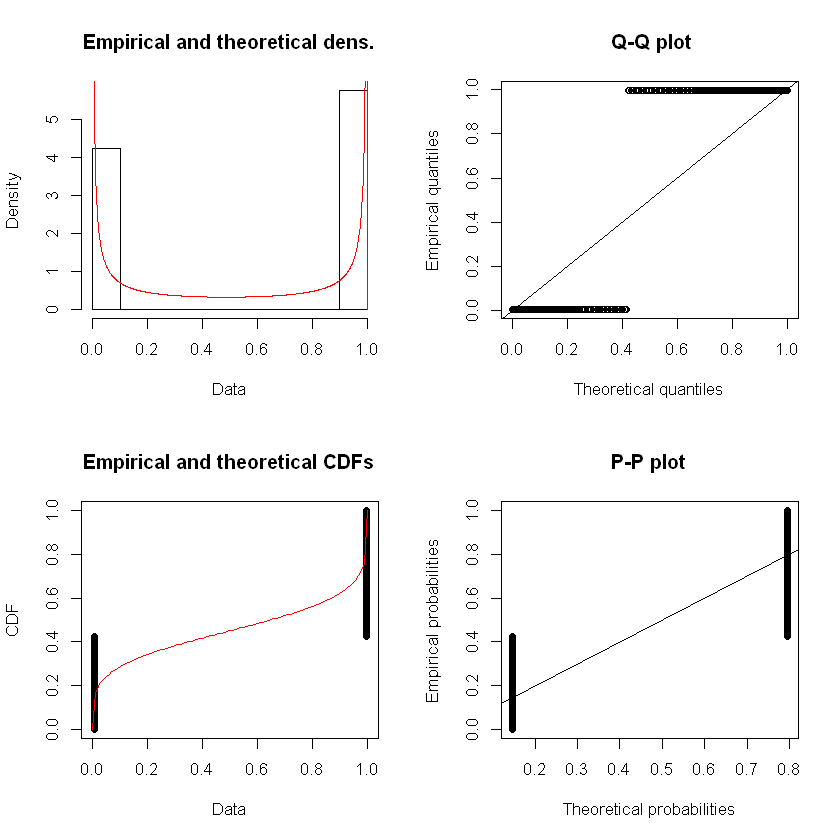
estimated sd: 1.000856

estimated skewness: -0.3086491

estimated kurtosis: 1.088718

looking at the observation we have no choice but to go with the beta distribution

**fitting the beta distribution**



# 13) valence difference

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.745035 max: 2.274041

median: 0.2645033

mean: -6.666671e-11

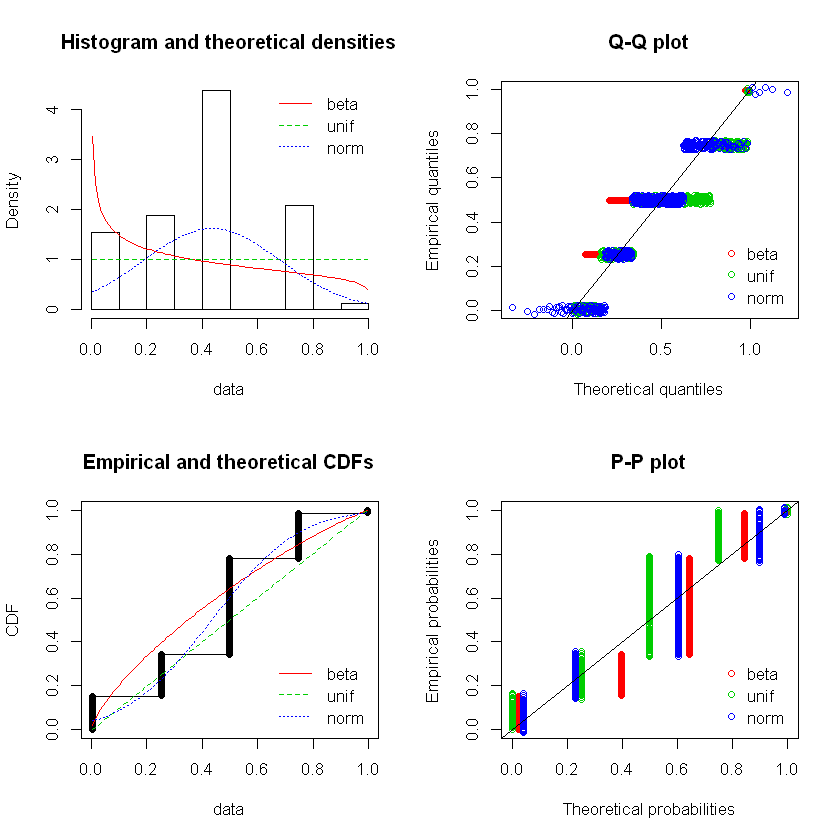
estimated sd: 1.000856

estimated skewness: -0.3193563

estimated kurtosis: 2.351189

the observation may fit to normal, uniform or beta distribution

**fitting the distributions**



BETA DIST-97.574050182686

UNIFORM DIST-7.64188709820641

NORMAL DIST-24.9852226573406

comparing the aic values we see that BETA DIST is best fit for the feature

# 14) Site1 Density

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.468652 max: 4.127357

median: 0.09502333

mean: -1.094017e-10

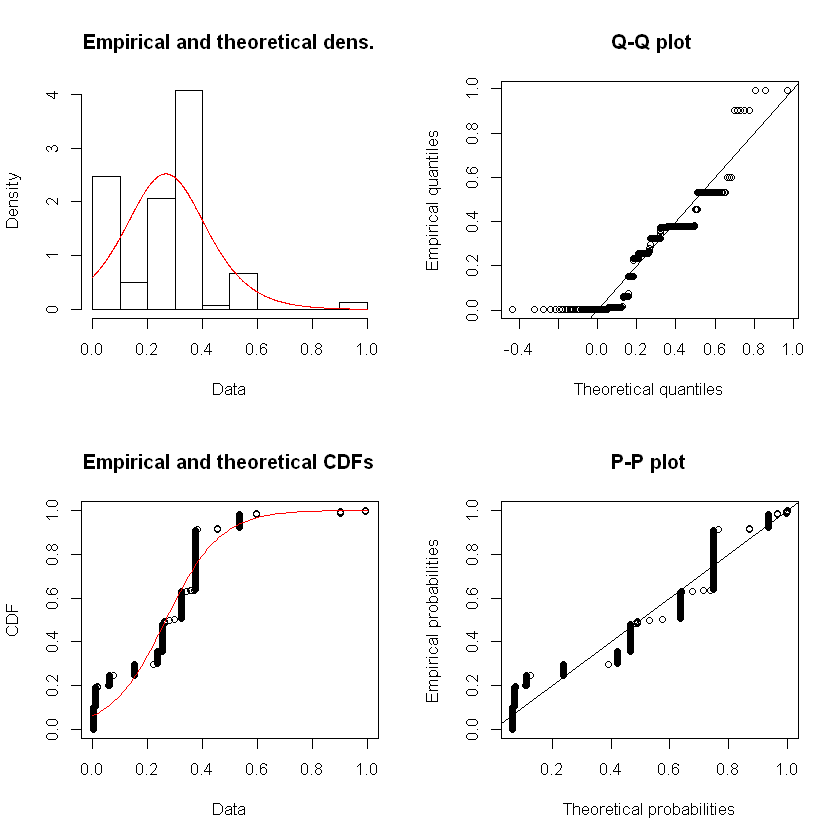
estimated sd: 1.000856

estimated skewness: 0.4281156

estimated kurtosis: 4.388066

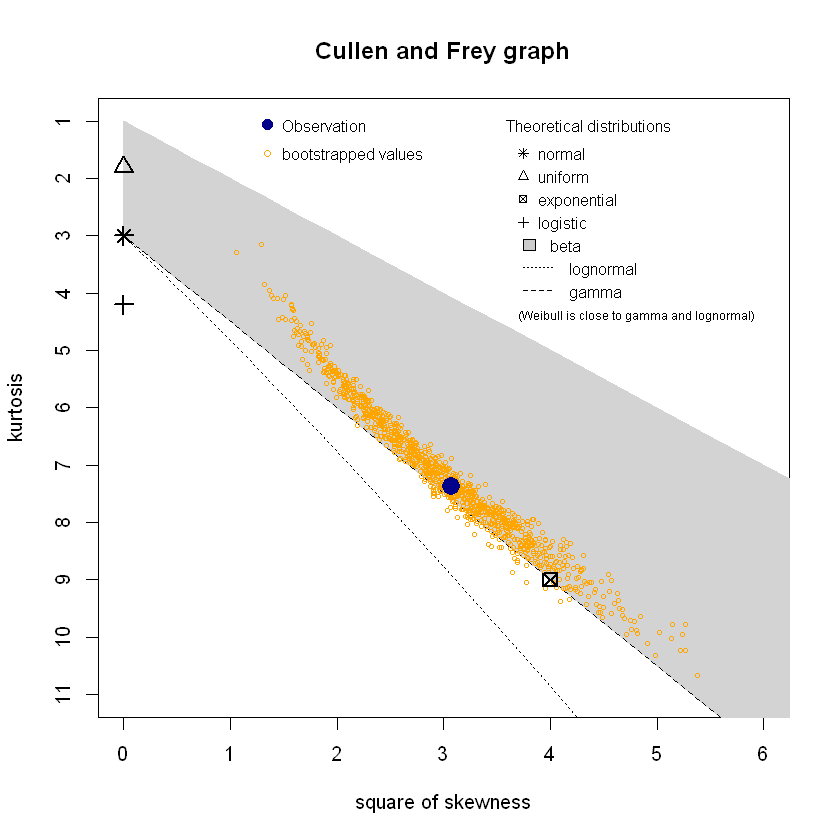
observation lies near logistic distribution

**fitting the logistic distribution**



# 15) Site1 HeatCapacityMass

# ****visualizing the skew kurtosis plot****



**summary statistics**

min: -1.279271 max: 4.995841

median: -0.3320846

mean: -5.982907e-11

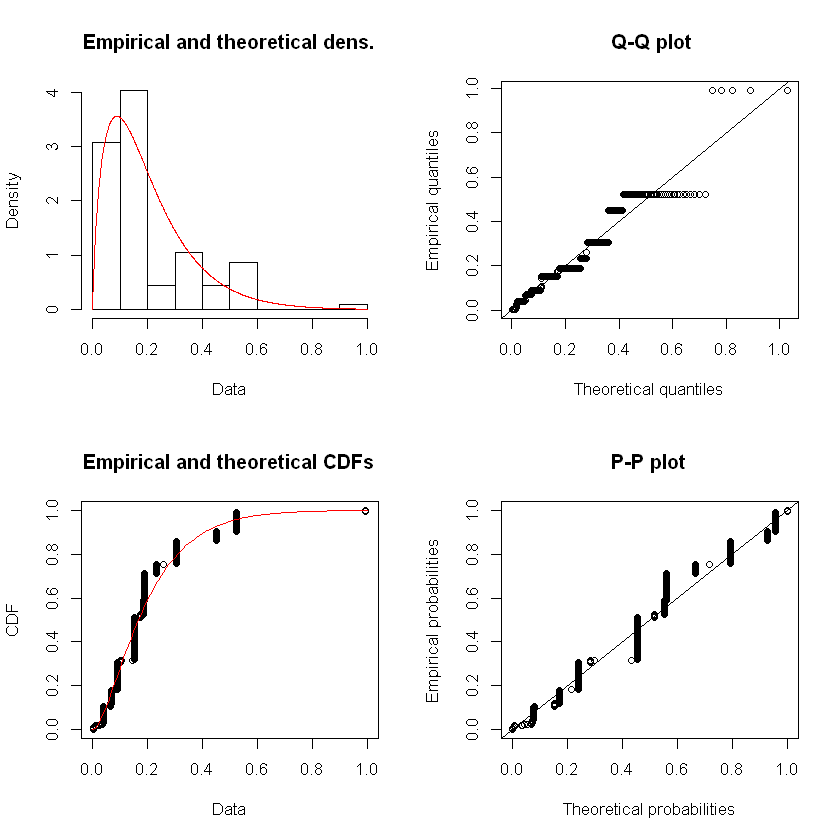
estimated sd: 1.000856

estimated skewness: 1.750919

estimated kurtosis: 7.361955

the observation clearly lies in the gamma distribution line

**fitting gamma distribution**



# 17) Site1 HeatFusion

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.740832 max: 3.664525

median: 0.05214307

mean: -4.444444e-11

estimated sd: 1.000856

estimated skewness: 0.3596201

estimated kurtosis: 2.688689

as per the graph the observation is close to uniform and normal distributions,it also lies in the beta region

**fitting distributions**

# 

1. BETA DIST = -363.557
2. UNIFORM DIST=-7.64
3. NORMAL DIST=-321.73

BETA DISTRIBUTION is the best fit as per the aic values

# 18) Site1 IsDBlock

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.392731 max: 0.718014

median: 0.718014

mean: -3.333333e-10

estimated sd: 1.000856

estimated skewness: -0.6764524

estimated kurtosis: 1.452285

# ****fitting beta distribution****

# 

# 19) Site1 IsTransitionMetal

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.392731 max: 0.718014

median: 0.718014

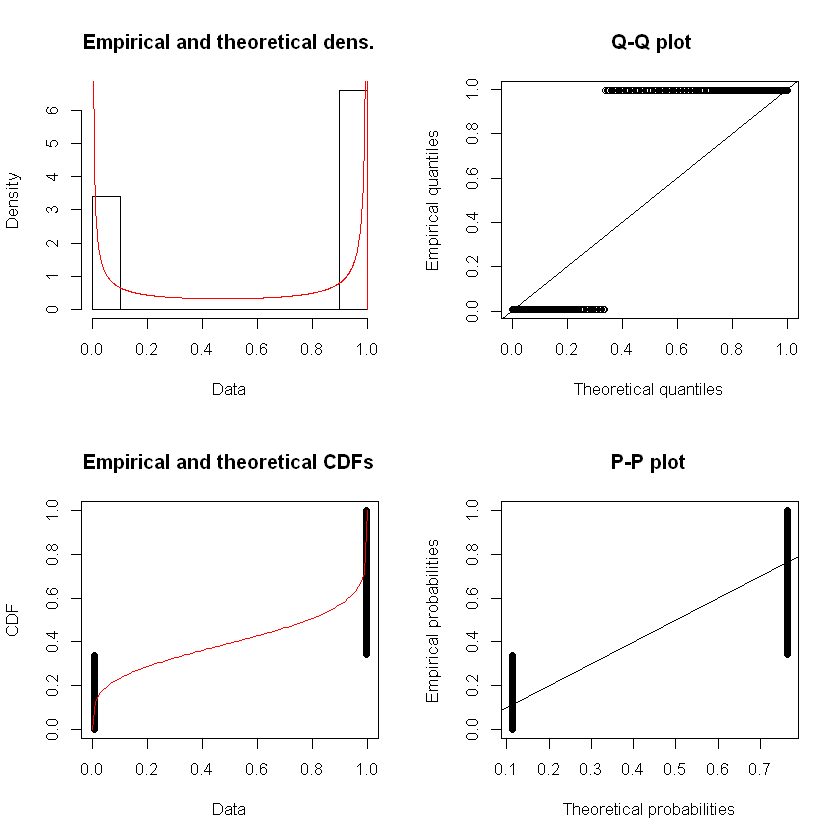
mean: -3.333333e-10

estimated sd: 1.000856

estimated skewness: -0.6764524

estimated kurtosis: 1.452285

**fitting the beta distribution**



# 20) Site1 NdValence

# ****visualizing the skew kurtosis plot****

# 

**summary statistics**

min: -1.106354 max: 1.31768

median: -0.6215474

mean: 9.74359e-11

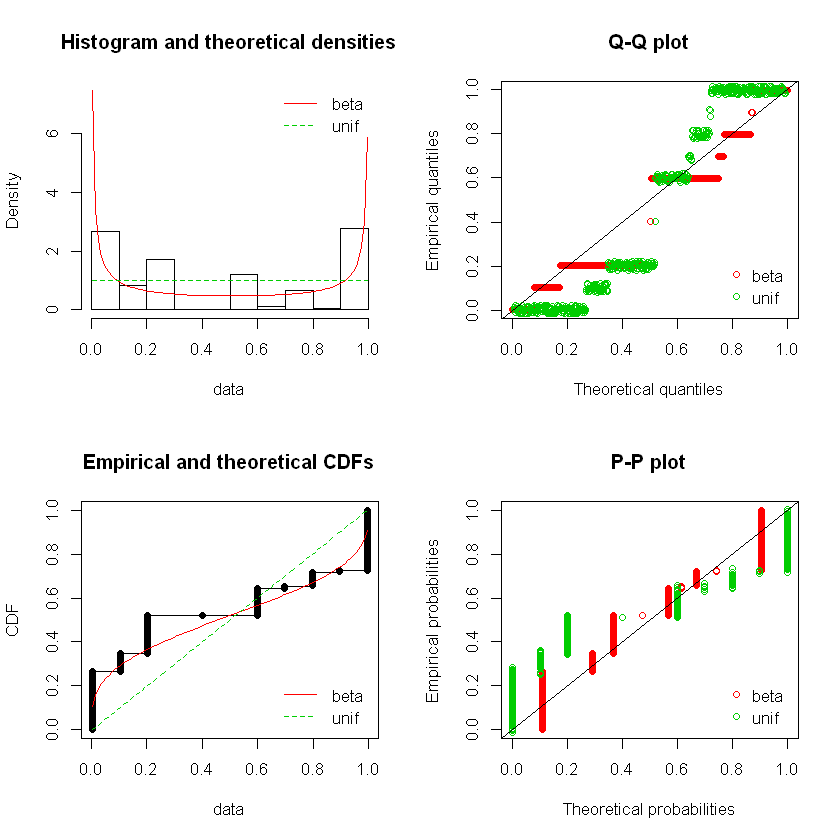
estimated sd: 1.000856

estimated skewness: 0.2285465

estimated kurtosis: 1.335521

the observation falls in the beta range close to uniform distribution

**fitting the distributions**



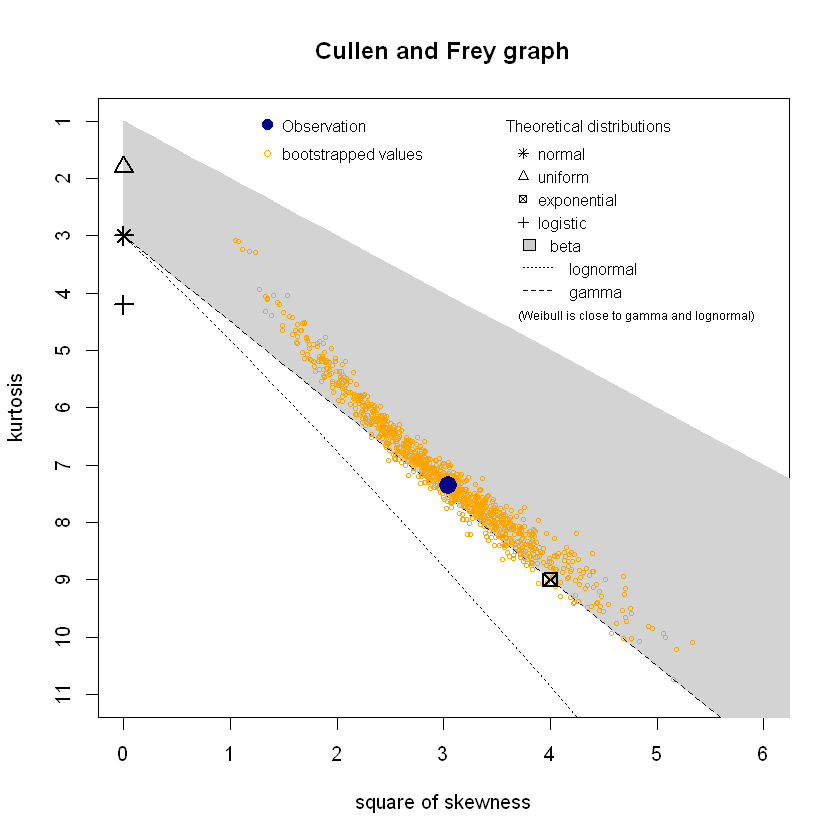
BETA DISTRIBUTION=-937.267450601306

UNIFORM DISTRIBUTION=-7.64188709820667

BETA DIST(AIC) has a lower value therefore it is a better fit

# 21) Site1 SpecificHeatCapacity

# ****visualizing the skew kurtosis plot****



**summary statistics**

min: -1.280819 max: 4.9969

median: -0.3300981

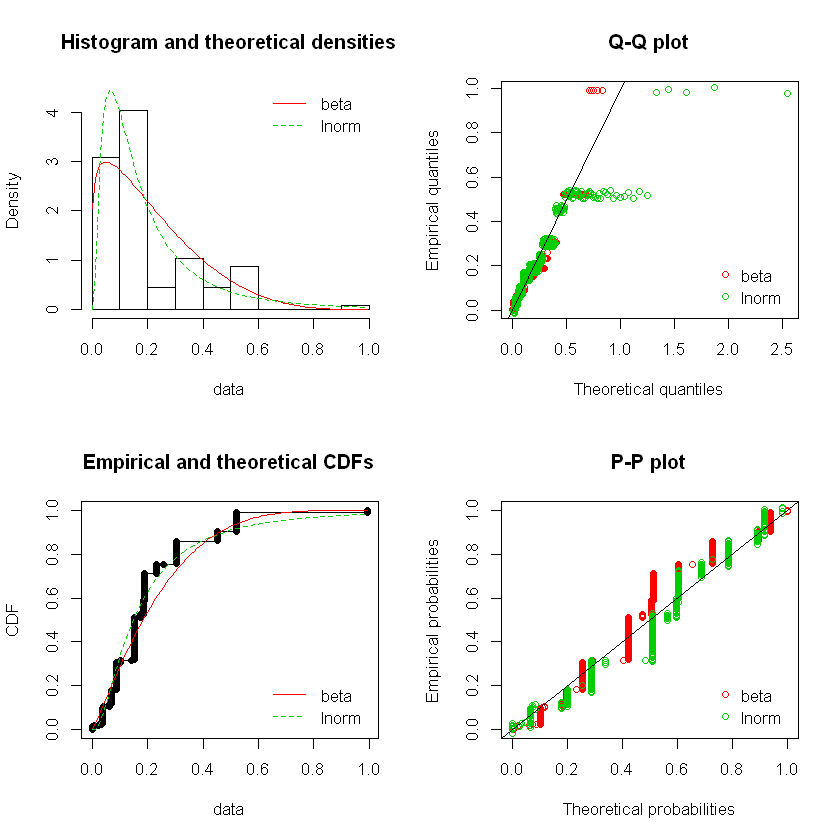
mean: -7.521366e-11

estimated sd: 1.000856

estimated skewness: 1.743029

estimated kurtosis: 7.352229

**fitting the distributions**



BETA DISTRIBUTION=-658.966198239053

LOGNORMAL DISTRIBUTION=-680.175837919245

LNORM DIST(AIC) which is lower and therefore is a better fit