STAT474/STAT574, Practice Test 2

1. A study of 190 first-year medical students asked the question, How old were you when you first smoked a cigarette? Responses were either the exact ages at which they started smoking, that they never smoked, or that they currently smoke but cannot remember when they started. The data is summarized below.

Age (t)	Number Who Started Smoking at Age t	Number of Age t Who Smoke Now but Do Not Know the Age They Started	Number of Age t Who Do Not Smoke
14	2	0	0
15	3	0	0
16	10	0	0
17	13	0	0
18	5	0	0
19	3	0	1
20	2	4	13
21	1	6	44
22	2	8	39
23	1	2	19
24	0	0	3
25	0	0	4
26	1	0	4
Total	43	20	127

(a) Is it possible to determine the youngest person interviewed in the study? If so, what is their age? If not, explain why not.

Ans. No it isn't possible. There was at least one 19 year old interviewed who said that they didn't smoke. However, someone could have been 18 and said that they started smoking at age 17, for example. We cannot tell from the data.

(b) What is the smallest left-censored time?

Ans. The smallest left-censored time is age 20.

2. Suppose the following table gives the estimated cumulative hazard rate

t_i	$\Delta\widetilde{H}(t_i)$
1	.0263
55	.0270
74	.0278
86	.0286

Assume this is the only data. Now answer the following:

(a) Estimate the hazard function at $t = 75$ using a bandwidth of $b = 30$ and using the kernel

$$K(x) = 1/2 \quad -1 \le x \le 1$$

Solution. We use the kernel density approach:

$$\begin{aligned}\widehat{h}(75) &= \frac{1}{b}\sum_{i=1}^{D} K\left(\frac{75-t_i}{b}\right)\Delta\widetilde{H}(t_i)\\ &= \frac{1}{30}\left[K\left(\frac{75-1}{30}\right)\Delta\widetilde{H}(1) + K\left(\frac{75-55}{30}\right)\Delta\widetilde{H}(55) + K\left(\frac{75-74}{30}\right)\Delta\widetilde{H}(74) + K\left(\frac{75-86}{30}\right)\Delta\widetilde{H}(86)\right]\\ &= \frac{1}{30}[0 + (1/2)(0.027) + (1/2)(0.0278) + (1/2)(0.0286)] = 0.00139\end{aligned}$$

(b) Estimate the hazard function at $t = 75$ using a bandwidth of $b = 10$ and using the kernel

$$K(x) = (15/16)(1-x^2); \quad -1 \le x \le 1$$

Solution.

$$\begin{aligned}\widehat{h}(75) &= \frac{1}{b}\sum_{i=1}^{D} K\left(\frac{75-t_i}{b}\right)\Delta\widetilde{H}(1)\\ &= \frac{1}{10}\left[K\left(\frac{75-1}{10}\right)\Delta\widetilde{H}(1) + K\left(\frac{75-55}{10}\right)\Delta\widetilde{H}(55) + K\left(\frac{75-74}{10}\right)\Delta\widetilde{H}(74) + K\left(\frac{75-86}{10}\right)\Delta\widetilde{H}(86)\right]\\ &= \frac{1}{10}[0 + 0 + (15/16)(1-(1/30)^2)(0.0278) + 0] = 0.0026\end{aligned}$$

(c) If you plot the estimated hazard using a kernel method and find that it is overly smoothed, how should you change the bandwidth chosen assuming that you don't change the kernel used?

Answer. Make the bandwidth smaller to have less smoothing.

3. Consider the following results from a parametric model for survival times.There isn't enough information to get a p-vaue, but based on these results, (a) are there any effects that you would suspect are not statistically signficant? (b) Which distribution appears to be the best explanation of the data?

	Exponential		*Weibull*		*Log Logistic*		*Log Normal*		*Generalized Gamma*	
	Estimate	*SE*	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*
μ	3.755	0.990	3.539	0.904	3.102	0.953	3.383	0.936	3.453	0.944
α_1	−0.146	0.460	−0.148	0.408	−0.126	0.415	−0.199	0.442	−0.158	0.431
α_2	−0.648	0.355	−0.587	0.320	−0.806	0.354	−0.900	0.363	−0.758	0.394
α_3	−1.635	0.399	−1.544	0.363	−1.766	0.426	−1.857	0.443	−1.729	0.449
α_4	−0.020	0.014	−0.017	0.013	−0.015	0.014	−0.018	0.014	−0.018	0.014
σ	1.000	0.000	0.885	0.108	0.715	0.086	1.264	0.135	1.104	0.257
θ									0.458	0.584
Log L	−108.50		−108.03		−108.19		−108.00		−107.68	
AIC	227.00		228.05		228.38		227.99		229.36	

Based on AIC, the exponential model is preferred because it has the smallest AIC.

4. Suppose the time until death from diagnosis in years for a patient with a certain type of cancer is lognormally distributed with $\lambda = 3$ and $\zeta = 2$. Recall that if X is lognormal with parameters λ and ζ, then $\log X$ is normally distributed with mean λ and standard deviation ζ. A standard normal table is provided at the back for $Z \leq 0$.

(a) Find the mean (not median) survival time.

Solution. The mean is $e^{\lambda+\zeta^2/2} = e^5 = 148.4$ years. Not bad! I made ζ and λ unrealistically large for this problem, which makes the variance quite large. Note that the median is $e^\lambda = e^3 = 20$ years, so this is a very skewed lognormal distribution.

(b) Find the probability that a patient survives more than 2 years.

$$1 - \Phi\left(\frac{\log 2 - \lambda}{\zeta}\right) = 1 - \Phi\left(\frac{.693 - 3}{2}\right) = 1 - \Phi(-1.1535) = 1 - 0.1251 = 0.8749$$

Solution.

(c) Find the probability that a patient survives more than 4 years given that the patient survives the first two years.

$$\begin{aligned}
P(X > 4 | X > 2) &= \frac{P(X > 4, X > 2)}{P(X > 2)} \\
&= \frac{P(X > 4)}{P(X > 2)} \\
&= \frac{1 - \Phi\left(\frac{\log 4 - \lambda}{\zeta}\right)}{0.8749} \\
&= \frac{1 - \Phi\left(\frac{1.386 - 3}{2}\right)}{0.8749} \\
&= \frac{1 - \Phi(-0.81)}{0.8749} \\
&= \frac{1 - 0.2090}{0.8749} \\
&= 0.904
\end{aligned}$$

Standard Normal Probabilities

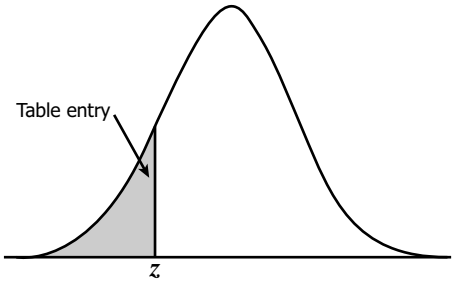

Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
–3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
–0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
–0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
–0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
–0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
–0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
–0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641