BIS 505b – Problem Set 6 **Due 11:59 PM Eastern, 5/7/2021**

Instructions: Follow the homework instructions outlined in the syllabus. Round your answers to 3 decimal places. Perform all tests at the α = 0.05-level and follow the steps of hypothesis testing. You may use **R** as a calculator to perform intermediate calculations or to compute p-values.

Assignment

Question 1: Displayed below are the survival times in months since diagnosis for 10 AIDS patients suffering from concomitant esophageal candidiasis, an infection due to candida yeast, and cytomegalovirus, a herpes infection that can cause serious illness. Right censored times are denoted by a plus (+) sign.

Survival (months)
0.5+, 1, 1, 1, 2, 5+, 8+, 9, 10+, 12+

a. [5] How many deaths were observed in this sample of patients?

5 deaths were observed in this sample of patients.

b. [10] Use the Kaplan-Meier method to estimate the survival function $S(t)$. Show your work.

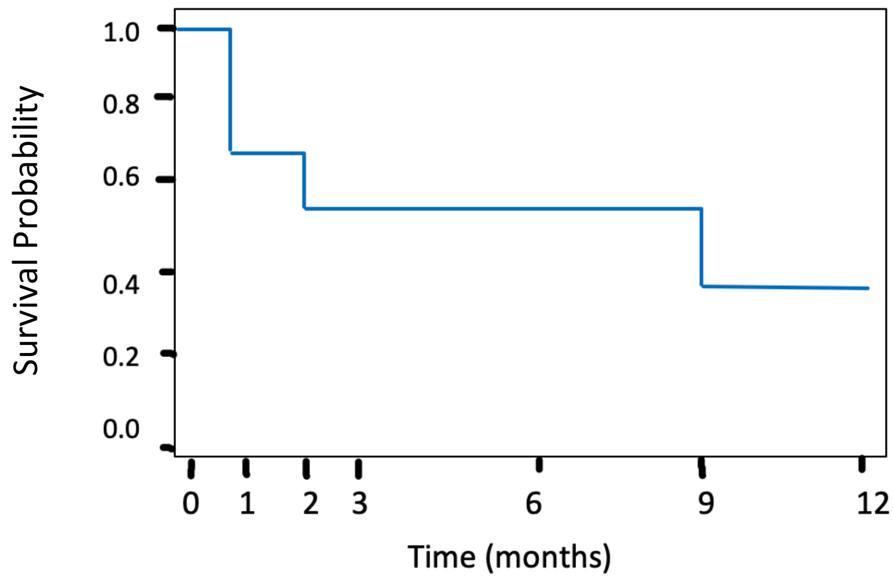
j	Ordered Failure Times ($t_{(j)}$)	Interval (I_j)	Number of failures (d_j)	Number of censored (c_j)	Number of at risk (n_j)	Conditional probability of surviving ($1 - d_j/n_j$)	Estimated survival probability ($\hat{S}(t_{(j)})$)
0	0	$[0, 1)$	0	1	10	$1 - 0/10 = 1$	1
1	1	$[1, 2)$	3	0	9	$1 - 3/9 = 0.667$	$1 \times 0.667 = 0.667$
2	2	$[2, 9)$	1	2	6	$1 - 1/6 = 0.833$	$1 \times 0.667 \times 0.833 = 0.556$
3	9	$[9, 12]$	1	2	3	$1 - 1/3 = 0.667$	$1 \times 0.667 \times 0.833 \times 0.667 = 0.371$

Estimated probability of survival at 1 month $\hat{S}(1) = 0.667$
Estimated probability of survival at 2 months $\hat{S}(2) = 0.556$
Estimated probability of survival at 9 months $\hat{S}(9) = 0.371$

c. [5] What is $\hat{S}(t)$, the estimated probability of survival at 1 month? What is the estimated probability of survival of survival at 5 months? At 6 months?

Estimated probability of survival at 1 month $\hat{S}(1) = 0.667$
Estimated probability of survival at 5 months $\hat{S}(5) = \hat{S}(2) = 0.556$
Estimated probability of survival at 6 months $\hat{S}(6) = \hat{S}(2) = 0.556$

d. [10] Plot the estimated Kaplan-Meier survival curve by hand. [Note: You may draw using shapes/lines in your word processing software]

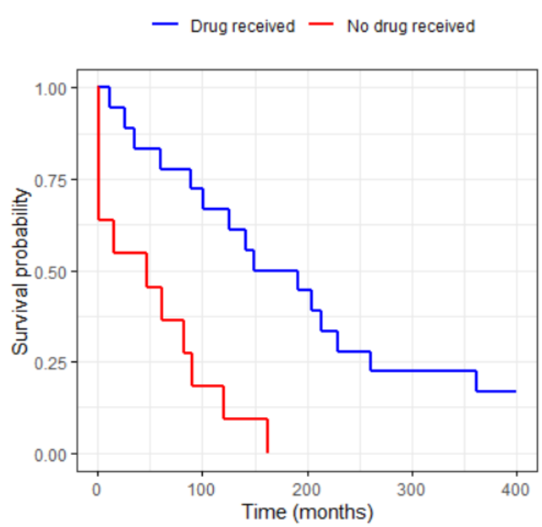


e. [5] What is the estimated median survival time?

$\hat{t}_{50} = min\{t|\hat{S}(t) \leq 0.5\} = 9$

The estimated median survival time is 9 months.

Question 2: In the 1980s, a study was conducted to examine the effects of the drug ganciclovir on AIDS patients suffering from disseminated cytomegalovirus infection. Two groups of patients were followed; 18 were treated with the drug, and 11 were not. Survival time (time to death) in months after diagnosis is analyzed. The Kaplan-Meier (product-limit) estimate of the survival curve for each group is shown below.



a. [5] Does it appear that individuals in one group survive longer than those in the other group? Which group has the better survival experience?

Yes, the survival curve of treatment group is consistently above the survival curve of control group. Treatment group has the better survival experience.

b. [5] Based on the survival curves above, roughly estimate the median survival time in each group.

The median survival time in treatment group is approximately 150 months while in control group is approximately 50 months.

c. [10] The test statistic from the log-rank test was found to equal 12.45. Based on this test statistic, perform the hypothesis test to determine if the survival experience is significantly different in the two arms of this study. Follow the steps of hypothesis testing using the reported test statistic.

(1) State the null and alternative hypotheses
$H_0: S_1(t) = S_2(t)$ for all t vs. $H_1: S_1(t) \neq S_2(t)$ for some t

(2) Specify the significance level, α = 0.05

(3) Compute the test statistic

$$\chi^2 = \frac{(O-E)^2}{V} = 12.45 \sim \chi_1^2$$

(4) Generate the decision rule
Given α = 0.05,
Reject H_0 if $\chi^2 \geq \chi^2_{1-\alpha}(1) = \chi^2_{0.95}(1) = 3.84$ or if $p \leq 0.05$

(5) Draw a statistical conclusion and state the conclusion in words in the context of the problem.

$$\chi^2 = 12.45 > 3.84 \ \rightarrow Reject\ H_0$$
$$or\ p = P(\chi^2 \geq 12.45) = 0.0004 < 0.05 \ \rightarrow Reject\ H_0$$

Conclusion: There is evidence to reject H_0 and conclude that the survival experience is significantly different for those receiving drugs and those without.

d. [5] If we were to estimate a hazard ratio from these data using a simple Cox proportional hazards model, would you expect the estimated hazard ratio comparing the drug group to the no drug group to be greater than 1 or less than 1? Why?

I would expect the estimated hazard ratio comparing the drug group to the no drug group to be less than 1. Because the treatment group has the better survival experience, which means that the treatment group has a lower rate of death, the estimated coefficient $\hat{\beta}$ in the simple Cox proportional hazards model will be less than 0, then the estimated hazard ratio $\widehat{HR} = e^{\hat{\beta}} < 1$.

Question 3. The data for this question come from an observational study of 686 patients with primary node positive breast cancer. The study looks at time from breast cancer diagnosis to recurrence or censoring. Hormone therapy, tumor size, and tumor grade are believed to impact recurrence free survival. A partial list of the variables in the dataset is given below:

- Time: time in days from diagnosis to breast cancer recurrence or censoring.
- Status: survival status (0 = censored, 1 = breast cancer recurrence).
- Hormone: hormone therapy (0 = No, 1 = yes). Note: “No” is the reference category.
- Size: Tumor size (mm). Treated as a continuous variable in the model.
- Grade: Tumor grade (1, 2, 3). Note: “1” is the reference category.

Partial computer results from a Cox PH model constructed using these data are given below:

	Estimate	Standard Error
Hormone	-0.347	0.126
Size	0.014	0.004
Grade 2	0.858	0.246
Grade 3	1.070	0.262

a. [5] Report the fitted Cox proportional hazards model.

$$log(\hat{h}(t,x)) = log(\hat{h}_0(t)) - 0.347\ Hormone + 0.014\ Size + 0.858\ Grade2 + 1.070\ Grade3$$

b. [10] Report and interpret the estimated hazard ratio associated with Hormone in the model above and provide a 95% confidence interval for the hazard ratio.

Controlling for all of the other variables in the model, the adjusted hazard ratio is given by the exponentiated slope $\widehat{HR} = e^{b_1} = e^{-0.347} = 0.707$ [95% CI $(0.552{,}0.905)$], which means that hormone therapy is associated with a 29.3% reduction in the recurrence rate.

95% CI for $log(HR), \beta$: $(-0.594, -0.100)$
$b \pm z_{1-\frac{\alpha}{2}}\, s_b = -0.347 \pm 1.96 \times 0.126 = (c_L = -0.594, c_L = -0.100)$
95% CI for HR, β: $(0.552{,}0.905)$
$(e^{-0.594}, e^{-0.100}) = (0.552{,}0.905)$

c. [10] Perform a hypothesis test of the slope coefficient (β) associated with Hormone in the model. Does the 95% confidence interval computed in part (b) support your conclusion? Explain.

(1) State the null and alternative hypotheses
H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$

(2) Specify the significance level, α = 0.05

(3) Compute the test statistic

$$z = \frac{b}{s_b} = \frac{-0.347}{0.126} = -2.754 \sim N(0,1)$$

(4) Generate the decision rule
Given α = 0.05,
Reject H_0 if $|z| \geq z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$ or if $p \leq 0.05$

(5) Draw a statistical conclusion and state the conclusion in words in the context of the problem.
$|z| = 2.754 > 1.96 \rightarrow Reject\ H_0$
$or\ p = 2 \times P(Z \geq 2.754) = 0.006 < 0.05 \rightarrow Reject\ H_0$

Conclusion: There is evidence to reject H_0 and conclude that there is a significant difference in hazard of recurrence between the patients with hormone therapy group and those without when controlling for tumor size, and tumor grade.

Does the 95% confidence interval computed in part (b) support your conclusion? Explain.
Yes, the 95% CI for hazard ratio is $(0.552, 0.905)$, which excludes 1, so we can reject H_0.

d. [5] Compute and interpret the estimated hazard ratio for the effect of a 10-mm increase in tumor size from the model above.

Controlling for all of the other variables in the model, as tumor size increases, the hazard of recurrence increases. The adjusted hazard ratio for the effect of a 10-mm increase in tumor size is given by the exponentiated slope $\widehat{HR} = e^{10b_2} = e^{10\times0.014} = 1.150$, which means a 10-mm increase in tumor size increases the rate of recurrence by 15.0%.

Question 4. A study is conducted to look at the efficacy of a homeopathic treatment for pain. Four patients were given the homeopathic treatment as soon as they reported onset of symptoms. The patients were closely observed by the staff for up to 60 minutes after the treatment was administered at the clinical research unit for self-reported pain relief. The study team recorded time to pain relief in minutes. The team wants to use survival analysis techniques to analyze these data.

a. [8] Help the study team translate their patient notes (below) into analyzable data where t_i is our time variable (in minutes) and δ_i is our event indicator (1 if event, 0 if right censored).

- Patient 1 had a family emergency and had to leave the observation room after 20 minutes. His pain was still present when he left.
- Patient 2 had an adverse reaction to the treatment and was withdrawn from the study at 28 minutes without pain relief.
- Patient 3 reported that his headache pain was gone after 36 minutes.
- Patient 4 did not experience pain relief during the observation period.

Patient $\boldsymbol{i}$	$\boldsymbol{t_i}$	$\boldsymbol{\delta_i}$
1	20+	0
2	28+	0
3	36	1
4	60+	0

b. [2] Are any of the patients “administratively censored” in this sample? If so, which patient(s)?

Yes, patient 4.