Quiz 2

Advance Business Analytics

Due February 5 at 11:59PM

[25 points]

1. Many disease-free individuals were enrolled in a study beginning January 1, 1970, and were followed for 30 years to assess the age at which they developed breast cancer. Individuals had clinical exams every 3 years after enrollment. For the selected individuals described below, discuss in detail, the type of censoring is represented (if any). [5]

(a) A healthy individual, enrolled in the study at age 30, did not develop breast cancer during the study.
(b) A healthy individual, enrolled in the study at age 40, was diagnosed with breast cancer at the fifth exam after enrollment (i.e., the disease started sometime between 12 and 15 years after enrollment).
(c) An individual, enrolled in the study at age 42, moved away from the community at age 55 and was never diagnosed with breast cancer during the period of observation.

2. A Cruize is interested in understanding its customers' repeat visit to the cruise. The firm surveys its customers and collects the attached data (cruize_visit.xlsx).
   Using python, can you plot the survival probability? Compute and comment on the confidence interval. [7]
   You want to compare whether promotion types make a difference in survival probability. First, plot the survival curves for each promotion type and then do a log rank test to compare promotion types. Show the calculation by hand. [13]