HYD298 - Homework #3

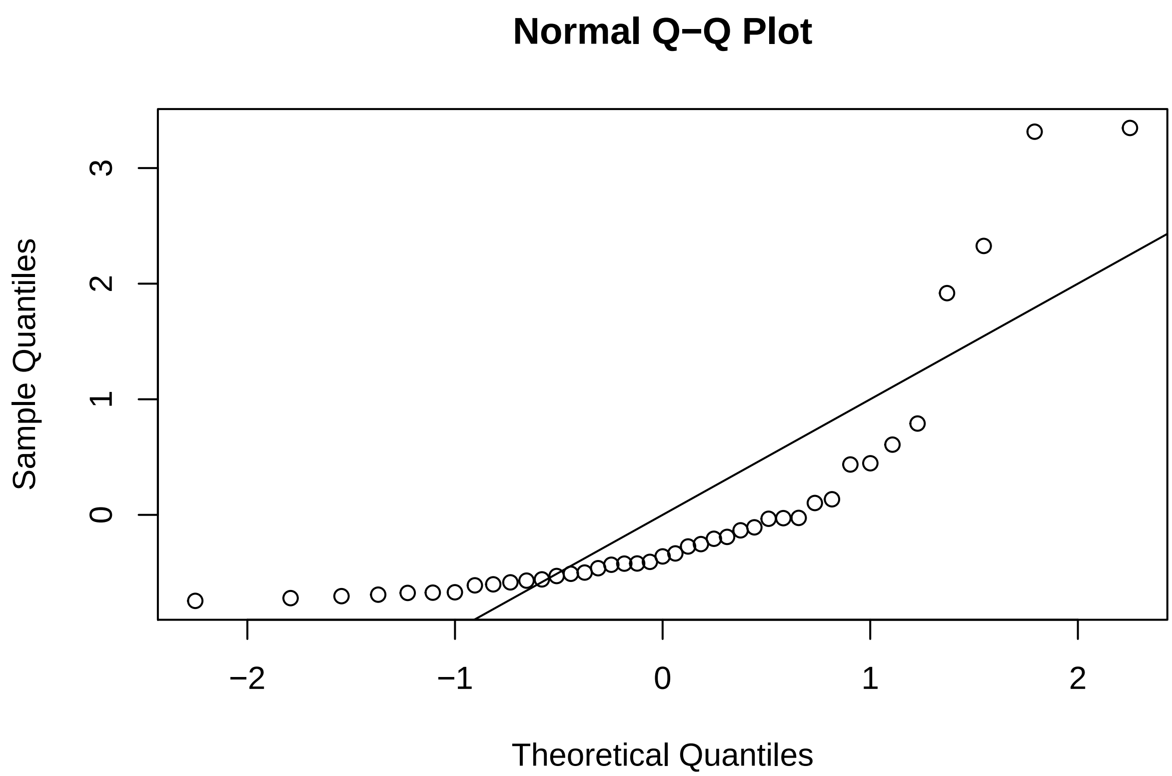
Andrew Brown

January 25, 2016

Fit a normal distribution to the Sebou River dataset

```
foo=read.csv("../HW1/sebou.csv")

#Fit a normal distribution
#normalize data with first and second moment'
xbar_norm=mean(foo$maxQ)
ssd_norm=sd(foo$maxQ)
z.norm<-(foo$maxQ-xbar_norm)/ssd_norm
#plot normalized data versus normal reference line
qqnorm(z.norm)
abline(0,1)
```



Clearly, the normal distribution is inappropriate for the Sebou River dataset. Confirm with a statistical test:

```
#Statistical test to confirm the visually observed misfit
shapiro.test(foo$maxQ)
```

```
##
```

```
##  Shapiro-Wilk normality test
##
## data:  foo$maxQ
## W = 0.66636, p-value = 2.173e-08
```

Reject null hypothesis of normally distributed data.

Normal distribution mean:

```
xbar_norm
```

```
## [1] 1004.488
```

Normal distribution variance:

```
ssd_norm^2
```

```
## [1] 866469.3
```

Quantiles 1% and 99%:

```
qnorm(mean=xbar_norm,sd=ssd_norm,p=c(0.01,0.99))
```

```
## [1] -1160.978  3169.953
```

Note that the 1% quantile is inappropriate for a dataset that necessarily is non-negative. Symmetry of the normal distribution results in tails that correspond to flows below zero.

Fit a 2-parameter Lognormal by moments and MLE

```
#Lognormal
lmaxq=log(foo$maxQ) #log-transform data vector
```

Maximum Likelihood Estimators (Mean):

```
mlemuhat=sum(lmaxq)/length(lmaxq)
mlemuhat
```

```
## [1] 6.648153
```

Maximum Likelihood Estimators (Variance and Std. Dev.):

```
mlevarhat=sum((lmaxq-mlemuhat)^2)/length(lmaxq)
mlevarhat
```

```
## [1] 0.4384729
```

```
sqrt(mlevarhat)
```

```
## [1] 0.6621729
```

Method of Moments (Variance):

```
lvarhat=log(1+ssd_norm^2/xbar_norm^2)
lvarhat
```

```
## [1] 0.6199011
```

Method of Moments (Mean):

```
lmuhat=log(xbar_norm)-0.5*lvarhat
lmuhat
```

```
## [1] 6.602282
```

The estimates bt MLE and MoM are

Quantiles 1% and 99%:

```
qlnorm(meanlog=lmuhat,sdlog=sqrt(lvarhat),p=c(0.01,0.99))
```

```
## [1]  117.9971 4600.4276
```

The lognormal distribution gives more physically realistic quantiles. There is not a 1:1 fit between the data and the lognormal but that is expected given the large uncertainty at the limits of the data range and that the 2parameter lognormal an approximation of a natural gradient. A log transformation produces a more appropriate fit. However, with small sample sizes (i.e. n=43) the results for predictive purposes may vary.

```r
z.lnorm<-(lmaxq-lmuhat)/sqrt(lvarhat)
#plot normalized data versus normal reference line
qqnorm(z.lnorm)
abline(0,1)
```