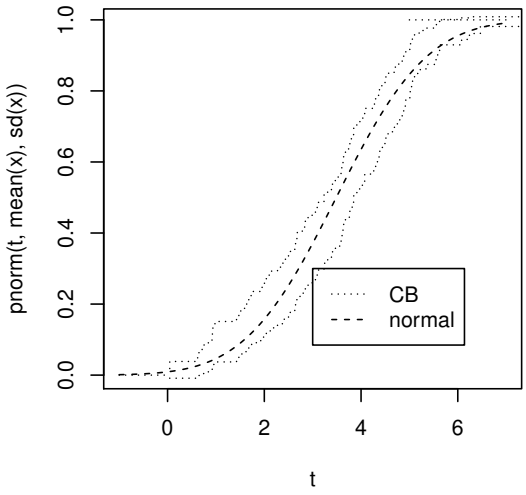
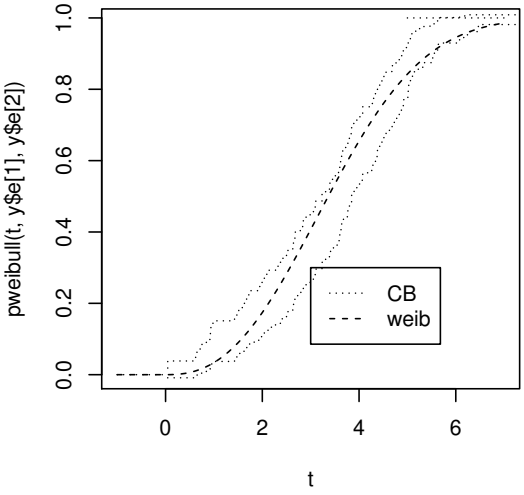
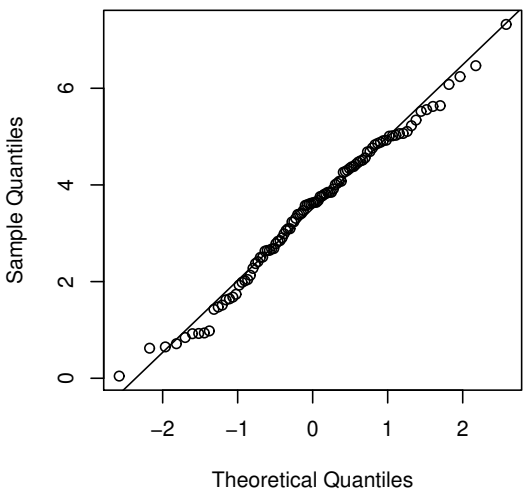
Math 534 Data Analysis, Homework 12

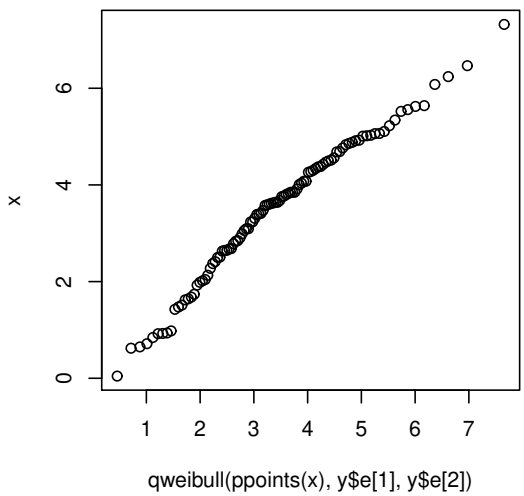
```
1. Find the outliers by boxplot for the given data set as follows.
c. Use qqplot to find out the possible distribution that the data
    are from (you may also check Gamma, Weibull, Uniform,
    Exponential, log-normal, etc.).
d. Carry out proper tests to justify the choice of distribution you
     selected in 2.c. e. What is the estimates of location and
    scale parameters in 2.d?
f. Repeat 2.c, 2.d and 2.e after deleting outliers.

library(MASS)
x=c(
4.496,3.094,5.065,4.882,3.640,3.758,5.017,2.500,6.079 ,3.411,
0.715,3.602,3.916,4.427, 3.235, 4.373, 1.679, 3.800, 3.820, 0.622,
5.640,2.683,2.646,5.624, 7.321, 2.630, 4.008, 5.522, 2.989, 1.513,
1.738,2.848,4.339,3.849, 2.507, 4.915, 4.076, 2.834, 5.342, 4.277,
1.926,2.373,3.388,3.234, 3.094, 3.471, 3.769, 3.627, 4.075, 4.925,
5.061,4.559,2.644,1.622, 2.785, 6.241, 4.302, 4.684, 5.106, 5.558,
50.33,0.045,4.690,0.980, 2.042, 1.985, 2.131, 4.470, 3.309, 6.467,
3.678,1.474,2.273,3.588, 3.641, 0.843, 4.765, 4.514, 3.398, 5.027,
3.057,0.937,3.846,2.674, 2.905, 3.619, 4.036, 4.831, 2.413, 5.009,
4.262,3.576,2.016,0.649, 0.927, 1.643, 4.859, 0.922, 3.847, 5.225,
1.426,4.381)
x=sort(x)
x=x[1:101]
sink("out")
makepsfile <- function() {
ps.options(horizontal = F)
ps.options(height=8.0, width=7.5)
postscript("weib.ps")
par(mfrow =c(2,2))
#plot(ecdf(x))
m=length(x)
x=sort(x)
t=(-10:70)/10
plot(t,pnorm(t,mean(x),sd(x)), type="l",lty=2)
s=1.96*sqrt(ppoints(x)*(1-ppoints(x))/m)
lines(x,ppoints(x)+s,type="S",lty=3)
lines(x,ppoints(x)-s,type="S",lty=3)
lines(c(-1,0),c(0,0),type="l",lty=3)
lines(c(5,7),c(1,1),type="l",lty=3)
leg.names=c("CB", "normal")
legend(3, 0.3, leg.names, lty=c(3,2),cex=1.0)
plot(t,pweibull(t,y$e[1],y$e[2]),type="l",lty=2)
s=1.96*sqrt(ppoints(x)*(1-ppoints(x))/m)
lines(x,ppoints(x)+s,type="S",lty=3)
lines(x,ppoints(x)-s,type="S",lty=3)
```

```
lines(c(-1,0),c(0,0),type="l",lty=3)
lines(c(5,7),c(1,1),type="l",lty=3)
leg.names=c("CB", "weib")
legend(3, 0.3, leg.names, lty=c(3,2),cex=1.0)
qqnorm(x)
qqline(x)
plot(qweibull(ppoints(x),y$e[1],y$e[2]),x)
dev.off()
}
(y=fitdistr(x,"weibull"))
summary(y)


makepsfile()

        shape       scale
  2.4751486   3.9007672
 (0.2030993) (0.1635049)

ks.test(x,  "pweibull", y$e[1], y$e[2])
        One-sample Kolmogorov-Smirnov test

data:  x
D = 0.098097, p-value = 0.2855
alternative hypothesis: two-sided

ks.test(x,  "pweibull", 2.5,4)

        One-sample Kolmogorov-Smirnov test

data:  x
D = 0.074868, p-value = 0.6231
alternative hypothesis: two-sided

y=fitdistr(x,"normal")
ks.test(x,  "pnorm", y$e[1], y$e[2])

        One-sample Kolmogorov-Smirnov test

data:  x
D = 0.068455, p-value = 0.7312
alternative hypothesis: two-sided
```


pnorm(t, mean(x), sd(x))
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
t
CB
normal


pweibull(t, y$e[1], y$e[2])
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
t
CB
weib


Normal Q−Q Plot
Sample Quantiles
0
2
4
6
−2
−1
0
1
2
Theoretical Quantiles


x
0
2
4
6
1
2
3
4
5
6
7
qweibull(ppoints(x), y$e[1], y$e[2])