TMA4275 Lifetime analysis
Spring 2020
Obligatory Exercise 2

Out: Friday March 13
In: Friday March 27 at (latest) 15.00.

Important information: *The instructions in this exercise tell you to use MINITAB. You are however free to use R (or another program that you prefer). An introduction to survival analysis in R can be downloaded from the course webpage (see "Statistical software"). Text-files with R-commands for survival analysis may also be downloaded from "Lecture plan and progress" on the course website. Please note that if you use another program than MINITAB, then the source code should be included in the report. You may write parts of the report by hand and submit a scanned version of this. The reports should be uploaded via Blackboard. Write your student number (and not your name) on the first page of the report. If two students have been working together, they must still upload individual reports. Write on the first page of the report the student number for your collaborator as well as your own.*

MINITAB-hint: In the MINITAB-window there is a column called `'Navigator'` to the left. Here are listed the various analyses that you perform. By right-clicking on an item in the list you can send the output directly to Word.

Exercise

On the web page "Data sets" on the course homepage you may download the Mintab Worksheet "Tiredata (.mwx)". The data are taken from the article "Regression approach to tire reliability analysis" by V.V. Krivtsov, D.E. Tanako and T.P. Davis, which can be downloaded from

`https://folk.ntnu.no/bo/TMA4275/2020v/CoxReliabilityPaper.pdf`

A text-file containing the data appropriate for use by R can also be downloaded under "Data sets". Note that the variable names are shortened in this file as compared to the Minitab-file.

The data (see Table 1 of the article) give survival times (**Survival**) for 34 tires, with censoring status given in the last column (**Censoring**). The covariates, which are explained in the article, are given in the first 7 columns of the table.

Note that Table 1 contains a variable **Wedge gauge** $\times$ **peel force** which one would think was given by the product of the variables **Wedge gauge** and **Peel force**. As you may see, this is not quite so, and the reason is not clear. However, to make the results obtained in this exercise comparable to those of the article, we will use the "product" covariate **Wedge gauge** $\times$ **peel force** with values as given in Table 1 of the article. (In the MINITAB worksheet this variable has been named **W** $\times$ **P**).

a) First one wants to find out whether **Tire age** alone is an important variable for tire reliability.

For this, introduce in the MINITAB worksheet a new variable **NewOld** which is 1 if **Tire age** $>$ 1.0 and is 0 otherwise. [You can use `Calc` $>$ `Calculator` and use for '`Expression`': (c1 $>$ 1)]

Use MINITAB to plot – in the same graph – the Kaplan-Meier estimates of the survival functions for each of the two values of **NewOld**. Disregard the other covariates in the data set. Do not include confidence curves in the plots.

What is your conclusion based on the plot? What are the estimated median lifetimes and expected lifetimes for the two groups? Comment on the differences also in light of MINITAB's reported result of a log-rank test.

b) The data are now to be analyzed by MINITAB using `Reliability/Survival` $>$ `Regression with Life Data`.

Perform an analysis with Weibull-regression. Define the model by writing

`c1 c2 c3 c4 c5 c6 c7`

in the box for `'Model'`.

Write down, in some form, the estimated model.

Which are the covariates with significant effects? (Look at p-values, use 5% significance level). Compare with Table 2 in the article (which, though, is based on a Cox regression analysis).

What is the estimated value of the shape parameter? Give a comment.

Write down the calculated value of the log likelihood.

c) Delete the non-significant covariates (5% significance level) and repeat the analysis in (b) using the reduced model.

Comment on the results and compare with Table 3 in the article.

Write down the calculated value of the log likelihood for this reduced model.

d) The article defines a *poor tire* to be one with **Wedge gauge** and **Interbelt gauge** both equal to 0.5 and **Peel force** equal to 1.0, while a *good tire* is one where **Wedge gauge** and **Interbelt gauge** are 1.2 and 1.0, respectively, while **Peel force** is 2.0. In both cases you shall calculate **W** $\times$ **P** as the corresponding product of **Wedge gauge** and **Peel force**.

The following questions can be answered by using the option `'Estimate'` appearing to the right in the `'Regression with Life Data'` box (see further instructions below):

What are, in the model of (c), the estimated lower quartile, median and upper quartile, respectively, for a poor tire and for a good tire?

What are the estimated probabilities of survival beyond times 1.0, 1.2 and 1.5, respectively, for a poor tire and for a good tire?

Instructions: In the `'Estimate'` box you need to fill out in `'Enter new predictor values'` the actual values of all the four predictors in the model (do not forget $\mathbf{W} \times \mathbf{P}$). Then you specify all the percentages for which you want the percentiles, and the times for which you want the survival probabilities.

e) The article compares models by means of log-likelihood values. You shall do a similar comparison for Weibull regression analyses.

[Recall from the lectures that you compare two models (where one is an extension of the other) by calculating 2 times the difference in log-likelihood value, and comparing this number to the chi-square distribution with degrees of freedom equal to the difference in number of covariates of the two models. If the calculated value is too high, then the largest model should be used. Otherwise, you may conclude that the smallest one is sufficient].

Compare the following models:

1. The (full) Weibull regression model in (b).
2. The (reduced) Weibull regression model in (c).
3. The Weibull model using no covariates. [Since MINITAB does not allow the `'Model'` box to be empty in survival regression, you will have to do a Weibull lifetime analysis without covariates using `Reliability/Survival > Distribution Analysis (Right Censoring)` to obtain the log-likelihood value].

Write down the relevant hypotheses and obtain a conclusion for each of the tests performed. Use significance level 5% for each test.

f) Write down a Cox-model corresponding to the reduced model in (c). What are the main differences between the Cox model and the Weibull model?

Compare the expression for the hazard rate in the Cox-model with the expression for the hazard rate you get by using Weibull regression.

Find estimates for the coefficients β_i in the Cox-model by using the results from the Weibull regression in (c). Compare them with the results in Table 3 of the article and comment.

Here you may check the results of the Cox analyses in the paper by doing the analysis in R. You may then use the following commands for the full model:

```
library(survival)
tiredata=read.table("https://folk.ntnu.no/bo/TMA4275/Data/
Tiredata.txt",header=T)
fit.tire=coxph(Surv(Survival,Censor==1)~Age+Wedge+Inter+EB2B
+Peel+Carbon+WxP, data=tiredata)
summary(fit.tire)
```

(You will not get exactly the same results as the ones in the paper. The differences might be due to different treatments of ties, or other sources like rounding errors or other small errors in the data.)

g) Consider in this point residual plots for the reduced model in (c).

Explain briefly how the plots are derived.

What do you conclude regarding model fit?

Fit lognormal, loglogistic and normal regression models to the reduced model and look at the corresponding residual plots. Do any of these models seem better suited than the Weibull model for these data? Give a comment.

h) Return to the analysis in (c) where you used Weibull regression. Using again the option `'Estimate'`, a column with estimated median survival times for each of the 34 tires is obtained by clicking on `'Use predictor values in data (storage only)'` and writing `'50'` in the box `'Estimate percentiles for percents'` and marking for `'Percentiles'` under `'Store percentiles'` below. Introduce a variable **PoorGood** into the MINITAB worksheet with the value 1 for the 17 tires with the lowest estimated median survival time and 2 for the rest. [This can be done in several ways, the most primitive one being to do it "by hand"].

Then do the following MINITAB-analysis:

`Reliability/Survival > Accelerated Life Testing`

with the variable **Survival** as response; **PoorGood** as accelerating variable; `'Relationship'` set to `'Linear'`; and `'Assumed distribution'` set to `'Weibull'`.

Then look at the resulting graph `'Probability Plot ...'`

How can you interpret this graph as a model checking tool for the model in (c)?

[See also page 12 of Slides 13. Why are the lines in the plot parallel?]