Problem 2

(a)

Compute

$$
\begin{aligned}
\hat{F}(t_i) &= \frac{\sum_{j=1}^{i} d_j}{130} \\
s.e.(\hat{F}(t_i)) &= \sqrt{\frac{\hat{F}(t_i) \times \left(1 - \hat{F}(t_i)\right)}{130}}
\end{aligned}
$$

t_i	$\hat{F}(t_i)$	$s.e.(\hat{F}(t_i))$
33	0.0077	0.0077
46	0.0154	0.0108
50	0.0231	0.0132
59	0.0308	0.0151
62	0.0385	0.0169
71	0.0462	0.0184
74	0.0538	0.0198
75	0.0615	0.0211
78	0.0769	0.0234
80	0.0769	0.0234

(b)

By Wald method :

$$
\begin{aligned}
\hat{F}_L(t_i) &= \hat{F}(t_i) \; - \; Z_{0.95}\sqrt{\frac{\hat{F}(t_i)\left(1 - \hat{F}(t_i)\right)}{130}} \\
\hat{F}_U(t_i) &= \hat{F}(t_i) \; + \; Z_{0.95}\sqrt{\frac{\hat{F}(t_i)\left(1 - \hat{F}(t_i)\right)}{130}}
\end{aligned}
$$

t_i	$\hat{F}(t_i)$	$\hat{F}_L(t_i)$	$\hat{F}_U(t_i)$	$\hat{F}_U(t_i) - \hat{F}_L(t_i)$
33	0.0077	-0.0049	0.0203	0.0252
46	0.0154	-0.0024	0.0331	0.0355
50	0.0231	0.0014	0.0447	0.0433
59	0.0308	0.0059	0.0557	0.0498
62	0.0385	0.0107	0.0662	0.0555
71	0.0462	0.0159	0.0764	0.0605

t_i	$\hat{F}(t_i)$	$\hat{F}_L(t_i)$	$\hat{F}_U(t_i)$	$\hat{F}_U(t_i) - \hat{F}_L(t_i)$
74	0.0538	0.0213	0.0864	0.0651
75	0.0615	0.0269	0.0962	0.0693
78	0.0769	0.0385	0.1154	0.0769
80	0.0769	0.0385	0.1154	0.0769

- 可以看出 interval 的長度隨著 $\hat{F}(t_i)$ 越接近 0.5 而越長，代表對該機率的預測越不精準

- 前兩個信賴區間的下界為負值，這是使用 Wald method 時候的一個缺點，我們可以選擇將那些下界設為零，或是先進行 logit transformation 後再計算 interval 即可避免此現象

(c)

$$\hat{p}_1 = \frac{1}{130}\ ,\ \hat{p}_2 = \frac{1}{129}\ ,\ \hat{p}_3 = \frac{1}{128} \quad \Rightarrow \quad 1 - \hat{p}_1 = \frac{129}{130}\ ,\ 1 - \hat{p}_2 = \frac{128}{129}\ ,\ 1 - \hat{p}_3 = \frac{127}{128}$$

and

$$\hat{S}(t_i) \ = \ \prod_{j=1}^{i} (1 - \hat{p}_j) \ \Rightarrow \ \hat{S}(t_1) = \frac{129}{130}\ ,\ \hat{S}(t_2) = \frac{128}{130}\ ,\ \hat{S}(t_3) = \frac{127}{130}$$

t_i	$\hat{F}(t_i)$	$\hat{S}(t_i)$	$1 - \hat{S}(t_i)$
33	0.0077	0.9923	0.0077
46	0.0154	0.9846	0.0154
50	0.0231	0.9769	0.0231

由此方法所計算出的 $\hat{F}(t_i) = 1 - \hat{S}(t_i)$ 和 $\hat{F}(t_i) = \frac{\sum_{j=1}^{i} d_j}{130}$ 所計算出的估計值一樣

(d)

如果將每次紀錄的間隔減少，可以增加我們資料的精準度，讓我們更明確的知道單位是在哪一個更精準的 interval censored data，但同時也會大幅的增加實驗成本，在此筆資料中的時間區間設計幾乎可以將每個 interval censored data 給單獨區分開來，由此可見我們每 1000 個 cycles 做一次紀錄已經足夠精準了，除了 $t_i = 78$ 的那個 interval，可以看到在該區間中有兩個 censored datas，如若我們將時間區間調整得更精準，則有可能區分出那兩筆資料的差別。

像是此筆資料的 interval censored data 存在著一個問題，就是我們無法準確得知單位到底是在時間區間中的哪一個點 fail，然而隨著時間區間的縮小，可以逐漸提升我們的準確度，所以要考慮實驗成本以及所需資料要有多準確來決定時間區間大小。

(e)

如果我們只從一、兩個 heats 中選出我們的單位，所做出的結果可能無法代表著整體的特徵，計算出的結果會被那一、兩個 heats 本身的變異所 bias，進而混淆我們想要的結果。

(f)

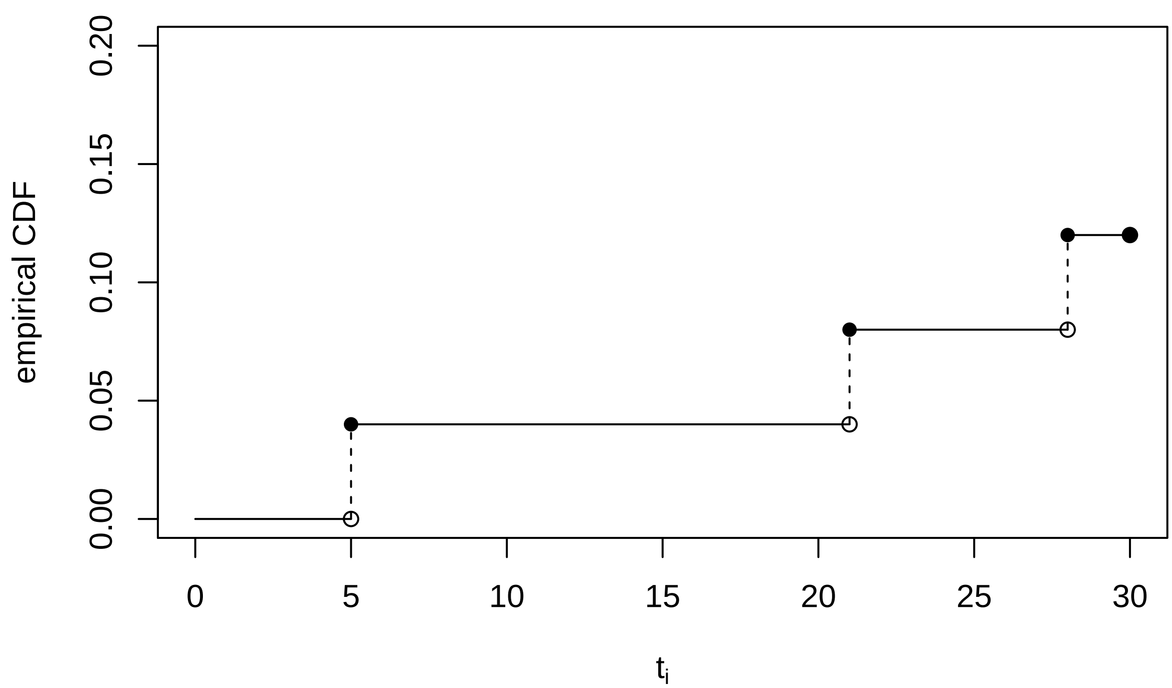
我們無法確定此實驗所使用的機台，是否會因為使用的次數而造成它本身的影響，例如：使用比較久之後的機台可能轉速下降，使得我們記錄到的次數低於實際旋轉的次數，導致資料紀錄上的錯誤，若是有提供給我們實驗順序的資訊，則可以將這些變數考慮進去。

Problem 3

(a)

$$\hat{F}(t_i) \;=\; \frac{\sum_{j=1}^{i} d_j}{25}$$

thousand cycles	d_i	r_i	$\hat{F}$
5	1	0	0.04
21	1	0	0.08
28	1	0	0.12
30	0	22	0.12



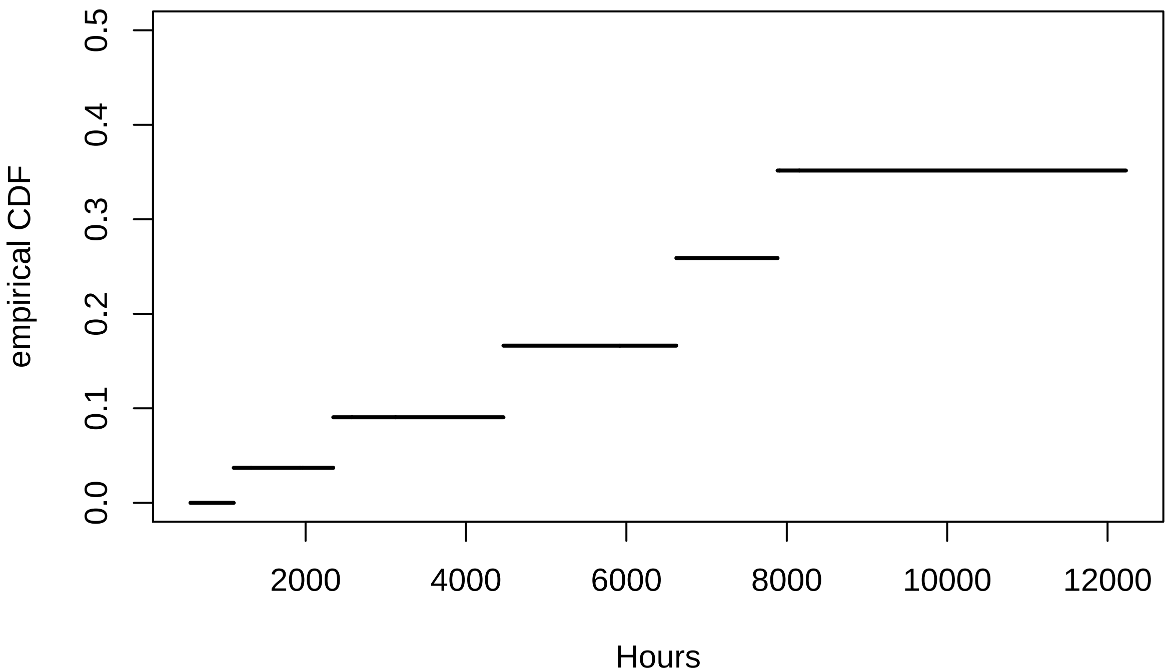
Problem 4

(a)

$$
\begin{aligned}
n_i &= n - \sum_{j=0}^{i-1} d_j - \sum_{j=0}^{i-1} r_j \quad , \quad i = 1, 2, ..., 15 \\
\hat{p}_i &= \frac{d_i}{n_i} \quad , \quad i = 1, 2, ..., 15 \\
\hat{S}(t_i) &= \prod_{j=1}^{i} \left(1 - \hat{p}_j\right) \quad , \quad i = 1, 2, ..., 15 \\
\hat{F}(t_i) &= 1 - \hat{S}(t_i) \quad , \quad i = 1, 2, ..., 15
\end{aligned}
$$

Hours	d_i	r_i	n_i	$\hat{p}_i$	$\hat{S}(t_i)$	$\hat{F}(t_i)$
564	0	3	30	0.0000	1.0000	0.0000
1104	1	0	27	0.0370	0.9630	0.0370
1321	0	2	26	0.0000	0.9630	0.0370
1933	0	3	24	0.0000	0.9630	0.0370
1965	0	3	21	0.0000	0.9630	0.0370
2345	1	0	18	0.0556	0.9095	0.0905
2578	0	2	17	0.0000	0.9095	0.0905
3122	0	3	15	0.0000	0.9095	0.0905
4467	1	0	12	0.0833	0.8337	0.1663
5918	0	2	11	0.0000	0.8337	0.1663
6623	1	0	9	0.1111	0.7410	0.2590
7885	1	0	8	0.1250	0.6484	0.3516
7912	0	2	7	0.0000	0.6484	0.3516
8156	0	3	5	0.0000	0.6484	0.3516
12229	0	2	2	0.0000	0.6484	0.3516

(b)



(c)

Use Greenwood's formula to compute the standard error of $\hat{F}(t_i)$

$$s.e.(\hat{F}(t_i)) \;=\; \sqrt{\hat{Var}\left(\hat{F}(t_i)\right)} \;=\; \left[\hat{S}(t_i)\right]^2 \sum_{j=1}^{i} \frac{\hat{p}_j}{n_j(1-\hat{p}_j)}$$

by Delta method and logit transformation, the 95% confidence interval limits are

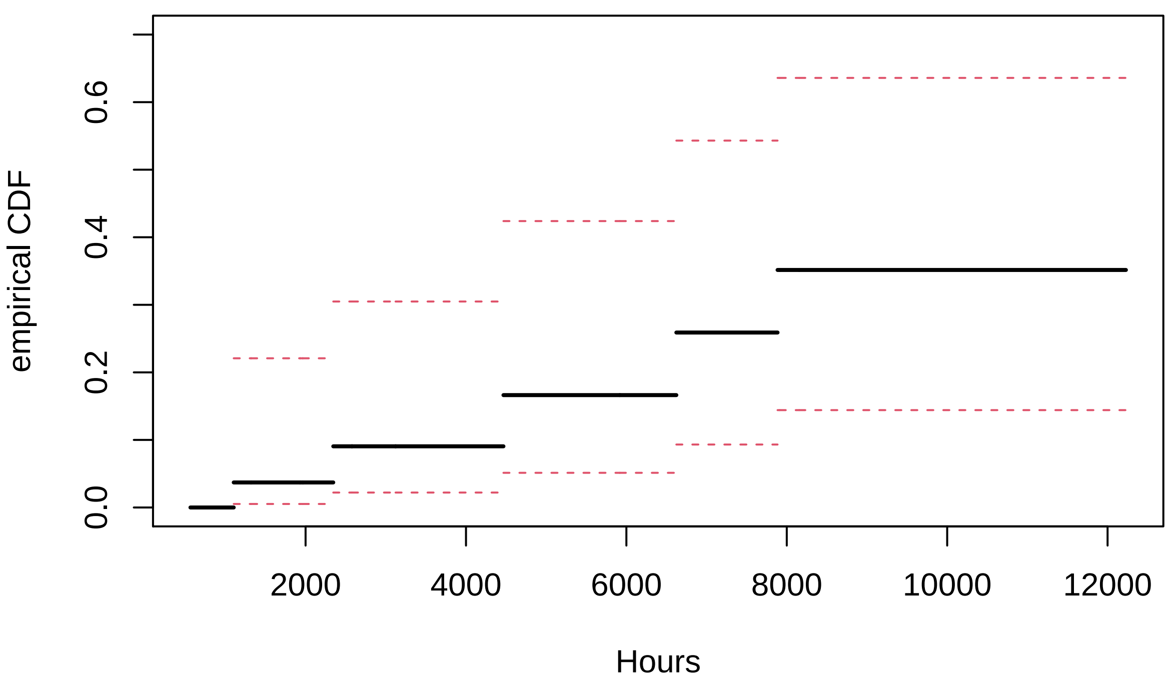
$$\hat{F}_L(t_i) \;=\; \frac{\hat{F}(t_i)}{\hat{F}(t_i) + \left(1-\hat{F}(t_i)\right) w_i}$$

$$\hat{F}_U(t_i) \;=\; \frac{\hat{F}(t_i)}{\hat{F}(t_i) + \left(1-\hat{F}(t_i)\right) / w_i}$$

where

$$w_i \;=\; exp\left[Z_{0.975}\ \frac{s.e.(\hat{F}(t_i))}{\hat{F}(t_i)\left(1-\hat{F}(t_i)\right)}\right]$$

Hours	$\hat{F}(t_i)$	$\hat{F}_L(t_i)$	$\hat{F}_U(t_i)$
564	0.0000	NaN	NaN
1104	0.0370	0.0052	0.2208
1321	0.0370	0.0052	0.2208
1933	0.0370	0.0052	0.2208
1965	0.0370	0.0052	0.2208
2345	0.0905	0.0221	0.3049
2578	0.0905	0.0221	0.3049
3122	0.0905	0.0221	0.3049
4467	0.1663	0.0513	0.4239
5918	0.1663	0.0513	0.4239
6623	0.2590	0.0932	0.5431
7885	0.3516	0.1441	0.6359
7912	0.3516	0.1441	0.6359
8156	0.3516	0.1441	0.6359
12229	0.3516	0.1441	0.6359



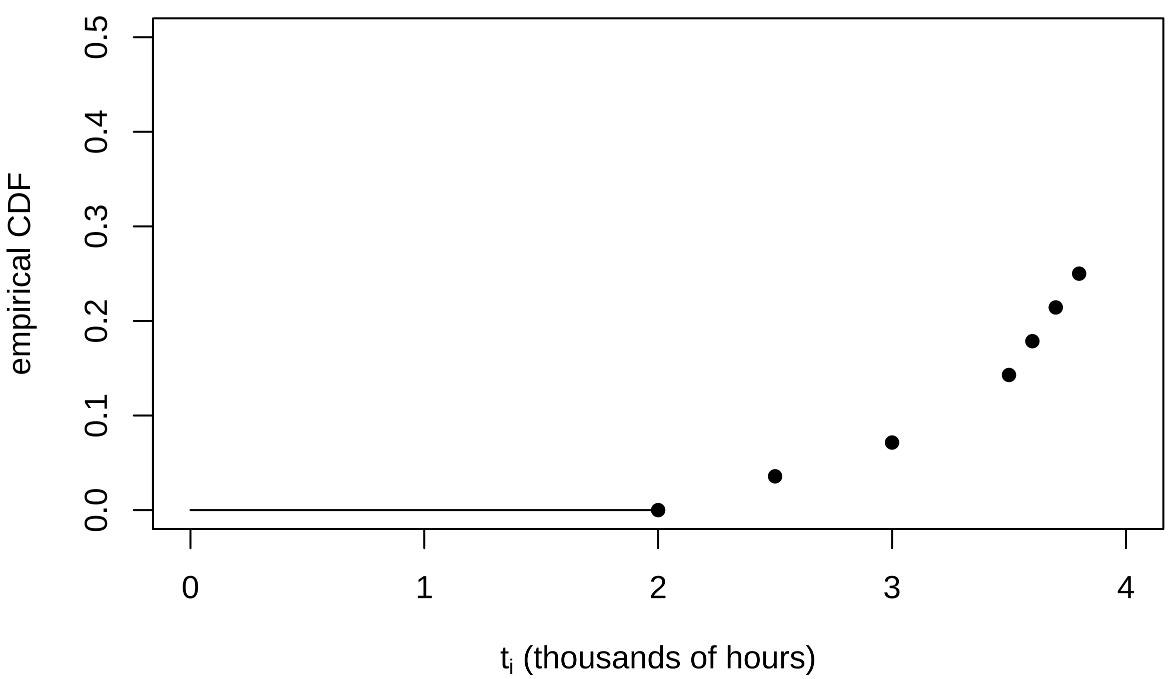
Problem 5

(b)

Denote d_i as the interval censored at (t_i, t_{i+1}), and compute

$$\hat{F}(t_i) = \frac{1}{28}\sum_{j=1}^{i} d_j$$

t_i	t_{i+1}	d_i	r_i	$\hat{F}(t_i)$
2.0	2.5	1	0	0.0000
2.5	3.0	1	0	0.0357
3.0	3.5	2	0	0.0714
3.5	3.6	1	0	0.1429
3.6	3.7	1	0	0.1786
3.7	3.8	1	0	0.2143
3.8	NA	0	21	0.2500



(c)

$$s.e.\left(\hat{F}(t_i)\right) \;=\; \sqrt{\frac{\hat{F}(t_i)\left(1-\hat{F}(t_i)\right)}{28}}$$

t_i	t_{i+1}	$\hat{F}(t_i)$	$s.e.(\hat{F}(t_i))$
2.0	2.5	0.0000	0.0000
2.5	3.0	0.0357	0.0351
3.0	3.5	0.0714	0.0487
3.5	3.6	0.1429	0.0661
3.6	3.7	0.1786	0.0724
3.7	3.8	0.2143	0.0775
3.8	NA	0.2500	0.0818

(d)

By Delta method and logit transformation, the 95% confidence interval limits are

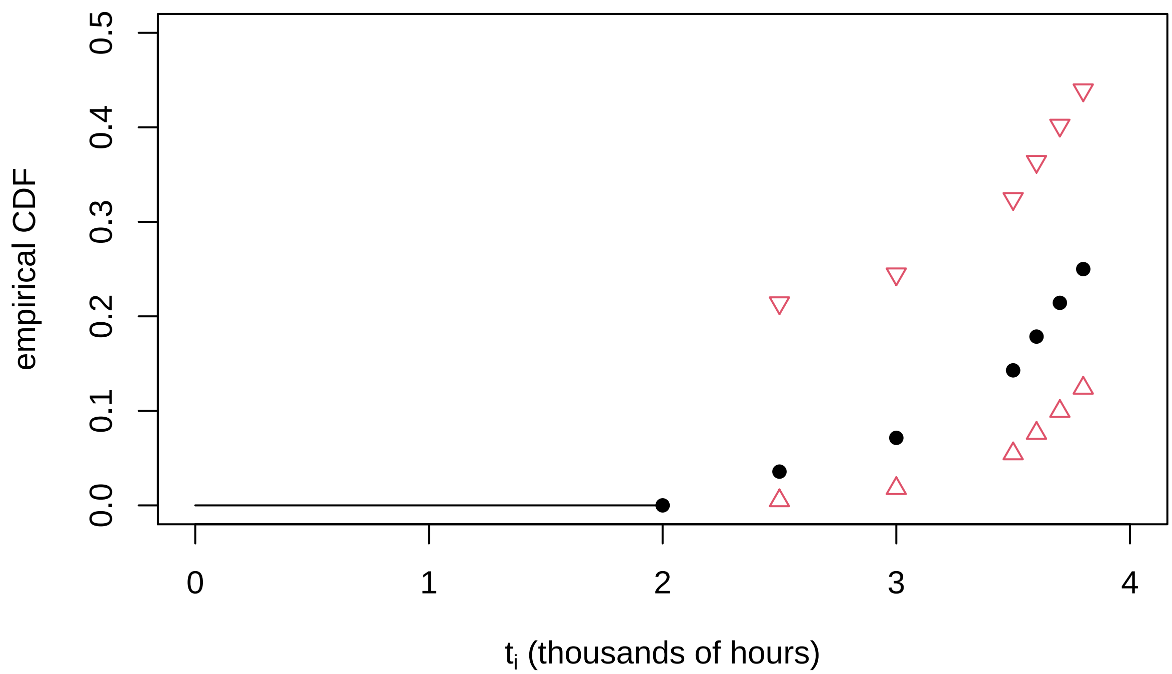
$$\hat{F}_L(t_i) \;=\; \frac{\hat{F}(t_i)}{\hat{F}(t_i)+\left(1-\hat{F}(t_i)\right)w_i}$$

$$\hat{F}_U(t_i) \;=\; \frac{\hat{F}(t_i)}{\hat{F}(t_i)+\left(1-\hat{F}(t_i)\right)/w_i}$$

where

$$w_i \;=\; exp\left[Z_{0.975}\,\frac{s.e.(\hat{F}(t_i))}{\hat{F}(t_i)\left(1-\hat{F}(t_i)\right)}\right]$$

t_i	t_{i+1}	$\hat{F}(t_i)$	$\hat{F}_L(t_i)$	$\hat{F}_U(t_i)$
2.0	2.5	0.0000	NaN	NaN
2.5	3.0	0.0357	0.0050	0.2142
3.0	3.5	0.0714	0.0179	0.2448
3.5	3.6	0.1429	0.0547	0.3245
3.6	3.7	0.1786	0.0763	0.3638
3.7	3.8	0.2143	0.0996	0.4021
3.8	NA	0.2500	0.1241	0.4395



(e)

Let $a = 0.01$, $b = 0.9 \Rightarrow e_{(a,b,1-\alpha/2)} = 3.21$. With no censoring the range of $t \in [t_L(a), t_U(b)]$ is given by the values of t for which $a \leq \hat{F}(t) \leq b$.

And by Delta method and logit transformation, the simultaneous approximate 95% confidence interval limits are

$$\hat{F}_L(t_i) = \frac{\hat{F}(t_i)}{\hat{F}(t_i) + \left(1 - \hat{F}(t_i)\right) w_i}$$

$$\hat{F}_U(t_i) = \frac{\hat{F}(t_i)}{\hat{F}(t_i) + \left(1 - \hat{F}(t_i)\right) / w_i}$$

where

$$w_i = exp\left[e_{(a,b,1-\alpha/2)} \frac{s.e.(\hat{F}(t_i))}{\hat{F}(t_i)\left(1 - \hat{F}(t_i)\right)}\right]$$

t_i	t_{i+1}	$\hat{F}(t_i)$	$\hat{F}_L(t_i)$	$\hat{F}_U(t_i)$
2.0	2.5	0.0000	NaN	NaN
2.5	3.0	0.0357	0.0014	0.4933

t_i	t_{i+1}	$\hat{F}(t_i)$	$\hat{F}_L(t_i)$	$\hat{F}_U(t_i)$
3.0	3.5	0.0714	0.0072	0.4478
3.5	3.6	0.1429	0.0286	0.4855
3.6	3.7	0.1786	0.0427	0.5145
3.7	3.8	0.2143	0.0585	0.5447
3.8	NA	0.2500	0.0759	0.5750

(f)

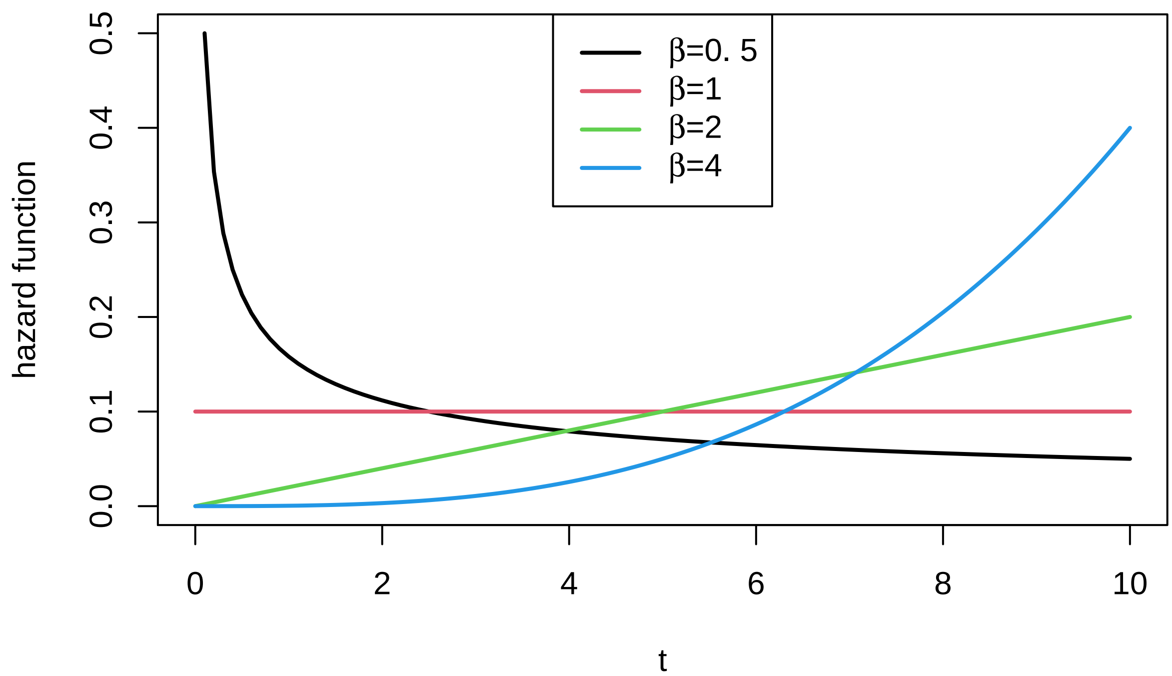
(1) The pointwise confidence intervals are useful for making a statement about $F(t_i)$ at one particular specified value of t_i (even thought it is common practice to plot a set of such intervals). However, simultaneous confidence band is useful over a range of value t.

(2) The overall coverage probability for the collection of pointwise intervals $(1-\alpha^*)$ is generally less than that for any individual interval $(1-\alpha)$ or the same level simultaneous confidence band $(1-\alpha)$.

(3) For any t_i, simultaneous confidence band is wider than the same level pointwise confidence interval.

Problem 7

(a)



(b)

計算在不同的 shape parameter β 和不同的時間 t 之下，hazard function 的數值變化

β	t=1	t=10
0.5	0.16	0.05
1.0	0.10	0.10
2.0	0.02	0.20
4.0	0.00	0.40

- 在 $t = 1$ 的情況下，hazard function 由大到小依次為：$\beta = 0.5$，$\beta = 1$，$\beta = 2$，$\beta = 4$
- 在 $t = 10$ 的情況下，hazard function 由大到小依次為：$\beta = 4$，$\beta = 2$，$\beta = 1$，$\beta = 0.5$，和上面的順序完全顛倒

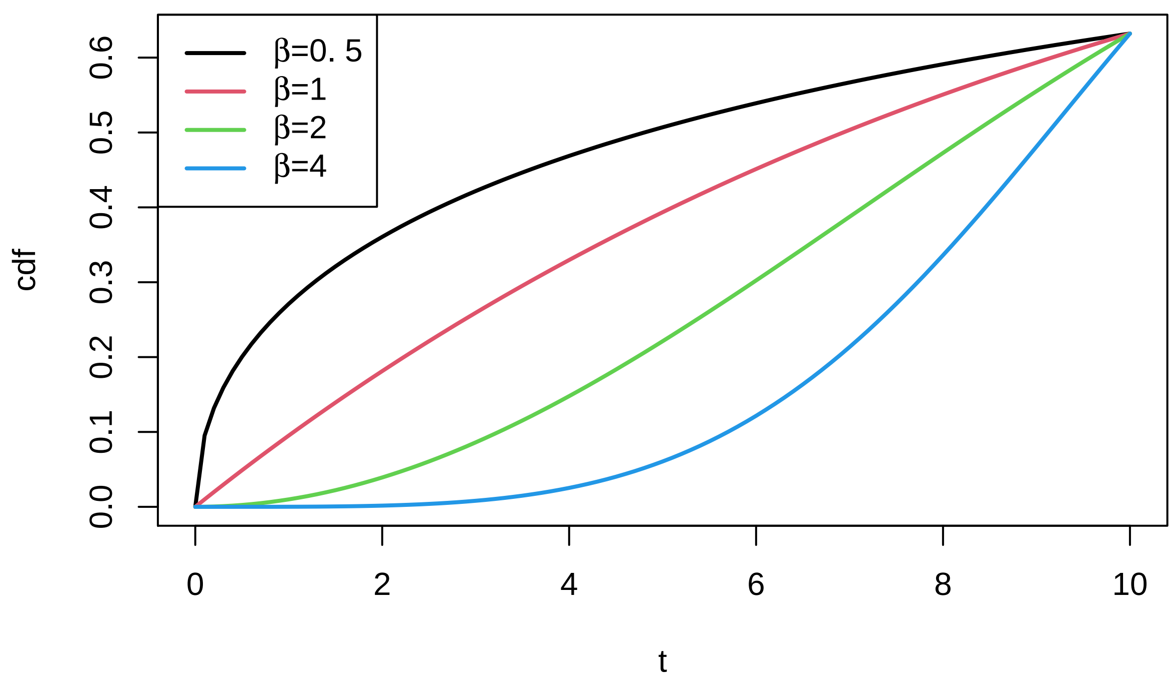
hazard function 所計算出的數值代表，在時刻 t 產品瞬間的死亡傾向，由 **(a)** 中的圖可知，Weibull distribution 的 hazard function 會隨著 β 的不同而呈現不一樣的趨勢變化：

(1) 若 $0 < \beta < 1$，hazard function 隨著時間遞減

(2) 若 $\beta = 1$，hazard function 不會隨著時間改變

(3) 若 $\beta > 1$，hazard function 隨著時間遞增

(c)



計算在不同的 shape parameter β 和不同的時間 t 之下，cdf 的數值變化

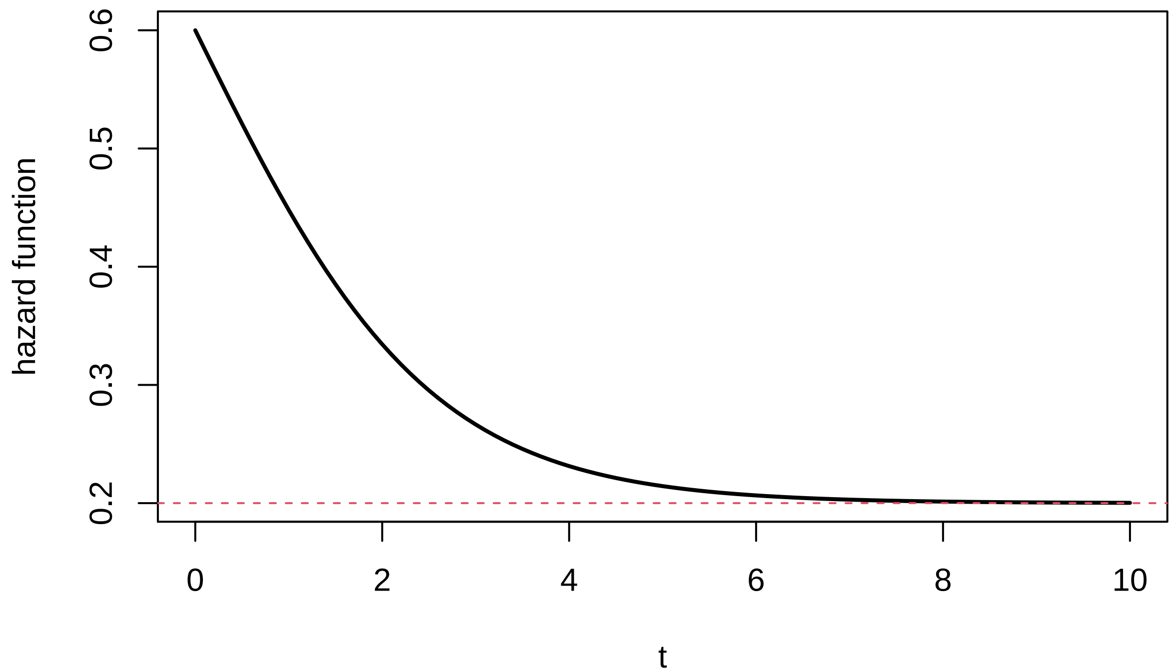
β	t=1	t=10
0.5	0.271	0.632
1.0	0.095	0.632
2.0	0.010	0.632
4.0	0.000	0.632

- $t = 1$ 時，$\beta = 0.5$ 所對應的 cdf 數值最大

- $t = 10$ 時，四種 β 所對應的 cdf 數值皆相同

因為在 $0 < t \leq 1$ 時，$\beta = 0.5$ 此種產品的 hazard function 也即死亡傾向大於其他三種 β 值的情況，所以很自然該種產品的 $F(1) = P(T \leq 1)$ 也會是最大的
然而藉由 Weibull distribution 的 cdf $F(t\,;\,\eta\,,\,\beta) = 1 - exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]$ 可知當 $t = \eta = 10$ 時，不論 β 數值為何，計算出的 cdf 數值皆相同

(d)



(e)

圖形呈現遞減，最後會趨近 0.2

此群產品可視為是一組 life time $\sim$ $Exp(\theta = 1)$ 和一組 life time $\sim$ $Exp(\theta = 5)$，這兩組產品的混合，而 $\theta = 1$ 的產品有較高的死亡傾向 $(h(t) = \frac{1}{\theta} = 1)$，會在實驗的初期拉高整體的死亡傾向平均，但隨著實驗進行，這種類型的產品會逐漸減少，最後 $\theta = 5$ 這類死亡傾向較低 $(h(t) = \frac{1}{5} = 0.2)$ 的產品會佔全部產品中較高的比例，造成整體的 hazard function 的數值逐漸趨近 0.2

(f)

exponential-mixture 的 hazard function 隨著時間遞減，最後趨近一個穩定數值 0.2，從 (e) 可知此產品會在實驗初期大量淘汰那些死亡傾向較高也即品質較差的產品，最後留下的大部分都是死亡傾向較低也即品質較高的產品，因此我們可以說此母體會隨著時間逐漸「改善」