1) (a) If $Z \sim \mathcal{N}(0, \sigma^2)$, show that

(i) the cdf of $|Z|$ is $G(x) = 2\Phi(x/\sigma) - 1$ where $\Phi(t)$ is the cdf of a $\mathcal{N}(0,1)$ random variable;

(ii) the τ quantile of the distribution of $|Z|$ is $G^{-1}(\tau) = \sigma\Phi^{-1}((\tau+1)/2)$.

i) Since $Z \sim N(0, \sigma^2)$ then cdf of abs of Z is

$$P(|Z| \leq x) = P(-x \leq Z \leq x)$$

$$= P(Z \leq x) - P(-x \leq Z)$$

$$= P(Z \leq x) - [1 - P(Z \leq x)]$$

$$= 2 \cdot P(Z \leq x) - 1$$

Here we know that $P(Z \leq x)$ is the cdf of $Z \sim N(0,1)$

Hence cdf of $|Z|$ is $2\Phi(\frac{x}{\sigma}) - 1$ ∎

ii) We know $G(x) = 2\Phi(\frac{x}{\sigma}) - 1$, and we want to find $G^{-1}(x)$

Let $y = 2\Phi(\frac{x}{\sigma}) - 1$ and solve for x.

Then $y + 1 = 2\Phi(\frac{x}{\sigma})$

$\frac{y+1}{2} = \Phi(\frac{x}{\sigma})$ inverting both sides

$\Phi^{-1}(\frac{y+1}{2}) = \frac{x}{\sigma}$

So, $\sigma\Phi^{-1}(\frac{y+1}{2}) = x$

Hence $G^{-1}(T) = \sigma\Phi^{-1}(\frac{y+1}{2})$ ∎

(b) Suppose that $Z_1, \cdots, Z_n$ are independent $\mathcal{N}(0, \sigma^2)$ random variables and define $W_i = |Z_i|$ for $i = 1, \cdots, n$ and the order statistics $W_{(1)} \leq W_{(2)} \leq \cdots \leq W_{(n)}$. The result of part (a) suggests that we could estimate σ using an order statistic $W_{(k)}$ as follows:

$$\widehat{\sigma}_k = \frac{W_{(k)}}{\Phi^{-1}((\tau_k + 1)/2)}$$

where (for example) $\tau_k = k/(n+1)$. If $\tau_k \to \tau \in (0,1)$ as $k, n \to \infty$ then

$$\sqrt{n}(\widehat{\sigma}_k - \sigma) \xrightarrow{d} \mathcal{N}(0, \gamma^2(\tau)).$$

Give an expression for $\gamma^2(\tau)$. For what value of τ is $\gamma^2(\tau)$ minimized? (You can determine the minimizing value of τ graphically.)

As shown in lecture, we know that:

$$Var(W_{(k)}) \approx \frac{\tau(1-\tau)}{n g^2(G^{-1}(\tau))} \qquad \text{where } \tau \in (0,1),\ g \text{ is pdf of } W_k,\ G \text{ is cdf of } W_k$$

So we have $\hat{\sigma}_k = \dfrac{W_k}{\Phi^{-1}\left(\frac{\tau_k+1}{2}\right)}$

Then $\sqrt{n}\left(\hat{\sigma}_k - \sigma\right) = \sqrt{n}\left(\dfrac{W_k}{\Phi^{-1}\left(\frac{\tau_k+1}{2}\right)} - \sigma\right)$

Then,

$$Var\left[\sqrt{n}\left(\frac{W_k}{\Phi^{-1}\left(\frac{\tau_k+1}{2}\right)} - \sigma\right)\right] = \frac{n}{\left[\Phi^{-1}\left(\frac{\tau_k+1}{2}\right)\right]^2} Var(W_k)$$

Which equals to,

$$Var\left[\sqrt{n}\left(\frac{W_k}{\Phi^{-1}\left(\frac{\tau_k+1}{2}\right)} - \sigma\right)\right] = \frac{n}{\left[\Phi^{-1}\left(\frac{\tau_k+1}{2}\right)\right]^2} \times \frac{\tau(1-\tau)}{n g^2(G^{-1}(\tau))}$$

$$= \frac{\tau(1-\tau)}{\left[\Phi^{-1}\left(\frac{\tau+1}{2}\right)\right]^2 \times g^2(G^{-1}(\tau))}$$

$$= \frac{\tau(1-\tau)}{\left(\frac{G^{-1}(\tau)}{\sigma}\right)^2 \times g^2(G^{-1}(\tau))} \qquad \text{Since } G^{-1} = \sigma\,\Phi^{-1}\left(\frac{\tau+1}{2}\right)$$

$$Var\left[\sqrt{n}\left(\hat{\sigma}_k - \sigma\right)\right] = \gamma^2(\tau) = \frac{\sigma^2\tau(1-\tau)}{(G^{-1}(\tau))^2 g^2(G^{-1}(\tau))}$$

Since we are minimizing the function wrt τ, we can keep σ^2 constant and ignore.

Then using R to compute and plot, we can use dhalfnorm for g = pdf of W_k, qhalfnorm for G^{-1} = inverse cdf of W_k

Then we minimize $\gamma^2(\tau)$ when $\tau = 0.86$ at $\gamma^2(\tau) = 0.77$

R code for graphical illustration attached below ■

(c) If $Z \sim \mathcal{N}(0,1)$, we have

$$P(|Z| > x) \leq \frac{2}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

and for larger x,

$$P(|Z| > x) \approx \frac{2}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

If $Z_1, \cdots, Z_n$ are independent $\mathcal{N}(0,1)$ random variables, use this approximation to show that for any $\epsilon > 0$

$$P\left(\max_{1\leq i\leq n} |Z_i| > (1+\epsilon)\sqrt{2\ln(n)}\right) \to 0$$

as $n \to \infty$. (Hint: Note that

$$P\left(\max_{1\leq i\leq n} |Z_i| > x\right) = P\left(\bigcup_{i=1}^{n} [|Z_i| > x]\right)$$

and use Bonferroni's inequality.)

Note that:

$$P\left(\max_{1\leq i\leq n} |Z_i| > (1+\varepsilon)\sqrt{2\ln(n)}\right) = P\left(\bigcup_{i=1}^{n}\left[|Z_i| > (1+\varepsilon)\sqrt{2\ln(n)}\right]\right)$$

Then using Bonferroni's inequality,

$$P\left(\bigcup_{i=1}^{n}\left[|Z_i| > (1+\varepsilon)\sqrt{2\ln(n)}\right]\right) \leq \sum_{i=1}^{n} P\left(|Z_i| > (1+\varepsilon)\sqrt{2\ln(n)}\right)$$

$$\leq \frac{2n}{(1+\varepsilon)\sqrt{2\ln(n)\times 2\pi}}\, e^{\left(-\frac{(1+\varepsilon)^2 \times 2\ln(n)}{2}\right)}$$

$$\leq \frac{n}{(1+\varepsilon)\sqrt{\ln(n)\pi}} \times n^{-(1+\varepsilon)^2}$$

$$\leq \frac{1}{(1+\varepsilon)\sqrt{\ln(n)\pi} \times n^{(1+\varepsilon)^2-1}}$$

Then $\displaystyle\lim_{n\to\infty} \frac{1}{(1+\varepsilon)\sqrt{\ln(n)\pi} \times n^{(1+\varepsilon)^2-1}}$

$$= \lim_{n\to\infty} \frac{1}{(1+\varepsilon)\sqrt{\ln(n)\pi} \times n^{\varepsilon^2+2\varepsilon}}$$

$= 0$ Since $\displaystyle\lim_{n\to\infty} (1+\varepsilon)\sqrt{\ln(n)\pi} \times n^{\varepsilon^2+2\varepsilon} = \infty$

Hence $P\left(\max_{1\leq i\leq n} |Z_i| > (1+\varepsilon)\sqrt{2\ln(n)}\right) = 0$ as $n \to \infty$ ∎

2) (a) A useful formula for the expected value of any non-negative random variable is

$$E(X) = \int_0^\infty (1 - F(x))\, dx.$$

If X is also continuous with pdf $f(x)$ then this formula can be derived as follows:

$$\begin{aligned} E(X) &= \int_0^\infty x f(x)\, dx \\ &= \int_0^\infty \int_0^x f(x)\, dt\, dx \\ &= \int_0^\infty \int_t^\infty f(x)\, dx\, dt \\ &= \int_0^\infty (1 - F(t))\, dt. \end{aligned}$$

If $h(x)$ is the hazard function of X, show that

$$E(X) = \int_0^1 \frac{1}{h(F^{-1}(\tau))}\, d\tau.$$

(Hint: Make the change of variables $u = F^{-1}(\tau)$.)

$E(X) = \int_0^1 \frac{1}{h(F^{-1}(\tau))} d\tau$ $\qquad\qquad \frac{1-\tau}{f F^{-1}}$

Let $F^{-1}(\tau) = u$

Then $du = \frac{1}{f(F^{-1}(\tau))} d\tau$

So $d\tau = f[F^{-1}(\tau)]\, du = F'(u)\, du$

Note that for $\tau \in (0,1) \Rightarrow F^{-1}(\tau) \in (0, \infty)$

Using change of variable we rewrite the following:

$$E(X) = \int_0^\infty \frac{1}{h(u)} F'(u)\, du$$

$$= \int_0^\infty \frac{F'(u)}{\frac{f(u)}{1-F(u)}}\, du$$

$$= \int_0^\infty (1-F(u)) \frac{F'(u)}{f(u)}\, du \qquad \text{but } F'(u) = f(u)$$

$$= \int_0^\infty (1-F(u))\, du$$

(b) Suppose that $X_{(k)}$ is the k-th order statistic where $k \approx \tau n$ (for some $\tau \in (0,1)$) and define $D_k = X_{(k)} - X_{(k-1)}$. From lecture, we know that the distribution of $n\,D_k$ is approximately Exponential with mean $1/f(F^{-1}(\tau))$. Use this fact to show that the distribution of $(n-k+1)D_k$ is approximately Exponential with mean $1/h(F^{-1}(\tau))$. (Hint: Note that (i) $h(F^{-1}(\tau)) = f(F^{-1}(\tau))/(1-\tau)$ and (ii) $(n-k+1) = n(n-k+1)/n \approx n(1-\tau)$ since $k/n \approx \tau$ and $1/n \approx 0$.)

we know that nD_k is distributed exponentially with n mean $\frac{1}{f(F^{-1}(\tau))}$

Since $\frac{k}{n} \approx \tau$ and $\frac{1}{n} \approx 0$

Then $(n-k+1) = \frac{n(n-k+1)}{n} \approx n(1-\tau)$

So we have $(n-k+1)D_k \approx n(1-\tau)D_k$

Then mean of $n(1-\tau)D_k$ is $(1-\tau) \times \frac{1}{f(F^{-1}(\tau))}$

$$= \frac{(1-\tau)}{f(F^{-1}(\tau))}$$

$$= \frac{1}{h(F^{-1}(\tau))}$$

Since constant multiplication of an exponential distribution is still exponentially distributed

$(n-k+1)D_k$ is exponentially distributed with mean $\frac{1}{h(F^{-1}(\tau))}$ ∎

STA355 A1

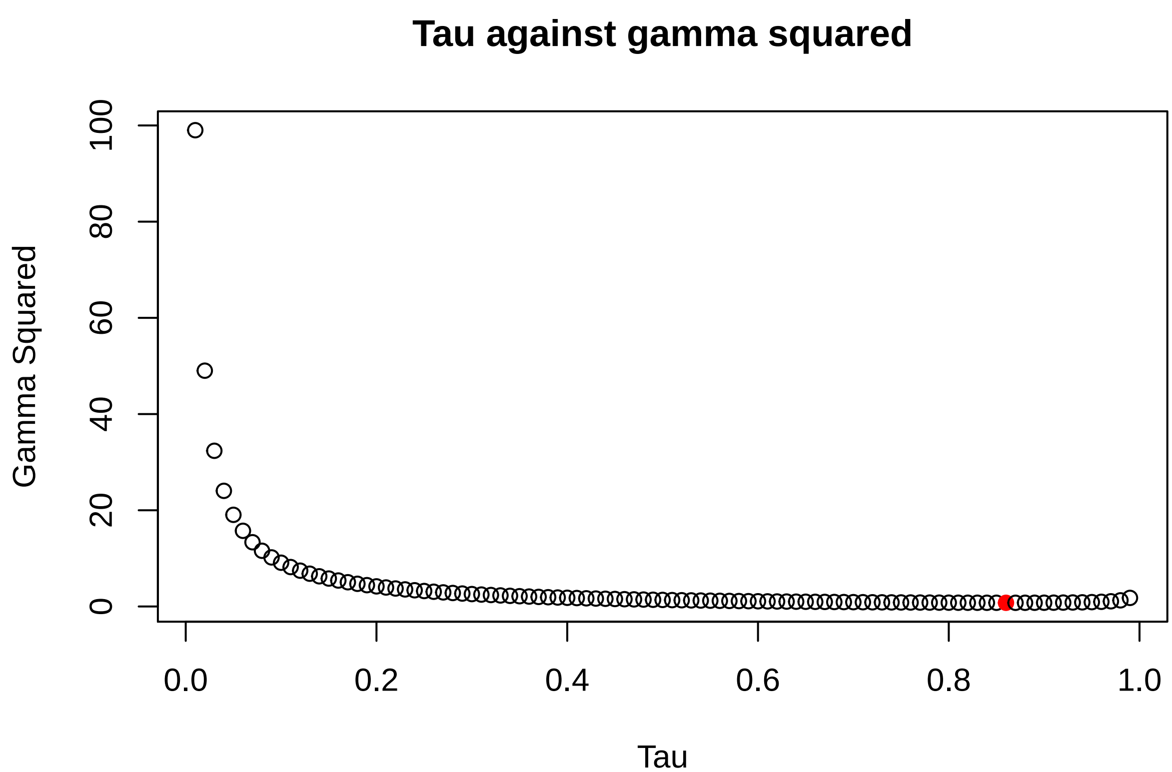
Harold Hyun Woo Lee

1/27/2021

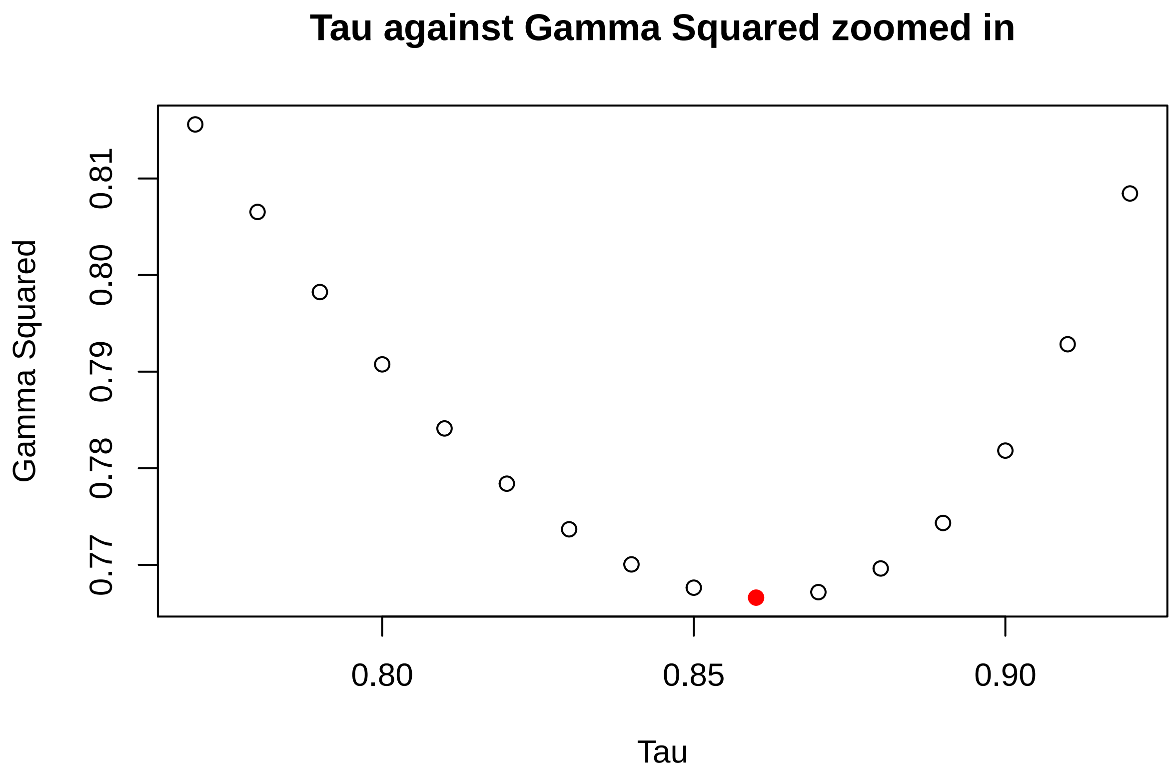
Question 1b)

```
library(fdrtool)
tau <- seq(from=0, to=1, by=0.01)
tau <- tau[2:100]
r <- (tau*(1-tau))/(((qhalfnorm(tau))^2)*((dhalfnorm(qhalfnorm(tau)))^2))

plot(tau, r, xlab="Tau", ylab="Gamma Squared",pch=ifelse(tau==0.86, 19, 1),
     col=ifelse(tau==0.86, "red", "black"), main="Tau against gamma squared")
```




```
plot(tau[77:92], r[77:92], pch=ifelse(tau[77:92]==0.86, 19, 1),
     col=ifelse(tau[77:92]==0.86, "red", "black"), xlab="Tau", ylab="Gamma Squared",
     main="Tau against Gamma Squared zoomed in")
```




```
# Find the minimum value of r squared
min(r)
```

```
## [1] 0.7666061
```

Here we can see that the $\tau = 0.86$ minimizes the gamma squared function at 0.77

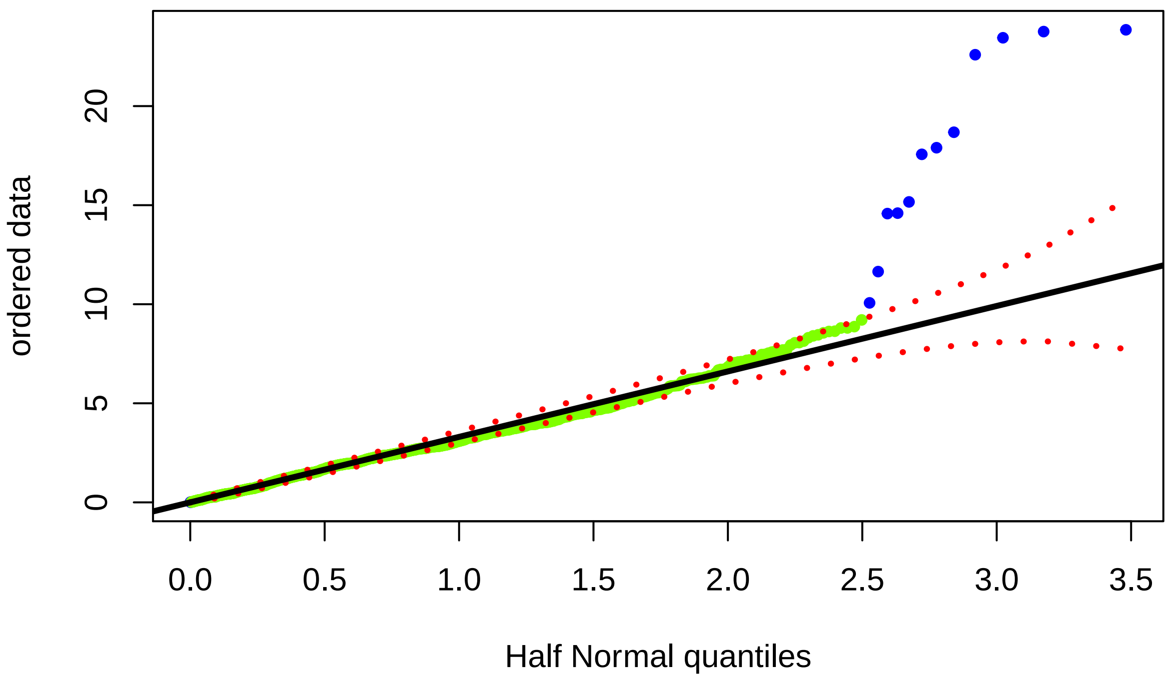
Question 1d)

```
dat <- scan("prob1-data.txt")


halfnormal <- function(x,tau=0.5,ylim) {
       sigma <- quantile(abs(x),probs=tau)/sqrt(qchisq(tau,1))
       n <- length(x)
       pp <- ppoints(n)
       qq <- sqrt(qchisq(pp,df=1))
# upper envelope
       upper <- sigma*(qq + 3*sqrt(pp*(1-pp))/(2*sqrt(n)*dnorm(qq)))
# lower envelope
       lower <- sigma*(qq - 3*sqrt(pp*(1-pp))/(2*sqrt(n)*dnorm(qq)))
# add upper and lower envelopes to plot
       if (missing(ylim)) ylim <- c(0,max(c(upper,abs(x))))
       plot(qq,sort(abs(x)),
          xlab="Half Normal quantiles",ylab="ordered data",pch=20,
```

```
    ylim=ylim, col=ifelse(sort(abs(x)) > 1.2*sigma*qq, "blue", "chartreuse"))
lines(qq,lower,lty=3,lwd=3,col="red")
lines(qq,upper,lty=3,lwd=3,col="red")
abline(a=0,b=sigma,lwd=3)
}
```

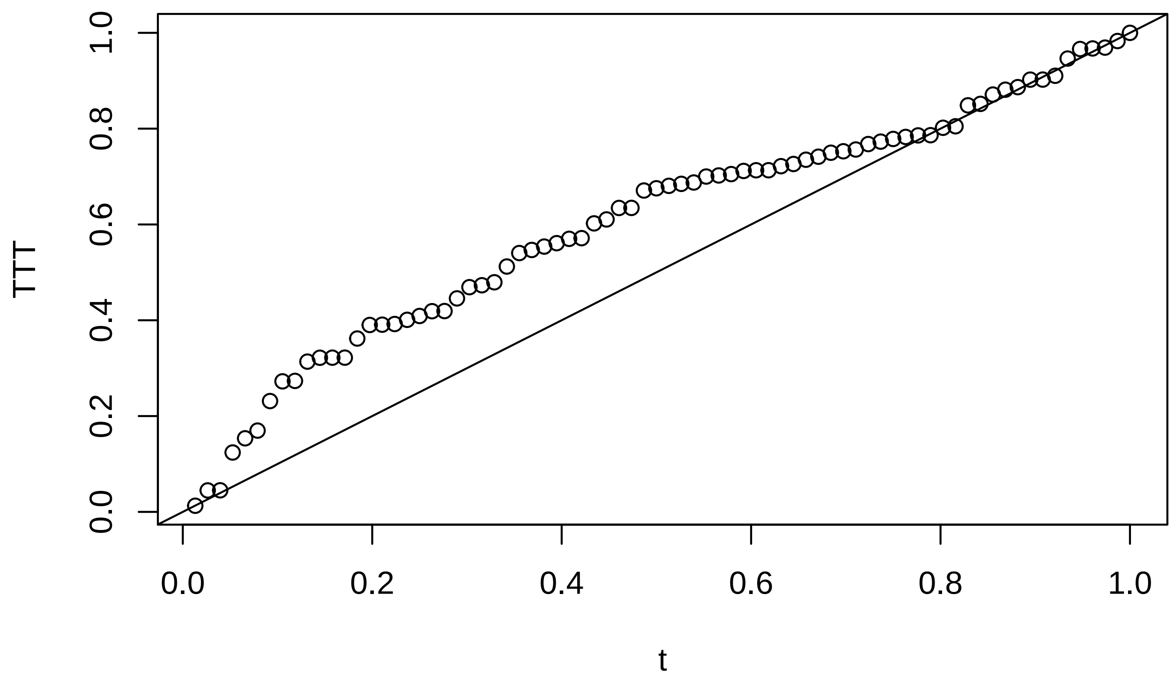
```
halfnormal(dat, tau=0.5)
```



For the above graph, I edited the half normal function so that points that are above the sigma slope line by 20% are colored in blue. The rest of the points that are more or less close to the straight line whose slope is sigma are colored in neon green. I've given it a weight of 1.2 to color the points by blue to give a sort of leeway to those points that are close to the slope but slightly above the slope. We can see that there are about 12 points whose means are non-zero.

Question 2c)

```
x <- scan("kevlar.txt")
x <- sort(x)
n <- length(x)
d <- c(n:1*c(x[1], diff(x)))
plot(c(1:n)/n, cumsum(d)/sum(x), xlab="t", ylab="TTT")
abline(0,1)
```



As shown above in TTT plot for the Kevlar 373/epoxy strands data, we can see that the hazard function is roughly increasing. We can note that the points are mostly above the 45 degree line and roughly sketch out the concave shape from $t = 0$ to $t = 0.8$.