STAT 417: Survival Analysis Methods

Computing Assignment 2: 1-15-2025

Due: 1-16-2025 11:59pm

Name: Andrew Kerr

Minitab instructions for the lab questions are provided at the end.

1. **[Chocolate Chips Part I].** Define a time-to-event random variable T = seconds until a chip completely melts.

(a) Under the assumption that T follows an exponential distribution with scale parameter λ = 95, use Minitab to find:

i. The probability that a randomly selected chip remains unmelted for longer than 60 seconds.

S(60) = **0.5318**

ii. The probability that a randomly selected chip takes less than 90 seconds to melt.

F(90) = **0.6122**

iii. The probability that a randomly selected chip will take between 35 and 45 seconds to melt.

P(35 < T < 45) = F(45) - F(35) = **0.06912**

(b) Under the assumption that T follows a log-normal distribution with location parameter μ = 4.3 and scale parameter σ = .2, use Minitab to find:

i. The probability that a randomly selected chip will take longer than 60 seconds to melt.

S(60) = **0.8481**

ii. The probability that a randomly selected chip takes less than 35 seconds to melt.

F(35) = **0.00009833**

iii. The probability that a randomly selected chip will take between 35 and 45 seconds to melt.

P(35 < T < 45) = F(45) - F(35) = **0.00672**

(c) Under the assumption that T follows a Weibull distribution with scale parameter λ = 77.0 and shape parameter β = 7.1, use Minitab to find:

i. The probability that a randomly selected chip will take less than 80 seconds to melt.

F(80) = **0.7307**

ii. The probability that a randomly selected chip takes longer than 70 seconds to melt.

S(70) = **0.6015**

iii. The probability that a randomly selected chip will take between 50 and 70 seconds to melt.

P(50 < T < 70) = F(70) - F(50) = **0.3529**

2. **[Chocolate Chips Part II].** Recall the chocolate chip melting activity. The times to melt, as well as the censoring status variable have been saved in the Minitab file MELT TIMES W2025 V2 (located in the Computing Assignment 2 folder ). Use this data set to answer the following questions. Note: the time variable is located in the column Seconds, and the censoring status variable is located in the column C. The event time is censored if it’s corresponding censoring value is 0.

(a) Under the assumption that T follows a lognormal distribution, use Minitab to construct a graph of the survival function S(t) and estimate the probability that a randomly selected chip remains unmelted for at least 50 seconds (you do not have to sketch the curve).

S(50) ≈ **0.8**

(b) Under the assumption that T follows an exponential distribution, use Minitab to produce a graph of the survival function S(t) to estimate the probability that a randomly selected chip remains unmelted for at least 50 seconds.

S(50) ≈ **0.5**

3. **[Fruit Flies].** This data set introduced by Partridge and Farquhar (1981), and further analyzed by Hanley (1983) and by Hanley and Shapiro (1994), was originally analyzed for the purpose of investigating the relationship between increased sexual activity of male fruit flies and their lifetime (in days). The data set is located in the Minitab file Fruitfly (located in the Computing Assignment 2 folder. The lifetimes of the male fruit flies are located in the column Longevity, and the censoring status values are located in the column Censor. Let T define the lifetime of a male fruit fly, i.e. the number of days that the fruit fly lived, and assume that T follows an exponential probability distribution. The column Partners contains the number of female mating partners each male fruit fly had during his lifetime and takes the value 0, 1, or 8. Use Minitab to produce the graphs of the three survival curves corresponding to each number of female partners and answer the following:

(a) Briefly comment on what the curves suggest about the relationship between number of partners and survival for male fruit flies.

As the number of partners increases, the survival rate for male fruit flies decreases for male fruit flies of the same longevity (i.e. more partners decreases the survival of the male fruit fly).

(b) The median lifetime is defined to be the value of T such that half the subjects have failed (experienced the event of interest), and half are still alive. Based on this definition, examine the curves and separately approximate the median lifetimes of the fruit flies who have had 0, 1, and 8 female partners. Is your answer consistent with the answer provided in part (a)?

median(0) ≈ 47 days

median(1) ≈ 45 days

median(8) ≈ 35 days

Minitab Instructions for Graphing a Probability Density Function:

1. Open Minitab software

2. On the top menu, follow: Graph → Probability Distribution Plot…

3. Double click “View Single” box

4. In the “Distribution” drop-down menu, select the distribution (e.g. lognormal)

5. Enter the given values of the parameters in the boxes. Note: If the “Threshold:” box is present, then leave that box as is.

6. Click “OK”

Minitab Instructions for Finding Probabilities Associated with a Given Distribution:

1. Open Minitab software

2. On the top menu, follow: Graph → Probability Distribution Plot…

3. Double click “View Probability” box

4. In the “Distribution” drop-down menu, select the distribution (e.g. lognormal)

5. Enter the given values of the parameters in the boxes. Note: If the “Threshold:” box is present, then leave as is.

6. Click the “Shaded Area” tab

7. Under the options for “Define Shaded Area By,” click the “X Value” button

8. Depending on what probability you are requesting, click “Right Tail”, “Left Tail”, etc.

9. Then enter the value(s) of X in the “X Value:” or “X Value 1:” and “X Value 2:” boxes

10. Click “OK”

11. Remark: An alternative way to find probabilities is to go to: Calc → Probability Distributions, and then follow the steps similar to those given above.

Minitab Instructions for Producing the Cumulative Distribution Function and Survival Function:

1. Open the Minitab data file. Note: F (t) and S(t) can only be constructed by using a data file with at least a time variable and a censoring status variable.

2. On the top menu, follow: Stat → Reliability/Survival → Distribution Analysis (Right Censoring) → Parametric Distribution Analysis

3. Enter the time variable in the “Variables:” box.

4. Select the probability distribution in the “Assumed Distribution” drop-down menu.

5. Optional: If you want to produce several curves for different values of a grouping variable (e.g. you have the grouping variable gender and you want to examine the survival curves for males and females on the same graph), check the “By variable:” box and enter the grouping variable (e.g. gender).

6. On the right, click the “Censor...” button. Then:

* Enter the censoring status variable in the “Use censoring column:” box by double clicking the variable. Note: you may have to first click inside the box to view all the variables in the file.
* In the “Censoring value:” box, enter the value of the indicator variable that corresponds to a censored event time (usually a value of 0 or 1). Note: in all the course files, a censoring status value of 0 will correspond to a censored event time.
* Click the “OK” box.

7. On the right, click the “Estimate...” button.

* Under the options for “Estimation Method,” select the “Maximum Likelihood” tab.
* Click the “OK” box.

8. On the right, click the “Graphs...” button.

* To produce the graph of the survival function S(t), check the box next to “Survival plot.” To produce the graph of the cumulative distribution function F (t), check the box next to “Cumulative failure plot.”
* Click the “OK” box.

9. Click “OK”