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: Onsdag 8. juni 2016

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 standardnormalfordelingen er gitt i vedlegget til slutt i oppgavesettet.

Målform/språk: bokmål

Antall sider: 5

Antall sider vedlegg: 3

Kontrollert av:

Dato Sign

Oppgave 1 *Graft versus host disease*

I en studie av pasienter med aplastisk anemi (en tilstand der beinmargen ikke produserer tilstrekkelig med nye blodceller), fikk 64 pasienter en beinmargstransplantasjon. Pasientene ble så randomisert til behandling med (i) methotrexate (MTX) og cyclosporin (CSP), eller (ii) bare methotrexate. For hver pasient registrerte man tiden (i dager) fra randomisering til forekomst av den livstruende komplikasjonen GVHD (= *Graft Versus Host Disease*). Vi vil her bare betrakte de 32 pasientene som fikk bare MTX, der 15 av dem opplevde GVHD.

Følgende tider til GVHD ble observert, der * betyr en sensurert observasjon.

9, 11, 12, 20*, 20, 20, 22, 25*, 25, 25, 27*, 28, 28, 30*, 31, 35, 35, 41*, 46, 49, 50*, 53*, 54*, 56*, 58*, 60*, 64*, 65*, 66*, 74*, 75*, 77*

a) Bruk den (redigerte) MINITAB-utskriften til slutt i dette punktet til å tegne Kaplan-Meier (KM) kurven for tiden til GVHD for disse pasientene.

Anta at vi er interessert i sannsynligheten for ikke å få GVHD i de første tre ukene (21 dager). Vis først hvordan denne sannsynligheten kan finnes fra MINITAB-utskriften, og vis deretter hvordan den beregnes fra de observerte dataene ovenfor. Beregn også standardfeilen for dette estimatet og finn et 95% konfidensintervall for sannsynligheten.

(*Vink:* Greenwoods formel er

$$\left(\hat{R}(t)\right)^2 \sum_{T_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)}. \qquad)$$

```
Kaplan-Meier Estimates

        Number  Number      Survival
Time   at Risk  Failed   Probability
   9        32       1      0,968750
  11        31       1      0,937500
  12        30       1      0,906250
  20        29       2      0,843750
  22        26       1      0,811298
  25        25       2      0,746394
  28        21       2      0,675309
  31        18       1      0,637792
  35        17       2      0,562758
```

```
  46        14        1      0,522561
  49        13        1      0,482364

Mean(MTTF)   51,3269
```

b) Bruk MINITAB-utskriften til å estimere den nedre kvartil og medianen i fordelingen for tid til GVHD. Hvorfor kan man ikke estimere den øvre kvartil fra denne analysen?

MINITAB-utskriften gir også et estimat for forventet tid til GVHD. Forklar kort hvordan dette estimatet er funnet. Hvilken egenskap ved forventningsverdien $E(T)$ for en generell levetid T er grunnlaget for dette estimatet?

c) Gi en kort forklaring på den generelle egenskapen til KM-estimatoren $\hat{R}(t)$, at $\hat{R}(t)$ er lik 0 ved siste feiltid hvis dette er den høyeste rapporterte tiden, mens $\hat{R}(t)$ har en strengt positiv verdi ved siste feiltid hvis den høyeste rapporterte tiden er en sensurering. I det siste tilfellet forblir estimatet $\hat{R}(t)$ lik denne verdien for alle større t.

Hvilket tilfelle har vi i den aktuelle situasjonen? Basert på KM-kurven, hvordan ville du anslå sannsynligheten for å være fri for GVHD for t større enn 90 dager, for eksempel?

I lys av det ovenfor, diskuter den praktiske gyldigheten av den estimerte forventede tid til GVHD i punkt b).

En måte å fortolke KM-kurven i den foreliggende situasjonen er at det er to grupper av pasienter i studien. Én gruppe vil aldri oppleve GVHD, mens resten av pasientene vil få GVHD på ett eller annet tidspunkt.

La X være tiden da en pasient får GVHD. La $X = \infty$ for de pasientene som ikke får GVHD, mens X er den faktiske tiden til GVHD for de resterende. Overlevelsesfunksjonen tilhørende X kan da representeres ved en såkalt *kureringsmodell* som:

$$P(X > t) = q + (1-q)R_0(t), \tag{1}$$

definert for alle $t > 0$. Her er q ($0 \leq q < 1$) sannsynligheten for ikke å få GVHD ("bli kurert"), mens $R_0(t)$ er overlevelsesfunksjonen for tiden til GVHD for pasientene som får GVHD ("ikke kurert").

d) Forklar kort hvorfor Kaplan-Meier kurven i denne Oppgaven kan anses som estimator for $P(X > t)$ i (1).

Hvordan vil du estimere q fra KM-estimatet?

La t_r være $100r$-prosentilen for fordelingen med overlevelsesfunksjon $R_0(t)$, definert ved

$$R_0(t_r) = 1 - r.$$

La x_p være $100p$-prosentilen for fordelingen til X gitt ved (1). Vis at

$$x_p = t_{\frac{p}{1-q}}$$

for $p < 1 - q$.

Oppgave 2 *Produktklager*

Følgende tabell gir data for rapporterte klager for et bestemt produkt, for enheter som ble solgt på en bestemt dag (dag nummer 1) på ti forskjellige steder. Data om klager ble samlet inn for disse enhetene for en periode på to uker (dager nummerert 1-14).

Lokasjon	Antall solgte	Dagnummer da klager ble mottatt
1	10	3, 4, 9
2	14	1,2, 7
3	7	2, 9
4	5	3
5	12	1, 4, 12
6	15	2, 2, 7
7	10	2, 5
8	13	1, 3, 6, 10
9	9	2, 3, 7
10	5	1
	Sum 100	Totalt antall klager: 25
		Sum av dagnummer for klager: 108

For en enkelt enhet av dette produktet, la $W(t)$ være forventet antall klager mottatt inntil tid t (dager) etter kjøpet. Det antas at alle enheter har den samme funksjonen $W(t)$. Merk at en enhet kan ha mer enn én klage i perioden på to uker.

a) Anta at du har data bare for Lokasjon 1. Hvordan vil du estimere funksjonen $W(t)$ i dette tilfellet?

Hva er estimatet for $W(14)$? Hvordan vil du fortolke dette tallet?

b) Estimer funksjonen $W(t)$ ved å bruke all informasjon i tabellen. Hva er navnet på den estimatoren du bruker?

Hva er nå estimatet for $W(14)$?

Angi en grov skisse av den estimerte kurven for $W(t)$. Hva kan du si om dens form, og hvordan vil du tolke denne formen? Er dette den formen du ville forvente?

c) For å sjekke på en formell måte om det er en trend i tidspunktene for klager, brukte man pooled Laplace-test.

Beregn testobservatoren ved å bruke alle de ti lokasjonene.

(*Vink:* Testobservatoren for pooled Laplace test forenkles når alle prosessene er observert på det samme tidsintervallet.)

Hva er nullhypotesen som testes ved denne testen? Hva er en naturlig alternativ hypotese i det aktuelle tilfellet?

Beregn p-verdien for testen og skriv ned din konklusjon basert på denne.

Oppgave 3 *Den log-logistiske fordeling*

Den *log-logistiske fordeling* (også kalt *Fisk-fordelingen*) kan parametriseres slik at pålitelighetsfunksjonen er gitt ved

$$R(t) = \frac{1}{1 + \left(\frac{t}{\theta}\right)^{\alpha}} \tag{2}$$

for $t > 0$, der $\theta > 0, \alpha > 0$.

a) Sjekk at funksjonen i (2) tilfredsstiller kravene til en pålitelighetsfunksjon.

Utled et uttrykk for tettheten $f(t)$ og vis at hasardraten $z(t)$ kan skrives

$$z(t) = \frac{\alpha t^{\alpha-1}}{\theta^{\alpha} + t^{\alpha}} \tag{3}$$

for $t > 0$.

b) Forklar hvordan uttrykket for $R(t)$ i (2) kan brukes til å konstruere et sannsynlighetsplott for å sjekke om et høyresensurert datasett kan antas å komme fra en log-logistisk fordeling.

Forklar også kort hvordan estimater for θ og α kan finnes fra plottet.

c) Vis at hasardraten $z(t)$ i (3) er avtagende i t når $\alpha \leq 1$.

Vis deretter ved derivasjon at for $\alpha > 1$ er $z(t)$ først voksende fra 0 til et maksimum og deretter avtagende til 0 når t går mot ∞.

Hvilken verdi av t gir maksimal $z(t)$?

Oppførselen til hasardraten for $\alpha > 1$ er analog i form med hasardraten for *lognormalfordelingen*. Den log-logistiske fordelingen blir ofte brukt istedenfor lognormalfordelingen i analyser.

Hvilke fordeler kan du peke på ved å bruke log-logistisk fordeling istedenfor lognormalfordelingen?

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