Stat538 HW6 - Accerleration Failure Time Models

Arthur Lui

28 February 2014

Introduction:

Knowing the relationship between time from onset of first symptoms to Myocardial Infarction (MI) can help doctors and patients intervene early to mitigate the severity of the effects of heart attacks. We will analyse data provided by Dr. Engler, which includes data of the variables:

- time: time from onset of symptoms to MI (in minutes)
- status: censoring indicator
- gender
- perf4: score on a lung capacity test administered at the hospital

After analyzing this data (using both parametric and non-parametric models), we will assess model assumptions.

Results:

Choosing a Distribution for the AFT Model

We first plotted the Kaplan Meier (K.M.) Survival Curve without any covarites. We then fit AFT models using the weibull, loglogistic, and lognormal distributions. The lognormal curve resembled the K.M. curve the closest (see Figure 1). So, lognormal model was chosen to model survival times.

A more rigorous way of determining which distribution best models the data is to plot the probability plots for each of the distributions. The plots reveal that the lognormal probability plot most resembles a straight line. This suggests that the lognormal distribution is a better choice for modelling survival times (see Figures 2-4). The lognormal probability plot provides other valuable information. For example, the slope (3) and the intercept (-3.5) provide estimates for σ^{-1} and $-\mu/\sigma$, respectively.

Fitting the Model with Covariates: PRINT XTABLE

After fitting the AFT model, we find that the median survival time for a male is 1.29 times that of females, for a particular perf4. That is, males have a higher survival rate that females at any particular perf4. Also, within each gender, the median (or, in general, any percentile) survival time is increased by 10% as perf4 increases by 1 unit. That is, fixing the gender, people with higher perf4 have higher survival rates (see Figure 5 and Table 1).

Table 1: Summary Table For Cox Model

	Value	Std. Error	z	p
(Intercept)	4.84	0.02	235.04	0.00
genderm	0.26	0.01	17.46	0.00
perf4	0.10	0.00	24.01	0.00
Log(scale)	-1.72	0.02	-81.98	0.00

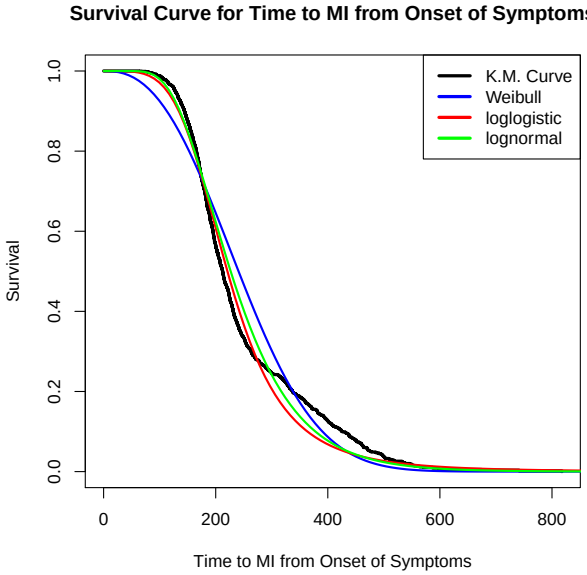

Figure 1: Plot of Parametric and Non-parametric Survival Curves. The lognormal curve looks closest to the K.M. curve. So, the lognormal distribution was chosen to model survival times.

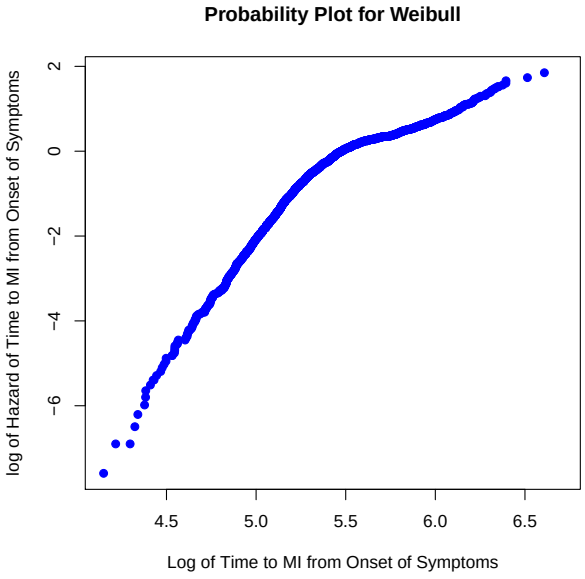

Figure 2: The line is not very straight. This means that the weibull model for survival times is not suitable.

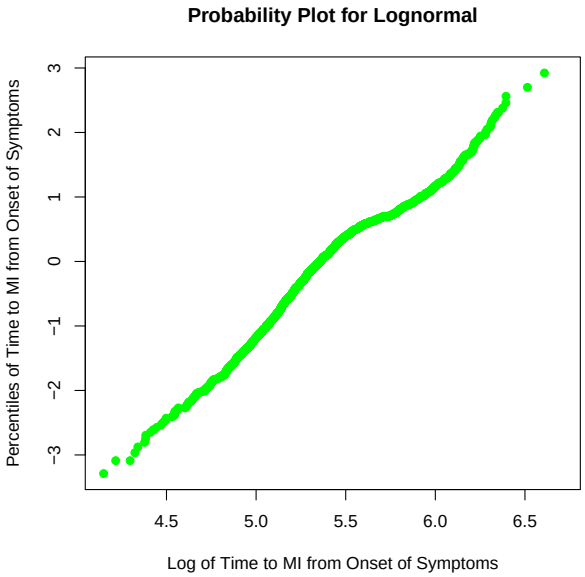

Figure 3: The line is reasonably straight. This means that the lognormal model for survival times is suitable and the slope (3) and intercept (-3.5) provide estimates for σ^{-1} and $-\mu/\sigma$, respectively.

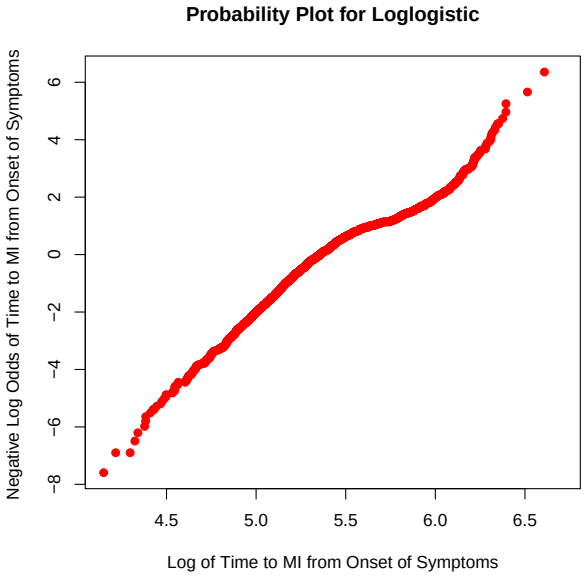

Figure 4: The line is not very straight. This means that the loglogistic model for survival times is not suitable.

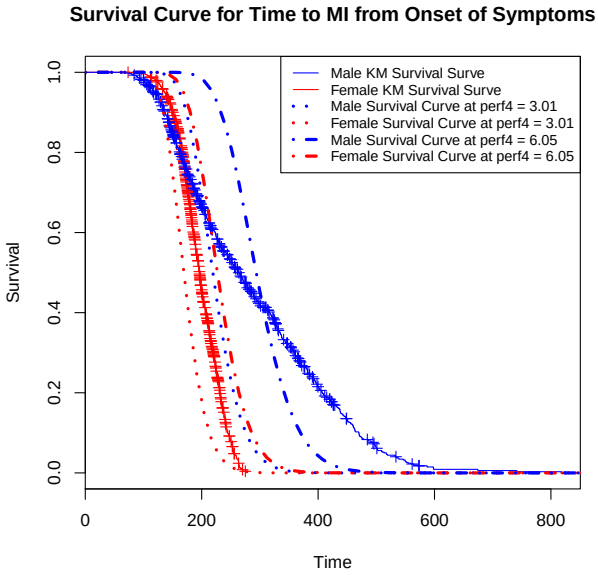

Figure 5: KM Curve for male and female at 1st and 3rd quartiles of perf4.

We can do a similar analysis by dichotomizing perf4. I chose the cut-off point for the dichotomization of perf4 to be 6.05 (75^{th} percentile). After fitting the AFT model, we find that the median survival time for a male is 1.29 times that of females, for a particular perf4. That is, males have a higher survival rate that females at any particular perf4. Also, within each gender, the median (or, in general, any percentile) survival time for the higher perf4 is 1.32 times that of the lower group. That is, fixing the gender, people with higher perf4 have higher survival rates (see Figure 6 and Table 2).

Table 2: Summary Table For Cox Model

	Value	Std. Error	z	p
(Intercept)	5.21	0.01	465.43	0.00
genderm	0.25	0.02	15.60	0.00
perf4	0.28	0.02	15.63	0.00
Log(scale)	-1.63	0.02	-77.41	0.00

Assessing Model Fit

To assess model fit, we plot the percentile-percentile plots for each gender. For males, the Q-Q plot line passes through the origin but is not straight. This suggests that the AFT model may not be a valid model for the two groups of data - males with perf4 < 6.05, and males with perf4 > 6.05 (75^{th} percentile survival time) (see Figure 7).

For females, the Q-Q plot line passes through the origin and is Figure straight. This suggests that the AFT model may be a valid model for the two groups of data, females with perf4 < 6.05, and females with perf4 > 6.05 (75^{th} percentile survival time). The slope for this line appears to be approximately 1.1, which is an estimate for the acceleration factor, ϕ^{-1} (see Figure 8).

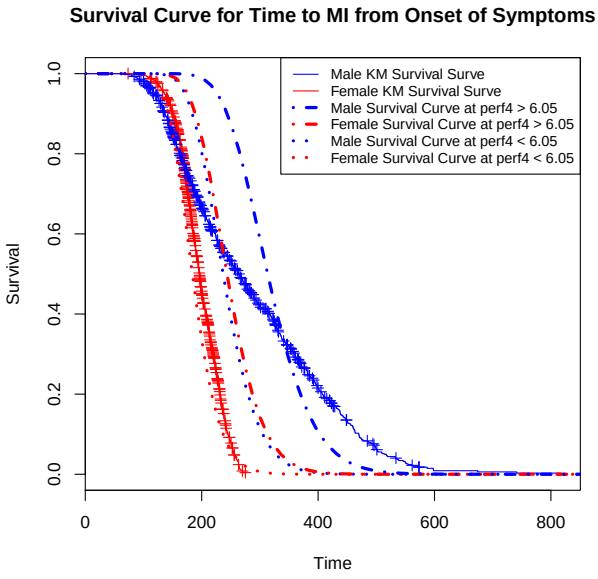

Figure 6: KM Curve for male and female and Dichotomized perf4.

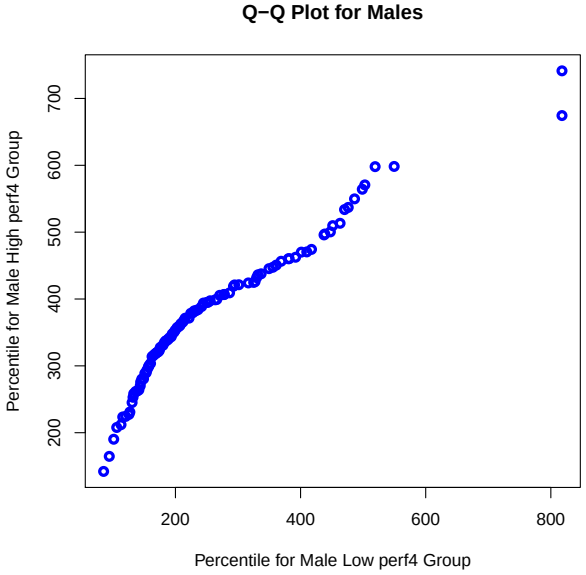

Figure 7: Percentile-percentile plot for males. The line passes through the origin but is not straight. This suggests that the AFT model may not be a valid model for the two groups of data, males with perf4 < 6.05, and males with perf4 > 6.05 (75^{th} percentile survival time).

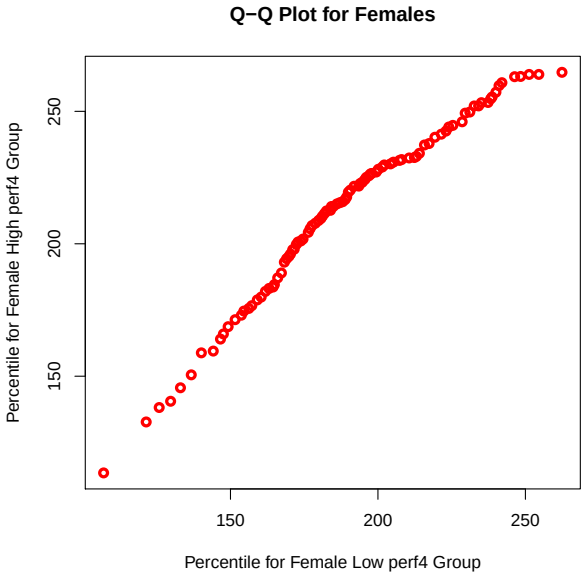

Figure 8: Percentile-percentile plot for females. The line passes through the origin and is reasonably straight. This suggests that the AFT model may be a valid model for the two groups of data, females with perf4 $<$ 6.05, and females with perf4 $>$ 6.05 (75^{th} percentile survival time). The slope for this line appears to be approximately 1.1, which is an estimate for the acceleration factor, ϕ^{-1}.

Last of all, we inspect the deviance residuals plot. The plot exhibits no patterns and the deviance residuals are symmetric about 0. This implies that the fitted model is appropriate (see Figure 9).

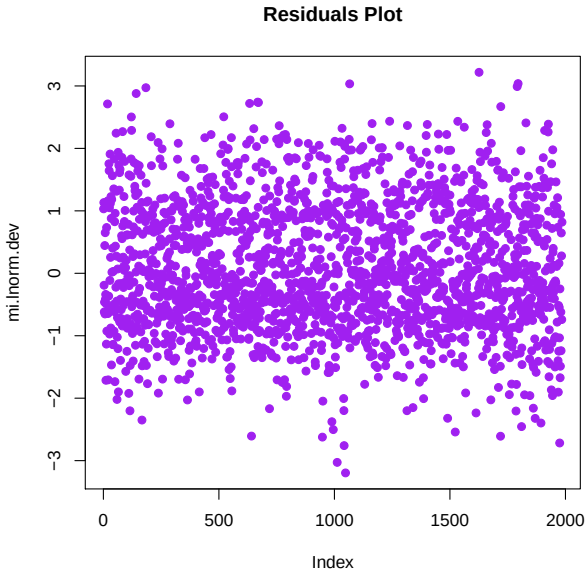

Figure 9: Deviance Residual plot. No patterns, and symmetric aboue 0. This implies that the fitted model is appropriate.

Appendix:

R Code:

```
rm(list=ls())

library(survival)

mi <- read.csv("../Data/mi.csv")
mi <- mi[sample(1:nrow(mi),nrow(mi),replace=F),] # Random sort because data
                                                 # is arranged by time and gender
# time:     Minutes to MI onset of symptoms
# status:   Cesoring Indicator
# gender:
# perf4:    score on a lung capacity test (surrogate for smoking status)


# survival functions
weib.surv <- function(x,gamma,lambda) {
  exp(-(x/lambda)^gamma)
}

loglog.surv <- function(x,beta,alpha) {
  1-(((x/alpha)^(-beta)) + 1)^(-1)
}

lognorm.surv <- function(x,mu,sigma) {
```

```
  pnorm((log(x)-mu)/sigma,lower=F)
}

#1: Fit AFT Models: (Weibull, loglogistic, lognormal)
mi.weib <- survreg(Surv(time,status)~1,data=mi,dist="weibull")
mi.weib.lambda <- exp(mi.weib$coef[[1]]) # lambda = scale = exp(Intercept)
mi.weib.gamma <- 1/mi.weib$scale         # gamma = shape = 1/scale

mi.llog <- survreg(Surv(time,status)~1,data=mi,dist="loglogistic")
mi.llog.lambda <- exp(mi.llog$coef[[1]]) # lambda = scale = exp(Intercept)
mi.llog.gamma <- 1/mi.llog$scale         # gamma = shape = 1/scale


mi.lnorm <- survreg(Surv(time,status)~1,data=mi,dist="lognormal")
mi.lnorm.mu <- mi.lnorm$coef[[1]]
mi.lnorm.sigma <- mi.lnorm$scale


#2: Plot Parametric and Non-parametric (KM) Curves:
mi.KM <- survfit(Surv(time,status)~1,data=mi,type="kaplan-meier")
plot.1 <- function(cex=1){
plot(c(0,mi.KM$time),c(1,mi.KM$surv),ylim=c(0,1),type="s",lty=1,
     xlab="Time to MI from Onset of Symptoms",ylab="Survival",
     main="Survival Curve for Time to MI from Onset of Symptoms",lwd=3)
curve(weib.surv(x,mi.weib.gamma,mi.weib.lambda),fr=0,to=900,add=T,col="blue",lwd=2)
curve(loglog.surv(x,mi.llog.gamma,mi.llog.lambda),fr=0,to=900,add=T,col="red",lwd=2)
curve(lognorm.surv(x,mi.lnorm.mu,mi.lnorm.sigma),fr=0,to=900,add=T,col="green",lwd=2)
legend("topright",legend=c("K.M. Curve","Weibull","loglogistic","lognormal"),
       col=c("black","blue","red","green"),lwd=3,cex=cex)
# It appears that the lognormal curve follows the K.M. curve the closest.
}


#3: Fit using Covariates:
mi.lnorm.cov <-   survreg(Surv(time,status)~.,dat=mi,dist="loglogistic",x=T)
  # Parameter Estimates:
  mu <- mi.lnorm.cov$coef[[1]]
  sigma <- mi.lnorm.cov$scale
  coef <- mi.lnorm.cov$coef[-1]

  # Summary Tables:
  summary(mi.lnorm.cov)
  # Plots:
  lnorm.surv.cov <- function(t,x,m=mu,s=sigma,cf=coef){
    eta <- rbind(x) %*% cf
    1- pnorm((log(t) - (eta+m))/s)
  }

  p4.25 <- round(quantile(mi$perf4,.25),2)
  p4.75 <- round(quantile(mi$perf4,.75),2)
  mi.KM.g <- survfit(Surv(time,status)~gender,data=mi,type="kaplan-meier")
```

```
  plot.2 <- function(cex=1) {
  plot(mi.KM.g,xlab="Time",ylab="Survival",
       main="Survival Curve for Time to MI from Onset of Symptoms",
       lwd=1,col=c(2,4))
  curve(lnorm.surv.cov(x,c(0,p4.25)),fr=0,to=1200,add=T,col='red',lwd=3,lty=3)
  curve(lnorm.surv.cov(x,c(1,p4.25)),fr=0,to=1200,add=T,col='blue',lwd=3,lty=3)
  curve(lnorm.surv.cov(x,c(0,p4.75)),fr=0,to=1200,add=T,col='red',lwd=3,lty=4)
  curve(lnorm.surv.cov(x,c(1,p4.75)),fr=0,to=1200,add=T,col='blue',lwd=3,lty=4)
  legend("topright",col=rep(c(4,2),3),lwd=c(1,1,3,3,3,3),lty=c(1,1,3,3,4,4),
         leg=c(
           paste(c("Male","Female"),"KM Survival Surve"),
           paste(c("Male","Female"),"Survival Curve at perf4 =",p4.25),
           paste(c("Male","Female"),"Survival Curve at perf4 =",p4.75)
           ),cex=cex)
  }

#4: Dichotomize perf4:
mi.d <- mi
p4.cut <- round(quantile(mi$perf,.75),2)
mi.d$perf4 <- ifelse(mi.d$perf>p4.cut,1,0)
mi.lnorm.cov.d <- survreg(Surv(time,status)~.,dat=mi.d,dist="loglogistic",x=T)
  # Parameter Estimates:
  mu.d <-    mi.lnorm.cov.d$coef[[1]]
  sigma.d <- mi.lnorm.cov.d$scale
  coef.d <-  mi.lnorm.cov.d$coef[-1]

  # Summary Tables:
  summary(mi.lnorm.cov.d)
  # Plots:
  lnorm.surv.cov.d <- function(t,x,m=mu.d,s=sigma.d,cf=coef.d){
    eta <- rbind(x) %*% cf
    1- pnorm((log(t) - (eta+m))/s)
  }

  plot.3 <- function(cex=1){
  plot(mi.KM.g,xlab="Time",ylab="Survival",
       main="Survival Curve for Time to MI from Onset of Symptoms",
       lwd=1,col=c(2,4))
  curve(lnorm.surv.cov.d(x,c(0,0)),fr=0,to=1200,add=T,col='red',lwd=3,lty=3)
  curve(lnorm.surv.cov.d(x,c(1,0)),fr=0,to=1200,add=T,col='blue',lwd=3,lty=3)
  curve(lnorm.surv.cov.d(x,c(0,1)),fr=0,to=1200,add=T,col='red',lwd=3,lty=4)
  curve(lnorm.surv.cov.d(x,c(1,1)),fr=0,to=1200,add=T,col='blue',lwd=3,lty=4)
  legend("topright",col=rep(c(4,2),3),lwd=c(1,1,3,3,3,3),lty=c(1,1,4,4,3,3),
         leg=c(
           paste(c("Male","Female"),"KM Survival Surve"),
           paste(c("Male","Female"),"Survival Curve at perf4 >",p4.cut),
           paste(c("Male","Female"),"Survival Curve at perf4 <",p4.cut)
           ),cex=cex)
  }
```

```
#5: Model Fit:
  # Assess AFT Assumption: Q-Q plot
  # One for each covariate

  mi.mal.lo <- mi.d[which(mi.d$perf4==0 & mi.d$gender=="m"),]
  mi.mal.hi <- mi.d[which(mi.d$perf4==1 & mi.d$gender=="m"),]
  mi.fem.lo <- mi.d[which(mi.d$perf4==0 & mi.d$gender=="f"),]
  mi.fem.hi <- mi.d[which(mi.d$perf4==1 & mi.d$gender=="f"),]

  glm.KM <- survfit(Surv(time,status)~1,data=mi.mal.lo)
  ghm.KM <- survfit(Surv(time,status)~1,data=mi.mal.hi)
  glf.KM <- survfit(Surv(time,status)~1,data=mi.fem.lo)
  ghf.KM <- survfit(Surv(time,status)~1,data=mi.fem.hi)

  glm.p <- ghm.p <- glf.p <- ghf.p <- NA
  #p <- seq(.1,1,by=.1)
  p <- 1:100/100
  for(i in 1:length(p)) {
    glm.p[i] <- min(glm.KM$time[glm.KM$surv <= (1-p[i])])
    ghm.p[i] <- min(ghm.KM$time[ghm.KM$surv <= (1-p[i])])
    glf.p[i] <- min(glf.KM$time[glf.KM$surv <= (1-p[i])])
    ghf.p[i] <- min(ghf.KM$time[ghf.KM$surv <= (1-p[i])])
  }

  m.index <- min(c(sum(glm.p < Inf), sum(ghm.p < Inf)))
  plot.male.qq <- function(){
    plot(glm.p[1:m.index],ghm.p[1:m.index],type="p",
         main="Q-Q Plot for Males",col="blue",lwd=3,
         xlab="Percentile for Male Low perf4 Group",
         ylab="Percentile for Male High perf4 Group")
  }

  f.index <- min(c(sum(glf.p < Inf), sum(ghf.p < Inf)))
  plot.female.qq <- function(){
    plot(glf.p[1:f.index],ghf.p[1:f.index],type="p",
         main="Q-Q Plot for Females",col="red",lwd=3,
         xlab="Percentile for Female Low perf4 Group",
         ylab="Percentile for Female High perf4 Group")
  }

#6: Assess Parametric Assumptiosn:
    #Probability plots
    summary(mi.KM)

    # Weibull
    plot.weib.prob <- function()
      plot(log(mi.KM$time), log(log(1/(mi.KM$surv))),pch=19,
           xlab="Log of Time to MI from Onset of Symptoms",
           ylab="log of Hazard of Time to MI from Onset of Symptoms",
           main="Probability Plot for Weibull",col="blue")
```

```
    # Lognormal
    plot.lnorm.prob <- function()
      plot(log(mi.KM$time), qnorm(1-mi.KM$surv),pch=19,
           xlab="Log of Time to MI from Onset of Symptoms",
           ylab="Percentiles of Time to MI from Onset of Symptoms",
           main="Probability Plot for Lognormal",col="green")

    # Loglogistic
    plot.llog.prob <- function()
      plot(log(mi.KM$time), log((1/(mi.KM$surv))-1),pch=19,
           xlab="Log of Time to MI from Onset of Symptoms",
           ylab="Negative Log Odds of Time to MI from Onset of Symptoms ",
           main="Probability Plot for Loglogistic",col="red")

#7: Assessing Overall Fit: Deviance Residuals
    mi.lnorm.dev <- residuals(mi.lnorm,type="deviance")
    plot.resid <- function()
      plot(mi.lnorm.dev,pch=19,col='purple',main="Residuals Plot")


# Plots:
plot.1()
plot.2()
plot.3()

plot.male.qq()   #KM
plot.female.qq() #KM

plot.weib.prob()  #AFT
plot.lnorm.prob() #AFT
plot.llog.prob()  #AFT

plot.resid()
```