STA355 A1
Ze Qian Wu
1002192254

1.a)

$$\int f(x;\lambda) = \lambda\, e^{-\lambda x}$$

$$\int_0^{\tau} f(\tau;\lambda) = \int_0^{\tau} \lambda e^{-\lambda\tau}$$

$$= -e^{-\lambda\tau}\Big|_0^{\tau}$$

$$= -e^{-\lambda(\tau)} + 1$$

$$= 1 - e^{-\lambda(\tau)}$$

$$F(\tau) = 1 - e^{-\lambda(\tau)}$$

$$F^{-1}(\tau) = 1 - e^{-\lambda(y)}$$

$$\tau - 1 = -e^{-\lambda(y)}$$

$$1 - \tau = e^{-\lambda(y)}$$

$$\ln(1-\tau) = -\lambda y$$

$$\ln(1-\tau)/(-\lambda) = y$$

$$\therefore F^{-1}(\tau) = -\lambda^{-1}\ln(1-\tau)$$

1.b)

Based on the graphs that the Rcode generate, the points on the exponential graph (x<-rgamma, x<-rlnorm, x<-rweibull) form a straight line and all the points are very close to the best fit line. The majority of the points are clustered around the left side of the line and points on the right side seem to be very spread apart. A few points on the upper right can be spotted in the graph as they seem to have extreme values comparing to the rest of the points. By changing the values of n for all three graphs, the shape of the distribution does not change, but the right side of the points seem to have a greater deviation to the straight line comparing to the larger sample size.

On the other hand, the points on the non-exponential graph(x<-rnorm) does not form a straight line. All the points lie in a curve which the two extreme ends lie under the best fit line and the middle part lies above the best fit line. By changing the values of n, the shape of the graph also stays the same in this case, with extreme value points on the right.

Comparing these four graphs, the rgamma, rlnorm, and rweibull graph show linearity as the rnorm graph shows non-linearity.

1.c)

By using the QQ plot from part b), an exponential distribution is a reasonable model for these data. The exponential graph from part b) forms approximately a straight line and shows linearity, whereas the data provided on blackboard also follows a straight line. In this case, it is appropriate to have exponential distribution to model the 199 observations provided.

2.a)

$f(x;\alpha,\sigma) = 1 - e^{\{-(\frac{x}{\sigma})^{\alpha}\}}$

$f(\tau;\alpha,\sigma) = 1 - e^{\left\{-\left(\frac{\tau}{\sigma}\right)^{\alpha}\right\}}$

$y = 1 - e^{\left\{-\left(\frac{\tau}{\sigma}\right)^{\alpha}\right\}}$

$\tau = 1 - e^{\left\{-\left(\frac{y}{\sigma}\right)^{\alpha}\right\}}$

$\tau - 1 = -e^{\left\{-\left(\frac{y}{\sigma}\right)^{\alpha}\right\}}$

$1 - \tau = e^{\left\{-\left(\frac{y}{\sigma}\right)^{\alpha}\right\}}$

$\ln(1-\tau) = -\left(\frac{y}{\sigma}\right)^{\alpha}$

$\ln(1-\tau)^{1/\alpha} = -\left(\frac{y}{\sigma}\right)$

$-\sigma \cdot \ln(1-\tau)^{\frac{1}{\alpha}} = y$

$\sigma\{\ln(1-\tau)\}^{\frac{1}{\alpha}} = y$

$\therefore F^{-1}(\tau;\alpha,\sigma) = \sigma\{\ln(1-\tau)\}^{\frac{1}{\alpha}}$

2.b)

$Var(X_{(k)}) = \dfrac{\tau(1-\tau)}{nf^2\left(F^{-1}(\tau)\right)}$

$nf^2\left(F^{-1}(\tau)\right) = n(\frac{\alpha}{\sigma})^2[-\ln(1-\tau)]^{\frac{2(\alpha-1)}{\alpha}}(1-\tau)^2$

$Var(X_{(k)}) = \dfrac{\tau(1-\tau)}{n(\frac{\alpha}{\sigma})^2[-\ln(1-\tau)]^{\frac{2(\alpha-1)}{\alpha}}(1-\tau)^2}$

$Var(X_{(k)}) = \dfrac{\tau}{n\left(\frac{\alpha}{\sigma}\right)^2[-\ln(1-\tau)]^{\frac{2(\alpha-1)}{\alpha}}(1-\tau)}$

$\because k \approx \tau n$

$\therefore Var(X_{(k)}) = \dfrac{k}{\left(\frac{\alpha}{\sigma}\right)^2[-\ln(1-\tau)]^{\frac{2(\alpha-1)}{\alpha}}(1-\tau)}$

2.c)

The mean of the exponential distribution as a function of τ, α, and σ

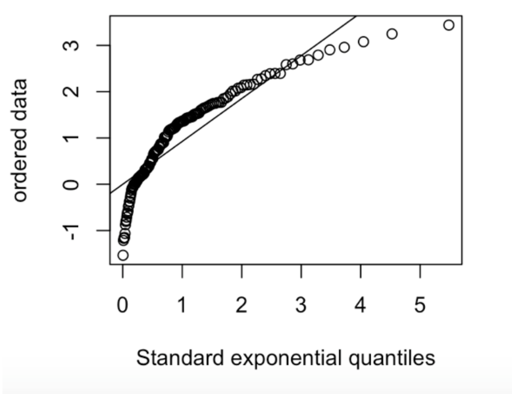
$nD_{kn} \xrightarrow{d} Exponential(f\left(F^{-1}(\tau)\right))$

$\therefore nD_{kn} \xrightarrow{d} Exp((\frac{\alpha}{\sigma})[-\ln(1-\tau)]^{\frac{\alpha-1}{\alpha}}(1-\tau)$

2.d) (Supplemental rcode is attached in appendix)
When $\alpha \neq 1$, there are two cases which α can be either greater or smaller than 1
When $\alpha > 1$, τ increases as mean increases, and τ decreases as mean decreases.
When $\alpha < 1$, τ increases as mean decreases, and τ decreases as mean increases.
For instance, plugging in $\alpha = 0.5$ and $\tau = 0.1$ would give us a mean that is greater than $\tau = 0.9$; and plugging in $\alpha = 5$ and $\tau = 0.1$ would give us a mean that is smaller than $\tau = 0.9$.
From the data above, we can estimate the mean by knowing the shape parameter and τ for Weibull distribution. Based on this information, we can apply the Weibull distribution to real life cases, such as comparing the failure rate for different brands of laptops. As $\tau \in [0,1]$ indicates the time of how long laptops last from day one until the day they break down, we can calculate the mean of the lifetimes of laptops. Weibull distribution allows us to take a close look at the difference between the failure rates, which helps us to make better decisions in real life.

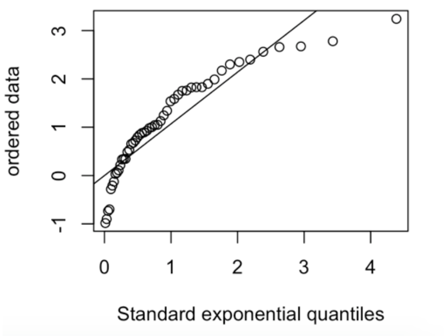
Appendix
1.b)

```
x <- rnorm(150,1)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  tau <- (c(1:n)-3/8)/(n+1/4)
  if (plot) {
    plot(-log(1-tau),x,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=x,quantiles=-log(1-tau))
    r
  }
}
qq.expo(x)
```

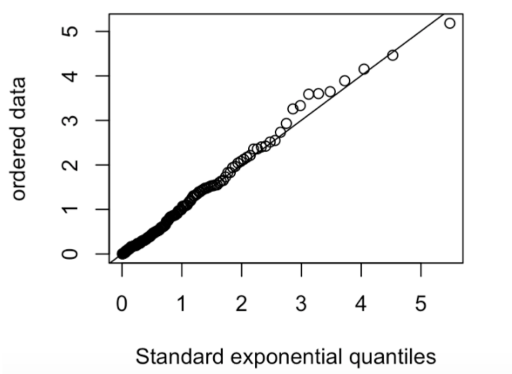



```
x <- rnorm(50,1)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  tau <- (c(1:n)-3/8)/(n+1/4)
  if (plot) {
    plot(-log(1-tau),x,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=x,quantiles=-log(1-tau))
```

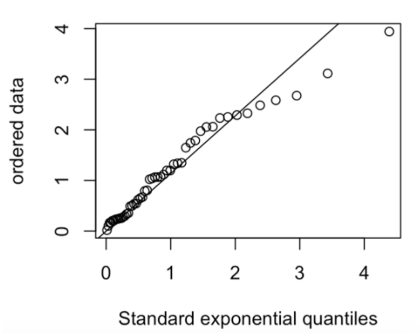
```
    r
  }
}
qq.expo(x)
```




```
x <- rgamma(150,1)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  tau <- (c(1:n)-3/8)/(n+1/4)
  if (plot) {
    plot(-log(1-tau),x,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=x,quantiles=-log(1-tau))
    r
  }
}
qq.expo(x)
```

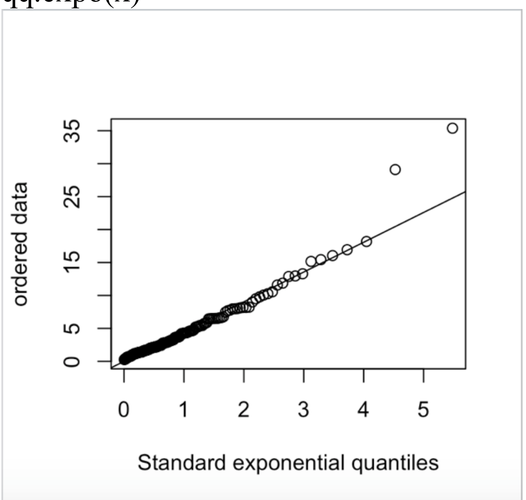



```
x <- rgamma(50,1)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  tau <- (c(1:n)-3/8)/(n+1/4)
  if (plot) {
    plot(-log(1-tau),x,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=x,quantiles=-log(1-tau))
    r
  }
}
qq.expo(x)
```




```
x <- rlnorm(150,1)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  tau <- (c(1:n)-3/8)/(n+1/4)
```

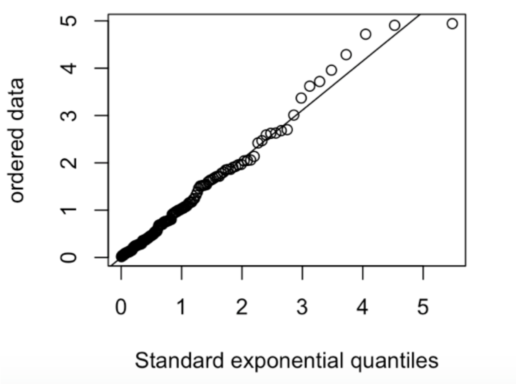
```
  if (plot) {
    plot(-log(1-tau),x,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=x,quantiles=-log(1-tau))
    r
  }
}
qq.expo(x)
```




```
x <- rlnorm(50,1)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  tau <- (c(1:n)-3/8)/(n+1/4)
  if (plot) {
    plot(-log(1-tau),x,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=x,quantiles=-log(1-tau))
    r
  }
}
qq.expo(x)

x <- rweibull(150,1)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
```

```
  n <- length(x) # length of x
  tau <- (c(1:n)-3/8)/(n+1/4)
  if (plot) {
    plot(-log(1-tau),x,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=x,quantiles=-log(1-tau))
    r
  }
}
qq.expo(x)
```

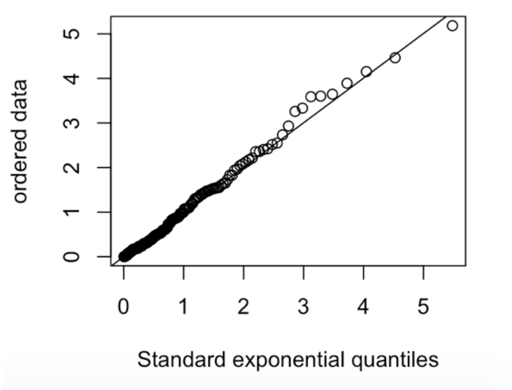



```
x <- rweibull(50,1)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  tau <- (c(1:n)-3/8)/(n+1/4)
  if (plot) {
    plot(-log(1-tau),x,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=x,quantiles=-log(1-tau))
    r
  }
}
qq.expo(x)
```

1.c)

```
x <- read.table("/Users/Claire/Desktop/aircon.txt")
attach(x)

qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  tau <- (c(1:n)-3/8)/(n+1/4)
  if (plot) {
    plot(-log(1-tau),x,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=x,quantiles=-log(1-tau))
    r
  }
}
qq.expo(x)
```



2.d)

```
x <- rweibull(150,5)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  k <- c(1:n-1)
  dk <- diff(k)
```

```
  tau <-0.1
  if (plot) {
    plot(n(1-tau)dk,k,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=k,quantiles=n(1-tau)dk)
    r
  }
}

qq.expo(x)
mean(x)
[1] 0.8932343

x <- rweibull(150,5)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  k <- c(1:n-1)
  dk <- diff(k)
  tau <-0.9
  if (plot) {
    plot(n(1-tau)dk,k,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=k,quantiles=n(1-tau)dk)
    r
  }
}

qq.expo(x)
mean(x)
[1] 0.9384572

x <- rweibull(150,0.2)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  k <- c(1:n-1)
  dk <- diff(k)
```

```
  tau <-0.1
  if (plot) {
    plot(n(1-tau)dk,k,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=k,quantiles=n(1-tau)dk)
    r
  }
}

qq.expo(x)
mean(x)
183.6961



x <- rweibull(150,0.2)
qq.expo <- function(x,plot=T) {
  x <- sort(x)  # sorts the data from smallest to largest
  n <- length(x) # length of x
  k <- c(1:n-1)
  dk <- diff(k)
  tau <-0.9
  if (plot) {
    plot(n(1-tau)dk,k,xlab="Standard exponential quantiles",
         ylab="ordered data")
    abline(a=0,b=mean(x)) # draws a line with slope equal to
    # the sample mean and intercept 0
  }
  else {
    r <- list(x=k,quantiles=n(1-tau)dk)
    r
  }
}

qq.expo(x)
mean(x)
[1] 70.68081
```