1. The following edited output from a Kaplan-Meier estimate of a survival curve using R. Assume that all censoring events are due to right-censoring and that there is no truncation. Also assume that there were 26 individuals at the start of the study.

```
> summary(km)
Call: survfit(formula = Surv(x$dur, x$status) ~ 1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    8     26       2    0.923  0.0523        0.826        1.000
   13     24       1    0.885  0.0627        0.770        1.000
   18     23       1    0.846  0.0708        0.718        0.997
   23     22       1    0.808  0.0773        0.670        0.974
   52     21       1    0.769  0.0826        0.623        0.949
   63     20       2    0.692  0.0905        0.536        0.895
   70     18       2    0.615  0.0954        0.454        0.834
   76     16       1    0.577  0.0969        0.415        0.802
  180     15       1    0.538  0.0978        0.377        0.769
  195     14       1    0.500  0.0981        0.340        0.734
  210     13       1    0.462  0.0978        0.305        0.699
  220     12       1    0.423  0.0969        0.270        0.663
  632     10       1    0.381  0.0960        0.232        0.624
  700      9       1    0.338  0.0942        0.196        0.584
 1296      7       1    0.290  0.0923        0.155        0.541
```

a. Find Y_1, Y_2, Y_3, and Y_4.

Y_i is the number at risk, so these values are $Y_1 = 26$, $Y_2 = 24$, $Y_3 = 23$, $Y_4 = 22$.

b. How many observations were censored?

Two observations were censored with times less than 1296, and 6 were censored with times greater than 1296, so 8 observations were censored.

c. Was the largest observation censored? Explain how you know.

Yes, we know because the survival estimate at 1286 is not 0, and also because there were 7 individuals at risk at time 1296 and only 1 experienced the event, so there must have been 6 observations with censored times larger than 1296.

d. What was the median survival time?

195, this is when the estimated survival curve is at 0.5.

e Describe how you could compute the mean survival time from this output. Discuss any complications with computing the mean survival time, but don't actually compute the mean survival time (this would just take too long).

To compute the mean survival time, you need the area under the survival curve. This is done by

$$Mean = \sum_{i=2}^{D} (t_i - t_{i-1})(\widehat{S}(t_i) - \widehat{S}(t_{i-1})$$

for times up to the last observation. If the last largest observation were a death time, then this would be sufficient, but since there are censored observations, this underestimates the mean. Efron's suggestion is to change the largest censored to a death time, so that the survival curve is estimated to be 0 beyond this, and then apply this summation. If the survival curve continues indefinitely at 0.29 after $t = 1296$, then the area under the curve is infinite unless the curve is truncated at some point.

2.10 A model used in the construction of life tables is a piecewise, constant hazard rate model. Here the time axis is divided into k intervals, $[\tau_{i-1}, \tau_i)$, $i = 1, \ldots, k$, with $\tau_o = 0$ and $\tau_k = \infty$. The hazard rate on the ith interval is a constant value, θ_i; that is

$$h(x) = \begin{cases} \theta_1 & 0 \leq x < \tau_1 \\ \theta_2 & \tau_1 \leq x < \tau_2 \\ \vdots & \\ \theta_{k-1} & \tau_{k-2} \leq x < \tau_{k-1} \\ \theta_k & x \geq \tau_{k-1} \end{cases}.$$

(a) Find the survival function for this model.

(b) Find the mean residual-life function.

(c) Find the median residual-life function.

3.5 Suppose that the time to death has a log logistic distribution with parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$. Based on the following left-censored sample, construct the likelihood function.

DATA: 0.5, 1, 0.75, 0.25-, 1.25-, where - denotes a left- censored observation.

For problem 2.10, from page 36, the survival function is related to the hazard by

$$S(t) = \prod_{t_j \leq t} [1 - h(t_j)]$$

Thus we have

$$S(x) = \begin{cases} (1-\theta_1) & 0 \leq x < \tau_1 \\ (1-\theta_1)(1-\theta_2) & \tau_1 \leq x < \tau_2 \\ \vdots & \\ \prod_{j=1}^{k-1}(1-\theta_j) & \tau_{k-2} \leq x < \tau_1 \\ \prod_{j=1}^{k}(1-\theta_j) & x \geq \tau_{k-1} \end{cases}$$

For the mean residual life, there isn't much to do except write down the formula from the book, also on page 36, that for $(x_{i+1} \leq x < x_i$,

$$\mathrm{mrl}(x) = \frac{(x_{i+1} - x)S(x_i) + \sum_{j \geq i+1}(x_{j+1} - x_j)S(x_j)}{S(x)}$$

where $S(x)$ is defined as above. The first term in the sum can be written as

$$\frac{(x_{i+1} - x_i)(S(x_i)}{S(x)} = \frac{x_{i+1} - x_i}{1 - \theta_i}$$

The median residual life given that you have lived until time x is the smallest value x_m such that $S(x_m | x_m > x) \leq 0.5$. For the conditional survival function, we find the smallest value x_m such that $\frac{S(x_m)}{S(x)} \leq 0.5$ This can also be written as $\inf\{t : S(t) \leq 0.5S(x)\}$, in otherwords, the smallest value of t such that $S(t)$ is less than half of $S(x)$.

The solution to 3.5 is in the back of the book.