Dept of Mathematics and Statistics
King Fahd University of Petroleum & Minerals

AS475: Survival Models for Actuaries
Dr. Mohammad H. Omar
Major 2 Exam Term 152 FORM A
Monday April 4 2016
6.30pm-7.50pm

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 80 minutes, GOOD LUCK, and you may begin now!

Question	Total Marks	Marks Obtained	Comments
1	6+4=10		
2	2+3+3+3=11		
3	3+2=5		
4	2+6+3=11		
5	2+3+3=8		
6	1+4=5		
Total	50		

Extra blank page

1. (6+4=10 points) In a survival study on life insurance policyholders,
   - People enter the study when they purchase insurance
   - Leave the study when they surrender the insurance
   - Death (event) occuring exactly at integer age is counted in the previous age band
   - The study began on 1/1/2011 and ended on 12/31/2013

   You observe the history of the 5 following people:

Date of Birth (DOB)	**Purchase Date**	**Date of Death (DOD)**	**Surrender Date**
6/1/1946	7/1/2010	–	11/1/2012
3/1/1947	9/1/2011	3/1/2012	–
4/1/1947	6/1/2011	5/1/2013	–
1/1/1948	6/1/2012	4/1/2013	–
7/1/1948	3/1/2011	–	–

a) Find the **exact exposure estimates** of $\hat{q}_{65}$ and $\hat{q}_{66}$.
b) Find the **actuarial exposure estimates** of $\hat{q}_{65}$ and $\hat{q}_{66}$.

2. (2+3+3+3=11 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.

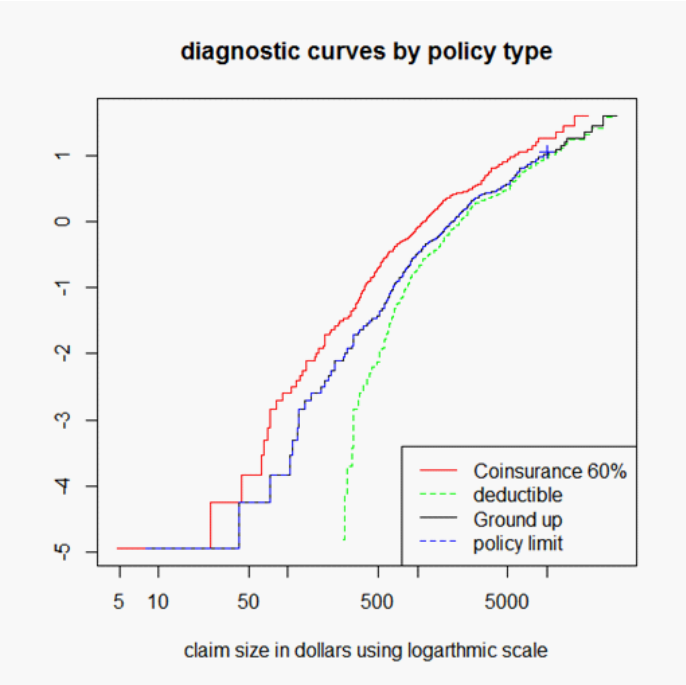
b) State the *null hypothesis* and Carry out a test for the *overall significance* of the model?

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

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

.

3. (3+2=5marks) For the general insurance claim data by Gray and Pitts (2012) described in question 2, 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.

4. (2+6+3=11 points) The dataset "vets.dat" considers survival times in days for 137 patients from the Veteran's Administration Lung Cancer Trial (cited in Kalbfleisch and Prentice, 2002). The exposure variable of interest is treatment status. Other variables of interest as control variables are cell type (four types, defined by dummy variables), performance status, and prior therapy status. Failure status defined by the status variable (0 if censored, 1 if died). A complete list of the variables is given below.

1: **Treatment** (standard = 1, test = 2)	6: Survival **time** (days)
2: Cell type **1** (**large** = 1, other = 0)	7: **Performance status** (0 = worst, ..., 100 = best).
3: Cell type **2** (**adeno** = 1, other = 0)	8: Prior **therapy** (none = 0, some = 10)
4: Cell type **3** (**small** = 1, other = 0)	9: **Status** (0 = censored, 1 = died)
5: Cell type **4** (**squamous** = 1, other = 0)	

Based on your earlier analysis, you have decided to stratify on the "*Prior therapy*" variable as given by the edited R output below:

$n = 137$, number of events = 128

	Coef.	**Std.Err.**	z	$p > \|z\|$	**exp(coef)**	**exp(-coef)**	**95% Conf. Int**	
Treatment		0.201358	1.356	0.175034	1.314009	0.7610		1.9498
Large cell	0.360776	0.282975	1.275	0.202330	1.434442	0.6971	0.8238	2.4978
Adeno cell	1.111310	0.296788		0.000181 ***	3.038337	0.3291		
Small cell	0.801184	0.269352	2.974	0.002935 **		0.4488	1.3142	3.7777
Perf status	-0.030671	0.005248	-5.844	5.09×10^{-9} ***	0.969795	1.0311	0.9599	0.9798
			Loglikelihood= −402.9143					

Stratified by "*Prior therapy*" Signif. codes: 0 = "***", 0.001 = "**", "*" = 0.01, "." = 0.05, " · " = 0.1

Rsquare= 0.342 (max possible= 0.998)

Likelihood ratio test= 57.34 on 5 df, $p = 4.31 \times 10^{-11}$

Wald test = 58.46 on 5 df, $p = 2.528 \times 10^{-11}$

Score (logrank) test = 61.61 on 5 df, $p = 5.635 \times 10^{-12}$

Answer the following questions.

a) What reason from the previous analysis would have driven you to do the above analysis by stratifying on "*Prior therapy*"?

b) Complete the missing blank cells above. Be sure to show your work

c) Write the stratified Cox model and write this model for

(i) No Prior **therapy** (None) and

(ii) Some Prior therapy (some)

5. (2+3+3=8 points) The dataset "vets.dat" in question (4) above is analyzed again. But this time, you decided to introduce the variable $\ln t \times$ps where ps="performance status". Below, you obtained edited R outputs for different Cox models that were fit to the dataset. A number of questions will be asked about these results.

	Variable	**Coef.**	**Std.Err.**	z	$p > \|z\|$	**exp(coef)**	**exp(-coef)**	**95% Conf. Int**	
1	Treatment	0.257313	0.200629	1.283	0.1997	1.293450	0.7731	0.8729	1.9166
	Large cell	0.392959	0.282233	1.392	0.1638	1.481358	0.6751	0.8520	2.5757
	Adeno cell	1.147673	0.294932	3.891	9.97×10^{-5} ***	3.150853	0.3174	1.7676	5.6166
	Small cell	0.819614	0.268809	3.049	0.0023 **	2.269624	0.4406	1.3401	3.8438
	Perf status	-0.031112	0.005167	-6.022	1.73×10^{-9} ***	0.969367	1.0316	0.9596	0.9792
		Log likelihood = -475.676							
2	Treatment	0.117301	0.165974	0.707	0.47972	1.124458	0.8893	0.8122	1.5568
	Large cell	0.358490	0.245124	1.462	0.14361	1.431167	0.6987	0.8852	2.3139
	Adeno cell	1.132089	0.281861	4.016	5.91×10^{-5} ***	3.102131	0.3224	1.7854	5.3899
	Small cell	0.888314	0.273413	3.249	0.00116 **	2.431026	0.4113	1.4225	4.1545
	Perf status	-0.080273	0.019188	-4.184	2.87×10^{-5} ***	0.922864	1.0836	0.8888	0.9582
	$\ln t \times$ps	0.013377	0.004660	2.870	0.00410 **	1.013467	0.9867	1.0043	1.0228
		Log likelihood = -470.4989							

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

a) For model 2, give an expression for the **hazard ratio** for the effect of the **treatment** variable adjusted for all **other predictors**.
b) Using model 2, compute the **estimated hazard ratio** for patients with **large cells** and **performance status** of **40** at **100** days

Also compute the **hazard expression** for patients with **large cells** and **performance status** of **50** at **120** days

c) Carry out an appropriate test of **hypothesis** to evaluate whether there is any significant **interaction** in model 2. What is your conclusion?

6. (4+1=5 points) You are given:

(i) A Cox proportional hazards (PH) model was used to study claim amounts on two groups of general insurance policies.

(ii) A single predictor variable z was used with $z = 0$ for a policy in Group A and $z = 1$ for a policy in Group B.

(iii) A sample of three policies was taken from each group. The losses were:

Group A:	75	125	320
Group B:	15	50	100

(iv) The baseline survival function obtained as $S_0(x) = \left(\frac{200}{200+x}\right)^{\alpha}, \quad x > 0, \quad \alpha > 0.$

Calculate the **maximum likelihood estimate** of the Cox PH model coefficient β.

(A) -0.92
(B) -0.40
(C) 0.40
(D) 0.92
(E) 2.51

Work Shown (4 points)

(Hint: $L = \prod_{j=1} f(x_j)$)

So Answer is ___

END OF TEST PAPER