NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4275 Lifetime analysis**

Academic contact during examination: Jarle Tufto

Phone: 99 70 55 19

Examination date: June 1, 2018

Examination time (from–to): 09:00–13:00

Permitted examination support material: *Tabeller og formler i statistikk*, Tapir Forlag, K. Rottmann: *Matematisk formelsamling*, Calculator Casio fx-82ES PLUS, CITIZEN SR-270X, CITIZEN SR-270X College or HP30S, one yellow A4-sheet with your own handwritten notes.

Other information:
Note that you should explain your reasoning behind your answers. You may write in English and/or Norwegian. You may write with a pencil.

Language: English

Number of pages: 6

Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☐	

Date Signature

Problem 1 Consider a lifetime variable T for which the cumulative hazard function is given by $Z(t) = \ln(t+1)$, $t > 0$.

a) Find the survival function $R(t) = P(T > t)$, the probability density function $f(t)$ of T, the hazard function $z(t)$ and the median survival time.

b) If possible, find ET and $E\ln(T+1)$.

i	y_i	δ_i
1	1	1
2	2	1
3	5	1
4	10	0
5	12	1
6	20	0
7	30	1
8	31	0

Table 1: Lifetime data in problem 2. Observations for which the censoring indicator $\delta_i = 0$ are right censored.

Problem 2 Consider the lifetime data listed in Table 1.

a) Compute the Kaplan-Meier estimate of the survival function $R(t)$. Also, if possible, compute an estimate of the median and mean survival time.

The cumulative distribution function of the standard logistic distribution is given by

$$F(t) = \frac{1}{1+e^{-t}}. \tag{1}$$

b) Derive the density function and the survival function of the standard logistic distribution.

Suppose we want to fit a parametric survival model belonging to the log location-scale family to the data in Table 1, constructing the model from the standard logistic distribution.

c) Write down the likelihood function for this parametric model for the right censored data listed in Table 1.

Using the `survreg` function in R, we maximise the likelihood and obtain the following results.

```
> model <- survreg(Surv(y, delta) ~ 1, dist = "loglogistic")
> summary(model)

Call:
survreg(formula = Surv(y, delta) ~ 1, dist = "loglogistic")
             Value Std. Error      z        p
(Intercept) 2.5967      0.665 3.9037 9.47e-05
Log(scale)  0.0143      0.369 0.0387 9.69e-01

Scale= 1.01

Log logistic distribution
Loglik(model)= -20.3   Loglik(intercept only)= -20.3
Number of Newton-Raphson Iterations: 4
n= 8
> vcov(model)
            (Intercept) Log(scale)
(Intercept)  0.44247667 0.03916445
Log(scale)   0.03916445 0.13610794
```

d) Compute an estimate of the upper 5%-quantile of the survival distribution based on the above parametric model. Also compute an approximate estimate of its standard error and a 95% confidence interval for the same quantile. Explain the reasoning behind any additional approximations you make in deriving the confidence interval.

e) Compute a total time on test (TTT) plot of the data and carry out the Barlow-Prochan test of the null hypothesis that the hazard function $z(t)$ is constant.

Problem 3 We are studying the lifetimes T (in days) of 1000 units of an electric component operating at different temperatures in the interval between 10 and 90°Celcius. The data is shown in Fig. 1. Note that the temperatures are centered around their mean of 50 °C in the plot and in further analysis.

We fit a Cox proportional hazard model in R as follows.

```
> model <- coxph(Surv(y, delta) ~ x)
> model
Call:
coxph(formula = Surv(y, delta) ~ x)

     coef exp(coef) se(coef)    z      p
x 0.10321   1.10872  0.00322 32.1 <2e-16

Likelihood ratio test=1460  on 1 df, p=0
n= 1000, number of events= 679
```

a) Write down the model in mathematical notation and state its assumptions. Briefly explain how the regression coefficients of the model are estimated.

b) An estimate of the baseline survival function $R_0(t)$ is shown in Fig. 2. This is survival function for units operating at the a mean-centered temperature of zero, that is, at 50°C. Based on the plot, compute an estimate of the probability that a unit operating at 70°C is functioning after 100 days of operation. Verify if your estimate roughly agrees with the observed data in Fig. 1.

c) Fig. 3 shows the Schoenfeld residuals for the temperature covariate plotted against the log of the observed failure times. How do these residuals behave if the assumptions of the Cox proportional model hold? Do the plot support the assumptions of the fitted model?

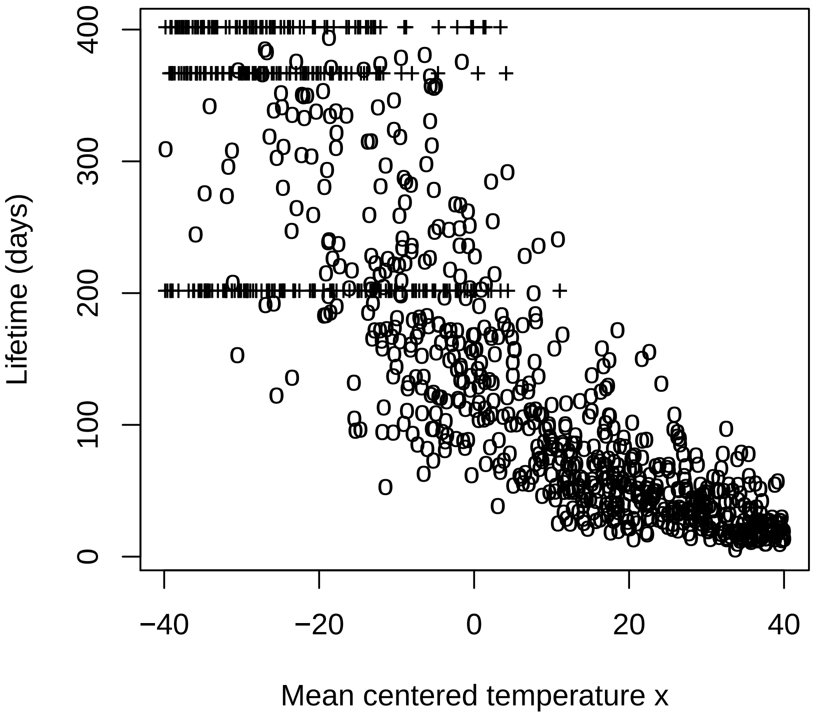

Figure 1: Data in problem 3. Circles represent failures and crosses right censoring events.

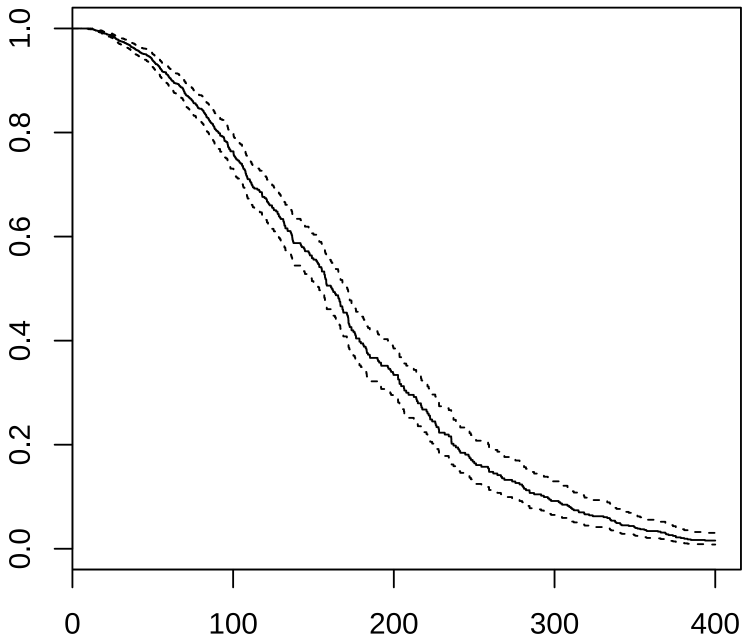

Figure 2: The estimated baseline survival function $\hat{R}_0(t)$.

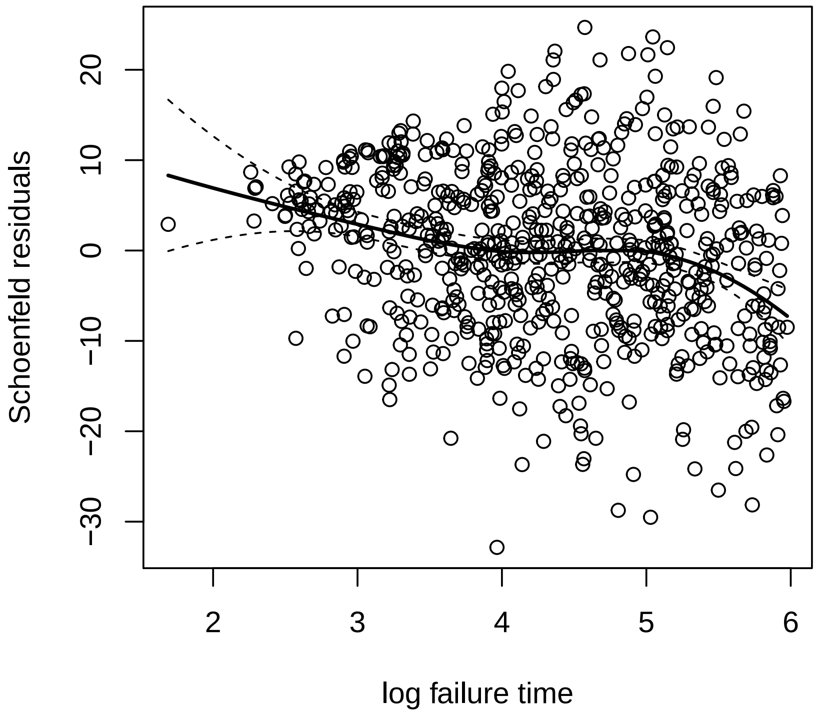

Figure 3: Schoenfeld residuals plotted against the log of observed failure times. Also shown is the mean of the residuals estimated using local polynomial regression (solid curve) and associated 95% confidence bands (dashed curves).