Dept of Mathematics and Statistics
King Fahd University of Petroleum & Minerals
AS476: Survival Models for Actuaries
Dr. Ridwan A. Sanusi
Major 2 Exam Term 231 FORM A
Thursday October 26 2023
6.30pm-8.30pm

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	2		
2	5		
3	6+3+3+3=15		
4	2+4+4+3=13		
5	2+3=5		
Total	40		

Extra blank page

1. (2 marks) A Stratified Cox Model that allows for interaction is given below. If the Likelihood ratio statistic for testing the significance of the interaction assumption follows a Chi-square. What is the degree of freedom of the Chi-square?

$$h_g(t,\mathbf{X}) = h_{0g}(t)\exp[\beta_{1g}X_1 + \beta_{2g}X_2 + \ldots + \beta_{pg}X_p]$$
$g = 1, 2, \ldots, k^*$, strata defined from Z^*

A) K*-1
B) P-1
C) P(K*-1)
D) K*(p-1)
E) PK* - 1

2. (5 marks) Proof that the general formular for a Kaplan-Meier survival probability at failure time t(f) is

$$\hat{S}(t_{(f)}) = \hat{S}(t_{(f-1)}) \times \hat{P}r(T > t_{(f)} | T \geq t_{(f)})$$

Hint: P(A.B) = P(B) . P(A|B).

Answer:

$$\hat{S}(t_{(f)}) = \prod_{i=1}^{f} \hat{P}r[T > t_{(i)} | T \geq t_{(i)}] = \hat{S}(t_{(f-1)}) \times \hat{P}r(T > t_{(f)} | T \geq t_{(f)})$$

Math proof:

Pr(A and B) = Pr(A) × Pr(B | A) always

A = "$T \geq t_{(f)}$" → A and B = B
B = "$T > t_{(f)}$"
Pr(A and B) = Pr(B) = $S(t_{(f)})$

No failures during $t_{(f-1)} < T < t_{(f)}$
$\Pr(A) = \Pr(T > t_{(f-1)}) = S(t_{(f-1)})$

$\Pr(B|A) = \Pr(T > t_{(f)} | T \geq t_{(f)})$

Thus, from Pr(A and B) formula,

Pr(A and B) = Pr(A) × Pr(B | A)
$$S(t_{(f)}) = S(t_{(f-1)}) \times \Pr(T > t_{(f)} | T \geq t_{(f)})$$

3. (6+3+3+3=15 points) Consider the failure data below and the hypothesis of equivalent survival curves,

		# failures		# in risk set		# expected		Observed - expected	
j	$t_{(j)}$	$m_{(1j)}$	$m_{(2j)}$	$n_{(1j)}$	$n_{(2j)}$	$e_{(1j)}$	$e_{(2j)}$	$m_{(1j)} - e_{(1j)}$	$m_{(2j)} - e_{(2j)}$
1	1	0	2	21	21	1	1	−1.00	1.00
2	2	0	2	21	19	1		−1.05	1.05
3	3	0	1	21	17	21/38	17/38		0.55
4	4	0	2	21	16	42/37	32/37	−1.14	1.14
5	5	0	2	21	14	42/35		−1.20	
6	6	3	0	21	12	21/11	12/11	1.09	−1.09
7	7	1	0	17	12	17/29	12/29	0.41	−0.41
8	8	0	4	16	12	16/7	12/7	−2.29	2.29
9	10	1	0	15	8		8/23		−0.35
10	11	0	2	13	8	13/7	16/21	−1.24	1.24
11	12	0	2	12	6	12/9		−1.33	
12	13	1	0	12	4	3/4	1/4	0.25	−0.25
13	15	0	1	11	4	11/15	4/15	−0.73	0.73
14	16	1	0	11	3	11/14	3/14	0.21	−0.21
15	17	0	1	10	3	10/13	3/13	−0.77	0.77
16	22	1	1	7	2	14/9	4/9	−0.56	0.56
17	23	1	1	6	1	12/7	2/7	−0.71	0.71
Totals		9	21			19.26	10.74		

(a) complete the blank spaces (provide sample calculation below)
(b) calculate the approximate chi-square statistics.
(c) calculate the log rank statistic. (Hint: $\widehat{Var}(O_2 - E_2)$ = 6:2570)
(d) decide at α = 0.10, if the two groups have equal survival curves (Hint: ChiSquareInv(0:90; 1) = 2.7055)

Solution

a)

		#failures		#in risk set		# expected		Observed - Expected	
j	$t_{(j)}$	$m_{(1j)}$	$m_{(2j)}$	$n_{(1j)}$	$n_{(2j)}$	$e_{(1j)}$	$e_{(2j)}$	$m_{(1j)} - e_{(1j)}$	$m_{(2j)} - e_{(2j)}$
1	1	0	2	21	21	$(21/42) \times 2 = 1$	$(21/42)2 = 1$	−1.00	1.00
2	2	0	2	21	19	$(21/40) \times 2 = 1$	$\mathbf{(19/40)2 = 19/20}$	−1.05	1.05
3	3	0	1	21	17	$(21/38) \times 1 = 21/38$	$(17/38)1 = 17/38$	**−0.55**	0.55
4	4	0	2	21	16	$(21/37) \times 2 = 42/37$	$(16/37)2 = 32/37$	−1.14	1.14
5	5	0	2	21	14	$(21/35)2 = 6/5$	$\mathbf{(14/35)2 = 4/5}$	−1.20	**1.20**
6	6	3	0	21	12	$(21/33)3 = 21/11$	$(12/33)3 = 12/11$	1.09	−1.09
7	7	1	0	17	12	$(17/29)1 = 17/29$	$(12/29)1 = 12/29$	0.41	−0.41
8	8	0	4	16	12	$(16/28)4 = 16/7$	$(12/28)4 = 12/7$	−2.29	2.29
9	10	1	0	15	8	$\mathbf{(15/23)1 = 15/23}$	$(8/23)1 = 8/23$	**0.35**	−0.35
10	11	0	2	13	8	$(13/21)2 = 26/21$	$(8/21)2 = 16/21$	−1.24	1.24
11	12	0	2	12	6	$(12/18)2 = 12/9$	$\mathbf{(6/18)2 = 2/3}$	−1.33	**1.33**
12	13	1	0	12	4	$(12/16)1 = 3/4$	$(4/16)1 = 1/4$	0.25	−0.25
13	15	0	1	11	4	$(11/15)1 = 11/15$	$(4/15)1 = 4/15$	−0.73	0.73
14	16	1	0	11	3	$(11/14)1 = 11/14$	$(3/14)1 = 3/14$	0.21	−0.21
15	17	0	1	10	3	$(10/13)1 = 10/13$	$(3/13)1 = 3/13$	−0.77	0.77
16	22	1	1	7	2	$(7/9)2 = 14/9$	$(2/9)2 = 4/9$	−0.56	0.56
17	23	1	1	6	1	$(6/7)2 = 12/7$	$(1/7)2 = 2/7$	−0.71	0.71
$\sum$		9	21			19.26	10.74	**−10.26**	**10.26**

b) χ^2 statistics $= \sum_{i=1}^{G} \frac{(O_i - E_i)^2}{E_i} = \frac{(-10.26)^2}{19.26} + \frac{(10.26)^2}{10.74} = 15.267060$

c)		# failures		# in risk set		$Var(O_2 - E_2)$ calculation	
j	$t_{(j)}$	$m_{(1j)}$	$m_{(2j)}$	$n_{(1j)}$	$n_{(2j)}$	$\frac{n_{(1j)}n_{(2j)}(m_{(1j)}+m_{(2j)})(n_{(1j)}+n_{(2j)}-m_{(1j)}-m_{(2j)})}{(n_{(1j)}+n_{(2j)})^2(n_{(1j)}+n_{(2j)}-1)}$	
1	1	0	2	21	21	$\frac{21(21)(0+2)(21+21-0-2)}{(21+21)^2(21+21-1)} =$	0.487805
2	2	0	2	21	19	$\frac{21(19)(0+2)(21+19-0-2)}{(21+19)^2(21+19-1)} =$	0.485962
3	3	0	1	21	17	$\frac{21(17)(0+1)(21+17-0-1)}{(21+17)^2(21+19-1)} =$	0.247230
4	4	0	2	21	16	$\frac{21(16)(0+2)(21+16-0-2)}{(21+16)^2(21+16-1)} =$	0.477234
5	5	0	2	21	14	$\frac{21(14)(0+2)(21+14-0-2)}{(21+14)^2(21+14-1)} =$	0.465882
6	6	3	0	21	12	$\frac{21(12)(3+0)(21+12-3-0)}{(21+12)^2(21+12-1)} =$	0.650826
7	7	1	0	17	12	$\frac{17(12)(1+0)(17+12-1-0)}{(17+12)^2(17+12-1)} =$	0.242568
8	8	0	4	16	12	$\frac{16(12)(0+4)(16+12-0-4)}{(16+12)^2(16+12-1)} =$	0.870748
9	10	1	0	15	8	$\frac{15(8)(1+0)(15+8-1-0)}{(15+8)^2(15+8-1)} =$	0.226843
10	11	0	2	13	8	$\frac{13(8)(0+2)(13+8-0-2)}{(13+8)^2(13+8-1)} =$	0.448073
11	12	0	2	12	6	$\frac{12(6)(0+2)(12+6-0-2)}{(12+6)^2(12+6-1)} =$	0.418301
12	13	1	0	12	4	$\frac{12(4)(1+0)(12+4-1-0)}{(12+4)^2(12+4-1)} =$	0.187500
13	15	0	1	11	4	$\frac{11(4)(0+1)(11+4-0-1)}{(11+4)^2(11+4-1)} =$	0.195556
14	16	1	0	11	3	$\frac{11(3)(1+0)(11+3-1-0)}{(11+3)^2(11+3-1)} =$	0.168367
15	17	0	1	10	3	$\frac{10(3)(0+1)(10+3-0-1)}{(10+3)^2(10+3-1)} =$	0.177515
16	22	1	1	7	2	$\frac{7(2)(1+1)(7+2-1-1)}{(7+2)^2(7+2-1)} =$	0.302469
17	23	1	1	6	1	$\frac{6(1)(1+1)(6+1-1-1)}{(6+1)^2(6+1-1)} =$	0.204082
Totals		9	21				6.2570

$\widehat{Var}(O_2 - E_2) = \sum_{j=1}^{17} \frac{n_{(1j)}n_{(2j)}(m_{(1j)}+m_{(2j)})(n_{(1j)}+n_{(2j)}-m_{(1j)}-m_{(2j)})}{(n_{(1j)}+n_{(2j)})^2(n_{(1j)}+n_{(2j)}-1)} = 6.2570$ (see calculation above)

logrank statistics $= \frac{(O_2 - E_2)^2}{\widehat{Var}(O_2 - E_2)} = \frac{(10.26)^2}{6.2570} = 16.823\,973.$

d) $\chi^2_{1,1-\alpha} = \text{ChiSquareInv}(0.90; 1) = 2.7055$. Both the logrank and the chi-square statistics are compared with this critical value. The test statistics are in the critical region, so we **reject** the null hypothesis of equal survival curves at the significance level $\alpha = 0.10$

4. (2+4+4+3 = 13 marks) For the claims on general insurance data (Gray and Pitts, 2012), there are n = 140 claims for each of four different possible types of policies as follows:

Variable #		Variable name	Coding
1		Claim amounts	loss are nonnegative
2	indicator	*C60*	60% $coinsurance = 1,\ other = 0$
3	variables	*Deductible*	$franchise\ Deductible$ (\$250) $= 1,\ other = 0$
4	for	*Ground up*	$Ground\ up = 1,\ other = 0$
5	*policy* type	*Limit*	$Limit$ (\$10000) $= 1,\ other = 0$
6		*Payment status*	Paid at loss amount= 1, censored at policy limit= 0

For this actuarial survival model, claim amounts is treated as the survival variable and the *payment* status of 1 is regarded as the **event**.

The following is an edited *R* survival analysis results on the claim amounts.

n = 544, missing = 16, number of events = 536

Variable	Coef.	Std.Err.	z	$p > \|z\|$	exp(coef)	exp(-coef)	95% Conf. Int	
Deductible	−0.4918	0.1241	−3.962	7.43×10^{-5} ***	0.6115	1.635	0.4795	0.7800
Ground up	−0.3768	0.1203	−3.132	0.00174 **	0.6861	1.458	0.5420	0.8685
Limit	−0.3645	0.1218	2.992	0.00277 **	0.6945	1.440	0.5470	0.8818
	Log likelihood = −2855.656							

Signif. codes: 0 = "***", 0.001 = "**", "*" = 0.01, "." = 0.05, " · " = 0.1

Likelihood ratio test= 17.31 on 3 df

Score (logrank) test = 18.55 on 3 df, p=0.0003394

Based on the output, answer the following:

a) Write the cox PH model represented by the output.

$$\begin{aligned} h(t, \mathbf{X}) &= h_0(t) \exp\left(\widehat{\beta}_1 Deductible + \widehat{\beta}_2 \triangleright Groundup \quad + \widehat{\beta}_3 Limit\right) \\ &= h_0(t) \exp\left(-0.4918 Deductible - 0.3768 \triangleright Groundup \quad - 0.3645 Limit\right) \end{aligned}$$

b) State the null hypothesis and Carry out a test for the overall significance of the model?

b) H_0: all $\beta_i = 0$.
LRT=17.31 compared with $\chi^2_{3,0.05} = 7.8147$. Reject H_0 at $\alpha = 0.05$. Conclude that the model with three predictors (deductible, ground up, and limit) is significant.

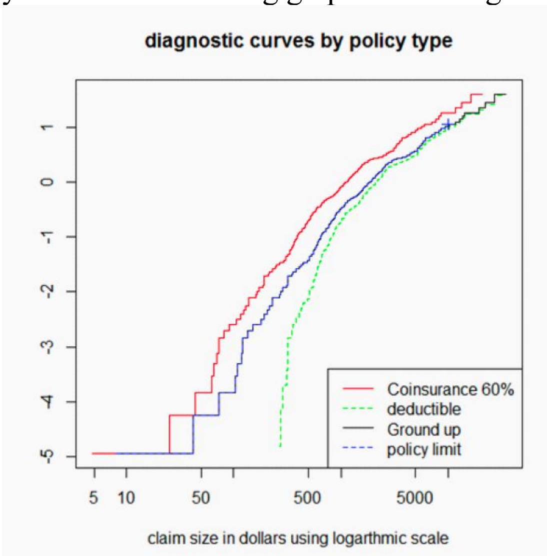
c) Test for the variable "*Deductible*" What is your conclusion at the 0.05 significance level?

c) H_0: all $\beta_1 = 0$.
Wald's $z = -3.962$ compared with $-z_{0.05} = -1.96$. Reject H_0 at $\alpha = 0.05$. Conclude that the model with deductible is significant.

d) Obtain the Hazard Ratio for "*Deductible*". And *Interpret* this Hazard Ratio.

d) Hazard Ratio for "*Deductible*"$= e^{\beta_1} = 0.6115$. This indicates that "deductible" has about 61% lower hazards than the default policy group, *C60* (60% coinsurance).

5. (2+3=5 marks) For the general insurance claim data by Gray and Pitts (2012) described in question 3,
you did the following graph to investigate the assumptions of the Cox PH model.



In addition, you conducted a summary correlation analysis between the schoenfeld residuals and the survival variable (claim amounts) for each predictor as follows:

	rho	chisq	p(chisq)
Deductible	0.1234	7.95	0.0048
Ground up	0.0552	1.62	0.2038
Limit	0.0604	1.93	0.1646
GLOBAL	NA	7.98	0.0464

a) What should be the likely label for the y-axis in the above graph? And what is the graph indicating?
b) Using the information provided, what can you conclude regarding the PH assumption for the variables used in the model? Explain briefly.

Solution

a) -log(-log(S(x))
The lines are not parallel. This may indicate more analysis is needed to see if the PH assumption is not met.
b) All other variables are meeting the *PH assumption* except for "*Deductible*" with p-value(*PH*)$= 0.0048 < 0.01$.