Dept of Mathematics and Statistics
King Fahd University of Petroleum & Minerals
AS476: Survival Models for Actuaries
Dr. Ridwan A. Sanusi
Final Exam Term 251
Wednesday, December 24, 2025
7.00 pm - 9.00 pm

Name……. ID#:_____ Serial #:___

Instructions.

1. Please turn off your cell phones and place them under your chair. Any student caught with mobile phones on during the exam will be considered under the cheating rules of the University.
2. If you need to leave the room, please do so quietly so not to disturb others taking the test. No two person can leave the room at the same time. No extra exam time will be provided for the time spent outside the room.
3. Only materials provided by the instructor can be present on the table during the exam.
4. Do not spend too much time on any one question. If a question seems too difficult, leave it and go on.
5. Use the blank portions of each page for your work. Extra blank pages can be provided if necessary. If you use an extra page, indicate clearly what problem you are working on.
6. ***Only answers supported by work will be considered. Unsupported guesses will not be graded.***
7. While every attempt is made to avoid defective questions, sometimes they do occur. In the rare event that you believe a question is defective, the instructor cannot give you any guidance beyond these instructions.
8. Mobile calculators, I-pad, or communicable devices are disallowed. Use regular scientific calculators, financial calculators, or SOA approved calculators only. ***Write important steps to arrive at the solution of the exam problems.***

The test is 120 minutes, GOOD LUCK, and you may begin now!

Question	Total Marks	Marks Obtained	Comments
1-12	12		
PART B			
1	2		
2	2		
3	2		
4	4		
5	4		
6	4		
Total	30		

Extra blank page

1. Survival analysis is a collection of statistical procedures for data analysis for which the outcome variable is time until an event occurs.
   a) TRUE
   b) FALSE

2. In using Age as the time scale, a key decision in any survival analysis is where to define the starting point for determining individual's "true" survival time, which we call time 0. Depending on the study, choices for time 0 might be:
   (a) the time the subject enters the study.
   (b) the time the subject begins treatment.
   (c) a point in calendar time.
   (d) the time of a seminal event (e.g., surgery), birth, or conception.
   (e) All of the above.

3. The following questions derive from the dataset vets.dat concerning the Veteran's Administration Lung Cancer Trial. Consider the following two edited printouts obtained from fitting a Cox PH model to these data. How do the printouts differ in terms of what the P(PH) information says about which variables satisfy or do not satisfy the PH assumption?

Cox regression
Analysis time _t:

survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]	*P(PH)*
Treatment	0.290	0.207	0.162	1.336	0.890	2.006	0.628
Large cell	0.400	0.283	0.157	1.491	0.857	2.594	0.033
Adeno cell	1.188	0.301	0.000	3.281	1.820	5.915	0.081
Small cell	0.856	0.275	0.002	2.355	1.374	4.037	0.078
Perf.Stat	−0.033	0.006	0.000	0.968	0.958	0.978	0.000
Dis.Durat.	0.000	0.009	0.992	1.000	0.982	1.018	0.919
Age	−0.009	0.009	0.358	0.991	0.974	1.010	0.198
Pr.Therapy	0.007	0.023	0.755	1.007	0.962	1.054	0.145

No. of subjects = 137 Log likelihood = −475.180

Cox regression
Analysis time _t:

survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf.	Interval]	*P(PH)*
Treatment	0.298	0.197	0.130	1.347	0.916	1.981	0.739
Small cell	0.392	0.210	0.062	1.481	0.981	2.235	0.382
Perf.Stat	−0.033	0.005	0.000	0.968	0.958	0.978	0.000
Dis.Durat.	−0.001	0.009	0.887	0.999	0.981	1.017	0.926
Age	−0.006	0.009	0.511	0.994	0.976	1.012	0.211
Pr.Therapy	−0.003	0.023	0.884	0.997	0.954	1.042	0.146

No. of subjects = 137 Log likelihood = −487.770

   a) The first printout indicates that the variables large cell, adeno cell, small cell, and performance status satisfy the PH assumption at the 0.10 level.
   b) The second printout indicates that small cell does not satisfy the PH assumption.
   c) The performance status variable satisfies the PH assumption.
   d) All of the above.
   e) None of the above.

4. Drawback of using the log–log approach for assessing the PH assumption is:
   a) How parallel is parallel?
   b) How to categorize a continuous variable?
   c) How to evaluate several variables simultaneously?

d) All of the above.
e) None of the above.

5. The acceleration factor comparing exposed and unexposed subjects, (E . 1 vs. E . 0), is a ratio of their median survival times (time to S(t) . 0.5), or. more generally the ratio of their times to any fixed value of S(t) = q.
   a) TRUE
   b) FALSE

6. Let $S_0(t)$ be the survival function for unexposed subjects (E = 0) and let $S_1(t)$ be the survival function for exposed subjects (E = 1). If $S_0(t) = S_1(3t)$ then the median survival time for the unexposed subjects is 3 times longer than the median survival time for the exposed subjects.
   a) TRUE
   b) FALSE

F: The median survival time for the unexposed is 1/3 of the median survival time for the exposed.

7. The Cox proportional hazards model is a parametric model.
   (a) TRUE
   (b) FALSE

F: The Cox model is a semiparametric model. The distribution of survival time is unspecified in a Cox model.

8. if the acceleration failure time (AFT) assumption holds in a Weibull model then the proportional hazards assumption also holds.
   (a) TRUE
   (b) FALSE

9. A competing risk is an event-type (i.e., failure status) that can occur simultaneously with another event of interest on the same subject.
   (a) TRUE
   (b) FALSE

F: Only one competing risk event can occur at a given time.

10. An example of competing risks survival data is a study in which patients receiving radiotherapy for head and neck cancer may either die from their cancer or from some other cause of death.
    (a) TRUE
    (b) FALSE

11. If all competing risks in a given study are different causes of death, then it is possible to have both competing risks and recurrent events in the same study.
    (a) TRUE
    (b) FALSE

    F: You can die only once.

12. Suppose patients with advanced-stage cancer may die after surgery before their hospital stay is long enough to get a hospital infection. Then such deaths from surgery reduce the hospital's burden of infection control.
    (a) TRUE
    (b) FALSE

PART B

1. (2 points) Obtain the empirical distribution function and the Nelson-Aalen estimate of the distribution function for the time to surrender using Data Set Dl. Assume that the surrender time is known for those who die.

Table 11.4 Data Set D1.

Policyholder	Time of death	Time of surrender
1	–	0.1
2	4.8	0.5
3	–	0.8
4	0.8	3.9
5	3.1	1.8
6	–	1.8
7	–	2.1
8	–	2.5
9	–	2.8
10	2.9	4.6
11	2.9	4.6
12	–	3.9
13	4.0	–
14	–	4.0
15	–	4.1
16	4.8	–
17	–	4.8
18	–	4.8
19–30	–	–

SOLUTION: EXERCISE 11.1, KPW

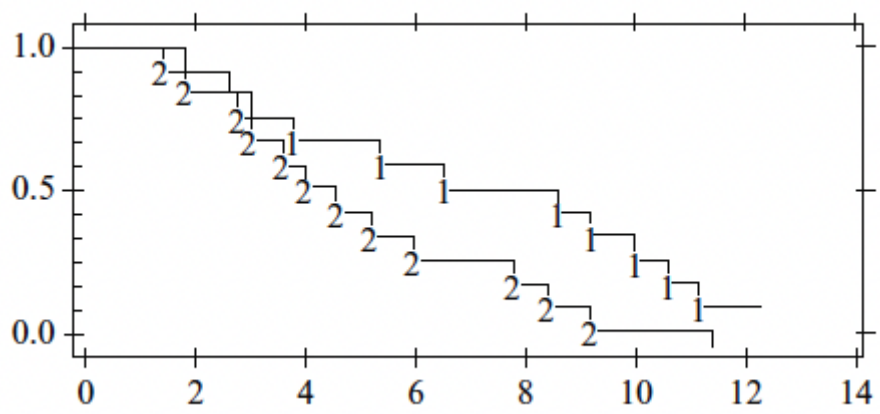
11.1 When all information is available, the calculations are in Table 11.1. As in Example 11.4, values apply from the current y-value to the next one.

Table 11.1 Calculations for Exercise 11.1.

j	y_j	s_j	r_j	$F_{30}(x)$	$\hat{H}(x)$	$\hat{F}(x)^*$
1	0.1	1	30	$1-29/30=0.0333$	$1/30=0.0333$	0.0328
2	0.5	1	29	$1-28/30=0.0667$	$0.0333+1/29=0.0678$	0.0656
3	0.8	1	28	$1-27/30=0.1000$	$0.0678+1/28=0.1035$	0.0983
4	1.8	2	27	$1-25/30=0.1667$	$0.1035+2/27=0.1776$	0.1627
5	2.1	1	25	$1-24/30=0.2000$	$0.1776+1/25=0.2176$	0.1956
6	2.5	1	24	$1-23/30=0.2333$	$0.2176+1/24=0.2593$	0.2284
7	2.8	1	23	$1-22/30=0.2667$	$0.2593+1/23=0.3027$	0.2612
8	3.9	2	22	$1-20/30=0.3333$	$0.3027+2/22=0.3937$	0.3254
9	4.0	1	20	$1-19/30=0.3667$	$0.3937+1/20=0.4437$	0.3583
10	4.1	1	19	$1-18/30=0.4000$	$0.4437+1/19=0.4963$	0.3912
11	4.6	2	18	$1-16/30=0.4667$	$0.4963+2/18=0.6074$	0.4552
12	4.8	2	16	$1-14/30=0.5333$	$0.6074+2/16=0.7324$	0.5192
13	5.0	14	14	$1-0/30=1.0000$	$0.7324+14/14=1.7324$	0.8231

$^*\hat{F}(x)=1-e^{-\hat{H}(x)}$.

2. (2 points) The KM curves for CHR data for groups 1 and 2 are plotted on the same graph. Comment on how these curves compare with each other.



SOLUTION: CHAPTER 2, KK; EXERCISE 1b

Any 2 of the 3 comments

- Group 1 appears to have consistently better survival prognosis than group 2.
- However, the KM curves are very close during the first 4 years, but are quite separate after 4 years,
- although they appear to come close again around 12 years.

3. (2 points) Consider the observations 2,500, 2,500, 2,500, 3,617, 3,662, 4,517, 5,000, 5,000, 6,010, 6,932, 7,500, and 7,500. No truncation is possible. First, determine the Nelson-Aalen estimate of the cumulative hazard rate function at 7,000 assuming all the observations are uncensored. Second, determine the same estimate, assuming the observations at 2,500, 5,000, and 7,500 were right censored.

SOLUTION: EXERCISE 12.11, KPW

12.11 With no censoring, the r-values are 12, 9, 8, 7, 6, 4, and 3, and the s values are 3, 1, 1, 1, 2, 1, and 1. Then

$$\hat{H}(7{,}000) = \tfrac{3}{12} + \tfrac{1}{9} + \tfrac{1}{8} + \tfrac{1}{7} + \tfrac{2}{6} + \tfrac{1}{4} + \tfrac{1}{3} = 1.5456.$$

With censoring, there are only five uncensored values with r values 9, 8, 7, 4, and 3, and all five s values are 1. Then

$$\hat{H}(7{,}000) = \tfrac{1}{9} + \tfrac{1}{8} + \tfrac{1}{7} + \tfrac{1}{4} + \tfrac{1}{3} = 0.9623.$$

4. (4 points) The following edited printout gives computer results for fitting a Cox PH model containing the three predictives Rx, log WBC, and Sex.

Cox regression Analysis time_t: survt	Coef.	Std. Err.	p > |z|	Haz. Ratio	[95% Conf. Interval]		P(PH)
Sex	0.263	0.449	0.558	1.301	0.539	3.139	0.042
log WBC	1.594	0.330	0.000	4.922	2.578	9.397	0.714
Rx	1.391	0.457	0.002	4.018	1.642	9.834	0.500
No. of subjects = 42			Log likelihood = −72.109				

i. Which of the variables in the model fitted above are time-independent and which are time-dependent?
ii. Based on this printout, is the PH assumption satisfied for the model being fit? Explain briefly.
iii. Suppose you want to use an extended Cox model to assess the PH assumption for all three variables in the above model. State the general form of an extended Cox model that will allow for this assessment.
iv. Suppose you wish to assess the PH assumption for the Sex variable using a heaviside function approach designed to yield a constant hazard ratio for less than 15 weeks of follow-up and a constant

hazard ratio for 15 weeks or more of follow-up. State two equivalent alternative extended Cox models that will carry out this approach, one model containing one Heaviside function and the other model containing two Heaviside functions.

SOLUTION: CHAPTER 6, KK; Practice EXERCISE 1-4

1. All three variables in the model are time-independent variables.
2. The computer results indicate that the Sex variables do not satisfy the PH assumption because the $P(PH)$ value is 0.042, which is significant at the 0.05 level.
3. $$h(t, \mathbf{X}(t)) = h_0(t)\exp[\beta_1(\text{sex}) + \beta_2(\log\ \text{WBC}) + \beta_3(Rx) + \delta_1(\text{sex})g_1(t) + \delta_2(\log\ \text{WBC})g_2(t) + \delta_3(Rx)g_3(t)]$$
   where the $g_i(t)$ are functions of time.
4. Model 1 (one heaviside function)

$$h(t, \mathbf{X}(t)) = h_0(t)\exp[\beta_1(\text{sex}) + \beta_2(\log\ \text{WBC}) + \beta_3(Rx) + \delta_1(\text{sex})g_1(t)]$$

where

$$g_1(t) = \begin{cases} 1 \text{ if } 0 \le t < 15 \text{ weeks} \\ 0 \text{ if } t \ge 15 \text{ weeks} \end{cases}$$

Model 2 (two heaviside functions):

$$h(t, \mathbf{X}(t)) = h_0(t)\exp[\beta_2(\log\ \text{WBC}) + \beta_3(Rx) + \delta_1(\text{sex})g_1(t) + \delta_2(\text{sex})g_2(t)]$$

where

$$g_1(t) = \begin{cases} 1 \text{ if } 0 \le t < 15 \text{ weeks} \\ 0 \text{ if } t \ge 15 \text{ weeks} \end{cases}$$

and

$$g_2(t) = \begin{cases} 0 \text{ if } t \ge 15 \text{ weeks} \\ 1 \text{ if } 0 \le t < 15 \text{ weeks} \end{cases}$$

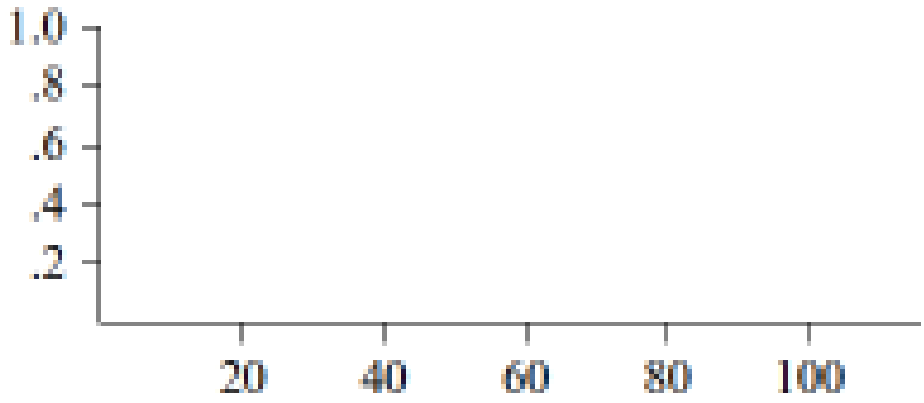
5. (4 points) Suppose that Allie (A), Sally (S), and Callie (C) are the only three subjects in the dataset shown below. All three subjects have two recurrent events that occur at different times.

ID	Status	Stratum	Start	Stop	tx
A	1	1	0	70	1
A	1	2	70	90	1
S	1	1	0	20	0
S	1	2	20	30	0
C	1	1	0	10	1
C	1	2	10	40	1

i. Fill in the following data layout describing survival (in weeks) to the first event (stratum 1). Recall that m_f and q_f denote the number of failures and censored observations at time $t_{(f)}$. The survival probabilities in the last column use the KM product limit formula.

$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)})$	$S_1(t_{(f)})$
0	3	0	0	[A, S, C]	1.00
10	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

ii. Plot the survival curve that corresponds to the data layout obtained in (i).



iii. Fill in the following data layout describing survival (in weeks) from the first to second event using the Gap Time approach:

$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)})$	$S_2(t_{(f)})$
0	3	0	0	[A, S, C]	1.00
10	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

iv. Fill in the following data layout describing survival (in weeks) to the second event using the Marginal approach:

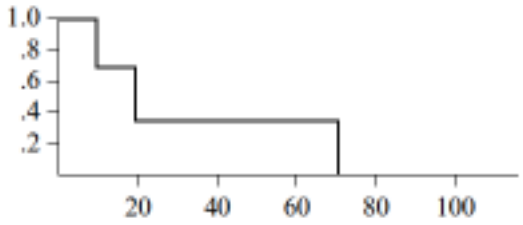
$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)})$	$S_2(t_{(f)})$
0	3	0	0	[A, S, C]	1.00
30	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

SOLUTION: CHAPTER 8, KK; Practice EXERCISE 16-18, 20

16.

$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)}]$	$S_1(t_{(f)})$
0	3	0	0	[A, S, C]	1.00
10	3	1	0	[A, S, C]	0.67
20	2	1	0	[A, S]	0.33
70	1	1	0	[A]	0.00

17. $\mathbf{S_1}(t)$



18.

$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)}]$	$S_2(t_{(f)})$ Gap Time
0	3	0	0	[A, S, C]	1.00
10	3	1	0	[A, S, C]	0.67
20	2	1	0	[A, C]	0.33
30	1	1	0	[C]	0.00

20.

$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)}]$	$S_2(t_{(f)})$ Marginal
0	3	0	0	[A, S, C]	1.00
30	3	1	0	[A, S, C]	0.67
40	2	1	0	[A, C]	0.33
90	1	1	0	[A]	0.00

6. (4 points) Thirteen different distributions were fit to a data The results of that process revealed six models with p-values above 0.01 for the chi-square goodness-of-fit test. Information about those models is given below:

Model	Number of parameters	Negative loglikelihood	χ^2	p-value
Negative binomial	2	5,348.04	8.77	0.0125
ZM logarithmic	2	5,343.79	4.92	0.1779
Poisson–inverse Gaussian	2	5,343.51	4.54	0.2091
ZM negative binomial	3	5,343.62	4.65	0.0979
Geometric–negative binomial	3	5,342.70	1.96	0.3754
Poisson–ETNB	3	5,342.51	2.75	0.2525

i. What is the best model and why?
ii. Consider the three-parameter model (Poisson-ETNB) and the two-parameter model (Poisson-inverse Gaussian), which is better and why?

SOLUTION: EXAMPLE 16.11, KPW

i. Based on the negative log-likelihood, chi-square test statistic, and p-value, Poison-ETNB is the best. However, If we perform a likelihood ratio test to compare the models, The likelihood ratio test indicates that the three-parameter model with the smallest negative loglikelihood (Poisson-ETNB) is not significantly better than the two-parameter Poisson-inverse Gaussian model. Hence, the Poisson-inverse Gaussian model would be the best and an excellent choice.

You can also compare their AIC using
AIC=2×(Negative loglikelihood)+2×(Number of parameters)

ii. The likelihood ratio test indicates that the three-parameter model with the smallest negative loglikelihood (Poisson-ETNB) is not significantly better than the two-parameter Poisson-inverse Gaussian model. The latter appears to be an excellent choice. Also, we choose Poisson-inverse Gaussian because the log-likelihood improvement (1.0) in Poisson-ETNB is small, not worth the extra parameter — simpler model is preferable when fit is nearly as good.