# University of Edinburgh, School of Mathematics

# Biostatistics (MATH11230), 2021/2022

## Vanda Inácio

In this supplementary file I show how to fit a Weibull distribution to a sample that contains event censored times. The `survreg` function fits a Weibull model with two parameters: intercept, say $\beta_0$ and a scale parameter $\sigma$. The two parameters we use $\lambda$ and $\alpha$ are related to $\beta_0$ and $\sigma$ by

$$\lambda = e^{-\beta_0/\sigma}, \qquad \alpha = \frac{1}{\sigma}.$$

We will analyse the `veteran` dataset available in the package survival. For more information about the dataset, please type `help(veteran)`.

```
require(survival)
fit_weibull <- survreg(Surv(time, status) ~ 1, dist = "weibull", data = veteran)
summary(fit_weibull)
```

```
##
## Call:
## survreg(formula = Surv(time, status) ~ 1, data = veteran, dist = "weibull")
##              Value Std. Error    z       p
## (Intercept) 4.7931     0.1078 44.46 <2e-16
## Log(scale) 0.1601     0.0669  2.39  0.017
##
## Scale= 1.17
##
## Weibull distribution
## Loglik(model)= -748.1   Loglik(intercept only)= -748.1
## Number of Newton-Raphson Iterations: 6
## n= 137
```

```
alpha <- 1/ fit_weibull$scale
lambda <- exp(- fit_weibull$coefficients/fit_weibull$scale)

y <- seq(0, 1000, by = 1)
est_surv <- exp(-lambda * y^alpha)

#Kaplan--Meier estimate
km_fit <- survfit(Surv(time, status) ~ 1, conf.type = "log-log", data = veteran)

plot(km_fit, conf.int = FALSE, lwd = 2)
lines(y, est_surv, col = 2, lwd = 2)
```
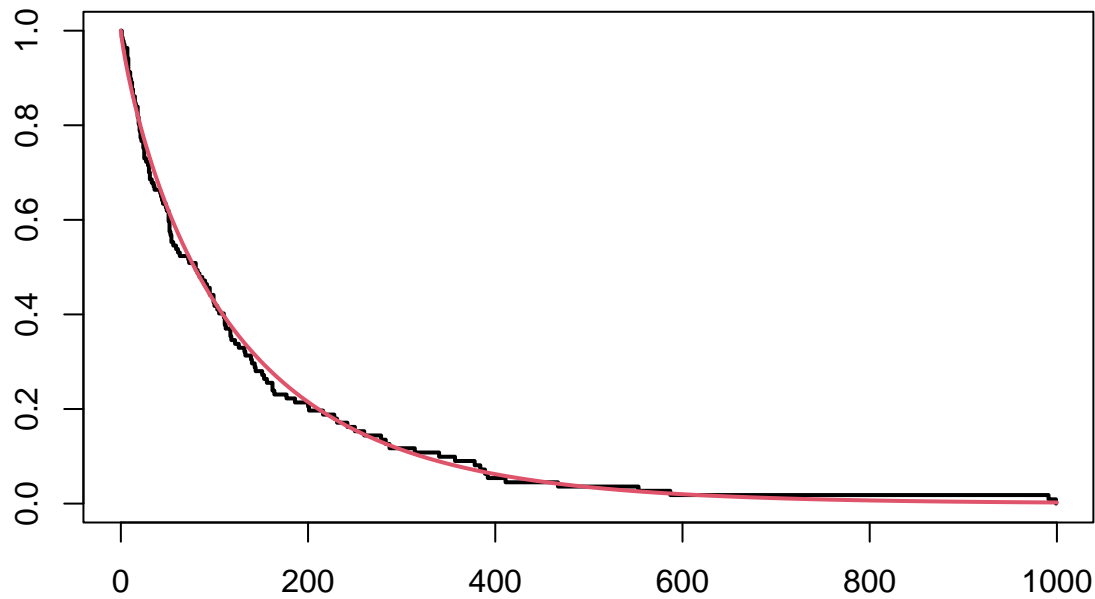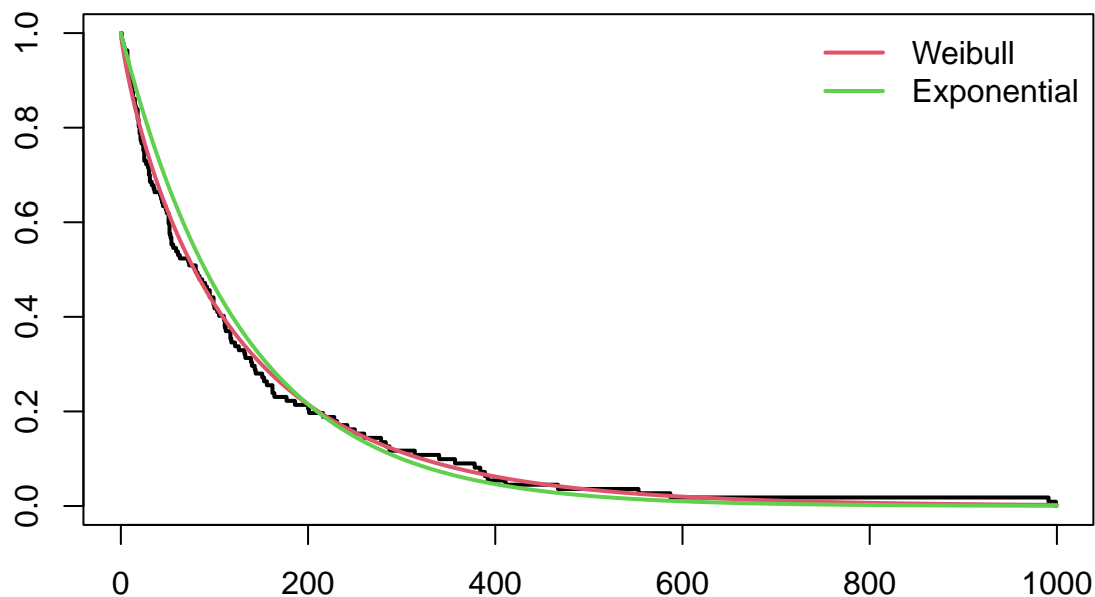
Let us compare with the fit of the exponential model.

```r
theta_est <- sum(veteran$status)/sum(veteran$time)
theta_est
```

```
## [1] 0.00768169
```

```r
est_surv_exp <- exp(-theta_est*y)
plot(km_fit, conf.int = FALSE, lwd = 2)
lines(y, est_surv, col = 2, lwd = 2)
lines(y, est_surv_exp, col = 3, lwd = 2)
legend("topright", col = c(2, 3),
       lty = c(1,1), lwd = c(2, 2),
       c("Weibull", "Exponential"), bty = "n")
```



Although the two fits are similar, when we include the pointwise confidence bands of the Kaplan-Meier estimate, we see that for small times, the exponential estimate of the survival distribution is not included in

the bands. Nevertheless, in this case, both parametric models seem to follow quite nicely the Kaplan-Meier estimate.

```r
plot(km_fit, conf.int = TRUE, lwd = 2)
lines(y, est_surv, col = 2, lwd = 2)
lines(y, est_surv_exp, col = 3, lwd = 2)
legend("topright", col = c(2, 3),
       lty = c(1,1), lwd = c(2, 2),
       c("Weibull", "Exponential"), bty = "n")
```