STAT 417: Survival Analysis Methods

Computing Assignment 3: Due Friday 1-24-2025 at 11:59pm

Name: Andrew Kerr

Instructions for constructing probability plots, parametric survival curves, and parametric hazard curves in Minitab are provided in the file "Graphing Functions of Survival Time and Computing Probabilities" in the Week 3 module on Canvas.

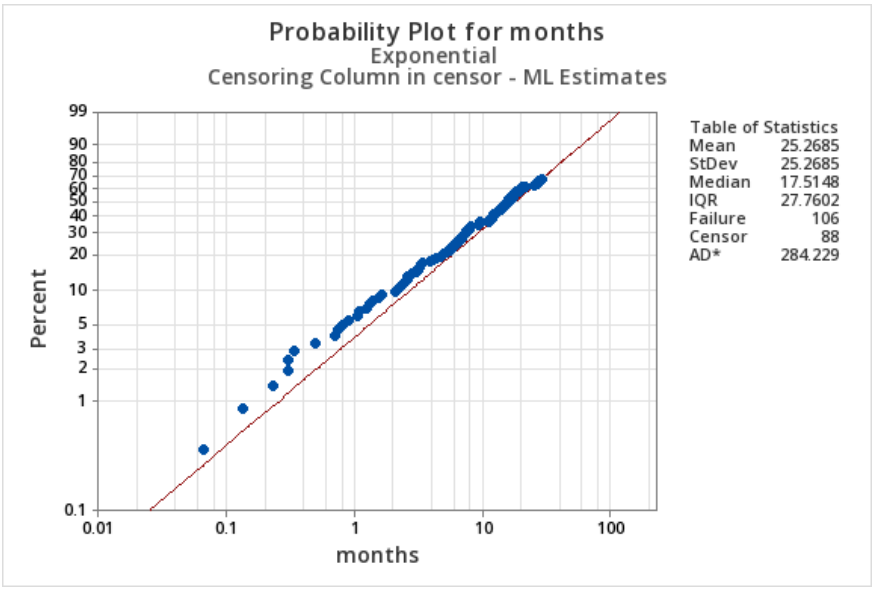
1. Henning and Frueh (1996) followed criminal activities of 194 inmates released from a medium security prison for 36 months. We can use the data from this study to investigate the time until the former inmates were re-arrested. If the former inmate had been re-arrested for a criminal act before 36 months (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 (located in the Canvas folder Computing Assignment 3):

- 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:

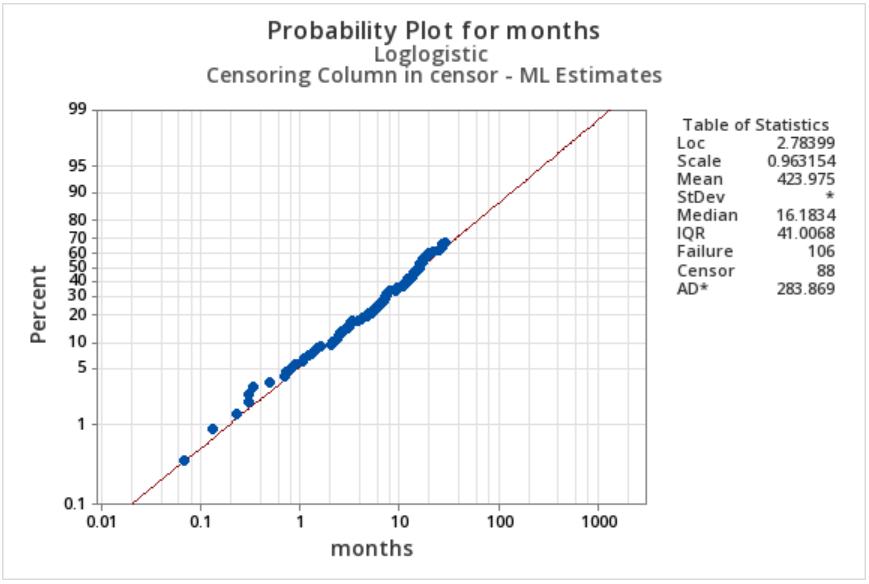
(a) Construct probability plots for each of following distributions, and report the maximum likelihood estimates of the parameters. Be sure to indicate if the parameters are location, scale, or shape:

i. Exponential:



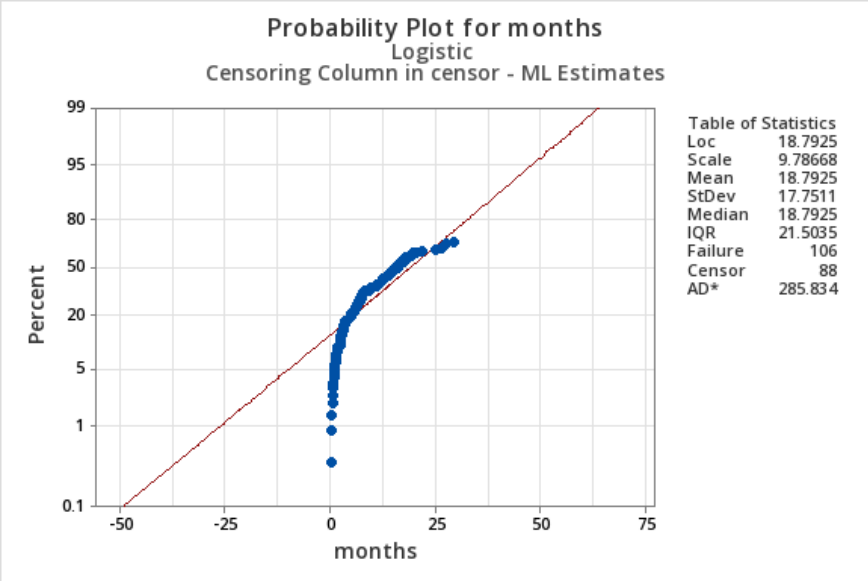
MLE Mean = 25.2685

ii. Loglogistic:



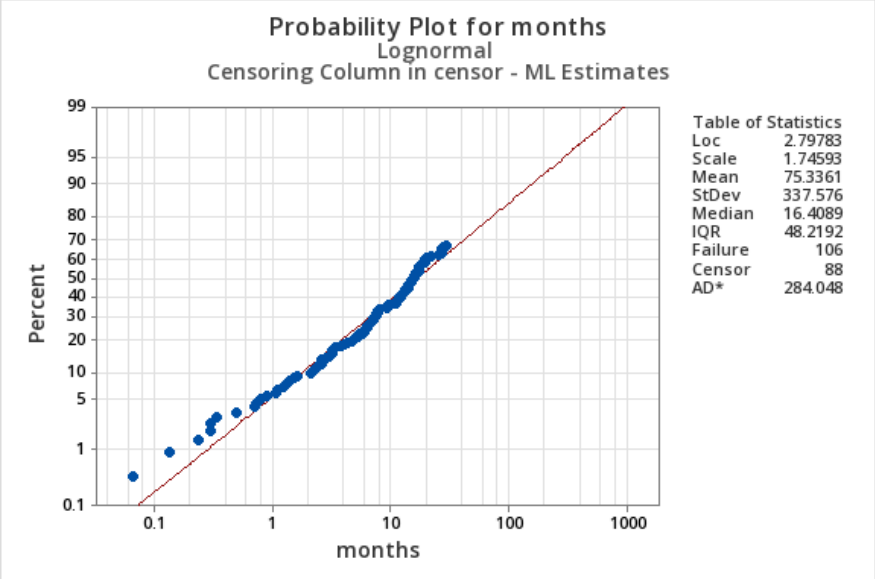
MLE Location = 2.78399
MLE Scale = 0.963154

iii. Logistic:



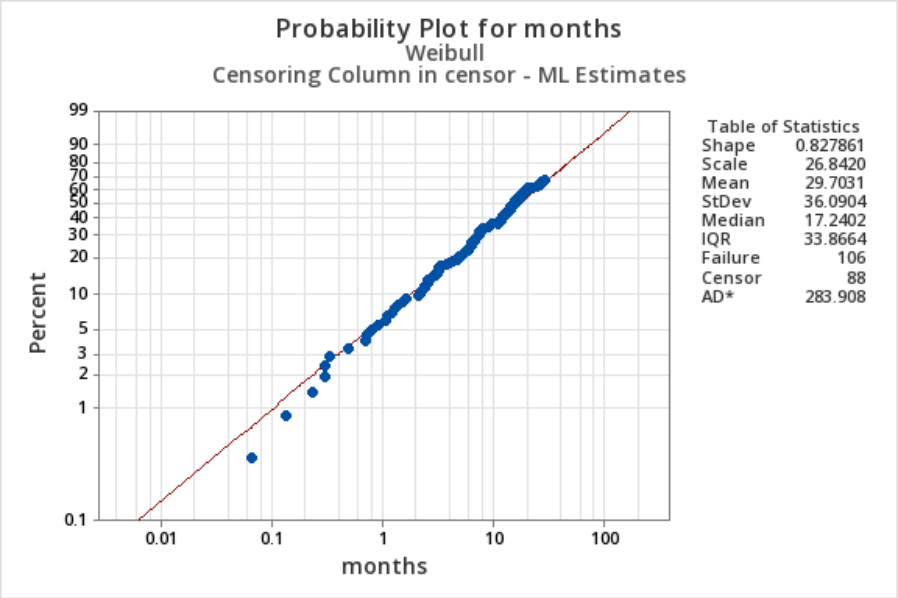
MLE Location = 18.7925
MLE Scale = 9.78668

iv. Lognormal:



MLE Location = 2.79783
MLE Scale = 1.74593

v. Weibull:



MLE Location = 0.827861
MLE Scale = 26.8420

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

Loglogistic fits the data the best with the smallest AD test statistic value of 283.869. Additionally, it passes the visual check with the point lying close to the line!

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

Logistic fits the data the worst with the largest AD test statistic value of 285.834. Looking at the probability plot visually, the points do not lay close to the line with the left tail turning away.

2. For the survival time distribution that fits the data the best in Problem (1), 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 computed mean time to failure is not very useful in the context of the rearrest study.

Mean: 423.975 months
Median: 16.1834 months

The computed mean time to failure is not very useful because we are assuming a logistic distribution, which is heavily right skewed. The mean is sensitive to this skew, leading to its high estimation. However, the median is not affected by the skew since it is set at the point where half the population has experienced the event.

(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.

20th percentile: 4.25788 months

20% of former inmates were re-arrested within roughly 4.26 months after being released from prison.

(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.

$P(T > 11) \approx 1 - 0.4 = 0.6$

(d) Now construct parametric survival curves for former inmates who did and did not commit person-related crimes (on the same graph). What do the survival curves suggest about former inmates who had committed person-related crimes versus former inmates who did not commit person-related crimes?

People who committed person-related crimes have a lower survival probability at all times t than people who did not commit person-related crimes, meaning that they are likely to be re-arrested sooner.

(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).

Non-personal:

Mean: 196.783 months

Median: 19.5715 months

Personal:

Mean: *

Median: 10.2493 months

Looking at the median times, people with personal-related crimes have a lower time to re-arrest than those with non-personal-related crimes which is consistent with (d).

(f) Construct the hazard curves (on the same graph) for former inmates who did and did not commit person-related crimes. What do these curves suggest about being re-arrested?

The curves both start with a peak into a sharp decline. That for personal-related crimes peaks higher, suggesting that people with this type of crime have a higher risk of re-arrest immediately after release. The decline of both groups suggest that, over time, the risk of re-arrest decreases, and since the two curves are practically on-top of each other during and after this decline, this suggests that the risk of being re-arrested is the same regardless of whether the crime was personal-related or not after immediate release.