TMA4275 Lifetime analysis
Spring 2006
Obligatory exercise 2
Out: Tuesday April 4
In: Friday April 28.

On the course homepage (under "Data files") you can download the dataset 'TMA4275VeteranLungCancer.MTW' which can be read into MINITAB via "Open Worksheet".

These are survival data on 137 advanced lung cancer patients as collected by the Veterans Administration Lung Cancer Study Group (from R. L. Prentice: *Exponential survivals with censoring and explanatory variables,* Biometrika, 1973). Patients were randomized according to one of two chemotherapeutic agents (*Treatment*: 1= standard, 2=test). Of particular interest was the possible differential effects of therapy on tumor cell type. Tumors are classified into one of four broad groups (*Celltype*: 1=squamous, 2=smallcell, 3=adeno, 4=large). Further covariates recorded when the patients were taken on study were *Performance status* (a measure of general medical status where 10-30 is completely hospitalized; 40-60 is partial confinement to hospital; 70-90 is able to care for oneself), *Months from diagnosis* to starting on study, *Age* in years, and *Prior therapy* (0=no, 10=yes).

The data may also be downloaded from

`http://lib.stat.cmu.edu/datasets/veteran`

From this file we quote the list

```
#  Variables
#       Treatment  1=standard,  2=test
#       Celltype   1=squamous,  2=smallcell,  3=adeno,  4=large
#       Survival in days
#       Status     1=dead, 0=censored
#       Performance status
#       Months from diagnosis
#       Age in years
#       Prior therapy  0=no, 10=yes
```

The first 15 cases look as follows:

1	1	72	1	60	7	69	0
1	1	411	1	70	5	64	10
1	1	228	1	60	3	38	0
1	1	126	1	60	9	63	10
1	1	118	1	70	11	65	10
1	1	10	1	20	5	49	0
1	1	82	1	40	10	69	10
1	1	110	1	80	29	68	0
1	1	314	1	50	18	43	0

1	1	100	0	70	6	70	0
1	1	42	1	60	4	81	0
1	1	8	1	40	58	63	10
1	1	144	1	30	4	63	0
1	1	25	0	80	9	52	10
1	1	11	1	70	11	48	10
.							
.							
.							
.							

a) First one wants to find out whether there are differences between the lifetime distributions of patients in the two treatment groups.

Using MINITAB you shall plot – in the same graph – the Kaplan-Meier estimates of the survival functions for each of the two treatments. 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 patients in each of the two groups? Comment on the differences.

Make similar plots for comparison of estimated survival curves for the four cell types. Describe briefly your conclusion.

b) The data are now to be analyzed using "Regression with Life Data" under "Reliability/Survival" in MINITAB.

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

```
Cell Treat PS Month Age Prior
```

in the box for "Model" and

```
Cell
```

in the box "Factors (optional)". This gives a model where "Treat, PS, Month, Age, Prior" are ordinary covariates, while "Cell" is considered as a factor with four levels. Remark that MINITAB from this implicitly defines three covariates x_1, x_2, x_3 to represent "Cell". More precisely are defined

$$\begin{aligned} x_1 &= 1 \text{ if the patient has Cell=2, and } = 0 \text{ otherwise} \\ x_2 &= 1 \text{ if the patient has Cell=3, and } = 0 \text{ otherwise} \\ x_3 &= 1 \text{ if the patient has Cell=4, and } = 0 \text{ otherwise} \end{aligned}$$

Write down a table for the connection between Celltype and the value of x_1, x_2, x_3 corresponding to the one given in Table 3.1 of Ansell & Phillips, page 60. Why don't we use Cell as an ordinary numerical covariate?

Write down the complete model that MINITAB uses for the data, when you let $x_4 - x_8$ be the covariates "Treat, PS, Month, Age, Prior".

Interpret the results from the MINITAB-session. What is the estimated median lifetime of a patient with the same covariates as patient number 1? Number 19?

c) Find the covariates which do not have significant effects (look at p-values). Is there a significant difference between the two treatments? Is the shape parameter of the Weibull model significantly different from 1? What can you conclude from this? (The possible change of distribution from Weibull to Exponential should not be made in this point, but may well be discussed and tried in point (e) below!)

Take out the non-significant covariates, except "Treat" which should be kept in the model, and repeat the analysis with MINITAB with the reduced model. (If one or more of $x_1 - x_3$ has significant effect, then keep all three in the model. As mentioned above you should not change to exponential distribution in the new analysis here).

Which of the covariates has the largest effect on survival? Is there now a significant difference between the two treatments?

d) Write down a Cox-model for the reduced model used in (c).

Compare the expression for the hazard rate $z(t;x)$ in the Cox-model with the model that was used in the Weibull regression.

Find estimates for the coefficients β_i in the Cox-model by using results from the Weibull regression in (c). (It turns out that in this way you will get results which are not much different from what one would get by maximizing Cox' partial likelihood).

Give a practical interpretation of each estimated coefficient β_i in the Cox-model (for example in the form of "relative risks").

e) MINITAB also includes the possibility of investigating a possible interaction between covariates. As indicated in the beginning of the exercise, in the lung cancer study it was of particular interest to investigate the interaction between treatment and celltype.

MINITAB does this if you add

`Cell*Treat`

in the "model" box while keeping Cell as before in the "Factor" box.

This corresponds to defining new covariates z_1, z_2, z_3 by

$$z_1 = x_1 \cdot x_4$$
$$z_2 = x_2 \cdot x_4$$
$$z_3 = x_3 \cdot x_4$$

MINITAB now estimates coefficients of z_1, z_2, z_3 in addition to the other coefficients.

Do this analysis in MINITAB using the same covariates as in the reduced model in (c). The analysis will show that only z_1 gets a p-value smaller than 0.05. (What is the numeric value here?) What do you conclude regarding interaction between treatment and celltype?

Write down an expression for the estimated relative risk for a patient with Cell=2 in the standard treatment group, compared to a patient with Cell=2 in the test treatment group, the other variables being the same.

Then compute the corresponding relative risks when Cell is, respectively, 1,3 and 4. Comment.

f) Discuss the model fit in the final model in (e). You may here use various types of plots and tests.

May other basic models than the Weibull be useful for this regression analysis?