Problem 1

a) The survival function of T becomes

$$R(t) = e^{-Z(t)} = e^{-\ln(t+1)} = \frac{1}{t+1},$$

the pdf

$$f(t) = -\frac{d}{dt}R(t) = \frac{1}{(t+1)^2},$$

and the hazard

$$z(t) = \frac{d}{dt}Z(t) = \frac{1}{t+1}.$$

At the median $q_{1/2}$, $R(q_{1/2}) = \frac{1}{2}$ such that $\frac{1}{1+q_{1/2}} = \frac{1}{2}$ and $q_{1/2} = 1$.

b) The expected survival time becomes

$$ET = \int_0^\infty R(t)dt = \int_0^\infty \frac{1}{t+1}dt = \ln(t+1)\Big|_0^\infty = \infty,$$

that is, it is not finite, and

$$E\ln(T+1) = EZ(T) = 1,$$

since $Z(T) \sim \exp(1)$.

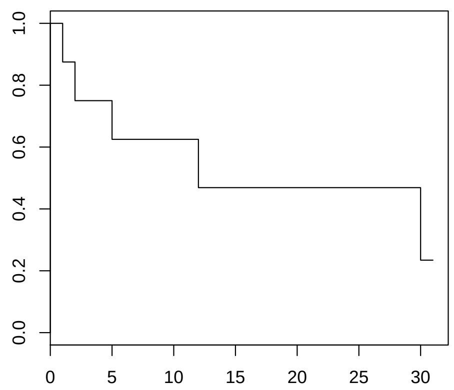
Problem 2

a) The Kaplan-Meier estimator is given by

$$\hat{R}(t) = \prod_{j:t_j \le t}\left(1 - \frac{d_j}{n_j}\right)$$

where t_j, $j = 1, \ldots, 5$ are the ordered distinct failure times, n_j the number at risk prior to those failure times and d_j the number failing.

j	t_j	n_j	d_j	$1 - \frac{d_j}{n_j}$	$\hat{R}(t), t_j \le t < t_{j+1}$
1	1	8	1	7/8	7/8
2	2	7	1	6/7	3/4
3	5	6	1	5/6	5/8
4	12	4	1	3/4	15/32
5	30	2	1	1/2	15/64



An estimate of mean survival time, $\widehat{ET} = \int_0^\infty \hat{R}(t)dt$ but beyond the last right censoring point $y_8 = 31$, $R(t)$ is not identifiable. If assuming that $R(t) = 0$ for $t > 31$, however, we obtain

$$\widehat{ET} = 1 + \frac{7}{8} + 3 \cdot \frac{3}{4} + 7\frac{5}{8} + 18 \cdot \frac{15}{32} + 1 \cdot \frac{15}{64} = \frac{1099}{64} = 17.17.$$

At the median survival time $q_{1/2}$, $R(q_{1/2}) = 1/2$. Based on our estimate of R, an estimate of the median is thus $\hat{q}_{1/2} = 12$.

b) The survival function of the standard logistic distribution becomes

$$R(t) = 1 - F(t) = 1 - \frac{1}{1 + e^{-t}},$$

and the density

$$f(t) = \frac{d}{dt}F(t) = \frac{e^{-t}}{(1 + e^{-t})^2} = \frac{1}{(1 + e^{-t})(1 + e^{t})}.$$

c) A log-location-scale model is constructed from the standard logistic distribution by assuming that

$$\ln T = \mu + \sigma U$$

where U is standard logistic with survival function R. The survival function of T is then

$$\begin{aligned} R_T(t) &= P(T > t) \\ &= P(\mu + \sigma U > \ln t) \\ &= P(U > \frac{\ln t - \mu}{\sigma}) \\ &= R(\frac{\ln t - \mu}{\sigma}) \\ &= 1 - \frac{1}{1 + e^{-\frac{\ln t - \mu}{\sigma}}}. \end{aligned}$$

Similarly,

$$F_T(t) = F(\frac{\ln t - \mu}{\sigma})$$

such that the pdf of T,

$$\begin{aligned} f_T(t) &= \frac{d}{dt}F_T(t) \\ &= \frac{d}{dt}F(\frac{\ln t - \mu}{\sigma}) \\ &= f(\frac{\ln t - \mu}{\sigma})\frac{1}{\sigma t} \\ &= \frac{1}{\sigma t(1+e^{-\frac{\ln t - \mu}{\sigma}})(e^{\frac{\ln t - \mu}{\sigma}})}. \end{aligned}$$

For right censored observations the likelihood is then

$$L(\mu,\sigma) = \prod_{i:\delta_i=1} f_T(y_i) \prod_{i:\delta_i=0} R_T(y_i) = \prod_{i:\delta_i=1} \frac{1}{\sigma t(1+e^{-\frac{\ln y_I - \mu}{\sigma}})(e^{\frac{\ln y_i - \mu}{\sigma}})} \prod_{i:\delta_i=0} 1 - \frac{1}{1+e^{-\frac{\ln y_i - \mu}{\sigma}}}$$

d) The quantile q_α, $\alpha = 0.05$, satisfies

$$P(T > q_\alpha) = \alpha$$

that is,

$$\begin{aligned} R_T(q_\alpha) &= \alpha \\ R(\frac{\ln q_\alpha - \mu}{\sigma}) &= \alpha \\ 1 - \frac{1}{1+e^{-\frac{\ln q_\alpha - \mu}{\sigma}}} &= \alpha \\ \frac{1}{1+e^{-\frac{\ln q_\alpha - \mu}{\sigma}}} &= 1 - \alpha \\ \frac{\ln q_\alpha - \mu}{\sigma} &= \operatorname{logit}(1-\alpha) \\ q_\alpha &= e^{\mu + \sigma \operatorname{logit}(1-\alpha)} \end{aligned}$$

Thus, an estimate of q_α is given by

$$\hat{q}_\alpha = e^{2.59 + 1.01 \ln(0.95/.05)} = 262.58.$$

The estimator

$$\widehat{\ln q_\alpha} = \hat{\mu} + e^{\widehat{\ln \sigma}} \operatorname{logit}(1-\alpha) = 5.57$$

is perhaps better approximated by a normal distribution than $\hat{q}_\alpha$ (which is restricted to positive values). To find its approximate variance we need

$$\begin{aligned} \frac{\partial \ln q_\alpha}{\partial \mu} &= 1, \\ \frac{\partial \ln q_\alpha}{\partial \ln \sigma} &= e^{\ln \sigma} \operatorname{logit}(1-\alpha) = 1.01 \ln(0.95/.05) = 2.97. \end{aligned}$$

Based on the delta method, using the output from `vcov(model)`, we then find that

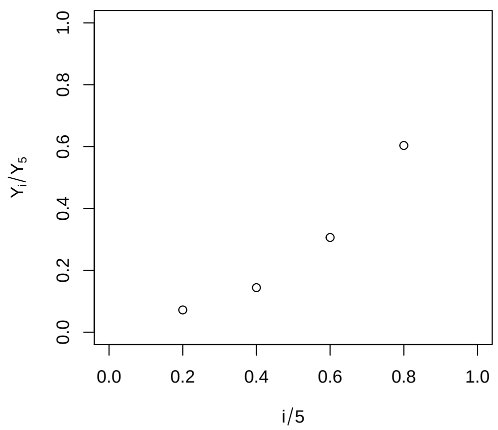
$$\mathrm{Var}(\widehat{\ln q_\alpha}) \approx (1)^2 0.4424 + (2.97)^2 0.1361 + 2 \cdot 1 \cdot 2.97 \cdot 0.03916 = 1.88.$$

An approximate 95% confidence interval for $\ln q_\alpha$ is then $5.57 \pm 1.96\sqrt{1.88} = (2.88, 8.25)$ and a corresponding interval for q_α is $(e^{2.88}, e^{8.25}) = (17.86, 3856.11)$, admittedly a very wide interval reflecting the small amount of data we have.

e) The total time on test at the time of each observed failure becomes as follows.

i	y_i	δ_i	n_i	$\mathcal{T}(y_i)$	j	Y_j	Y_j/Y_5
1	1	1	8	$0+8\cdot 1=8$	1	8	0.07
2	2	1	7	$8+7\cdot 1=15$	2	15	0.14
3	5	1	6	$15+6\cdot 3=33$	3	33	0.30
4	10	0	5	$33+5\cdot 5=58$			
5	12	1	4	$58+4\cdot 2=66$	4	66	0.60
6	20	0	3	$66+3\cdot 8=90$			
7	30	1	2	$90+2\cdot 10=110$	5	110	
8	31	0	1	$110+1\cdot 1=111$			

which gives the following TTT-plot.



The convex shape indicates a decreasing failure rate.

The test statistic of the Barlow-Prochan test becomes

$$Z = \frac{\sum_{i=1}^{4} Y_j/Y_5 - 4/2}{\sqrt{4/12}} = \frac{1.13-2}{.57} = -1.54,$$

which is not smaller than the lower critical value $-z_{0.025} = -1.96$. We can thus not reject the null hypothesis that the data comes from an exponential distribution.

Problem 3

a) The model assumes that the hazard function for each the lifetime T_i of component $i = 1, 2, \dots, n$ is given by

$$z(t; x_i) = z_0(t)e^{\beta x_i}$$

where x_i is the covariate value (mean-centered temperature) for component i.

The regression coefficient β is estimated by maximising the partial likelihood

$$L(\beta) = \prod_{j=1}^{r} \frac{e^{\beta x_{i_j}}}{\sum_{i \in R_j} e^{\beta x_i}}$$

where i_j is the component that failed at the j'th failure and R_j is the set of individual at risk prior to the j'th failure.

b) From Fig. 2, the baseline survival after 100 days is estimated to $\hat{R}_0(100) = 0.78$. The estimated survival at $t = 100$ days for a component with $x_i = 20$ is then

$$\hat{R}(t; x_i = 20) = R_0(t)^{e^{\beta x_i}} = 0.78^{e^{0.10 \cdot 20}} = 0.78^{7.87} = 0.1415. \tag{1}$$

This agrees well with the observed data in Fig. 1—for units operating around 70 degrees ($x = 20$), most failures (roughly about 85%) occur before 100 days.

c) If the proportional hazard assumption holds, the Schoenfeld residuals have expected values of zero. Based on the confidence bands of loess estimated mean there is some indication that proportional effect of x (temperature) on the hazard is not constant across time but instead slightly decreasing for components that become very old (for $\log t > 5.3$, that is, $t > 200$ days).