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

AS475: Survival Models for Actuaries
Dr. Mohammad H. Omar
Major 2 Exam Term 142 FORM A
TSunday April 5 2015
7.00pm-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 time will be provided for the time missed outside the classroom.
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 or financail calculators only. Write important steps to arrive at the solution of the following problems.

The test is 90 minutes, GOOD LUCK, and you may begin now!

Question	Total Marks	Marks Obtained	Comments
1	4+6=10		
2	5		
3	2+4+6+2=14		
4	3+3=6		
5	1+4=5		
Total	40		

Extra blank page

1. (4+6=10 points) From a sample, you are given that the mean is 35000, the standard deviation is 75000, the median is 10000, and the 90th percentile is 100000. You are using the Weibull distribution to model this data. The weibull distribution has the following properties:

$$\begin{aligned} f(x) &= \frac{(x/\theta)^{\alpha-1}}{\theta\Gamma(\alpha)}e^{-x/\theta}, \qquad \alpha>0,\ \theta>0,\ x>0 \text{ with} \\ E[X^k] &= \theta^k\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}, \qquad k>-\alpha. \end{aligned}$$

a) Using the percentile matching method, estimate the parameters of a Weibull distribution.
b) using the method of moments, estimate the parameters of the Weibull distribution.

2. (5 points) . Ten lives are subject to the survival function

$$S(t) = \left(1 - \frac{t}{\beta}\right)^{1/2}, \quad 0 \le t \le \beta,$$

where t is time since birth. There are 10 lives observed from birth. At time 10, 2 of the lives die and the other 8 are withdrawn from observation. Determine the maximum likelihood estimate of β.

3. (2+4+6+2=14 points) The data for this question contain survival times of 65 multiple myeloma patients (Krall et al., 1975). A partial list of the variables in the dataset is given below:

1:	observation number	2:	survival **time** (in months) from time of diagnosis
3:	survival **status** (0 = alive, 1 = dead)	4:	**platelets** at diagnosis (0 = abnormal, 1 = normal)
5:	**age** at diagnosis (years)	6:	**sex** (1 = male, 2 = female)

Below, we provide edited computer results for several different Cox models that were fit to this dataset. A number of questions will be asked about these results.

MODEL	**Variable**	**Coef.**	**Std.Err.**	$p > \|z\|$	**Haz.Ratio**	**[95% Conf.**	**Interval]**
1	Platelets	0.470	2.854	0.869	1.600	0.006	429.689
	Age	0.000	0.037	0.998	1.000	0.930	1.075
	Sex	0.183	0.725	0.801	1.200	0.290	4.969
	Platelets × age	-0.008	0.041	0.850	0.992	0.915	1.075
	Platelets × sex	-0.503	0.804	0.532	0.605	0.125	2.924
		Log likelihood = -153.040					
2	Platelets	-0.725	0.401	0.071	0.484	0.221	1.063
	Age	-0.005	0.016	0.740	0.995	0.965	1.026
	Sex	-0.221	0.311	0.478	0.802	0.436	1.476
		Log likelihood = -153.253					
3	Platelets	-0.706	0.401	0.078	0.493	0.225	1.083
	Age	-0.003	0.015	0.828	0.997	0.967	1.027
		Log likelihood = -153.509					
4	Platelets	-0.705	0.397	0.076	0.494	0.227	1.075
	Sex	-0.204	0.307	0.506	0.815	0.447	1.489
		Log likelihood = -153.308					
5	Platelets	-0.694	0.397	0.080	0.500	0.230	1.088
		Log likelihood = -153.533					

a) For model 1, give an expression for the **hazard ratio** for the effect of the **platelet** variable adjusted for **age** and **sex**.
b) Using your answer in part a for model 1, compute the **estimated hazard ratio** for **40** year old **male**.

Also compute the estimated hazard ratio for a **50** year old **female**.

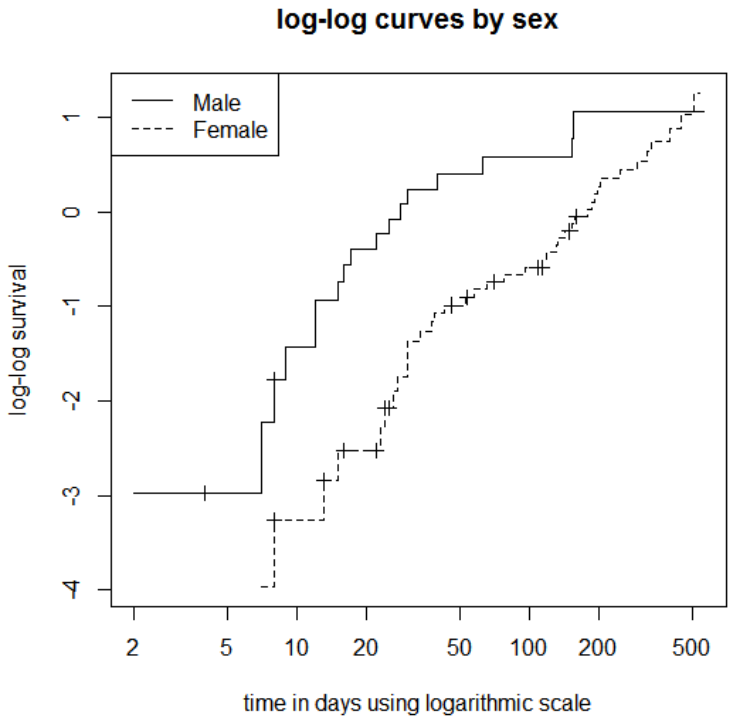
c) Carry out an appropriate test of **hypothesis** to evaluate whether there is any significant **interaction** in model 1. What is your conclusion?
d) Which of the five models, if any, do you think is the **best model** and why?

.

4. (3+3=6 points) The dataset "kidney.dat" (McGilchrist and Aisbett, 1991) considers survival times (recurrence times to infection) in days for 76 kidney patients using portable dialysis equipment. Catheters may be removed for reasons other than infection, in which case the observation is censored. Each patient has exactly 2 observations. The exposure variable of interest are age, sex, and disease type (four types, defined by dummy variables). Failure status defined by the status variable (0 if censored, 1 if died). A complete list of the variables is given below.

1:	**ID**	4:	Age (in years)
2:	Survival **time**	5:	sex (1=male, 2=female)
3:	Event Status (0 = infection, 1 = censored)	6:	disease type (0=GN, 1=AN, 2=PKD, 3=Other)

The following printout is obtained from fitting a log-log of KM curves and Cox PH model to these data.



Cox regression	**Coef.**	**Std.Err.**	$p > \|z\|$	**Haz.Ratio**	**[95% Conf. Interval]**		$P(PH)$
Age	0.00318	0.0111	0.78	1.003	0.98160	1.02525	0.7933
Sex	−1.48314	0.3582	0.000035	0.227	0.11245	0.45792	0.0822
DiseaseGN	0.08796	0.4064	0.83	1.092	0.49234	2.42178	0.8423
DiseaseAN.	0.35079	0.3997	0.8	1.420	0.64881	3.10869	0.7829
DiseasePKD	−1.43111	0.6311	0.023	0.239	0.06939	0.82353	0.8051

Likelihood ratio test=17.6 on 5 df, p =0.00342.

a) Using the information provided, what can you conclude about whether the **PH assumption** is satisfied for the variables used in the model? Explain briefly.

b) Using a significance level of 0.05, what are your **recommendation**s regarding the following variables?

i) Age

ii) Sex

iii) Disease type.

.

5. (4+1=5 points) You are given the following sample of claim sizes:

1500 3500 1800 4800 3900 6000 3800 5500 4200 3000

The underlying distribution is assumed to be gamma with $\alpha = 12$ and θ unknown.
What is the **variance** of the maximum likelihood estimator $\widehat{\theta}$?

a) $\frac{\theta^2}{1200}$ b) $\frac{1200}{\theta^2}$ c) $\frac{\theta^2}{120}$ d) $\frac{120}{\theta^2}$ e) $\frac{38000}{12}$

Work Shown (4 points)

Answer is _____.

END OF TEST PAPER