NTNU
Kunnskap for en bedre verden

Institutt for matematiske fag

Eksamensoppgave i **TMA4275 Levetidsanalyse**

Faglig kontakt under eksamen: Bo Lindqvist

Tlf: 975 89 418

Eksamensdato: Lørdag 31. mai 2014

Eksamenstid (fra–til): 09:00-13:00

Hjelpemiddelkode/Tillatte hjelpemidler: C: Godkjent kalkulator. Ett gult ark (A4 med stempel) med dine egne formler og notater.

Annen informasjon:
Tabeller over χ^2-fordelingen med 1 frihetsgrad og standardnormalfordelingen er gitt i vedlegget til slutt i oppgavesettet.

Målform/språk: bokmål

Antall sider: 6

Antall sider vedlegg: 4

Kontrollert av:

Dato Sign

Oppgave 1 *Utskriving fra sykehus*

På en intensivavdeling ved et sykehus er man interessert i om forekomst eller ikke forekomst av lungebetennelse ved innleggelse vil ha innvirkning på lengden av sykehusoppholdet, T, dvs. tid fra innleggelse til utskriving fra sykehusenheten.

Nedenfor er tider (i dager) til utskriving fra sykehuset for 8 pasienter *uten* lungebetennelse ved innleggelse ($x = 0$), og 7 pasienter *med* lungebetennelse ved innleggelse ($x = 1$).

Tid til utskriving (dager):

Ikke lungebetennelse ved innleggelse ($x = 0$): 2, 3+, 6, 6, 10, 11, 12+, 23

Lungebetennelse ved innleggelse ($x = 1$): 4+, 9, 12+, 17, 24, 26+, 32

Her betyr + en høyresensurert observasjon. (Høyresensurering skjedde i de tilfeller der en pasient fortsatt var på sykehuset ved slutten av studien, eller døde på sykehuset.)

Det ble besluttet å analysere dataene ved hjelp av en Weibull regresjonsmodell med den ene kovariaten x definert ovenfor, som representerer status for lungebetennelse ved innleggelse.

Modellen for utskrivingstid, T, for en pasient med lungebetennelse-status x, er da

$$\ln T = \beta_0 + \beta_1 x + (1/\alpha)W, \tag{1}$$

der W er standard Gumbel-fordelt.

a) La U være Weibull-fordelt med formparameter α og skalaparameter θ, slik at pålitelighetsfunksjonen til U er

$$R_U(u) = e^{-\left(\frac{u}{\theta}\right)^\alpha} \text{ for } u > 0.$$

Vis at hasardratefunksjonen for U kan skrives

$$z_U(u) = \frac{\alpha u^{\alpha-1}}{\theta^\alpha} \text{ for } u > 0.$$

Det kan vises (du blir ikke bedt om å gjøre dette) at T definert ved (1) er Weibull-fordelt med formparameter α og skalaparameter

$$\theta = e^{\beta_0 + \beta_1 x}.$$

Bruk dette til å sette opp et uttrykk for hasardratefunksjonen for utskrivingstiden T for (i) en pasient *uten* lungebetennelse ved innleggelse; (ii) en pasient *med* lungebetennelse ved innleggelse.

Vis at den relative risiko (med hensyn på utskrivingstid) for en pasient uten lungebetennelse i forhold til en pasient med lungebetennelse er gitt ved $e^{\alpha\beta_1}$.

b) Det følgende er en (redigert) MINITAB utskrift for en Weibull regresjonsmodell med de gitte dataene.

```
Regression Table

                        Standard                       95,0% Normal CI
Predictor       Coef       Error        Z        P       Lower     Upper
Intercept    2,53917    0,217702        *        *         *         *
x           0,741551    0,344049        *        *         *         *
Shape        1,91178    0,461130        *        *         *         *

Log-Likelihood = -36,045

Distribution:   Weibull
```

Skriv ned den estimerte modellen for T for en pasient med lungebetennelsestatus x.

Er det en signifikant forskjell mellom utskrivingstidene for pasienter uten og med lungebetennelse ved innleggelse? Formuler dette problemet som et testingsproblem for en parameter i modellen, og gi konklusjonen når signifikansnivået settes til 5 %.

Beregn estimatet for den relative risiko for en pasient uten lungebetennelse i forhold til en pasient med lungebetennelse (se punkt a)). Hva er den praktiske fortolkningen av denne verdien i den gitte situasjonen?

Hvordan vil du tolke den estimerte verdien for formparameteren?

c) Bruk informasjon fra MINITAB-utskriften ovenfor til å beregne et 95 % konfidensintervall for parameteren β_1.

Dataanalytikeren beregnet også et 95 % konfidensintervall for parameteren α basert på MINITAB-utskriften. Gjør beregningen og skriv ned resultatet for intervallet.

En årsak til analytikerens interesse for det siste konfidensintervallet, var ønsket om å undersøke om en eksponensiell regresjonsmodell kunne være tilstrekkelig for dataene som en forenkling av Weibull-modellen.

Formuler en passende nullhypotese og den tilsvarende alternative hypotese for dette problemet. Bruk det beregnede konfidensintervallet for α til å gi en konklusjon. Hva er signifikansnivået for den testen som brukes?

En kollega av analytikeren foreslo å gjøre en ny MINITAB-analyse, der en *eksponensiell* regresjonsmodell blir tilpasset til dataene. Utskriften fra denne kjøringen vises nedenfor. Forklar hvordan man kan bruke opplysninger fra denne kjøringen sammen med informasjon fra den tidligere kjøringen til å konstruere en alternativ test for modellvalgproblemet. Bruk signifikansnivå 5 % i testingen. Sammenlign konklusjonen med den som allerede er oppnådd.

```
Regression Table

                     Standard                        95,0% Normal CI
Predictor      Coef     Error     Z      P        Lower     Upper
Intercept   2,49870  0,408248  6,12  0,000      1,69855   3,29885
x          0,935287  0,645497  1,45  0,147    -0,329864   2,20044
Shape             1

Log-Likelihood = -38,728

Distribution:   Exponential
```

d) Man utførte til slutt en logrank test for å sammenligne på en *ikke-parametrisk* måte de to gruppene av pasienter med hensyn til fordelingen av utskrivingstiden.

Skriv ned den relevante nullhypotesen og den alternativ hypotesen for en slik test, og forklar hvordan testobservatoren beregnes, uten nødvendigvis å gjøre alle de mellomliggende beregningene.

Beregn den endelige testobservatoren ved å bruke at de beregnede forventede antall utskrivinger under nullhypotesen er henholdsvis 3.20 og 6.80 for pasienter uten og med lungebetennelse ved innleggelse.

Hva blir konklusjonen av testen når signifikansnivået er 5 %? Hvordan passer denne konklusjonen med konklusjonen på det tilsvarende problemet i punkt b)? Gi en kommentar.

Oppgave 2 *Ikke-parametrisk estimering med venstresensurerte data*

15 enheter av et bestemt mekanisk utstyr ble satt på test i et laboratorium. Feil ble automatisk registrert, men på grunn av en svikt i overvåkningsapparatet, ble de eksakte feiltidene for enhetene som sviktet i løpet av de første to timene, ikke registrert. 8 enheter sviktet i løpet av de 2 timene, mens de nøyaktig registrerte feiltidene (timer) for de resterende 7 enhetene var:

$$2.1,\ 2.5,\ 3.9,\ 5.6,\ 16.2,\ 22.5,\ 28.8.$$

a) Hva betyr det at en levetid T er henholdsvis *høyre*-sensurert, *venstre*-sensurert eller *intervall*-sensurert? Gi et enkelt eksempel for å illustrere hver av disse tre typene sensureringer.

Forklar hvorfor dataene fra laboratorietesten beskrevet ovenfor kan betraktes som et venstresensurert datasett.

b) La T være tid til feil for en enhet av det beskrevne mekaniske utstyret. Betrakt transformasjonen

$$V = 1/T.$$

Verifiser at denne transformasjonen, når den brukes på dataene gitt ovenfor, endrer datasettet fra et venstresensurert sett til et *høyre*sensurert sett, med 7 nøyaktig observerte "levetider"

$$\begin{aligned} 1/28.8 &= \mathbf{0.035},\ 1/22.5 = \mathbf{0.044},\ 1/16.2 = \mathbf{0.062},\ 1/5.6 = \mathbf{0.179}, \\ 1/3.9 &= \mathbf{0.256},\ 1/2.5 = \mathbf{0.400},\ 1/2.1 = \mathbf{0.476} \end{aligned}$$

og 8 høyresensurerte observasjoner $1/2.0 = \mathbf{0.500}$.

Beregn Kaplan-Meier estimatet $\hat{R}_V(v)$ for $R_V(v) = P(V > v)$ ved å bruke de transformerte dataene (fete tall).

c) Verifiser sammenhengen

$$R_T(t) = 1 - R_V(1/t) \text{ for alle } t > 0,$$

der $R_T(t) = P(T > t)$ og det antas at T har en kontinuerlig fordeling.

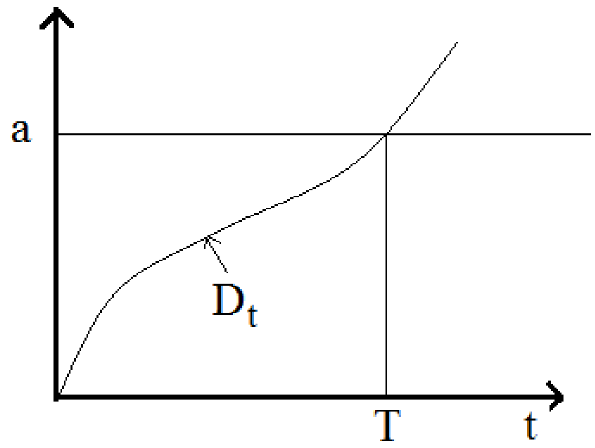
Bruk dette til å beregne et estimat $\hat{R}_T(t)$ for $R_T(t)$ og tegn det tilsvarende plottet.

Hvordan vil du estimere MTTF for utstyret med levetid T? (Du trenger ikke gjøre den fullstendige beregningen).

Oppgave 3 *Modellering av degradering og feil*

Betrakt igjen det mekaniske utstyret i Oppgave 2. I denne oppgaven skal vi først etablere en parametrisk modell for utstyrets levetid under laboratorietesten, og deretter skal vi estimere parameteren i denne modellen basert på dataene gitt i Oppgave 2.

Anta at degraderingen (slitasjen) av utstyret på tidspunkt $t \geq 0$ kan måles ved D_t, der $D_0 = 0$. Ettersom tiden går, utvikler D_t seg tilfeldig (stokastisk) på en strengt voksende måte, helt til den når et fast kritisk nivå $a > 0$, da utstyret feiler. La T være feiltiden (se figuren nedenfor).



a) Bruk figuren til å forklare at for alle $t \geq 0$,

$$P(T > t) = P(D_t < a). \tag{2}$$

Anta så at D_t-kurven er lineær i t, gitt som

$$D_t = Bt \tag{3}$$

for alle $t \geq 0$, der stigningstallet B er tilfeldig og eksponensialfordelt med hasardrate $\theta > 0$, dvs. $P(B \leq b) = 1 - e^{-\theta b}$ for $b > 0$.

Vis at den kumulative fordelingsfunksjonen til T under denne forutsetningen er

$$F(t) = P(T \leq t) = e^{-\frac{a\theta}{t}} \text{ for } t > 0.$$

I det følgende antas at (3) holder, og at det kritiske nivået a er satt til 1 (dette kan alltid oppnås ved å skalere θ tilsvarende).

Fordelingen til feiltiden T tilhører da den parametriske én-parameter familien med kumulativ fordelingsfunksjon

$$F(t;\theta) = P(T \leq t) = e^{-\frac{\theta}{t}} \text{ for } t > 0, \tag{4}$$

der $\theta > 0$ er parameteren. Denne fordelingen kalles *den inverse eksponensielle fordeling*.

b) La $t_p(\theta)$ være $100p$-persentilen for fordelingen (4), for $0 < p < 1$, slik at $F(t_p(\theta);\theta) = p$.

Finn et uttrykk for $t_p(\theta)$ for $0 < p < 1$.

Vis spesielt at medianen til T er $\theta/\ln 2$.

Finn også et uttrykk for interkvartilavstanden for T.

c) Anta nå at feildataene for utstyret som er gitt i begynnelsen av Oppgave 2, følger den parametriske modellen (4).

Sett opp likelihoodfunksjonen for dataene og finn sannsynlighetsmaksimeringsestimatet for θ.

Estimer median levetid for utstyret.

Tabell over kumulative sannsynligheter i χ^2-fordelingen med 1 frihetsgrad.

x	$P(X \leq x)$	x	$P(X \leq x)$
0.01	0.079656	2.7	0.899652
0.02	0.112463	2.8	0.905736
0.03	0.137510	2.9	0.911420
0.04	0.158519	3.0	0.916735
0.05	0.176937	3.1	0.921708
0.06	0.193504	3.2	0.926362
0.07	0.208663	3.3	0.930720
0.08	0.222703	3.4	0.934804
0.09	0.235823	3.5	0.938631
0.10	0.248170	3.6	0.942220
0.20	0.345279	3.7	0.945588
0.30	0.416118	3.8	0.948747
0.40	0.472911	3.9	0.951714
0.50	0.520500	4.0	0.954500
0.60	0.561422	4.1	0.957117
0.70	0.597216	4.2	0.959576
0.80	0.628907	4.3	0.961888
0.90	0.657218	4.4	0.964061
1.00	0.682689	4.5	0.966105
1.10	0.705734	4.6	0.968028
1.20	0.726678	4.7	0.969837
1.30	0.745787	4.8	0.971540
1.40	0.763276	4.9	0.973143
1.50	0.779329	5.0	0.974653
1.60	0.794097	5.1	0.976074
1.70	0.807712	5.2	0.977413
1.80	0.820288	5.3	0.978675
1.90	0.831922	5.4	0.979863
2.00	0.842701	5.5	0.980984
2.10	0.852701	5.6	0.982040
2.20	0.861989	5.7	0.983035
2.30	0.870626	5.8	0.983974
2.40	0.878665	5.9	0.984859
2.50	0.886154	6.0	0.985694
2.60	0.893136	6.1	0.986482

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 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

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