# BIOS 845 Homework #2

Date assigned: 02/15/2018
Due Date: 02/28/2018 by 11:59 pm (Blackboard clock time);

Instructions:

1. To receive full credit, show all work. Please make your work legible.
2. Total points for this homework are 100.
3. Do not forget to write your name on the homework.
4. Insert page numbers on all pages and also total # of pages submitted.
5. Homework can be typed or hand-written. Provide SAS code wherever necessary.
6. Use the BLACKBOARD drop box to turn in the homework (preferably as pdf) or bring it to class on 02/28/2018.

**Question # 1:** **25 points**

a. Solve Exercise 7.4 from textbook.

b. Solve Exercise 7.12 from textbook.

**Question # 2:** **45 points**

Researchers wish to explore the efficacy of triple-drug combinations of antiretroviral therapy for treatment of HIV-infected patients. Because of limitations on potency and the continuing emergence of drug resistance seen with the use of currently available antiretroviral agents in monotherapy and two-drug regimens, triple combination regimens should represent a more promising approach to maximize antiviral activity, maintain long-term efficacy, and reduce the incidence of drug resistance. Towards this end, investigators performed a randomized study comparing AZT + zalcitabine (ddC) versus AZT + zalcitabine (ddC) + saquinavir. The data, time from administration of treatment (in days) until the CD4 count reached a pre-specified level, is given below for the two groups:

AZT + zalcitabine (ddC): 4+, 6, 11, 12, 32, 35, 38+, 39, 45, 49, 75, 80, 84, 85, 87, 102, 180+

AZT + zalcitabine (ddC) + saquinivir: 2, 3, 4, 12, 22, 48, 51+, 56+, 80, 85, 90, 94+, 160, 171, 180, 180+, 238

a. Conduct a log-rank test to help answer the research question.

b. Based on the problem statement, do you think there is any justification to use a different weight function in analyzing this data? If so, choose a suitable weight function and re-conduct the analysis.

c. Fit an exponential regression model to these data with a single dichotomous covariate representing treatment group. Test the hypothesis of no effect of therapy on the time to obtain the pre-specified CD4 level. What are your conclusions?

d. Fit an Weibull regression model to these data with a single dichotomous covariate representing treatment group. Test the hypothesis of no effect of therapy on the time to obtain the pre-specified CD4 level. What are your conclusions?

e. Perform a likelihood ratio test to determine if the exponential regression model fits the data as good as the Weibull model. What are your conclusions?

f. Generate log-survival and log-log survival plots for these data. Based on these plots, what can you infer about the validity of the models that you fit in parts (c) and (d)? Do your conclusions based on the plots agree with your conclusions from the likelihood ratio test in part (e)?

g. Fit a lognormal regression model and log-logistic regression model to these data with a single dichotomous covariate representing treatment group. Report the log-likelihood values for each model. Are your conclusions regarding whether there was a significant effect of treatment dependent on the distributional assumptions?

h. Based on your understanding of the topic of 'Parametric Regression', select the "best" fitting parametric regression model to assess the effects of treatment on the time to obtain the pre-specified CD$ level. Justify your choice of a "best" model and discuss the shape of the hazard based on your choice.

i. Based on your "best" model chosen in part (h), plot the estimated survival curves for the two treatments on the same plot.

**Question # 3:** **15 points**

a. Let X be a random variable distributed as $Weibull(\gamma, \beta)$ with its pdf shown below.

$$f(x|\gamma,\beta) = \frac{\gamma}{\beta} x^{\gamma-1} \exp(-x^{\gamma}/\beta) \qquad x \geq 0, \quad \gamma > 0, \quad \beta > 0$$

Solve the following:
{i} Derive and identify the distribution of $X^{\gamma}$
{ii} Derive and identify the distribution of $c$X where c is a positive constant.
{iii} Derive and identify the distribution of $X^{1/\gamma}$
{iv} Derive an expression for the median and mode of X.

b. Let X be a random variable distributed as $Gamma(\alpha, \beta)$ with its pdf shown below.

$$f(x|\alpha,\beta) = \frac{x^{\alpha-1}\exp(-x/\beta)}{\Gamma\alpha\ \beta^{\alpha}} \qquad x \geq 0, \quad \alpha > 0, \quad \beta > 0$$

Derive and identify the distribution of $\frac{X^{1/\beta}}{\theta}$ for $\theta > 0$

**Question # 4:** **15 points**

The dataset below pertains to an animal carcinogenesis study and contains litter-matched time-to-event data. In this study, three rats from each of 50 litters were matched, where one rat was treated with a putative carcinogen and the other serving as controls. The outcome is the time to tumor occurrence or censoring in weeks. 'E' represents the Exposed rats. 'C1' and 'C2' represents the control rats. * represents censored times.

| Litter | $E$ | $C1$ | $C2$ | Litter | $E$ | $C1$ | $C2$ |
|---|---|---|---|---|---|---|---|
| 1 | 101.0* | 49.0 | 104.0* | 26 | 89.0* | 104.0* | 104.0* |
| 2 | 104.0* | 102.0* | 104.0* | 27 | 78.0* | 104.0* | 104.0* |
| 3 | 104.0* | 104.0* | 104.0* | 28 | 104.0* | 81.0 | 64.0 |
| 4 | 77.0* | 97.0* | 79.0* | 29 | 86.0 | 55.0 | 94.0* |
| 5 | 89.0* | 104.0* | 104.0* | 30 | 34.0 | 104.0* | 54.0 |
| 6 | 88.0 | 96.0 | 104.0* | 31 | 76.0* | 87.0* | 74.0* |
| 7 | 104.0 | 94.0* | 77.0 | 32 | 102.8 | 73.0 | 83.9 |
| 8 | 95.9 | 104.0* | 104.0* | 33 | 101.9 | 104.0* | 80.0* |
| 9 | 82.0* | 77.0* | 104.0* | 34 | 79.9 | 104.0* | 73.0* |
| 10 | 70.0 | 104.0* | 77.0* | 35 | 45.0 | 79.0* | 104.0* |
| 11 | 88.9 | 91.0* | 90.0* | 36 | 94.0 | 104.0* | 104.0* |
| 12 | 91.0* | 70.0* | 92.0* | 37 | 104.0* | 104.0* | 104.0* |
| 13 | 39.0 | 45.0* | 50.0 | 38 | 104.0* | 101.0 | 94.0* |
| 14 | 102.9 | 69.0* | 91.0* | 39 | 76.0* | 84.0 | 78.0 |
| 15 | 93.0* | 104.0* | 103.0* | 40 | 80.0 | 80.9 | 76.0* |
| 16 | 85.0* | 72.0* | 104.0* | 41 | 72.0 | 95.0* | 104.0* |
| 17 | 104.0* | 63.0* | 104.0* | 42 | 72.9 | 104.0* | 66.0 |
| 18 | 104.0* | 104.0* | 74.0* | 43 | 92.0 | 104.0* | 102.0 |
| 19 | 81.0* | 104.0* | 69.0* | 44 | 104.0* | 98.0* | 73.0* |
| 20 | 67.0 | 104.0* | 68.0 | 45 | 55.0* | 104.0* | 104.0* |
| 21 | 104.0* | 104.0* | 104.0* | 46 | 49.0* | 83.0* | 77.0* |
| 22 | 104.0* | 104.0* | 104.0* | 47 | 89.0 | 104.0* | 104.0* |
| 23 | 104.0* | 83.0* | 40.0 | 48 | 88.0* | 79.0* | 99.0* |
| 24 | 87.0* | 104.0* | 104.0* | 49 | 103.0 | 91.0* | 104.0* |
| 25 | 104.0* | 104.0* | 104.0* | 50 | 104.0* | 104.0* | 79.0 |

Ignoring the data for 'C2', conduct an analysis to help determine whether exposure to carcinogen significantly affects time to tumor progression.

**Question # 5 (Bonus):** **5 points**

Read the paper "Calculation of scores for a Wilcoxon generalization applicable to data subject to arbitrary right censorship" by Nathan Mantel (see pdf in folder titled 'Extra Reading'). Confirm that you understood how the calculations are done in the table on Page #246 of this paper.

GOOD LUCK ☺☺