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

AS475: Survival Models for Actuaries

Dr. Mohammad H. Omar
Final Exam   Term 152   FORM B
Tuesday May 18 2016
7.00pm-9.30pm

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

Question	Total Marks	Marks Obtained	Comments
1	$3+5+3+4=15$		
2	$4+4=8$		
3	$2+2+6=10$		
4	$3+4+3=10$		
5	$4+6=10$		
6	$6+6=12$		
7	$4+2+2+2=10$		
8	5		
9	$1+4=5$		
10	$1+4=5$		
Total	90		

Extra blank page

1. (3+5+3+4=15 points) Using data from the *Veteran's Administration Lung Cancer Trial,* you fitted a $log-logistic$ AFT model The outcome is *time to death* (in days). The exposure of interest is *treatment status* TX ($standard$ = 1, $test$ = 2). The control variables are performance status ($PERF$), disease duration (DD), AGE, and prior therapy ($PRIORTX$). The output is shown below.

log-logistic survival regression - accelerated failure time form
LR Chi2 (5) = 61.31
Log-likelihood = -200.196 Prob > chi2 = 0.0000

_t	*Coef.*	*Std.Err*	*z*	$P > \|z\|$
tx	−0.0973566	0.33540282	−0.2903	0.772
perf	0.0723285	0.00831384	8.6998	0.000
dd	0.00760878	0.01724958	0.4411	0.659
age	0.01561968	0.01668474	0.9362	0.349
priortx	0.00590508	0.04064202	0.1453	0.884
_cons	1.347464	0.6964462	1.9348	0.053
ln(1/p)	-0.4831854	0.0743015	-6.5030	0.000
1/p	0.6168149	0.0458303		

(a) State the **AFT log-logistic model** in terms of $\widehat{S(t)}$
(b) With a 95% confidence interval, **estimate** the acceleration factor γ comparing the *test* and *standard* treatment ($TX = 2$ vs. $TX = 1$). Interpret your answer.
(c) The AFT log-logistic model is also a proportional odds model. Use the output to estimate the odds ratio (odds of **death**) comparing the test and standard treatment. Also estimate the survival odds ratio comparing the test and standard treatment.
(d) The *Akaike Information Criterion* (AIC) is defined as $AIC = -2(Log\ likelihood) + 2p$. AIC is a method designed to compare the fit of different models, where smaller AIC suggesting better fit. Here, three models as follows are compared using the same 5 predictors presented earlier:

Model	**Frailty**	**Number of parameters**	**Log likelihood**
1) Log-logistic	No	7	-200.196
2) Weibull	No	7	-206.204
3) Weibull	Yes	8	-200.193

Which of the three models above should be selected based on the AIC criterion only?

2. (4+4=8 points) The *addicts.dat* dataset contain the following variables:

Clinic	1 (methadone clinic 1) or 2
Jail	1 (prison record) vs 0 (no record)
Dose	methadone dose (mg/day)

The data are analyzed with two models as follows:

Model A
Weibull regression
accelerated failure-time form
Gamma frailty
Log-likelihood $=-260.74854$

_t	$Coef.$	SE	z	$P>\|z\|$
Clinic	0.698	.158	4.42	0.000
Jail	0.145	.558	0.26	0.795
dose	0.027	.006	4.60	0.000
Jail×dose	-0.006	.009	-0.69	0.492
_const	3.977	.376	10.58	0.000
/ln_p	0.315	.068	4.67	0.000
p	1.370467			
1/p	0.729678			
theta	.00000002		.0000262	

LRT of theta=0
chibar2(01)=0.00
Prob>=chibar2=1.00

Model B
Weibull regression
accelerated failure-time form

Log-likelihood $=-260.74854$

_t	$Coef.$	SE	z	$P>\|z\|$
Clinic	0.698	.158	4.42	0.000
Jail	0.145	.558	0.26	0.795
dose	0.027	.006	4.60	0.000
Jail×dose	-0.006	.009	-0.69	0.492
_const	3.977	.376	10.58	0.000
/ln_p	0.315	.068	4.67	0.000
p	1.370467			
1/p	0.729678			

a) Using Model A, **test** for the **effect of adding a gamma frailty** component in the model.

b) Use Model B to **estimate** the *median* survival *time* for a patient with $Clinic = 2$ who has a *Jail* record and receives a methadone *dose* of 45 mg/day. [Hint: use the relationship that $t = [-\ln S(t)]^{1/p} \times \left(1/\lambda^{1/p}\right)$ for a Weibull model.]

3. (2+2+6=10 points) The following is a data layout for *Right*-, *Left*- and *Interval*-Censored survival data.

ID	LOWER	UPPER	SMOKE
Barry	2	2	1
Gary	7	-	0
Harry	5	5	0
Carrie	-	2	0
Larry	4	7	1

SMOKE is the only predictor ($1 = Smoker$, $0 = Nonsmoker$).

You try to fit the Weibull parametric survival regression model to the above data.

a) Write the *Weibull PH model*.

b) Write the *Weibull survival model.*

c) Obtain the *Weibull likelihood function* and discuss how you would get the *Maximum likelihood estimates* of the parameters in the model.

4. (3+4+3=10 points) For Data set below,

21	61.5	86.25	94.5	116.25	127.5	182.25	220.5	255	288
342.75	510	641.25	657.75	730.5	896.25	1005	1413	1918.5	11807.25

use the *method of moments* to estimate parameters for the following distribution models:

a) exponential distribution

b) gamma distribution and

c) Pareto distribution.

5. (4+6=10 points) For the following data,

28	82	115	126	155	170	243	294	340	384
457	680	855	877	974	1195	1340	1884	2558	15743

use *percentile matching* to estimate parameters for the following distribution models:

a) exponential distribution (hint: use the 50th percentile) and

b) Weibull distribution (hint: use the 30th and 80th percentile).

6. (6+6=12 points) Complete the table below and use the **Kolmogorov-Smirnov** test to determine if a **Weibull** distribution model is appropriate for the data below. (Hint: use the MLE $\widehat{\alpha} = 0.9959$ and $\widehat{\theta} = 631.4749$)

x	$F_n(x-)$	$F_n(x)$	$F^*(x)$	Maximum diff.
61.5	0.0000	0.0526		
86.25	0.0526	0.1053		0.076043785
94.5	0.1053	0.1579	0.139997207	
116.25	0.1579	0.2105		0.041293014
120.75	0.2105	0.2632		
182.25	0.2632	0.3158	0.25179948	
220.5	0.3158	0.3684		0.072600542
255	0.3684	0.4211	0.333237022	
288	0.4211	0.4737		
342.75	0.4737	0.5263		
510	0.5263	0.5789	0.554403217	
641.25	0.5789	0.6316		0.05884824
657.75	0.6316	0.6842	0.647052166	
730.5	0.6842	0.7368		0.051503337
894.75	0.7368	0.7895	0.757049579	
1005	0.7895	0.8421		
1413	0.8421	0.8947		0.050397564
1918.5	0.8947	0.9474		
2607	0.9474	1.0000	0.983501728	0.036101728

7. (4+2+2+2=10 points) Suppose that Bonnie (B) and Lonnie (L) are the only two subjects in the dataset shown below, where both subjects have two **recurrent** events that occur at different times (in weeks).

ID	Status	Stratum	Start	Stop
B	1	1	0	12
B	1	2	12	16
L	1	1	0	20
L	1	2	20	23

(a) Fill in the empty cells in the following data layout describing survival time to the **first event (stratum 1)**:

$t_{(f)}$	n_f	m_f	q_f	$R(t_{(f)})$
0	2	0	0	{B.L}
12				
20				

(b) For the **Stratified CP** approach, which of the following is correct .
i. **Lonnie** is in the risk set when Bonnie gets her *second* event.
ii. Bonnie is in the risk set when **Lonnie** gets her *second* event.
iii. Neither is in the risk set for the other's *second* event.

(c) For the **Gap Time** approach, which of the following is correct.
i. **Lonnie** is in the risk set when Bonnie gets her *second* event.
ii. Bonnie is in the risk set when **Lonnie** gets her *second* event.
iii. Neither is in the risk set for the other's second event.

(d) For the **Marginal** approach, which of the following is correct.
i. **Lonnie** is in the risk set when Bonnie gets her *second* event.
ii. Bonnie is in the risk set when **Lonnie** gets her *second* event.
iii. Neither is in the risk set for the other's second event.

8. (5 points) Consider the accident data in the table below, which is taken from Thyrion [110]. For the 9461 automobile insurance policies studied, the number of accidents under the policy is recorded in the table. Also recorded in the table is the observed value of the quantity that should be linear if the $(a, b,0)$ class is appropriate.

Number of accidents, k	0	1	2	3	4	5	6	7	8+	Total
Number of policies, n_k	7840	1317	239	42	14	4	4	1	0	9461
$k\frac{n_k}{n_k - 1}$		0.17	0.36	0.53	1.33	1.43	6.00	1.75		

The following members of the $(a, b, 0)$ class were fitted to the data.

Model	**Number of parameters**	**Negative log likelihood**
Poisson	1	5490.78
Negative binomial	2	5348.04

Using a Likelihood ratio test, you tested the following hypotheses:

H_0: the data came from a Poisson distribution

H_a: The data came from a Negative Binomial distribution.

Comment on the result and validity of this test.

9. (4+1=5 points) The claim payments on a sample of ten insurance policies are:

2	3	3	5	5^+	6	7	7^+	9	10^+

where + indicates that the loss exceeded the policy limit (or right censored).

Using the Kaplan-Meier Product-Limit estimator, calculate the probability that the **loss** on a policy **exceeds 8**.

a) 0.20
b) 0.24
c) 0.30
d) 0.36
e) 0.40

Final answer (1 point)
Work shown (4 points)

So the answer is (__).

10. (4+1=5 points) An insurance company is considering adding an ordinary deductible of 50 to one of their coverages. To estimate the savings, they model losses by using a Gamma kernel with $\alpha = 1$ applied to the following 5 randomly selected losses from this year:

100 300 350 500 600

Assuming that there is no inflation, what is the estimated **loss elimination ratio (LER)** from adding the deductible? (*Hint:* Loss Elimination Ratio $LER = \dfrac{E[X \wedge d]}{E[X]} = \dfrac{E[\min(X,d)]}{E[X]}$)

a) 0.120
b) 0.123
c) 0.126
d) 0.130
e) 0.134

Final answer (1 point)
Work shown (4 points)

So the answer is (__).

END OF TEST PAPER

9. (5 points) Consider the accident data in the table below, which is taken from Thyrion [110]. For the 9461 automobile insurance policies studied, the number of accidents under the policy is recorded in the table. Also recorded in the table is the observed value of the quantity that should be linear if the $(a, b, 0)$ class is appropriate.

Number of accidents, k	0	1	2	3	4	5	6	7	8+	Total
Number of policies, n_k	7840	1317	239	42	14	4	4	1	0	9461
$k\frac{n_k}{n_k - 1}$		0.17	0.36	0.53	1.33	1.43	6.00	1.75		

The following members of the $(a, b, 0)$ class were fitted to the data.

Model	**Number of parameters**	**Negative log likelihood**
Poisson	1	5490.78
Negative binomial	2	5348.04

a) Using a Likelihood ratio test, Louai tested the following hypotheses:
H_0: the data came from a Poisson distribution
H_a: The data came from a Negative Binomial distribution.
Comment on the result and validity of this test.

b) Members of the $(a, b, 1)$ class were also fit to the data. The results of that process revealed five potential models with p-values above 0.01 for the chi-square goodness of fit test. Information about these models is given in the table below.

Model	**Number of parameters**	**Negative log likelihood**	χ^2	p-value
Negative binomial (NB)	2	5348.04	8.77	0.0125
ZM logarithmic	2	5343.79	4.92	0.1779
Poisson-inverse Gaussian	2	5343.51	4.54	0.2091
ZM **Negative binomial**	3	5343.62	4.65	0.0979
Geometric-**Negative binomial**	3	5342.70	1.96	0.3754
Poisson-ET**NB**	3	5342.51	2.75	0.2525

Determine the **most suitable** model.