

# Parametric

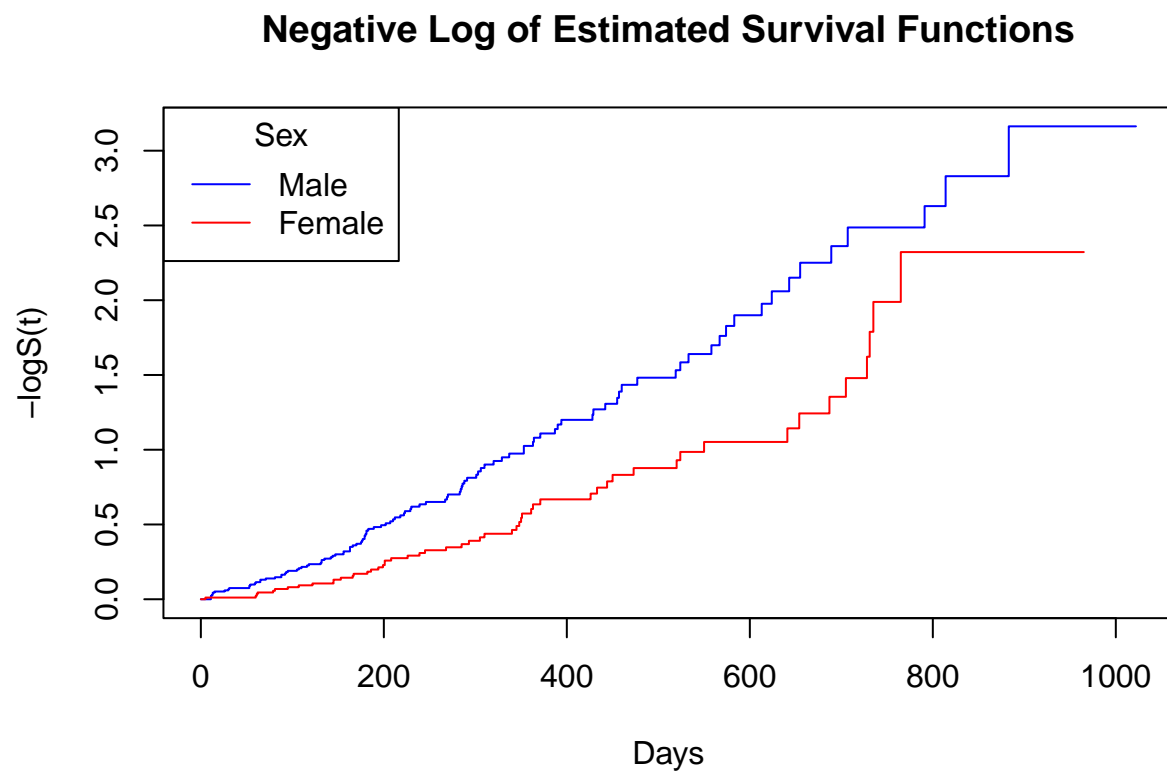
Jibei Zheng jz3425

```
lung_df = lung %>%  
  mutate_at(c(1, 3, 5, 6), .funs = ~as.factor(.))
```

## Model Checking

Plot  $-\log\hat{S}(t)$

```
lung_fit = survfit(Surv(time, status == 2) ~ sex,  
  data = lung_df)  
  
plot(lung_fit, col = c("blue", "red"),  
  fun = "cumhaz", xlab = "Days", ylab = "-logS(t)",  
  main = "Negative Log of Estimated Survival Functions")  
legend("topleft", legend = c("Male", "Female"),  
  title = "Sex", col = c("blue", "red"), lty = 1)
```

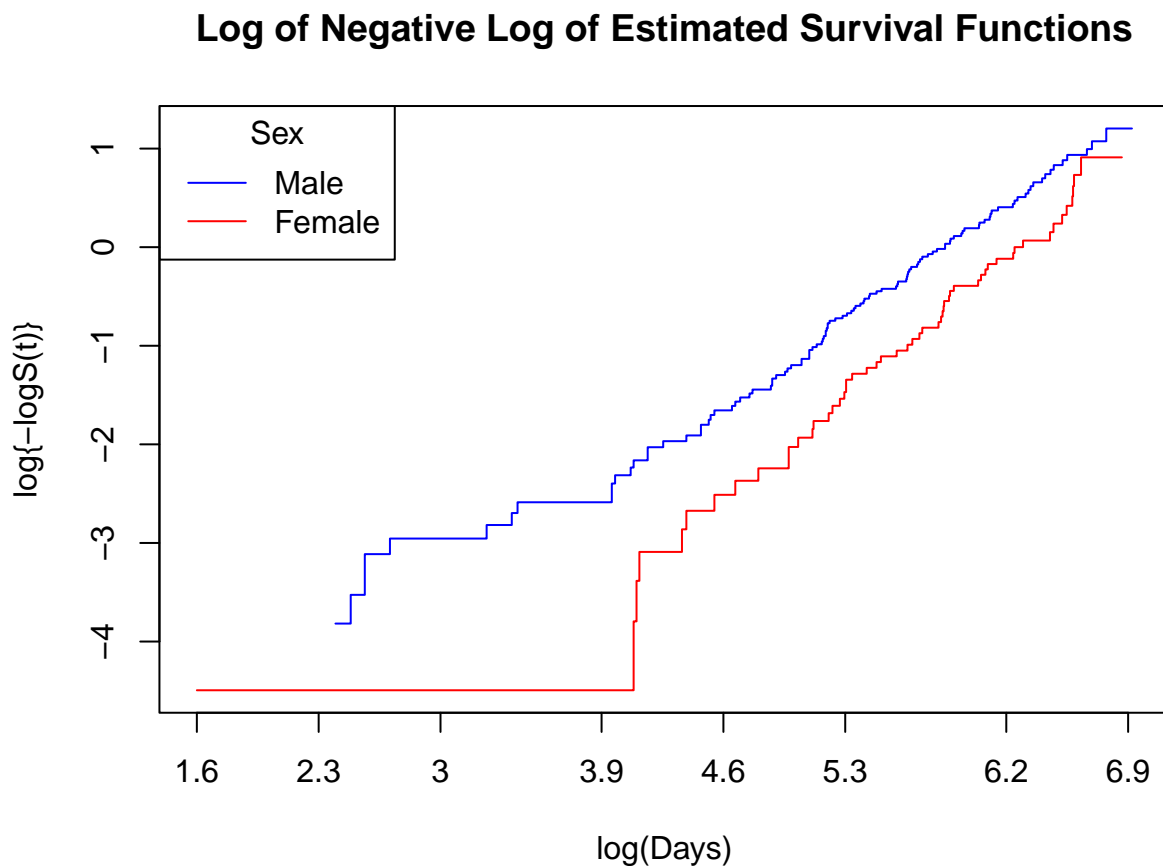


The curve for males is close to a straight line, while the curve for females is obviously non-linear, indicating a better choice of the Weibull distribution.

Plot  $\log(-\log S(t))$

```
x_tick = c(5, 10, 20, 50, 100, 200, 500, 1000)
log_x = round(log(x_tick), 1)

plot(lung_fit, col = c("blue", "red"), fun = "cloglog",
     xlab = "log(Days)", ylab = "log{-logS(t)}",
     xaxt = "n", main = "Log of Negative Log of Estimated Survival Functions")
axis(1, at = x_tick, labels = log_x)
legend("topleft", legend = c("Male", "Female"),
      title = "Sex", col = c("blue", "red"), lty = 1)
```



The slope of the male curve is close to 1, while the slope of the female curve is larger than 1, also indicating a Weibull distribution.

### Fit exponential and Weibull model

```
#parametric survival function
fit_exp = flexsurvreg(Surv(time, status == 2) ~ sex,
```

```

data = lung_df, dist = "exp")
fit_web <- flexsurvreg(Surv(time, status == 2) ~ sex,
data = lung_df, dist = "weibull")

```

```

#exp parameter estimation and CI
fit_exp

```

```

## Call:
## flexsurvreg(formula = Surv(time, status == 2) ~ sex, data = lung_df,
##   dist = "exp")
##
## Estimates:
##      data mean  est      L95%      U95%      se      exp(est)
## rate      NA  0.002865  0.002381  0.003448  0.000271      NA
## sex2  0.394737 -0.500399 -0.827169 -0.173628  0.166723  0.606289
##      L95%      U95%
## rate      NA      NA
## sex2  0.437285  0.840609
##
## N = 228, Events: 165, Censored: 63
## Total time at risk: 69593
## Log-likelihood = -1157.6, df = 2
## AIC = 2319.199

```

```

#Weibull parameter estimation and CI
fit_web

```

```

## Call:
## flexsurvreg(formula = Surv(time, status == 2) ~ sex, data = lung_df,
##   dist = "weibull")
##
## Estimates:
##      data mean  est      L95%      U95%      se      exp(est)  L95%
## shape      NA   1.324   1.173   1.495   0.082      NA      NA
## scale      NA  359.301 312.034 413.729 25.857      NA      NA
## sex2    0.395   0.396   0.145   0.646   0.128   1.485   1.156
##      U95%
## shape      NA
## scale      NA
## sex2    1.907
##
## N = 228, Events: 165, Censored: 63
## Total time at risk: 69593
## Log-likelihood = -1148.652, df = 3
## AIC = 2303.303

```

```

#plot km, exp fitted and web fitted
plot(fit_web,
lwd = 2, lwd.obs = 1,
col = c("blue", "red"), col.obs = c("skyblue", "pink"),
xlab = "Days", ylab = "Survival Probability",
main = "KM and Parametric Est with 95% CI")

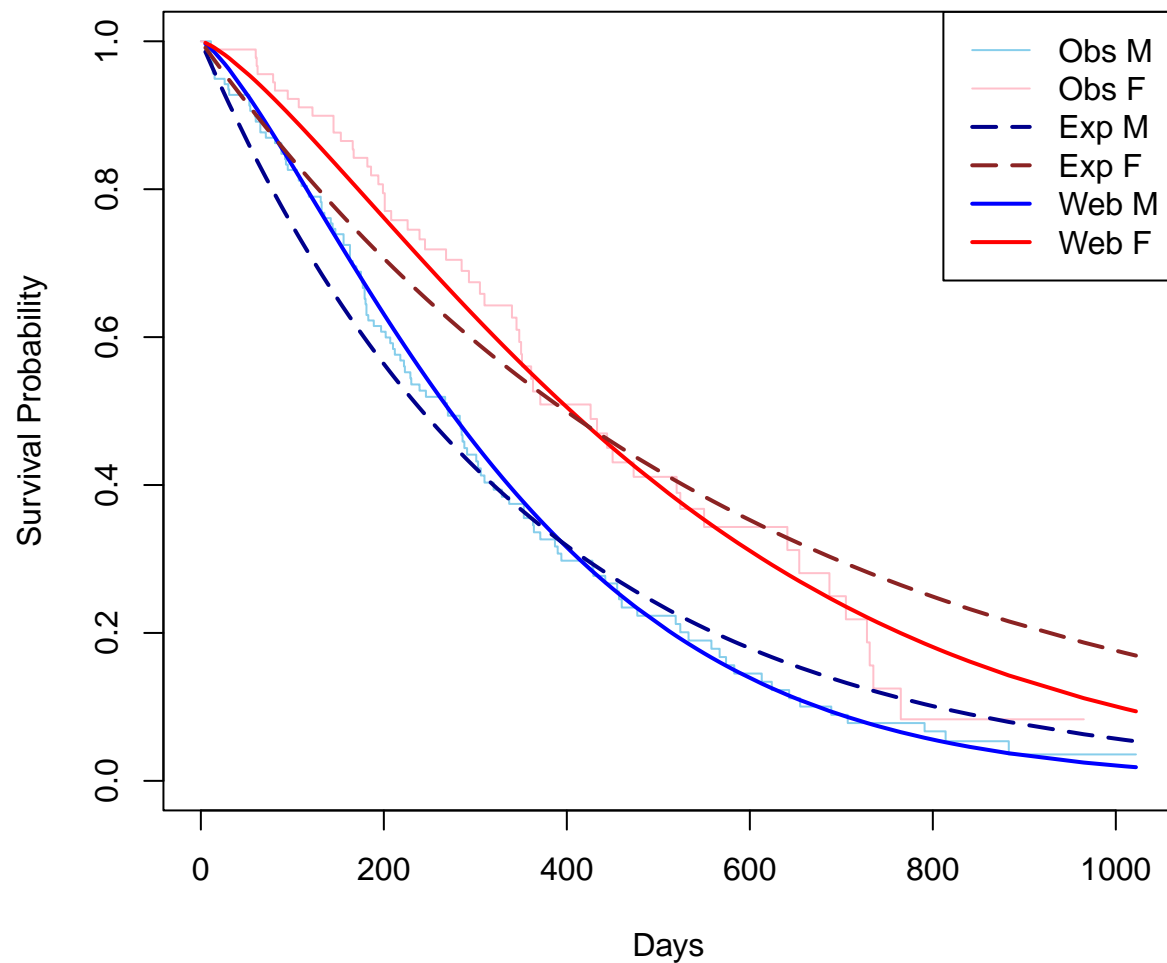
```

```

plot(fit_exp, add = TRUE,
     lwd = 2, col = c("blue4", "brown4"), lty = "longdash")
legend("topright", legend = c("Obs M", "Obs F", "Exp M", "Exp F", "Web M", "Web F"),
     col = c("skyblue", "pink", "blue4", "brown4", "blue", "red"),
     lty = c("solid", "solid", "longdash", "longdash", "solid", "solid"),
     lwd = c(1,1,2,2,2,2))

```

## KM and Parametric Est with 95% CI



From the plot we can see that fitting a Weibull distribution is actually more precise than an exponential distribution.