STAT 592 SECOND MIDTERM EXAMINATION SOLUTION

Problem 1. Researchers are interested in determining the correct dosage of oral penicillin for strep throat. Because of potential significant complications, if strep throat is detected, it must be treated adequately with antibiotics. Fourteen patients with strep throat are enrolled in the study. They take one pill of penicillin daily and report on which day the symptoms disappeared. The data are

2, 2, 3, 3, 4+, 4, 4, 5, 5, 5+, 6, 6, 8, 9

The observations were censored for two patients who were hospitalized and their participation in the study was discontinued.

(a) Estimate the survival function using the Kaplan-Meier method. Show your work. You can verify your calculations using the SAS output.

time, t_i	**# at risk, n_i**	**# died, d_i**	**survival rate, $1 - d_i/n_i$**	**estimator of $S(t)$, $\hat{S}(t), t_i \le t \le t_{i+1}$**
0	**14**	**0**	**1 -0/14 = 1.00**	**1.00**
2	**14**	**2**	**1-2/14 = 0.8571**	**(1)(0.8571) = 0.8571**
3	**12**	**2**	**1-2/12 = 0.8333**	**(0.8571)(0.8333) = 0.7143**
4	**10**	**2**	**1-2/10 = 0.8**	**(0.7143)(0.8) = 0.5714**
5	**7**	**2**	**1-2/7 = 0.7143**	**(0.5714)(0.7143) = 0.4082**
6	**4**	**2**	**1-2/4 = 0.5**	**(0.4082)(0.5) = 0.2041**
8	**2**	**1**	**1 – 1/2 = 0.5**	**(0.2041)(0.5) = 0.1020**
9	**1**	**1**	**1-1/1=0**	**(0.1020)(0)=0**

(b) What is the estimated probability that it will take longer than 2 days to recover? Longer than 7 days? Explain.

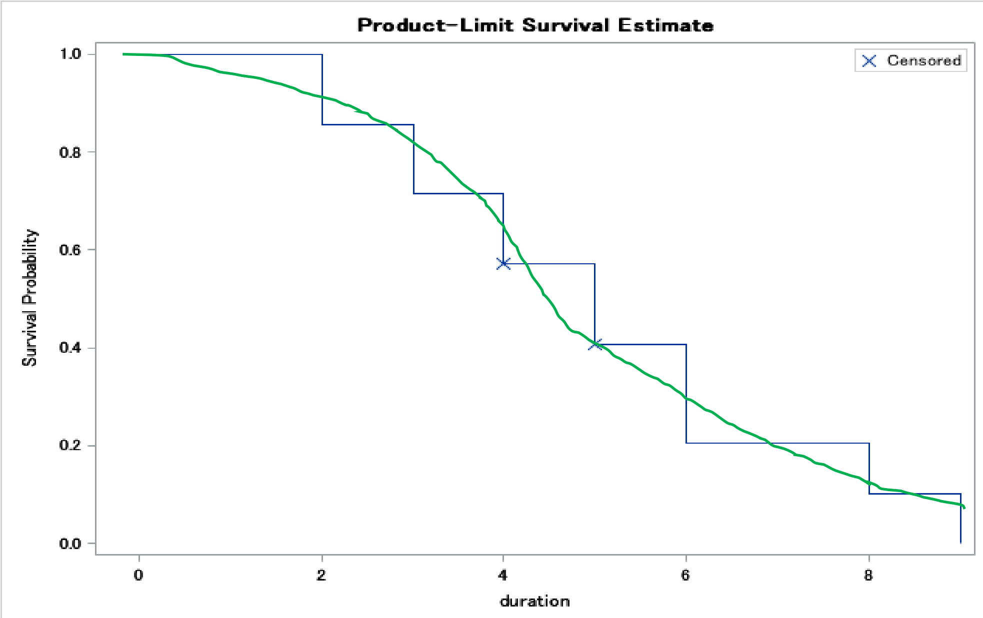
$\hat{S}(2) = 0.8571$, and $\hat{S}(7) = \hat{S}(6) = 0.2041$.

(c) Which parametric model of the survival time distribution, exponential or Weibull, fits the data better? Argue using three approaches: (1) quantitatively, using the results of the deviance test, (2) verbally, using the setting of the trial, and (3) visually, using the Kaplan-Meier survival curve in the output.

(1) The deviance test statistic is $\chi^2 = -2\big(-15.0218 - (-9.3119)\big) = 11.4198$. The *p*-value is $P(\chi^2(1) > 11.4198) < 0.001$, thus, we accept the alternative hypothesis and conclude that a Weibull survival curve has a better fit than an exponential curve.

(2) Oral penicillin is not going to take effect right away. So, the survival curve is going to be at a high level for a while, and then, once the medication accumulates in the body, more patients start recovering from strep throat and the curve will go down.

(3) The survival curve stays flat in the beginning, then drops, and then levels out. This behavior is best described by the Weibull model.



Problem 2. A retrospective study was performed in 77 male patients with nasopharyngeal cancer diagnosed and treated in a specialized clinic. The prognostic impact of smoking (yes/no), and type of treatment (chemo/radiotherapy) on patients' survival (in years) were evaluated using the Kaplan-Meier survival curves and log-rank tests. Study the SAS code and output to answer the questions below.

(a) Is there a significant difference in survival between smokers and non-smokers? Present the results of the log-rank test and describe the relative positions on the survival curves.

The *p*-value in the log-rank test is 0.0002, indicating that the survival curves are statistically significantly different. On the graph, we can see that one curve lies clearly underneath the other.

(b) Is there a significant difference in survival between chemo- and radiotherapy patients? Present the results of the log-rank test and describe the relative positions on the survival curves.

The *p*-value in the log-rank test is 0.1681, indicating that the survival curves are not statistically significantly different. On the graph, we can see that the curves lie close to each other, even touching and intersecting.

Problem 3. The data set from Problem 2 was used to fit the Cox proportional hazards model. See the SAS code and output.

(a) Write down the fitted model. Specify all parameters.

$\hat{S}(t) = [\hat{S}_0(t)]^{\hat{r}}, \hat{r} = \exp(0.05592(age - 60.8442) + 1.05656smoker + 0.39599chemotherapy), t \geq 0$, **where the estimate of the baseline survival function $\hat{S}_0(t)$ is given in the last two columns in the output.**

(b) Predict the probability of survival past 2.8 years for a 52-year-old smoker on radiotherapy.

$$\hat{S}(2.8) = [0.95289]^{\exp(0.05592(52-60.8442)+1.05656)} = 0.9188.$$

Problem 4. Investigators in a dental research clinic are testing a new revolutionary dental implant. For each subject who enrolls in the trial, they record age at the baseline, gender, and the number of implanted teeth. The survival time is the time (in months) until a complication with an implant develops. The study continues for twenty-four months. The subjects who do not have any complications are censored at the end of the trial. Use the attached SAS output to answer the questions below.

(a) Which parametric regression model for the survival function fits the data better, exponential or Weibull? Explain by giving appropriate numbers.

$\chi^2 = -2\ (\ln L(\text{exponential}) - \ln L(\text{Weibull})) = -2\ (-24.1406 - (-23.9182)) = 0.44$, **p-value=** $P(\chi^2(1) > 0.4448) > 0.25$, **therefore, the exponential model is appropriate.**

(b) Write down the fitted model. Estimate all parameters.

$\hat{S}(t) = \exp(-\hat{\lambda}\, t),\ t > 0$, **where** $\hat{\lambda} = \exp(-(3.1621 - 0.0220age - 0.0380nteeth - 0.7541male))$.

(c) Interpret the estimated significant parameters.

The estimated mean survival time $\widehat{ET} = \frac{1}{\hat{\lambda}} = \exp(3.1621 - 0.0220age - 0.0380nteeth - 0.7541male)$. **Age and gender are significant predictors. For a one-year increase in age, the estimated mean time until a complication occurs changes by** $(\exp(-0.0220) - 1) \cdot 100\% = -2.18\%$ **, that is, decreases by 2.18%. The estimated mean time until a complication for males is** $\exp(-0.7541) \cdot 100\% = 47.04\%$ **of that for females.**