DISTRIBUTION PARAMETERIZATION FROM SURVIVAL ANALYSIS in R

This document is a guide to interpret distribution parameters obtained from R survival analysis using SURVREG and FLEXSURVREG packages and to use them properly within TreeAge Pro's distributions parameters.

Please note that SURVREG and FLEXSURVREG generate parameters which are presented differently and often need to be further transformed with exp( ) or log( ) expressions in order to be equivalent to each other.

In the examples below a following steps were performed in R.

1. Generated 100,000 samples from a particular distribution with given input parameters.
2. Feed the 100,000 samples into SURVREG and FLEXSURVREG (no censoring) to obtain the estimates of the parameters for the given distribution.
3. The estimated parameters must match the input parameters from step 1, but often need to be transformed.
    a. Some distributions have different parameterization in TreeAge Pro and in R, appropriate transformation of parameters for TreeAge Pro (**TP**) are shown in square green shaded boxes.

LOGNORMAL DISTRIBUTION

```
LOGNORMAL SAMPLES from R function:

Y = rlnorm(100000,meanlog = 1.1, sdlog = 1.2)
```

The Results from the SURVREG function

```
Call:
survreg(formula = Surv(time, psurv) ~ 1, data = myData2, dist = "lognormal")

Coefficients:
(Intercept)
    1.10127

Scale= 1.201226

Loglik(model)= -270355.1   Loglik(intercept only)= -270355.1
n= 100000
```

Mu (Mean of Logs) = 1.10127

Sigma (std. dev. of logs) = 1.201226

```
*******************************
```

The Results from the FLEXSURVREG function

```
Call:
flexsurvreg(formula = Surv(time, psurv) ~ 1, data = myData2,
    dist = "lognormal")

Estimates:
         est      L95%     U95%      se
meanlog  1.10127  1.09383  1.10872  0.00380
sdlog    1.20123  1.19597  1.20650  0.00269

N = 100000,  Events: 100000,  Censored: 0
Total time at risk: 617840.2
Log-likelihood = -270355.1, df = 2
AIC = 540714.3
```

Mu (Mean of Logs) = 1.10127

Sigma (std. dev. of logs) = 1.20123

```
**********************
Mean of Samples = 6.178402, Standard Deviation = 10.883834
```

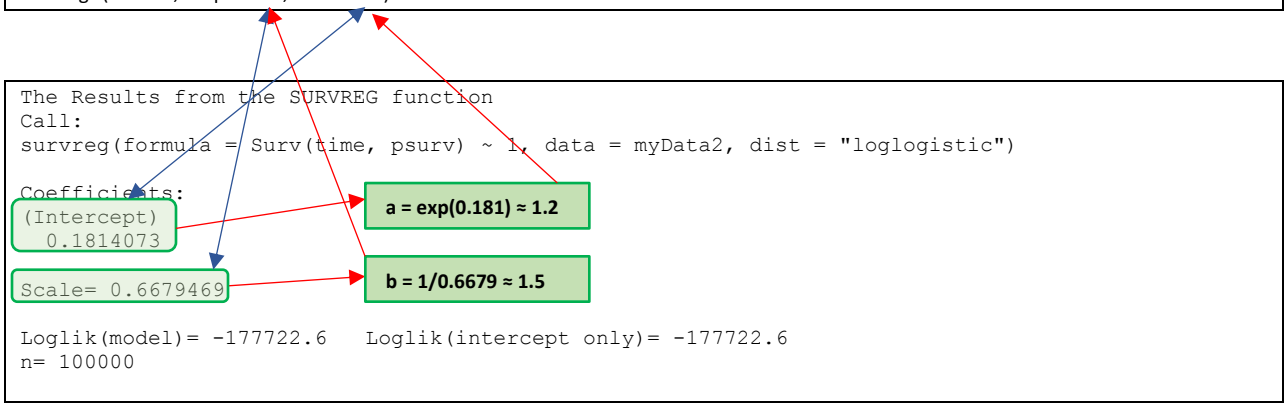
TP Parameters mu (Mean of Logs) = 1.1 sigma (std. dev. of logs) = 1.2

LOGLOGISTIC DISTRIBUTION - notice the confusing implementation of R parameterizations of SURVREG output!

****

LOGLOGISTIC SAMPLES from R function:

Y = rllogis(100000, shape = 1.5, scale = 1.2)



```
The Results from the SURVREG function
Call:
survreg(formula = Surv(time, psurv) ~ 1, data = myData2, dist = "loglogistic")

Coefficients:
(Intercept)
  0.1814073

Scale= 0.6679469

Loglik(model)= -177722.6   Loglik(intercept only)= -177722.6
n= 100000
```

a = exp(0.181) ≈ 1.2

b = 1/0.6679 ≈ 1.5

****

```
The Results from the FLEXSURVREG function
Call:
flexsurvreg(formula = Surv(time, psurv) ~ 1, data = myData2,
    dist = "llogis")

Estimates:
       est      L95%     U95%     se
shape  1.49706  1.48932  1.50483  0.00396
scale  1.19889  1.19032  1.20752  0.00439

N = 100000,  Events: 100000,  Censored: 0
Total time at risk: 301627.5
Log-likelihood = -177722.6, df = 2
AIC = 355449.3
```

b ≈ 1.5

a ≈ 1.2

****

Mean of Samples = 3.016275, Standard Deviation = 58.476523

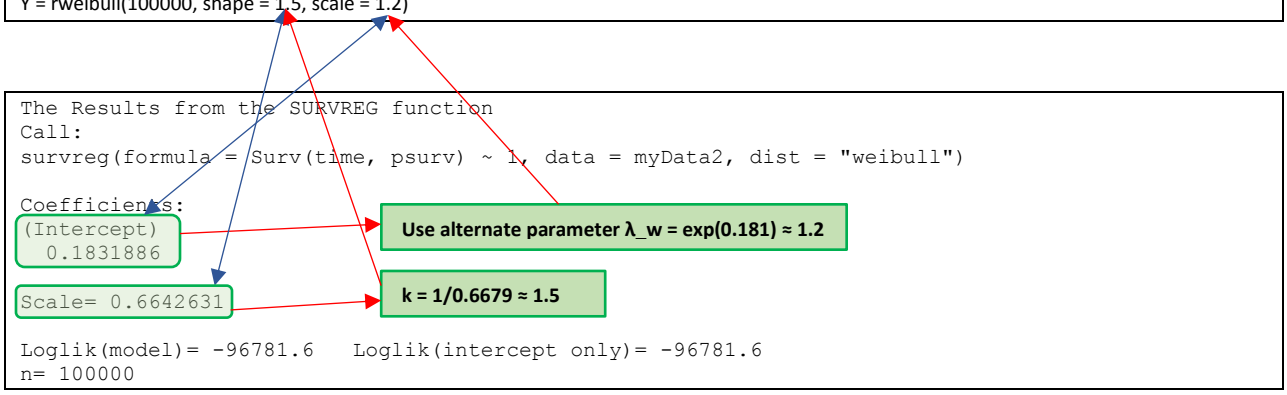
TP Parameters a = 1.2 b = 1.5

WEIBULL DISTRIBUTION - notice the confusing implementation of R parameterizations of SURVREG output!

****

WEIBULL SAMPLES from R function:

Y = rweibull(100000, shape = 1.5, scale = 1.2)



```
The Results from the SURVREG function
Call:
survreg(formula = Surv(time, psurv) ~ 1, data = myData2, dist = "weibull")

Coefficients:
(Intercept)
  0.1831886

Scale= 0.6642631

Loglik(model)= -96781.6   Loglik(intercept only)= -96781.6
n= 100000
```

Use alternate parameter λ_w = exp(0.181) ≈ 1.2

k = 1/0.6679 ≈ 1.5

****

```
The Results from the FLEXSURVREG function
Call:
flexsurvreg(formula = Surv(time, psurv) ~ 1, data = myData2,
    dist = "weibull")

Estimates:
       est      L95%     U95%     se
shape  1.50543  1.49817  1.51272  0.00371
scale  1.20104  1.19585  1.20626  0.00266

N = 100000,  Events: 100000,  Censored: 0
Total time at risk: 108371.7
Log-likelihood = -96781.64, df = 2
AIC = 193567.3
```

k ≈ 1.5

Use alternate parameter λ_w ≈ 1.2

****

Mean of Samples = 1.083717, Standard Deviation = 0.733614

TP Parameters Alternative Parameters λ_w = 1.2 k = 1.5

EXPONENTIAL DISTRIBUTION - notice the confusing implementation of R parameterizations of SURVREG output!

****

EXPONENTIAL SAMPLES from R function:

Y = rexp(100000, rate = 1.2)

```
The Results from the SURVREG function
Call:
survreg(formula = Surv(time, psurv) ~ 1, data = myData2, dist = "exponential")

Coefficients:
(Intercept)
 -0.1873548

Scale fixed at 1

Loglik(model)= -81264.5   Loglik(intercept only)= -81264.5
n= 100000
```

λ = exp(- (-0.181)) ≈ 1.2

****

```
The Results from the FLEXSURVREG function
Call:
flexsurvreg(formula = Surv(time, psurv) ~ 1, data = myData2,
    dist = "exp")

Estimates:
      est      L95%     U95%     se
rate  1.20606  1.19860  1.21355  0.00381

N = 100000,  Events: 100000,  Censored: 0
Total time at risk: 82914.95
Log-likelihood = -81264.52, df = 1
AIC = 162531
```

λ ≈ 1.2

****
Mean of Samples = 0.829149, Standard Deviation = 0.828808

TP Parameter λ = 1.2

GENERALIZED GAMMA DISTRIBUTION - not supported by R SURVREG package.

****

GENERALIZED GAMMA (ORIGINAL) SAMPLES from R function:

Y = rgengamma.orig(100000, shape = 1.2, scale=0.9, k=1.3)

```
The Results from the FLEXSURVREG function

Call:
flexsurvreg(formula = Surv(time, psurv) ~ 1, data = myData2,
    dist = "gengamma.orig")

Estimates:
       est     L95%    U95%     se
shape  1.1867  1.1616  1.2123  0.0129
scale  0.8879  0.8522  0.9250  0.0186
k      1.3211  1.2750  1.3690  0.0240

N = 100000,  Events: 100000,  Censored: 0
Total time at risk: 107271
Log-likelihood = -98912.74, df = 3
AIC = 197831.5
```

c ≈ 1.2

Beta ≈ 0.9

Alpha ≈ 1.3

****

Mean of Samples = 1.063595, Standard Deviation = 0.7829048

TP Parameters c = 1.2, Alpha =1.3, Beta = 0.9