1. Henning and Frueh (1996) followed criminal activities of 194 inmates released from a medium security prison for 36 months. We can use the data from this study to investigate the time until the former inmates were re-arrested. If the former inmate had been re-arrested for a criminal act before 36 months (after initial prison release) had passed, then that former inmate's event time was complete. If the former inmate had not been re-arrested for a criminal act after 36 months had passed, or had completely dropped out of the study, then that former inmate's event time was right censored. Measurements on the following variables are available in the Minitab file `Rearrest`:

   - `months`: months until re-arrest
   - `censor`: censoring status indicator variable (0 = censored event time)
   - `personal`: a dichotomous variable identifying former inmates who had a history of person-related crimes (`personal`=1), i.e. those with one or more convictions for offenses such as aggravated assault or kidnapping.
   - `property`: a dichotomous variable indicating whether former inmates were convicted of a property-related crime (`property`=1)
   - `cenage`: the "centered" age of individual, i.e. the difference between the age of the individual upon release and the average age of all inmates in the study.

   Use this data set to answer the following questions:

   (a) Construct probability plots for each of following distributions, and report the maximum likelihood estimates of the parameters and the Anderson-Darling test statistic. Be sure to indicate if the parameters are location, scale, or shape:

      i. **Exponential**:

      $\hat{\lambda} = 25.27$, *AD = 284.229*

      ii. **Loglogistic**:

      location $\hat{\alpha} = 2.78$, *scale* $\hat{\beta} = .96$, *AD = 283.869*

      iii. **Logistic**:

      $\hat{\alpha} = 18.79$ *(location) and* $\hat{\beta} = 9.79$ *(scale), AD = 285.834*

      iv. **Lognormal**:

      $\hat{\mu} = 2.798$ *(location) and* $\hat{\sigma} = 1.746$ *(scale), AD = 284.048*

      v. **Weibull**:

      $\hat{\beta} = .828$ *(shape) and* $\hat{\lambda} = 26.84$ *(scale), AD = 283.908*

   (b) Is there a distribution that appears to fit the data well? If so, which distribution? Briefly explain.

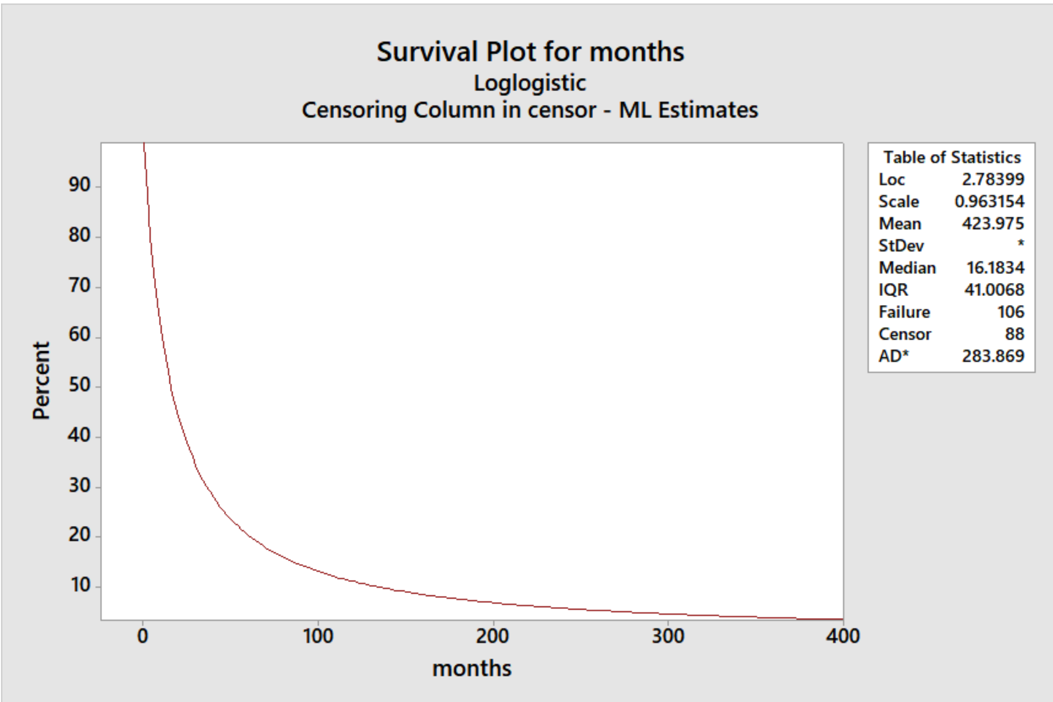
   Weibull or loglogistic appears to fit the data the best since the points fall close to the diagonal line assuming either of these distributions. The loglogistic distribution yields the smallest Anderson-Darling statistic.

(c) Which distribution appears to fit the data the worst? Briefly explain.

Logistic distribution has the worst fit since the points depart from the diagonal line the most, and has the largest Anderson-Darling statistic.

2. Utilizing the loglogistic distribution, answer the following questions:

(a) Construct a parametric survival curve for former inmates. Scroll up through the Minitab Session Window to find the **mean** and **median** time to be re-arrested, and report these values. Briefly explain why the mean time to failure does not make sense in the context of the rearrest data.



Mean(MTTF)=423.975 months, Median = 16.1834 months

Mean time does not make sense because the longest event times are 36 months.

(b) Examine the "Table of Percentiles" in the Minitab Session Window, and report the 20th percentile and interpret its value in the context of the problem.

$t_{20} = 4.25788$.

This means that 20% of former inmates have been rearrested by 4.26 months and 80% have not.

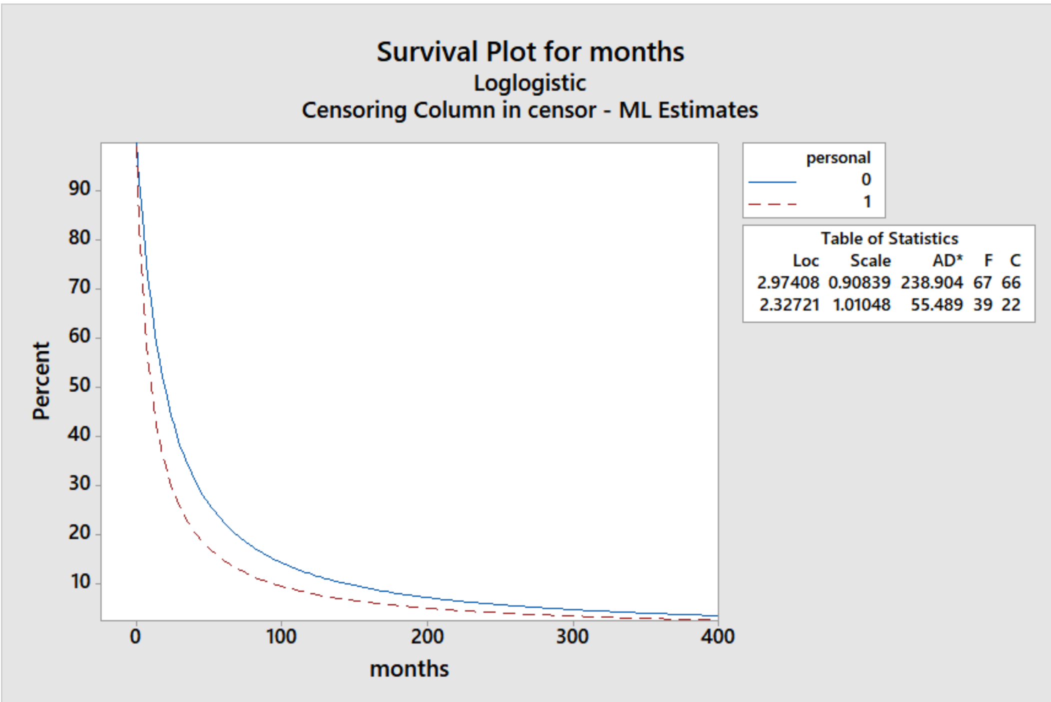
(c) Examine the "Table of Percentiles" and (approximately) find the probability that a randomly selected released inmate takes longer than 11 months to be rearrested.

$t_{40} \approx 11$.

$P(T < 11) \approx 0.4$.

$P(T > 11) \approx 0.6$.

(d) Now construct parametric survival curves for former inmates who did and did not commit person-related crimes (on the same graph; you may want to manually scale the x-axis to have a lower limit). What do the survival curves suggest about former inmates who had committed person-related crimes versus former inmates who did not commit non-person-related crimes?



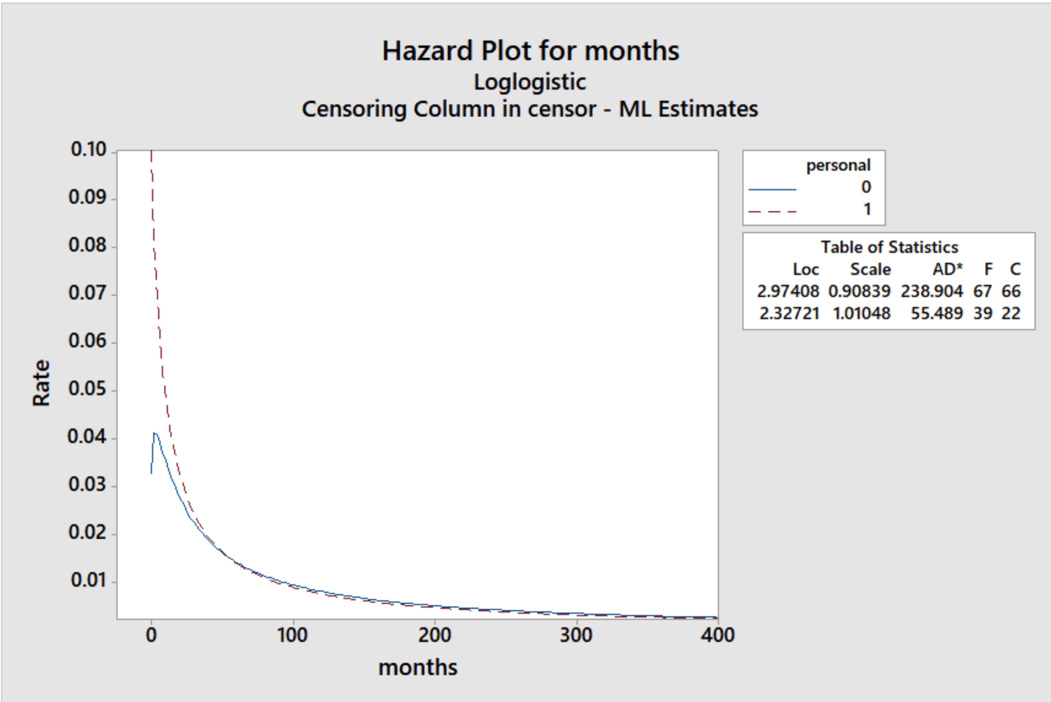
Former inmates with a history of person-related crimes typically are re-arrested sooner than those former inmates who have not committed these types of crimes. The proportions of former inmates who had committed person-related crimes who had not been re-arrested any time after their release from prison are generally smaller than the proportions for those who had not committed person-related crimes.

(e) Scroll up through the Minitab Session Window to report the **mean** and **median** time to be re-arrested for both groups of former inmates (there will be separate analyses for `personal`=0 and `personal`=1). Verify that these results are consistent with your answer in part (d).

Former inmates who committed person-related crimes: median = 10.25 months and the mean does not exist (For a loglogistic random variable, the mean only exists when the scale parameter value is less than 1).

Former inmates who had not committed person-related crimes: median = 19.57 months *and the mean = 196.78 months.*

(f) Construct the hazard curves (on the same graph; you may want to manually scale the x-axis to have a lower limit) for former inmates who did and did not commit person-related crimes. What do these curves suggest about being re-arrested?



Given that re-arrest has not occurred for at least t months, the risk of being rearrested is higher for former inmates who committed person-related crimes, especially during earlier months after being released from prison. For former inmates who did not commit person-related crimes, the hazard rate increases slightly after being released, and then decreases with time.

3. Consider the age at first drink of alcohol data located in the Minitab file `Firstdrink`. Construct probability plots to investigate which probability distribution (if any) you would suggest to model the time-to-event variable. Which probability distribution (if any) did you choose? Briefly explain.

   Since all the distributions fit the age at first drink data poorly, I would not suggest using any of the distributions to model the event times. A nonparametric method would be more appropriate.