ASTR 600 – HW #8
Truncation and Censoring

You the front desk manager at a business and are responsible for keeping track of when employees show up to work. The work day begins promptly at 9 am, but sometimes employees arrive early or late. As long as you are at your desk, you can mark the time that each individual arrives. If you are late to work, however, you can only count the number of people that are already at the office without knowing what time they arrived. You need to be able to make meaningful reports to your boss about the overall punctuality of the company's employees.

The data files store the time each employee arrived relative to the target time of 9 am in minutes. Therefore a value of -15 indicates that the individual arrived at 8:45 am and a value of 15 indicates an arrival time of 9:15 am.

1. On your first day, you are determined to do a good job and are the first to arrive at the office. Thus you are able to note the exact time of everyone's arrival (see data file 1.) Should you report to your boss that the company on the whole tends to arrive early or late, and by how much? Consider the statistics we have covered this semester (e.g. Student's t-test, KS test, mean, median, etc.) and justify your choice of statistics. You may want to inspect the distribution beforehand.

2. Waking up early is tiring, so you start wondering how late you can show up while still getting accurate results for your boss. Modify data1 as if you had arrived at work at 8:45 am (hint: use R's replace function to set an upper limit based on your arrival time.) Using the same statistics as you did in #1, how has your report to your boss changed? Try several later arrival times. At what point is the mean employee arrival time of the company late? In your report, include a histogram of the recorded employee arrival times at this point. How is this different if you use a more robust measure of center?

3. Another business has moved into the building with you. They are a hipster tech start up and don't care what hours their employees work most of the time, except for once in a while when they have a 9 am all-company meeting. Being the curious statistician you are, you wonder how timely their employees are to these meetings. After several weeks, you have obtained data on the average arrival time of each employee at your company and the tech company. However, some individuals consistently arrive before you do and thus you only have an upper limit on their arrival time. Data sets 2 and 3 contain the average arrival time of each employee at your company and the tech company respectively and a tag of whether that value is a limiting value (1 indicates limiting value, 0 indicates true measured value.) Calculate the Kaplan-Meier estimator of this data for each company and include plots of each in your homework. Also choose a censored 2-sample test and determine whether one company is significantly earlier or later than the other. Justify your choice of statistic.