

1. **Chocolate Chips.** The melting times from the chip activity are provided in the Minitab file `MELT_TIMES W2018`. We will assume that the times are a random sample of all chocolate chip melting times. The event times are not directly usable in their current form in the Minitab worksheet. Construct two columns of data corresponding to the time-to-event variable,  $T$ , and the censoring variable,  $C$ . (*Note: Don't copy and paste the first column because it is designated as "text" column - just re-enter the data into a new column.*) Once the data are properly arranged, answer the following questions:
  - (a) Briefly explain why an observed melting time of 85 seconds is right censored.
  - (b) Compute the average melting time using all the chip data. Do you think this value over-estimates or under-estimates the true (population) average chip melting time? Briefly explain.
  - (c) Would it be more appropriate to add or subtract 5 seconds to the right censored times? Adjust the right censored times, accordingly, by adding or subtracting 5 seconds to the incomplete times, and recompute the sample average melting time. Verify your answer to the second part of Part (b).
  - (d) Even though you appropriately adjusted the times upward or downward, the correct amount of time to adjust each of the incomplete observations is unknown. Now remove the right censored chip melting times and recompute the average melting time. How does removing the incomplete observations affect the mean chip melting time found in Part (b)?

- (e) Summarize the effects of 1) treating the right censored data as complete and 2) removing right censored observations from the data set on the average chip melting time.

2. **Age at First Bike Ride.** Recall the survey question that asked at which age you learned to ride a bike. These ages are provided in the Minitab file **Bike Ages W2018**. We will assume that the times are a random sample of all ages that Cal Poly students learned to ride a bike. Note that a bike age with a “-” to its right is a left censored event time. The event times are not directly usable in their current form in the Minitab worksheet. Construct two columns of data corresponding to the time-to-event variable,  $T$ , and the censoring variable,  $C$ . Once the data are properly arranged, answer the following questions:

- (a) Describe the beginning of time, time metric, and time-to-event random variable.
- (b) Briefly explain what it means for an age to be left-censored (in the context of the problem).
- (c) Use all the data to calculate the sample mean age when students first learned to ride a bike. Do you think this value over-estimates or under-estimates the true mean age at which students first learned to ride a bike? Briefly explain.
- (d) Would it be more appropriate to add or subtract 2 years to the left-censored observations? Adjust the left-censored ages, accordingly, by adding or subtracting 2 years to the incomplete times, and recompute the sample average age. Verify your answer to the second part of Part (c).

- (e) Remove the left-censored observations (ages) from the data set and recalculate the mean age of the sample. How does this value compare to your answer in Part (c)?
- (f) Describe the fundamental difference between the incomplete observations in Problems (1) and (2). (State something more meaningful than “Problem (1) has right censored data and Problem (2) has left censored data”.)