First Semester Examination
Academic Session 2018/2019

December 2018/January 2019

MSG460 - Survival Analysis
(Analisis Mandiri)

Duration : 3 hours
[Masa : 3 jam]

Please check that this examination paper consists of TWENTY-ONE (21) pages of printed material before you begin the examination.

[Sila pastikan bahawa kertas peperiksaan ini mengandungi DUA PULUH SATU (21) muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]

Instructions : Answer **FOUR (4)** questions.

*[**Arahan :** Jawab **EMPAT (4)** soalan.]*

In the event of any discrepancies, the English version shall be used.

[Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi Bahasa Inggeris hendaklah digunapakai].

Question 1

(a) What is Survival Analysis? Give two examples to illustrate your answer.

[20 marks]

(b) Why is it important to perform Survival Analysis?

[20 marks]

(c) Give a step-by-step outline of analyzing survival data.

[30 marks]

(d) When a censoring mechanism is employed, will the percentage of censoring affect the results of survival analysis? Justify your answer.

[30 marks]

Soalan 1

(a) Apakah Analisis Mandiri? Berikan dua contoh untuk mengilustrasi jawapan anda.

[20 markah]

(b) Mengapakah penting melaksanakan Analisis Mandiri?

[20 markah]

(c) Berikan langkah-demi-langkah untuk menganalisa data mandiri.

[30 markah]

(d) Apabila mekanisme tapisan digunakan, adalah peratus tapisan mempengaruhi keputusan analisis mandiri? Justifikasikan jawapan anda.

[30 markah]

Question 2

(a) The following is a dataset from Shouki and Pause (1999) for times (in months) until breast cancer recurrence in 73 female patients (see **Table 1**)**.** Much research has recently been directed at early detection through mass screening. The time to cancer recurrence are recorded. Only 17 observations were uncensored observed times until cancer recurrence and 56 observations were censored.

Table 1. Time to breast cancer recurrence data for 73 female patients.

Patient No.	Time (month)	Patient No.	Time (month)
1	130+	38	48+
2	136+	39	89+
3	117+	40	95+
4	50	41	91+
5	106+	42	47
6	103+	43	75+
7	86	44	49+
8	63+	45	66+
9	120+	46	65+
10	121+	47	22+
11	108	48	73+
12	121+	49	67+
13	109+	50	75+
14	111+	51	71+
15	60	52	80+

16	106+	53	25+
17	108+	54	67+
18	105+	55	74+
19	9+	56	64
20	108	57	64+
21	62+	58	41
22	106+	59	70
23	95	60	57
24	94+	61	59+
25	19+	62	53+
26	103	63	69+
27	60+	64	55+
28	91+	65	58
29	70+	66	68+
30	65+	67	60+
31	91+	68	126
32	86	69	127+
33	90+	70	126+
34	87+	71	102+
35	89	72	122+
36	89	73	100+
37	92+		

(i) Obtain a summary of the distribution of the data set in **Table 1.**
(ii) Interpret your results. What are your conclusions?
(iii) Find $\hat{S}(120)$. Explain your answer.

[50 marks]

(b) The following are the survival times (days) of street lights at various locations around the city: where '+' denotes censoring caused by traffic accidents destroying entire lighting units.

$$36, 37, 38, 38+, 78, 111, 112, 114+, 162, 189, 198, 237, 489+$$

(i) Determine an actuarial estimator of the survival function.

(ii) Determine the Product Limit (PL) estimator of the survival function.

(iii) When do you use actuarial and PL estimates in survival analysis? Explain.

[50 marks]

Soalan 2

*(a) Berikut adalah satu set data dari Shouki dan Pause (1999) bagi masa (dalam bulan) sehingga berulangnya kanser payudara bagi 73 pesakit wanita (lihat **Jadual 1**). Kebelakangan ini banyak kajian telah dijalankan ke arah pengesanan awal melalui saringan beramai-ramai. Masa sehingga kanser berulang dicatat. Hanya 17 cerapan adalah masa tak tertapis dan 56 cerapan adalah tertapis.*

Jadual 1. Data masa berulangnya kanser payudara bagi 73 pesakit wanita.

No. Pesakit	*Masa (bulan)*	*No. Pesakit*	*Masa (bulan)*
1	*130+*	*38*	*48+*
2	*136+*	*39*	*89+*
3	*117+*	*40*	*95+*
4	*50*	*41*	*91+*
5	*106+*	*42*	*47*
6	*103+*	*43*	*75+*
7	*86*	*44*	*49+*
8	*63+*	*45*	*66+*

9	*120+*	*46*	*65+*
10	*121+*	*47*	*22+*
11	*108*	*48*	*73+*
12	*121+*	*49*	*67+*
13	*109+*	*50*	*75+*
14	*111+*	*51*	*71+*
15	*60*	*52*	*80+*
16	*106+*	*53*	*25+*
17	*108+*	*54*	*67+*
18	*105+*	*55*	*74+*
19	*9+*	*56*	*64*
20	*108*	*57*	*64+*
21	*62+*	*58*	*41*
22	*106+*	*59*	*70*
23	*95*	*60*	*57*
24	*94+*	*61*	*59+*
25	*19+*	*62*	*53+*
26	*103*	*63*	*69+*
27	*60+*	*64*	*55+*
28	*91+*	*65*	*58*
29	*70+*	*66*	*68+*
30	*65+*	*67*	*60+*
31	*91+*	*68*	*126*
32	*86*	*69*	*127+*
33	*90+*	*70*	*126+*

…7/-

34	*87+*	*71*	*102+*
35	*89*	*72*	*122+*
36	*89*	*73*	*100+*
37	*92+*		

(i) *Dapatkan ringkasan taburan set data dalam **Jadual 1.***

(ii) *Tafsirkan keputusan anda. Apakah kesimpulan anda?*

(iii) *Cari* $\hat{S}(120)$. *Jelaskankan jawapan anda.*

[50 markah]

(b) *Berikut adalah jangka hayat (hari) bagi lampu jalan di berbagai lokasi di sekitar bandar raya: yang mana '+' menandakan penapisan yang disebabkan oleh kemalangan jalan raya yang merosakkan seluruh unit lampu.*

$$36, 37, 38, 38+, 78, 111, 112, 114+, 162, 189, 198, 237, 489+$$

(i) *Tentukan penganggar aktuari bagi fungsi mandiri.*

(ii) *Tentukan penganggar Had Produk (PL) bagi fungsi mandiri.*

(iii) *Bilakah anda gunakan anggaran aktuari dan PL dalam analisis mandiri? Terangkan.*

[50 markah]

Question 3

(a) What are the reasons for modeling survival data? Describe the Proportional Hazards Model which is a basic model for survival data.

[20 marks]

(b) Are there differences between the Accelerated Failure Time (AFT) Model and the usual Linear Regression Model? If there are differences, what are they?

[20 marks]

(c) Surgical treatment of ovarian cancer may be followed by a course of chemotherapy. Consider a clinical trial which involved 26 women with minimal residual disease and who had experienced surgical excision of all tumour masses greater than 2 cm in diameter. Following surgery, the patients were further classified according to whether the residual disease was completely or partially excised. The age of the patient and their performance status were also recorded at the start of the trial. The response variable was the survival time in days following randomization to one or other of the two chemotherapy treatments: cyclophosphamide alone (single) and cyclophosphamide combined with adriamycin (combined). The variables in the data set are therefore as follows:

TIME: Survival time in days
CENS: Censoring indicator (0 = censored, 1 = uncensored)
TREAT: Treatment (1 = single, 2 = combined)
AGE: Age of patient in years
RDISEASE: Extent of residual disease (1 = incomplete, 2 = complete)
PERF: Performance status (1 = good, 2 = poor)

The data are given in **Table 2**.

Table 2. Survival Times of 26 Ovarian Cancer Patients.

Patient	TIME	CENS	TREAT	AGE	RDISEASE	PERF
1	156	1	1	66	2	2
2	1040	0	1	38	2	2
3	59	1	1	72	2	1
4	421	0	2	53	2	1
5	329	1	1	43	2	1
6	769	0	2	59	2	2
7	365	1	2	64	2	1

…9/-

8	770	0	2	57	2	1
9	1227	0	2	59	1	2
10	268	1	1	74	2	2
11	475	1	2	59	2	2
12	1129	0	2	53	1	1
13	464	1	2	56	2	2
14	1206	0	2	44	2	1
15	638	1	1	56	1	2
16	563	1	2	55	1	2
17	1106	0	1	44	1	1
18	431	1	1	50	2	1
19	855	0	1	43	1	2
20	803	0	1	39	1	1
21	115	1	1	74	2	1
22	744	0	2	50	1	1
23	477	0	1	64	2	1
24	448	0	1	56	1	2
25	353	1	2	63	1	2
26	377	0	2	58	1	1

(i) In modelling the data in **Table 2**, fit a range of models to these data. Which model is most satisfactory? Give your justification.

(ii) Then identify which prognostic factors are associated with the survival times of the patients. What is your conclusion?

(iii) Is there a difference in the effect of the two chemotherapy treatments on the hazard of death? If there is a treatment difference, is it consistent over age?

…10/-

(iv) Obtain the estimated median survival time for patients of a given age on a given treatment.

[60 marks]

Soalan 3

(a) Apakah sebab-sebab bagi pemodelan data survival? Huraikan Model Bahaya Berkadaran yang merupakan model asas bagi data survival.

[20 markah]

(b) Adakah terdapat perbezaan antara Model Masa Kegagalan Terpecut (AFT) dan Model Regresi Linear yang biasa? Jika ada, apakah perbezaannya?

[20 markah]

(c) Rawatan surgeri bagi kanser ovari boleh diikuti dengan satu kursus kemoterapi Pertimbangkan suatu cubaan klinikal yang melibatkan 26 orang wanita dengan penyakit reja minimal dan yang mengalami pemotongan surgeri bagi semua tumor yang melebihi garis pusat 2 sentimeter. Selepas pembedahan, pesakit-pesakit dikelaskan lagi mengikut sama ada penyakit reja itu dipotong secara lengkap atau tidak lengkap. Umur pesakit dan prestasinya dicatat pada awal percubaan. Pemboleh ubah sambutan ialah masa survival dalam hari mengikut proses rawakan bagi dua rawatan chemoterapi. Pemboleh ubah dalam set data adalah seperti berikut:

TIME: Masa mandiri dalam hari

CENS: Petunjuk tapis (0 = tertapis, 1 = tak tertapis)

TREAT: Rawatan (1 = tunggal, 2 = kombinasi)

AGE: Umur pesakit dalam tahun

RDISEASE: Sejauh mana penyakit reja merebak
(1 = tidak lengkap, 2 = lengkap)

PERF: Status prestasi (1 = baik, 2 = lemah)

*Data diberi dalam **Jadual 2**.*

***Table 2.** Masa Mandiri bagi 26 Pesakit Kanser Ovari.*

Patient	*TIME*	*CENS*	*TREAT*	*AGE*	*RDISEASE*	*PERF*
1	*156*	*1*	*1*	*66*	*2*	*2*
2	*1040*	*0*	*1*	*38*	*2*	*2*

3	*59*	*1*	*1*	*72*	*2*	*1*
4	*421*	*0*	*2*	*53*	*2*	*1*
5	*329*	*1*	*1*	*43*	*2*	*1*
6	*769*	*0*	*2*	*59*	*2*	*2*
7	*365*	*1*	*2*	*64*	*2*	*1*
8	*770*	*0*	*2*	*57*	*2*	*1*
9	*1227*	*0*	*2*	*59*	*1*	*2*
10	*268*	*1*	*1*	*74*	*2*	*2*
11	*475*	*1*	*2*	*59*	*2*	*2*
12	*1129*	*0*	*2*	*53*	*1*	*1*
13	*464*	*1*	*2*	*56*	*2*	*2*
14	*1206*	*0*	*2*	*44*	*2*	*1*
15	*638*	*1*	*1*	*56*	*1*	*2*
16	*563*	*1*	*2*	*55*	*1*	*2*
17	*1106*	*0*	*1*	*44*	*1*	*1*
18	*431*	*1*	*1*	*50*	*2*	*1*
19	*855*	*0*	*1*	*43*	*1*	*2*
20	*803*	*0*	*1*	*39*	*1*	*1*
21	*115*	*1*	*1*	*74*	*2*	*1*
22	*744*	*0*	*2*	*50*	*1*	*1*
23	*477*	*0*	*1*	*64*	*2*	*1*
24	*448*	*0*	*1*	*56*	*1*	*2*
25	*353*	*1*	*2*	*63*	*1*	*2*
26	*377*	*0*	*2*	*58*	*1*	*1*

*(i) Dalam pemodelan data dalam **Jadual 2**, suaikan beberapa model kepada data ini. Model manakah yang paling memuaskan? Beri justifikasi anda.*

(ii) Kemudian kenalpasti faktor-faktor prognostik yang berkenaan dengan masa mandiri bagi pesakit. Apakah kesimpulan anda?

(iii) Adakah terdapat perbezaan dalam kesan bagi dua rawatan kemoterapi pada hazard kematian? Jika terdapat perbezaan rawatan, adakah ia konsisten dengan umur?

(iv) Dapatkan aggaran median masa mandiri bagi pesakit diberi umur dan rawatan.

[60 markah]

Question 4

(a) The data in **Table 3** gives the remission times (in weeks) for patients subjected to two different treatments. Twenty patients were assigned to each of the two treatments; '+' denotes 'right censoring'.

Table 3. Remission times (in weeks) for twenty patients

Treatment A	Treatment B
1 3 3 6 7 7 10 12 14 15 18 19 22 26 28+ 29 34 40 48+ 49+	1 1 2 2 3 4 5 8 8 9 11 12 14 16 18 21 27+ 31 38+ 44

(i) Compare the survival prospects of the two groups by interpreting the graphs of their survival functions.
(ii) Obtain estimates of the probability of survival beyond 12 weeks.
(iii) What is your conclusion based on the insight from parts (i) and (ii)?

[50 marks]

(b) Much cancer research has recently been directed at early detection through mass screening and an assessment of other variables that are predictive of remission times. In a study of breast cancer recurrence in female patients, only 17 observations were uncensored observed times until cancer recurrence and 56 observations were censored. **Table 4** displays the grouped data obtained from this study.

Table 4. Breast cancer recurrence data

Remission time (in months)	Number of patients	Number of recurrences	Number censored
0 - 50	73	2	5
50 - 70	66	5	15
70 - 80	46	1	6
80 - 90	39	4	3
90 - 100	32	1	8
100 - 120	23	3	11
120 -140	9	1	8

(i) What is the time to event in this study?

(ii) Why are some observations censored in this study?

(iii) What are the possible censoring mechanisms in this study?

(iv) Perform a suitable analysis on the grouped data in this study.

(v) Does the high proportion of censoring affect the estimates of survival probabilities? If it does, then explain.

(vi) What is the probability of survival of breast cancer patients beyond 120 months based on the results obtained in part (iv)?

[50 marks]

Soalan 4

*(a) Data dalam **Jadual 3** memberikan masa remitan (dalam minggu) untuk pesakit yang diberi dua rawatan yang berbeza. Dua puluh pesakit telah ditugaskan untuk setiap dua rawatan; '+' menandakan 'tapisan kanan.*

***Jadual 3**. Masa remitan (dalam minggu) untuk dua puluh pesakit*

Rawatan A	***Rawatan B***
1 3 3 6 7 7 10 12 14 15 18 19 22 26 28+ 29 34 40 48+ 49+	*1 1 2 2 3 4 5 8 8 9 11 12 14 16 18 21 27+ 31 38+ 44*

(i) Bandingkan prospek mandiri kedua-dua kumpulan dengan mentafsir graf fungsi mandiri mereka.

(ii) Dapatkan anggaran kebarangkalian mandiri melebihi 12 minggu.

(iii) Apakah kesimpulan anda berdasarkan pandangan dari bahagian (i) dan (ii)?

[50 markah]

*(b) Banyak penyelidikan kanser baru-baru ini telah diarahkan pada pengesanan awal melalui pemeriksaan besar-besaran dan penilaian pembolehubah lain yang meramalkan masa remitan. Dalam satu kajian kanser payudara berulang di kalangan pesakit wanita, hanya 17 pemerhatian adalah masa yang tidak tertapis sehingga berulangnya kanser dan 56 pemerhatian adalah tertapis. **Jadual 4** memaparkan data berkumpulan yang diperoleh daripada kajian ini.*

***Jadual 4**. Data kanser payudara berulang*

Masa Peremitan (dalam bulan)	***Bilangan pesakit***	***Bilangan berulang***	***Bilangan tertapis***
0 - 50	*73*	*2*	*5*
50 - 70	*66*	*5*	*15*
70 - 80	*46*	*1*	*6*

80 - 90	*39*	*4*	*3*
90 - 100	*32*	*1*	*8*
100 - 120	*23*	*3*	*11*
120 -140	*9*	*1*	*8*

(i) Apakah masa kepada peristiwa dalam kajian ini?

(ii) Kenapakah sesetengah pemerhatian tertapis dalam kajian ini?

(iii) Apakah mekanisme penapisan yang mungkin dalam kajian ini?

(iv) Laksanakan analisis yang sesuai atas data berkumpulan dalam kajian ini.

(v) Adakah kadaran penapisan yang tinggi mempengaruhi anggaran kebarangkalian mandiri? Jika ya, jelaskan.

(vi) Apakah kebarangkalian mandiri pesakit kanser payudara yang melebihi 120 bulan berdasarkan keputusan yang diperoleh dalam bahagian (iv)?

[50 markah]

APPENDIX

Summary of Reliability Formulae

$$F(t) = \int_0^t f(t)dt$$

$$R(t) = 1 - F(t)$$

$$f(t) = \frac{dF(t)}{dt} = -\frac{dR(t)}{dt}$$

$$h(t) = \frac{f(t)}{R(t)}$$

$$H(t) = \int_0^t h(t)dt$$

$$R(t) = e^{-H(t)}$$

$$H(t) = -\ln R(t)$$

$$MTTF = \int_0^\infty tf(t)dt = \int_0^\infty R(t)dt$$

APPENDIKS

Ringkasan Rumus-Rumus Kebolehpercayaan

$$F(t) = \int_0^t f(t)dt$$

$$R(t) = 1 - F(t)$$

$$f(t) = \frac{dF(t)}{dt} = -\frac{dR(t)}{dt}$$

$$h(t) = \frac{f(t)}{R(t)}$$

$$H(t) = \int_0^t h(t)dt$$

$$R(t) = e^{-H(t)}$$

$$H(t) = -\ln R(t)$$

$$MTTF = \int_0^\infty tf(t)dt = \int_0^\infty R(t)dt$$

APPENDIX (contd)

Summary of Reliability Formulae (contd)

Lifetime following an **Exponential Distribution:**

$$f(t) = \lambda e^{-\lambda t}$$
$$F(t) = 1 - e^{-\lambda t}$$
$$R(t) = e^{-\lambda t}$$
$$h(t) = \lambda$$
$$H(t) = \lambda t$$
$$MTTF = \frac{1}{\lambda}$$

Lifetime following a **Weibull Distribution:**

$$f(t) = \beta \alpha^{-\beta} t^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]$$

$$F(t) = 1 - e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$$

$$R(t) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$$

$$h(t) = \beta \alpha^{-\beta} t^{\beta-1}$$

$$H(t) = \left(\frac{t}{\alpha}\right)^{\beta}$$

$$MTTF = \alpha \Gamma\left(1 + \frac{1}{\beta}\right)$$

$$\text{Design Life} = t_R = \alpha(-\ln R)^{1/\beta}$$

APPENDIKS (sambung)

Ringkasan Rumus-Rumus Kebolehpercayaan (sambung)

Masahayat mengikut ***Taburan Eksponen:***

$$f(t) = \lambda e^{-\lambda t}$$
$$F(t) = 1 - e^{-\lambda t}$$
$$R(t) = e^{-\lambda t}$$
$$h(t) = \lambda$$
$$H(t) = \lambda t$$
$$MTTF = \frac{1}{\lambda}$$

Masahayat mengikut ***Taburan Weibull:***

$$f(t) = \beta \alpha^{-\beta} t^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right]$$

$$F(t) = 1 - e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$$

$$R(t) = e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$$

$$h(t) = \beta \alpha^{-\beta} t^{\beta-1}$$

$$H(t) = \left(\frac{t}{\alpha}\right)^{\beta}$$

$$MTTF = \alpha \Gamma\left(1 + \frac{1}{\beta}\right)$$

$$\text{Design Life} = t_R = \alpha(-\ln R)^{1/\beta}$$

APPENDIX (contd)

Summary of Reliability Formulae (contd)

Lifetime following a **Normal Distribution:**

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(t-\mu)^2}{\sigma^2} \right]$$

$$F(t) = \Phi\left(\frac{t-\mu}{\sigma} \right)$$

$$R(t) = 1 - \Phi\left(\frac{t-\mu}{\sigma} \right)$$

$$h(t) = \frac{f(t)}{1 - \Phi\left(\frac{t-\mu}{\sigma} \right)}$$

Lifetime following a **Lognormal Distribution:**

$$f(t) = \frac{1}{\sqrt{2\pi}st} \exp\left[-\frac{1}{2s^2} \left(\ln \frac{t}{t_{median}} \right)^2 \right]$$

$$F(t) = \Phi\left(\frac{1}{s} \ln \frac{t}{t_{median}} \right)$$

$$R(t) = 1 - \Phi\left(\frac{1}{s} \ln \frac{t}{t_{median}} \right)$$

$$MTTF = t_{median} \exp\left(\frac{s^2}{2} \right)$$

$$t_R = t_{median} \exp\left(s z_{1-R} \right)$$

APPENDIKS (sambung)

Ringkasan Rumus-Rumus Kebolehpercayaan (sambung)

*Masahayat mengikut **Taburan Normal:***

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \frac{(t-\mu)^2}{\sigma^2} \right]$$

$$F(t) = \Phi\left(\frac{t-\mu}{\sigma} \right)$$

$$R(t) = 1 - \Phi\left(\frac{t-\mu}{\sigma} \right)$$

$$h(t) = \frac{f(t)}{1 - \Phi\left(\frac{t-\mu}{\sigma} \right)}$$

*Masahayat mengikut **Taburan Lognormal:***

$$f(t) = \frac{1}{\sqrt{2\pi}st} \exp\left[-\frac{1}{2s^2} \left(\ln \frac{t}{t_{median}} \right)^2 \right]$$

$$F(t) = \Phi\left(\frac{1}{s} \ln \frac{t}{t_{median}} \right)$$

$$R(t) = 1 - \Phi\left(\frac{1}{s} \ln \frac{t}{t_{median}} \right)$$

$$MTTF = t_{median} \exp\left(\frac{s^2}{2} \right)$$

$$t_R = t_{median} \exp\left(s z_{1-R} \right)$$

- ooo00ooo -