# HYD298 - Homework #6

*Andrew Brown*

*February 17, 2016*

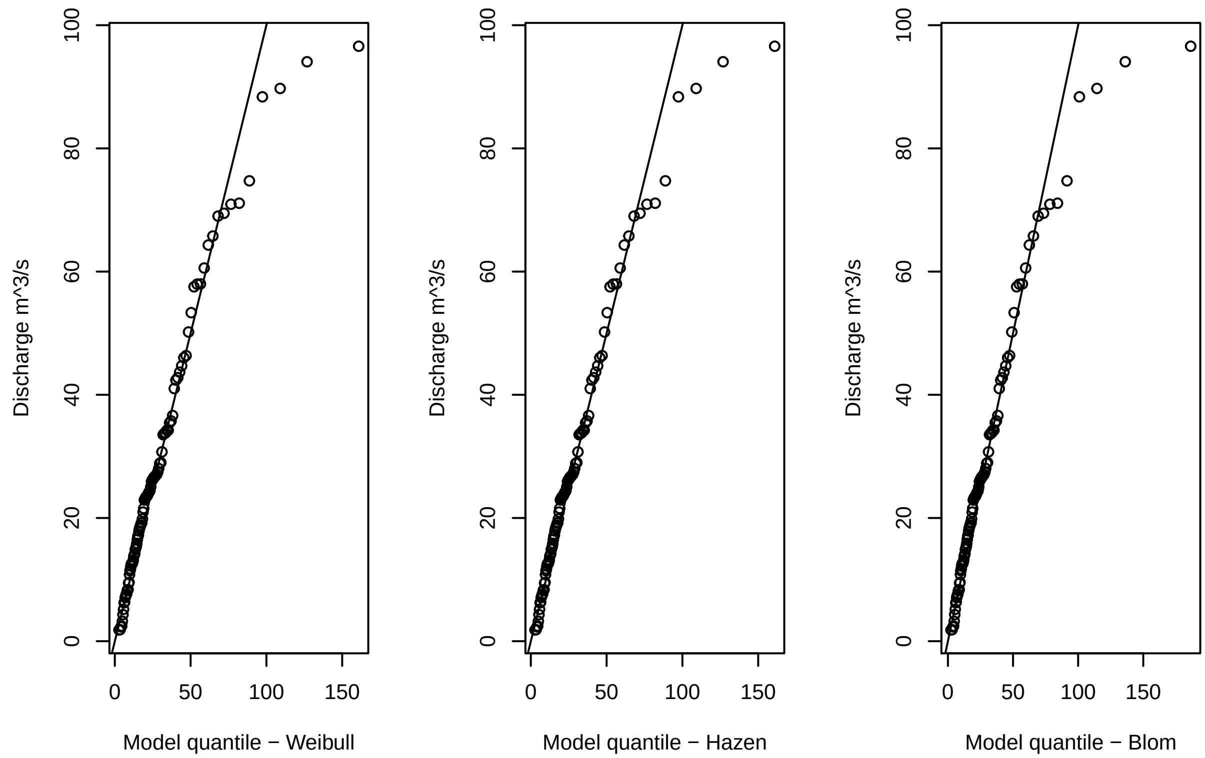
## Q 7.16

Define functions to calculate Weibull, Hazen and Blom positions and the variance of the ith largest observation:

```
ppwei=function(i,n)  { i/(n+1) }
pphaz=function(i,n)  { (i-(0.5))/(n) }
ppblom=function(i,n) { (i-(3/8))/(n+(1/4)) }
varpi=function(i,n)  { (i*(n-i-1))/(((n+1)^2)*(n+2)) }
```

Generate a random lognormal with n=100 observations logmean=3 logsd=1; sort it from largest->smallest:

```
lll=rlnorm(n=100,meanlog=3,sdlog=1)
lll=sort(lll,decreasing=TRUE)
nn=length(lll)
wei=ppwei(seq(1,length(lll)),length(lll))
haz=pphaz(seq(1,length(lll)),length(lll))
blom=ppblom(seq(1,length(lll)),length(lll))
sddd=(varpi(seq(1,length(lll)),length(lll)))
pwei=ppwei(1:nn,nn)
phaz=pphaz(1:nn,nn)
pblom=ppblom(1:nn,nn)
qiwei=sort(qlnorm(pwei,meanlog=mean(log(lll)),sdlog=sd(log(lll))),decreasing=TRUE)
qihaz=sort(qlnorm(pwei,meanlog=mean(log(lll)),sdlog=sd(log(lll))),decreasing=TRUE)
qiblom=sort(qlnorm(pblom,meanlog=mean(log(lll)),sdlog=sd(log(lll))),decreasing=TRUE)
```




```
mqwei=max(qiwei)
mqhaz=max(qihaz)
mqblo=max(qiblom)
data.frame(Weibull=mqwei,Hazen=mqhaz,Blom=mqblo,Actual=max(lll))
```

```
##    Weibull    Hazen     Blom   Actual
## 1 160.8872 160.8872 186.4083 96.58276
```

Comparison of the expected value for the largest observation using the three plotting positions and the actual largest extreme value. In general, these plotting positions fail to predict the magnitude of the 100 year flood (largest of 100 values in dataset). In general, plotting positions do not differ greatly from one another. However, they are all crude estimates of the exceedence probabilities of extreme values (e.g. smallest or largest obs). Further, with generation of multiple random lognormal datasets, they do not consistently under or overestimate the larges/smallest value.

```
data.frame(Min=0.29/100,Max=1.38/(100+2),Weibull=1-max(pwei),Hazen=1-max(phaz),Blom=1-max(pblom),StDev=
```

```
##      Min        Max    Weibull Hazen        Blom        StDev
## 1 0.0029 0.01352941 0.00990099 0.005 0.006234414 9.418531e-05
```

The actual exceedance probability for the largest observation lies between the minimum and maximum values in the table above. The corresponding probabilities estimated by the different plotting positions and the standard deviation are shown to illustrate the range of different values that may be obtained. Note that the differences in plotting positions, though within the theoretical range, exceed the standard deviation for that point by two orders of magnitude.

## Q 7.17

```
sebou=read.csv("../HW1/sebou.csv")
sebou$maxQ=sort(sebou$maxQ)
n=length(sebou$maxQ)
i=1:n
iii=2:(n-1)
```

```
ks.test(x=sebou$maxQ,"pnorm")
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  sebou$maxQ
## D = 1, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```
calpha=0.895/(sqrt(n)-0.01+(0.85/sqrt(n))) #for alpha = 0.05
up=((i-1)/n)+calpha
down=(i/n)-calpha
up[which(up>1)]=1
down[which(down<0)]=0
ks.hi=qnorm(up,mean=mean(sebou$maxQ),sd=sd(sebou$maxQ))
ks.lo=qnorm(down,mean=mean(sebou$maxQ),sd=sd(sebou$maxQ))

ks.hi.l=qlnorm(up,meanlog=mean(log(sebou$maxQ)),sdlog=sd(log(sebou$maxQ)))
ks.lo.l=qlnorm(down,meanlog=mean(log(sebou$maxQ)),sdlog=sd(log(sebou$maxQ)))
```

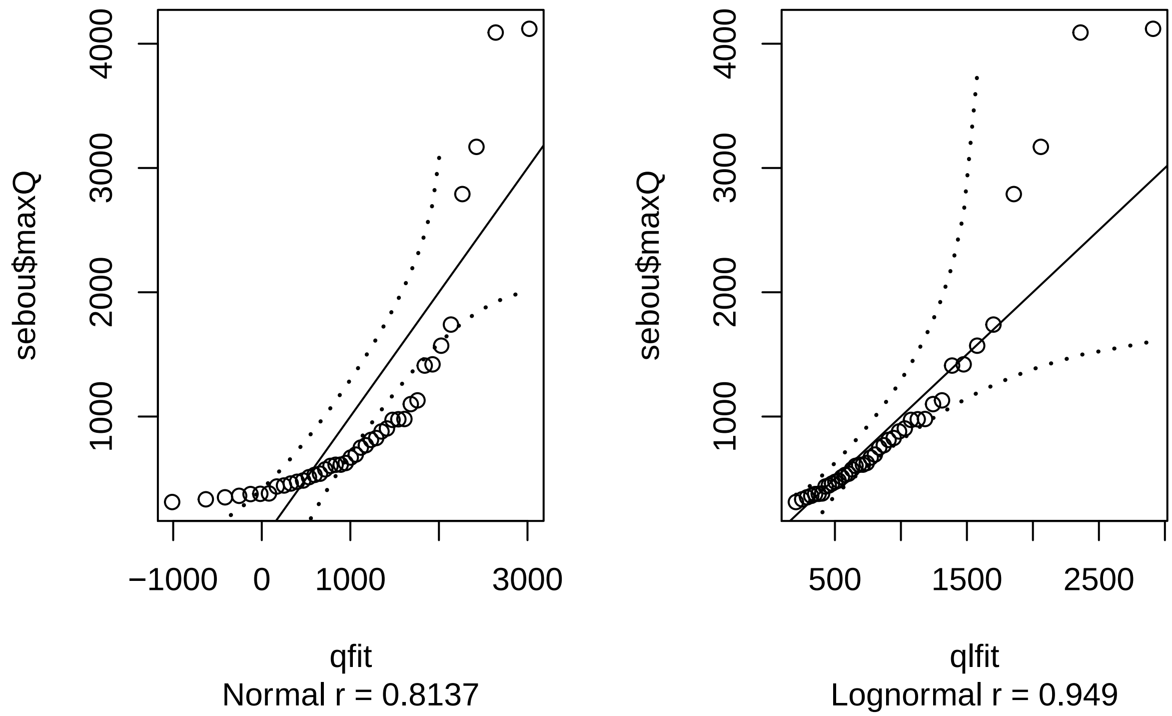
Reject null hypothesis of normally distributed maximum annual flows for Sebou R. data. Calculate critical value of test significance for 95% confidence interval and the upper and lower KS bounds for plotting in next step.

```
#7.17c - Normal
qfit=qnorm(ppblom(1:n,n),mean=mean(sebou$maxQ),sd=sd(sebou$maxQ))

#7.17d - Lognormal
qlfit=qlnorm(ppwei(1:n,n),meanlog=mean(log(sebou$maxQ)),sdlog=sd(log(sebou$maxQ)))

#7.17g
par(mfrow=c(1,2))
pi1=1-((0.5)^(1/n))
pi2n1=(iii-0.3175)/(n+0.365) #Filliben 1975 plotting position
piin=(0.5)^(1/n)
pii=c(pi1,pi2n1,piin)
cor1=cor(sebou$maxQ,qfit)
cor2=cor(sebou$maxQ,qlfit)
plot(qfit,sebou$maxQ,sub=paste("Normal r =",signif(cor1,digits=4)))
lines(qfit,ks.hi,lty=3,lwd=2)
lines(qfit,ks.lo,lty=3,lwd=2)
abline(0,1)
plot(qlfit,sebou$maxQ,sub=paste("Lognormal r =",signif(cor2,digits=4)))
lines(qlfit,ks.hi.l,lty=3,lwd=2)
```

```r
lines(qlfit,ks.lo.l,lty=3,lwd=2)
abline(0,1)
```




After consulting the full table of percent points from Filliben (1975) for n=41, the value of r at the 5% level is 0.972. Niether of the correlation coefficients exceed this value. T The normal is generally not a good fit at all, especially due to the issues with symmetry and necessarily positive values for discharge. Lack of normality is confirmed using the Komolgorov-Smirnov test and visually by plotting the 95% confidence interval (shown here as dotted lines). The lognormal is likely adequate for the mid-range of flows but becomes inaccurate at very high flows. Further, the KS upper bounds are infinite (taking inverse of cdf of a probability greater than 1) for the high flows and the KS lower bounds negative infinity (inverse cdf of p<0) for the low flows which illustrates the uncertainty these models have in those ranges.

For selecting a plotting position the choice is somewhat arbitrary. Ideally you will choose a position that is quantile unbiased for the distribution of interest.