**Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag**

Contact during exam:
Bo Lindqvist 975 89 418

EXAM IN TMA4275 LIFETIME ANALYSIS

Monday 27 May 2013

Time: 09:00–13:00

Aids:
Approved calculator.
One yellow sheet (A4 with stamp) with your own formulae and notes.

Grading: 17 June 2013

ENGLISH

Tables of the χ^2-distribution with 1 degree of freedom, and of the standard normal distribution are given at the end of the exam.

Problem 1 *Treatments for prostatic cancer*

A randomized controlled clinical trial to compare treatments for prostatic cancer was begun in 1967 by the Veteran's Administration Cooperative Urological Research Group. Two of the treatments used in the study were a placebo and 1.0 mg of diethylstilbestrol (DES), both administered daily by mouth. The time origin of the study is the date on which a patient was randomised to a treatment, and the end-point is the death of the patient from prostatic cancer. The survival times of patients who died from other causes, or who were lost during the follow-up process were regarded as censored.

The data used in this exercise are from a certain subset of the patients, with cancer having reached a certain stage. The data are given at the end of the exercise.

The following variables are recorded for each patient:

Treat: *Treatment group.* 0 = placebo; 1 = DES.

Time: *Survival time*, possibly censored (months).

Status: *Censoring status.* 0 = censored; 1 = dead.

Age: *Age* (years).

Shb: *Serum haemoglobin level* (gm/100 ml).

Size: *Size of tumour* (cm^2).

Index: *Gleason index* (a combined index of tumour stage and grade; the more advanced the tumour, the greater the value of the index).

a) Use the data to compute the Kaplan-Meier estimator separately for the placebo patients (Treat=0) and the patients who received DES (Treat=1), without taking the other covariates into account. (Note that the DES group has only one observed death).

Draw the two curves in the same figure.

Which preliminary conclusion can be drawn from this figure?

What is the estimated median survival time and what is the estimated expected survival time for the placebo patients? Why can't the median survival time be computed for the DES patients? Can the estimated expected survival time for the DES patients be computed? If so, what is the estimate?

The four prognostic variables Age, Shb, Size and Index may have an effect on the survival times, and one therefore performs a Weibull-regression using these covariates in addition to the Treat variable. The resulting output from MINITAB is as follows:

```
Regression Table

                       Standard                      95,0% Normal CI
Predictor       Coef      Error      Z      P        Lower        Upper
Intercept    8,97979    3,32964   2,70  0,007      2,45383      15,5058
Treat       0,425215   0,459101   0,93  0,354    -0,474606      1,32504
Age       -0,0070739  0,0231046  -0,31  0,759   -0,0523580    0,0382103
Shb       -0,0505088   0,148834  -0,34  0,734    -0,342217     0,241200
Size      -0,0368758  0,0170023  -2,17  0,030   -0,0701996   -0,0035520
Index      -0,275063   0,119843  -2,30  0,022    -0,509952   -0,0401752
Shape        2,73540   0,955769                    1,37913      5,42547

Log-Likelihood = -31,341
```

b) Write down the complete model behind this output. Use the covariate names given in the output (see description in the beginning of the exercise). Also write down an expression for the median survival time of a patient as a function of the parameters of the model and the observed covariates.

Give an interpretation of the estimated regression coefficients with respect to the influence of the corresponding covariate on the survival time of a patient.

What is the estimated relative increase in median survival time if a patient recieves the DES treatment instead of the placebo treatment? (Compute first the quotient between estimated median survival time under DES and under placebo).

Which of the explanatory variables have significant effect? Does the output show a significant effect of using DES? (Use significance level 5% when investigating significant covariates).

It was decided to use only Treat, Size and Index as covariates in the next analyses. In order to investigate the role of treatment (Treat), one did one analysis with Treat in the model and one without Treat in the model. Here are the results by MINITAB:

```
                          Standard                     95,0% Normal CI
Predictor         Coef       Error      Z      P        Lower        Upper
Intercept      7,73138     1,45451   5,32  0,000      4,88059      10,5822
Treat         0,434133    0,463267   0,94  0,349    -0,473854      1,34212
Size        -0,0370421   0,0174203  -2,13  0,033   -0,0711853   -0,0028989
Index        -0,269215    0,116184  -2,32  0,020    -0,496931   -0,0414985
Shape          2,69158    0,939226                    1,35826      5,33377

Log-Likelihood = -31,434

                          Standard                     95,0% Normal CI
Predictor         Coef       Error      Z      P        Lower        Upper
Intercept      8,21515     1,54411   5,32  0,000      5,18876      11,2415
Size        -0,0446092   0,0179154  -2,49  0,013   -0,0797227   -0,0094958
Index        -0,289745    0,124149  -2,33  0,020    -0,533072   -0,0464182
Shape          2,71880    0,956222                    1,36459      5,41690

Log-Likelihood = -32,015
```

c) Let β_1 be the coefficient of the variable Treat in the model corresponding to the first of the two outputs above. To test whether there is an effect of DES, one considers the following hypotheses:

$$H_0: \ \beta_1 = 0 \text{ vs. } H_1: \ \beta_1 \neq 0.$$

There are (at least) two ways of formally testing this based on the above output; using "Coef" or using "Log-Likelihood".

You are asked to describe both ways, and find the respective p-values (for a p-value taken directly from the MINITAB output, you need to give an explanation of how it arises).

Row	Treat	Time	Status	Age	Shb	Size	Index
1	0	2	0	76	10,7	8	9
2	0	14	1	73	12,4	18	11
3	0	23	0	68	12,5	2	8
4	0	24	0	71	13,7	10	9
5	0	26	1	72	15,3	37	11
6	0	36	1	72	16,4	4	9
7	0	42	1	57	13,9	24	12
8	0	43	0	60	13,6	7	9
9	0	51	0	61	13,5	8	8
10	0	52	0	73	11,7	5	9
11	0	58	0	64	16,2	6	9
12	0	59	0	77	12,0	7	10
13	0	61	0	75	13,7	10	12
14	0	62	0	63	13,2	3	8
15	0	65	0	67	13,4	34	8
16	0	67	0	70	14,7	7	9
17	0	67	0	71	15,6	8	8
18	0	69	1	60	16,1	26	9
19	1	5	0	74	15,1	3	9
20	1	16	0	73	13,8	8	9
21	1	28	0	75	13,7	19	10
22	1	45	0	72	11,0	4	8
23	1	50	1	68	12,0	20	11
24	1	51	0	65	14,1	21	9
25	1	51	0	65	14,4	10	9
26	1	54	0	51	15,8	7	8
27	1	55	0	74	14,3	7	10
28	1	57	0	72	14,6	8	10
29	1	60	0	77	15,6	3	8
30	1	61	0	60	14,6	4	10
31	1	64	0	74	14,2	4	6
32	1	65	0	51	11,8	2	6
33	1	66	0	70	16,0	8	9
34	1	66	0	70	14,5	15	11
35	1	67	0	73	13,8	7	8
36	1	68	0	71	14,5	19	9
37	1	70	0	72	13,8	3	9
38	1	70	0	71	13,6	2	10

Problem 2 *Valve seat replacements*

In a fleet of diesel engines it is assumed that the replacement of valve seats for a single engine follows a homogeneous Poisson process (HPP) with (constant) intensity λ. The intensity is assumed to have the same value for all engines.

a) Consider first one such engine. Let $N(t)$ be the number of replacements in the time interval from 0 to $t > 0$.

What is the distribution of $N(t)$ for a given time t?

What is the expected number of replacements in the time interval from 0 to t?

What is the probability that there is exactly one valve seat replacement in the time interval from 0 to t?

What is the distribution of the time to the first replacement for an engine? Show how you can find this distribution from the knowledge of the distribution of $N(t)$ for $t > 0$.

In the following, assume that there are m engines in the fleet, where the jth engine is observed in the time interval $(0, \tau_j)$, with N_j replacements occurring in this interval ($j = 1, \ldots, m$).

b) Derive the likelihood function $L(\lambda)$ under the given model assumptions and the given observed counts of valve seat replacements. Show that this function can be expressed as

$$L(\lambda) = \lambda^{\sum_{j=1}^m N_j} e^{-\lambda \sum_{j=1}^m \tau_j},$$

where ln is the natural logarithm.

Find the maximum likelihood estimator $\hat{\lambda}$ for λ.

c) Derive an estimator for the standard deviation of $\hat{\lambda}$, and an approximate 95% confidence interval for λ.

Calculate $\hat{\lambda}$, the standard deviation estimate and the confidence interval when $m = 3$ and the data are given as follows:

j	τ_j	N_j
1	240	3
2	300	1
3	450	2

Problem 3 *The log-logistic distribution*

Let T be the lifetime of a certain equipment, modeled by the log-location-scale family, so that

$$\ln T = \mu + \sigma W,$$

where μ and σ are parameters ($\sigma > 0$), while W has the standard logistic distribution with cumulative distribution function (CDF)

$$\Phi(w) = \frac{e^w}{1+e^w} \quad \text{for} \quad -\infty < w < \infty.$$

T is now said to have the *log-logistic distribution* with parameters μ and σ.

a) Show that the CDF of T is given by

$$F(t) = \frac{e^{\frac{\ln t - \mu}{\sigma}}}{1 + e^{\frac{\ln t - \mu}{\sigma}}} \quad \text{for } t > 0. \tag{1}$$

Let t_p be the $100p$-percentile of the distribution of T, for $0 < p < 1$, i.e. the time such that $F(t_p) = p$. Show that

$$\ln t_p = \mu + \sigma \ln \frac{p}{1-p}.$$

What is the median of the distribution of T? Also, find an expression for the interquartile range of T, i.e., the difference between the 75 and 25 percentiles of T.

Assume below that there is given a right censored data set $(Y_i, \delta_i),\ i = 1, \ldots, n$, from a sample of the given equipment. Here Y_i is the observed time while δ_i is the censoring status; $\delta_i = 1$ if Y_i is a lifetime; $\delta_i = 0$ if Y_i is a censoring time.

b) Explain how the expression for $F(t)$ in (1) can be used to construct a probability plot for checking whether the above data set can be assumed to come from a log-logistic distribution.

Explain also briefly how estimates of μ and σ can be found from the plot.

c) Find the cumulative hazard function $Z(t)$ of T.

What is the distribution of the random variable $Z(T)$? (You need not prove this).

Explain how this result can be used to check whether the right censored data set can be assumed to come from a log-logistic distribution. Do you know the name of this method?

Table of cumulative probabilities of χ^2-distribution with 1 degree of freedom.

Chi-Square with 1 DF

x	P(X <= x)
0,1	0,248170
0,2	0,345279
0,3	0,416118
0,4	0,472911
0,5	0,520500
0,6	0,561422
0,7	0,597216
0,8	0,628907
0,9	0,657218
1,0	0,682689
1,1	0,705734
1,2	0,726678
1,3	0,745787
1,4	0,763276
1,5	0,779329
1,6	0,794097
1,7	0,807712
1,8	0,820288
1,9	0,831922
2,0	0,842701
2,1	0,852701
2,2	0,861989
2,3	0,870626
2,4	0,878665
2,5	0,886154
2,6	0,893136
2,7	0,899652
2,8	0,905736
2,9	0,911420
3,0	0,916735
3,1	0,921708
3,2	0,926362
3,3	0,930720
3,4	0,934804
3,5	0,938631
3,6	0,942220

Standard normal distribution

Standard normalfordeling

$$\Phi(z) = P(Z \leq z)$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.7	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Standard normal distribution

Standard normalfordeling

$$\Phi(z) = P(Z \leq z)$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

Critical values of standard normal distribution

Kritiske verdier i standard normalfordelingen

$$P(Z > z_\alpha) = \alpha$$

α	z_α
.2	0.842
.15	1.036
.1	1.282
.075	1.440
.05	1.645
.04	1.751
.03	1.881
.025	1.960
.02	2.054
.01	2.326
.005	2.576
.001	3.090
.0005	3.291
.0001	3.719
.00005	3.891
.00001	4.265
.000005	4.417
.000001	4.753