Survival Analysis of Bop It Game Time

What drives Bop It successes?

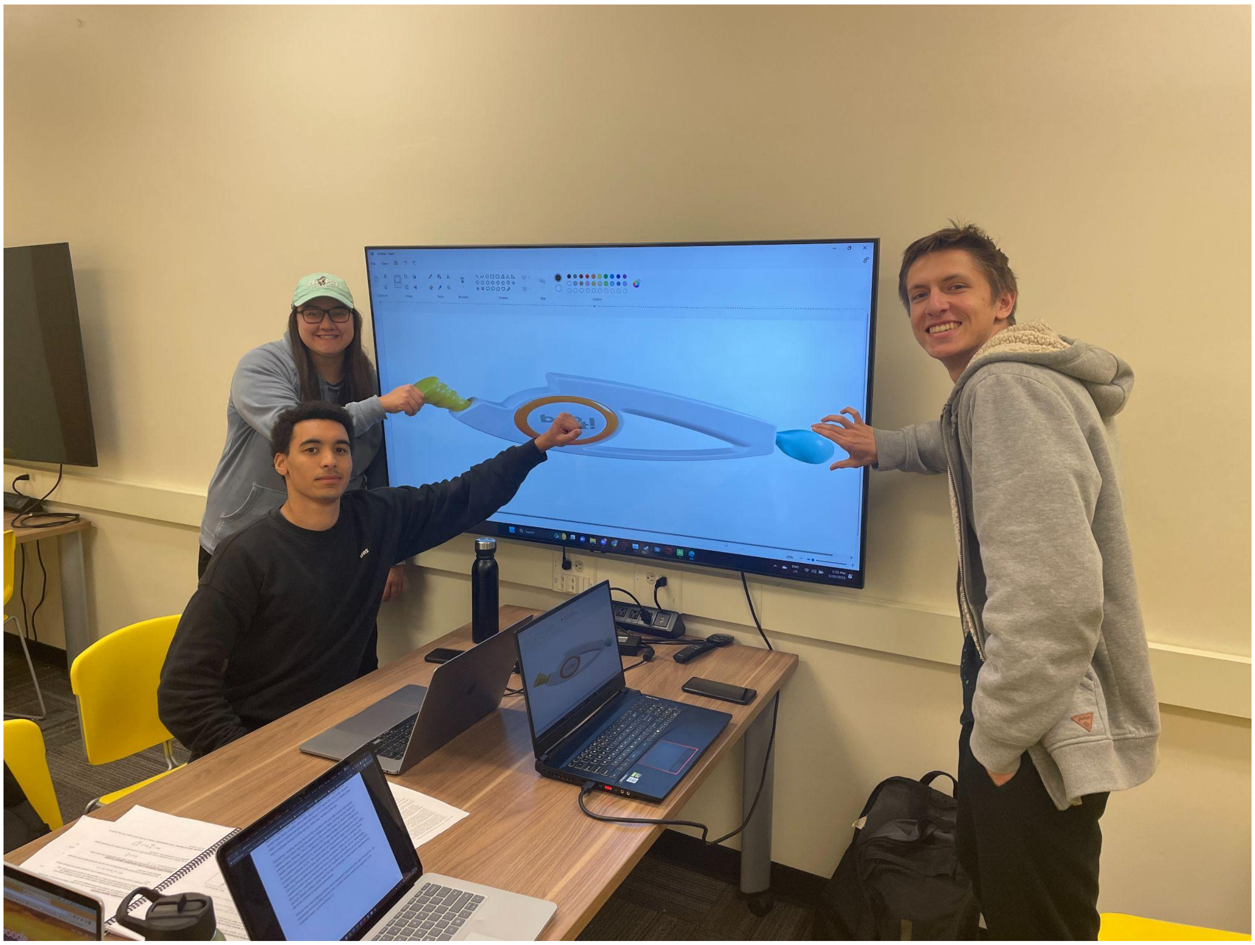
Zachary Felix, Jacob Perez, and Rachel Roggenkemper

College of Science and Math
Statistics
California Polytechnic University San Luis Obispo

Dedicated to
Rao and Blackwell,
Dr. Beth Chance,
and Dr. Bret Holladay

Acknowledgements

We would like to acknowledge our professor, Dr. Jeffrey Sklar. We appreciate your effort in making this class educational and enjoyable throughout the quarter. You have a gift in teaching that is very inspiring and valued. We admire all of your hard work you put into teaching this course, and we all thoroughly enjoyed being your students. We genuinely believe that if you put a fraction of the effort you do into teaching into learning how to play Bop It, you can beat your high score of 1! Please enjoy our report.



Data Ethics

Data is extremely powerful and can be used to educate others on decision making. It is extremely important to consider those affected by the conclusions drawn from data. For example, when working with data for diseases such as COVID-19, it is extremely important to understand that every observation in the data represents a loved one who has sadly passed away. These considerations should help shape the conclusions drawn from the data as well as the presentation of the conclusions. Due to the objective of the study being to analyze the game time of Bop It trials to perform some hands-on analysis of time-to-event data using and implementing the methods learned, the data ethics side of our report is much less trivial than studies for heavy topics like COVID-19 or lung cancer. However, there are still many important considerations for us. Firstly, when collecting our data, we made sure to explain our project in depth to all participants so they could make an informed decision about whether they were willing and comfortable to participate. Secondly, we made it our mission to support all of our participants, regardless of how well they did in their trial run. We did not want anyone to feel bad if they did not get a good score. With these considerations in mind, please view our conclusions as a practice exercise in using survival analysis with time-to-event data and not as a true model to predict one's Bop It game time. As famous British statistician George Box says, "All models are wrong but some are useful."

I. Introduction:

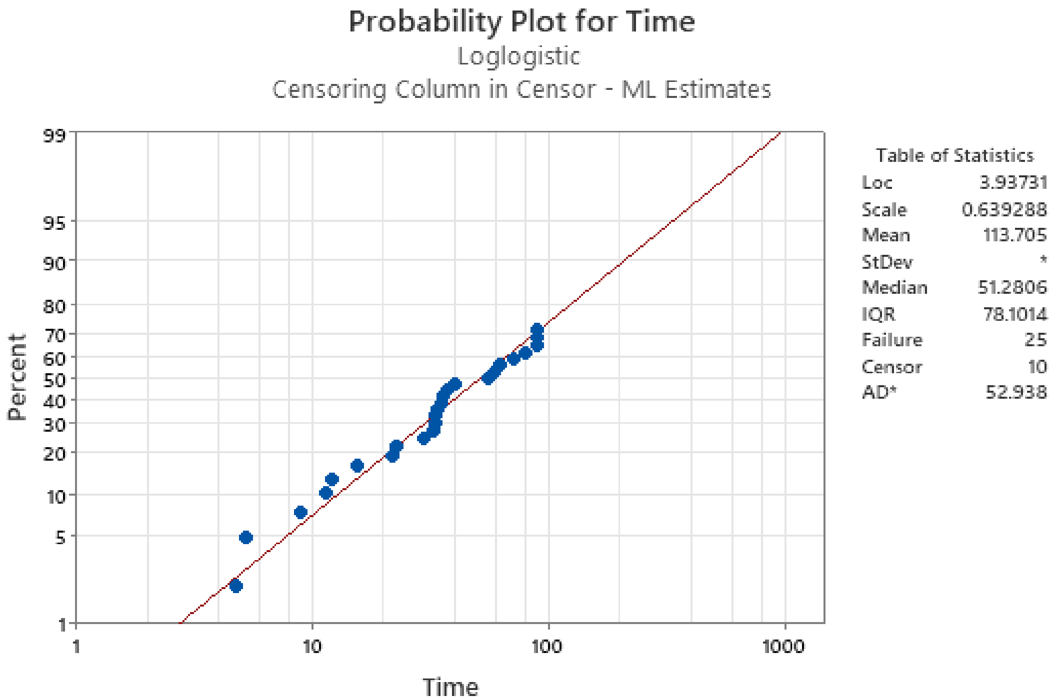
Our project is using data about Bop It, which is a handheld game where the player follows a series of commands issued through voice recordings produced by a speaker by the toy. It has multiple inputs including pressable buttons, pull handles, and twisting cranks. As the player progresses, the pace of the game increases. To play the game, one must respond in time to the commands he or she hears. Every time one responds correctly, they score 1 point, but if they are too slow or respond incorrectly, the game will end.

We collected our own data by asking fellow classmates, friends, and professors who were willing to participate in our study. If they were willing to participate, we explained the concept behind our project, asked them for their age and whether they played with a Bop It before. If they were unfamiliar with the mechanism, we explained which command meant.

Our time-to-event variable is time from the start of the game until the end of the game, and we measure this in seconds. Right censoring occurs because we only allow participants up to 90 seconds to play the game, and if their game has not ended by 90 seconds, they have a right censored tine. There are three explanatory variables we are taking into account. The first explanatory variable is sex, which is a categorical variable with the values “Female” or “Male”. The second explanatory variable is age, which is a continuous variable measured in years. The third explanatory variable is familiarity, which means whether or not they have played with a Bop It before, which is a categorical variable with values “Yes” or “No”.

II. Parametric Survival Analysis:

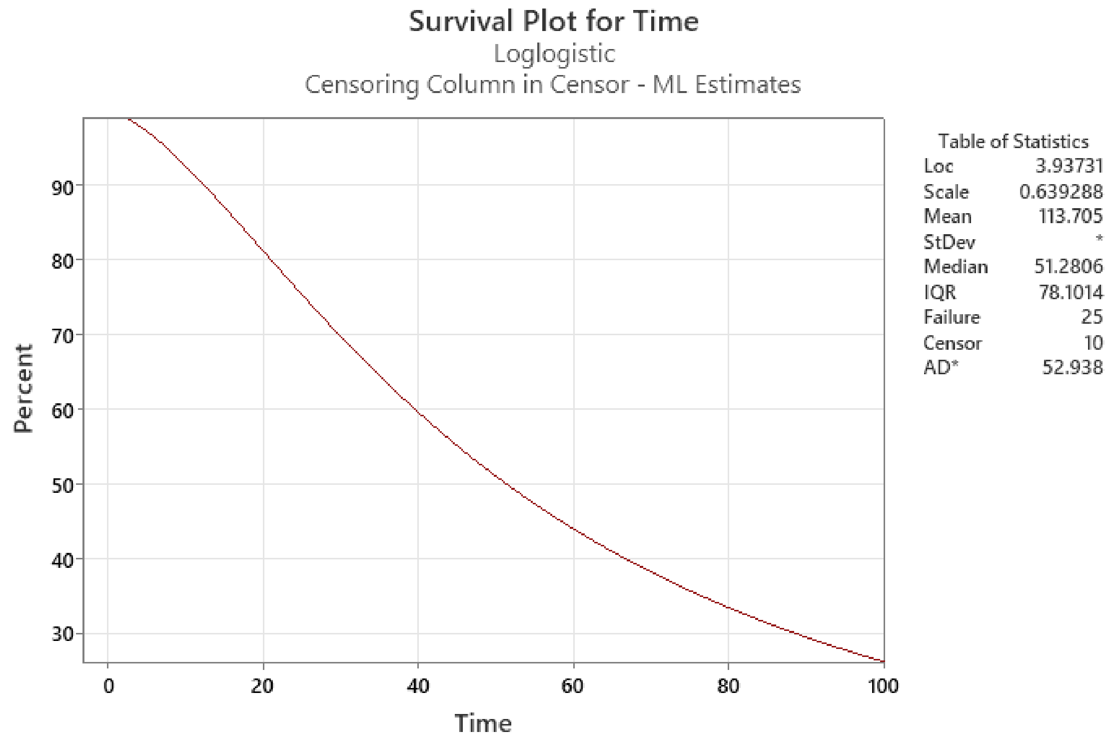
Figure 2.1: Loglogistic Probability Plot



After careful consideration, we ended up choosing a loglogisitc distribution to fit our data. From the five distributions that we fit to the data (exponential, Weibull, lognormal, logistic, and loglogistic), the loglogistic had the lowest Anderson-Darling test statistic, which means that our data fits it best of the five distributions that we attempted. Of note, the probability plots for loglogistic, lognormal, and Weibull distributions appeared to be very similar (the data were quite close to the diagonal line in all three). The Weibull and lognormal distributions even produced the same Anderson-Darling test statistic of 52.959. However, the loglogisitc distribution yielded an Anderson-Darling test statistic of 52.938, so it is the best fit for our data.

Using the MLE output from Minitab, our location parameter (alpha) for the loglogistic distribution is 3.937, and our scale parameter (beta) is 0.639.

Figure 2.2: Survival and Hazard Plots of All Individuals



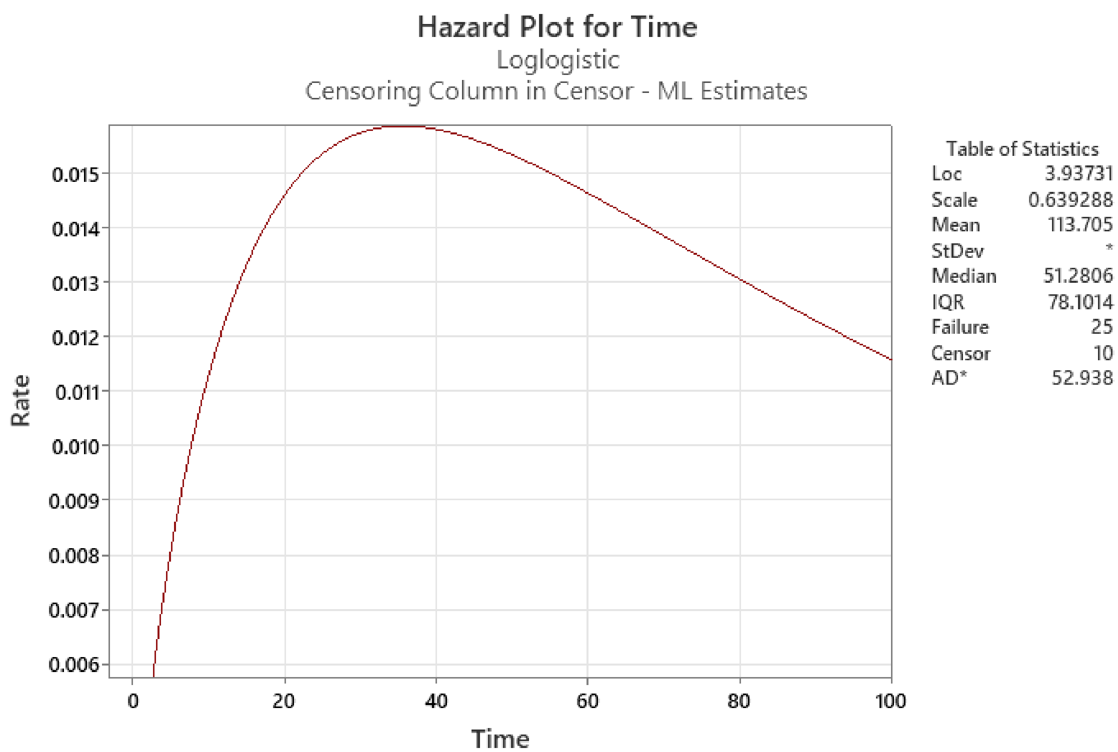
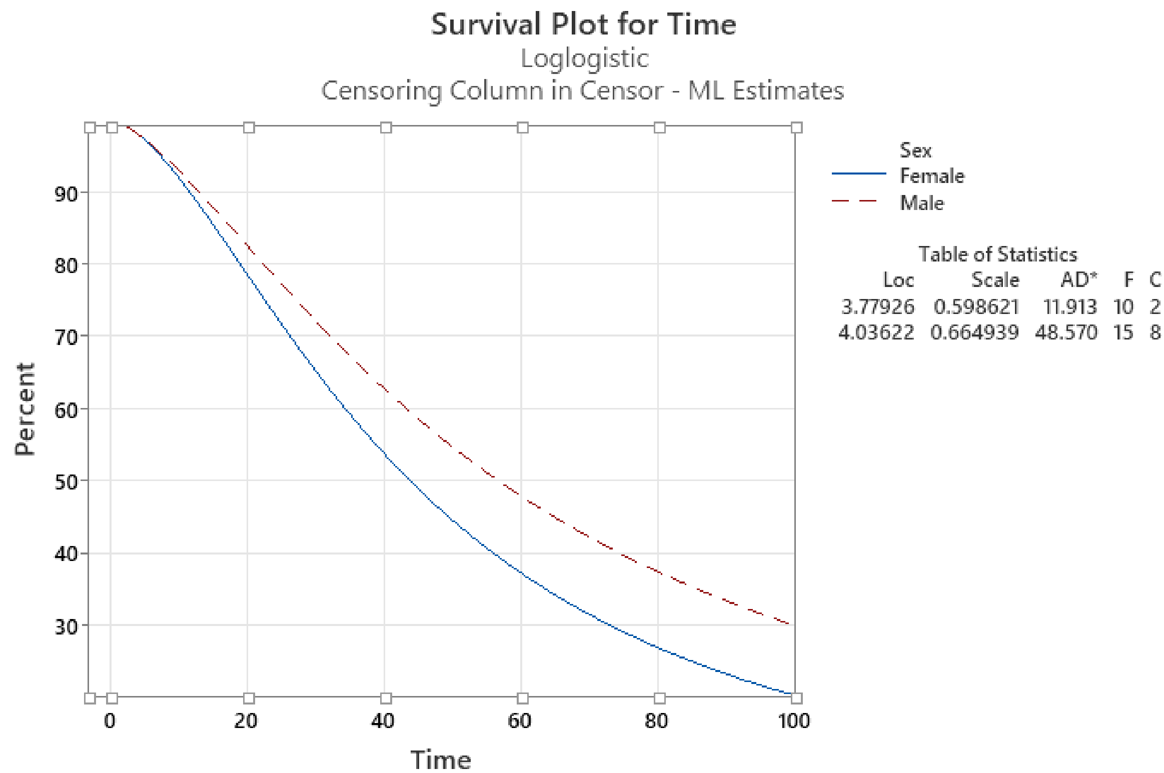


Figure 2.3: Survival and Hazard Plots of Individuals Based on Sex



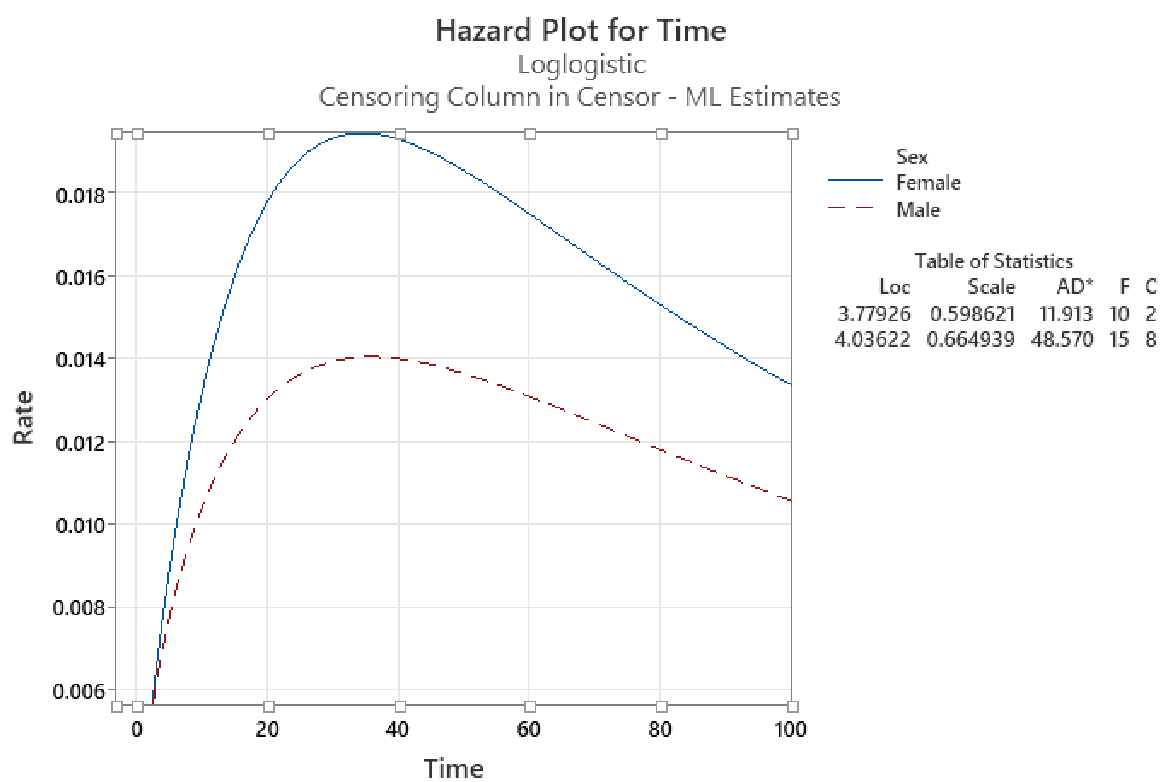
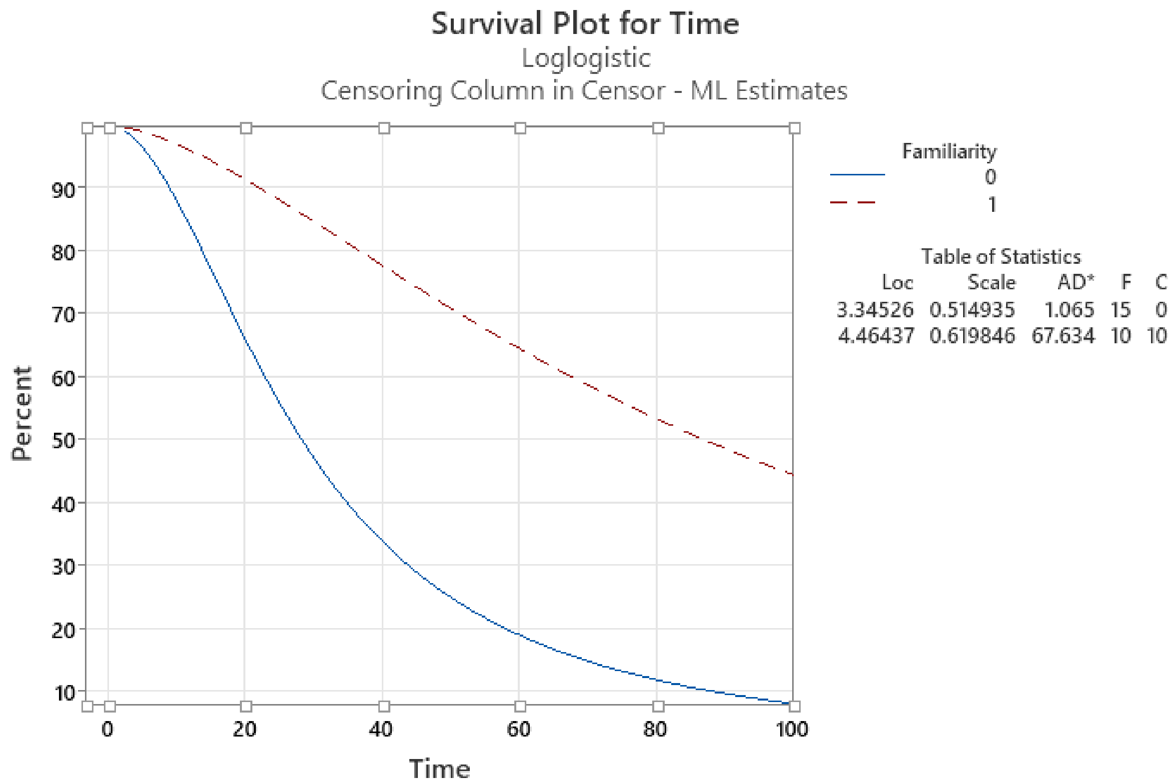
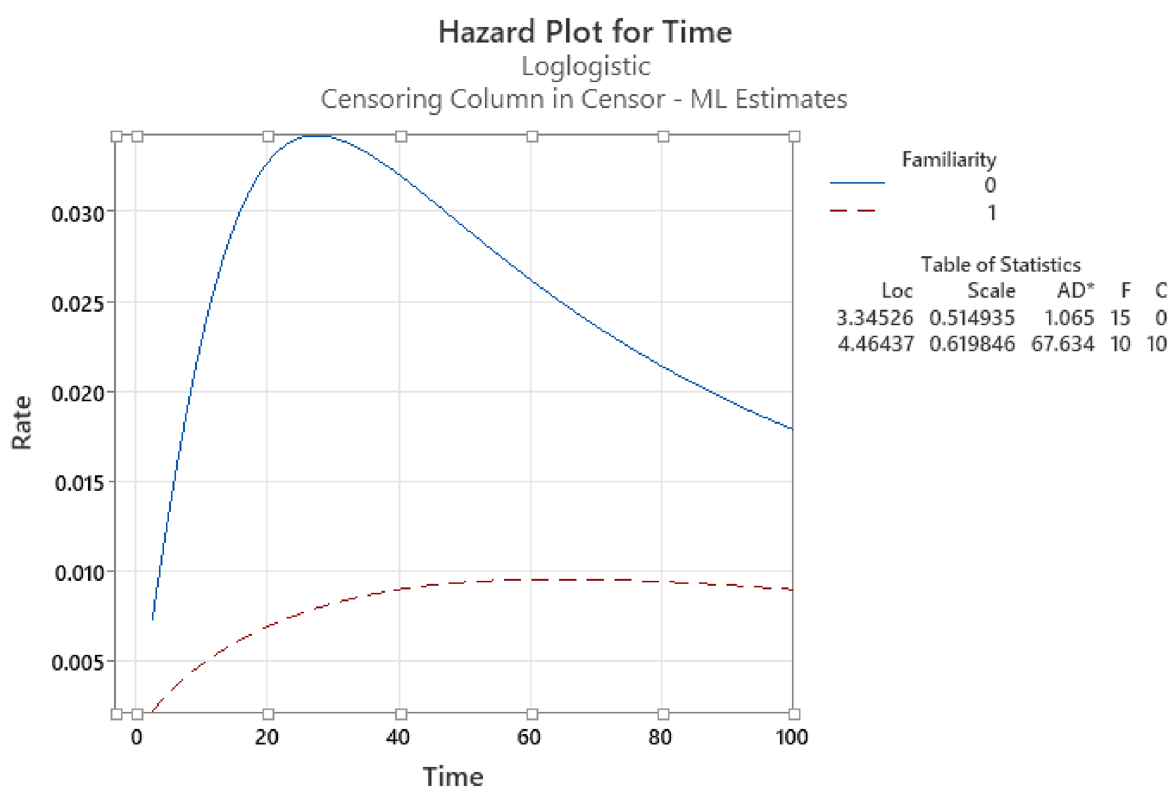


Figure 2.4: Survival and Hazard Plots of Individuals Based on Familiarity





The estimated mean survival time for all individuals is 113.7 seconds, and the estimated mean survival time for all individuals is 51.28 seconds. Looking at Figure 2.2, the survival plot of time for all individuals appears to be decreasing at a somewhat constant rate. Of note, the curve decreases at a faster rate at around 25 seconds, then begins to decrease at a slower rate at about 35 seconds. For the hazard plot, the hazard of failing the Bop-It increases (at a decreasing rate) until a time of about 37 seconds, then decreases at a relatively constant rate until a time of 100 seconds.

After grouping by sex, the estimated mean survival time for females is 86.46 seconds, and the estimated mean survival time for males is 136.131 seconds. The estimated median survival time for females is 43.78 seconds, and the estimated median survival time for males is 56.61 seconds. Observing Figure 2.3, the survival plot of time for males is consistently higher at a given time than it is for females. This means that males typically take longer to fail at Bop-It than females do. For the hazard plot, the hazards of both groups both increase/decrease at the same times, but the rate of increase/decrease is different between males and females. Females have larger rates of increase and decrease than males do, and females consistently have a higher hazard of failure than males as well.

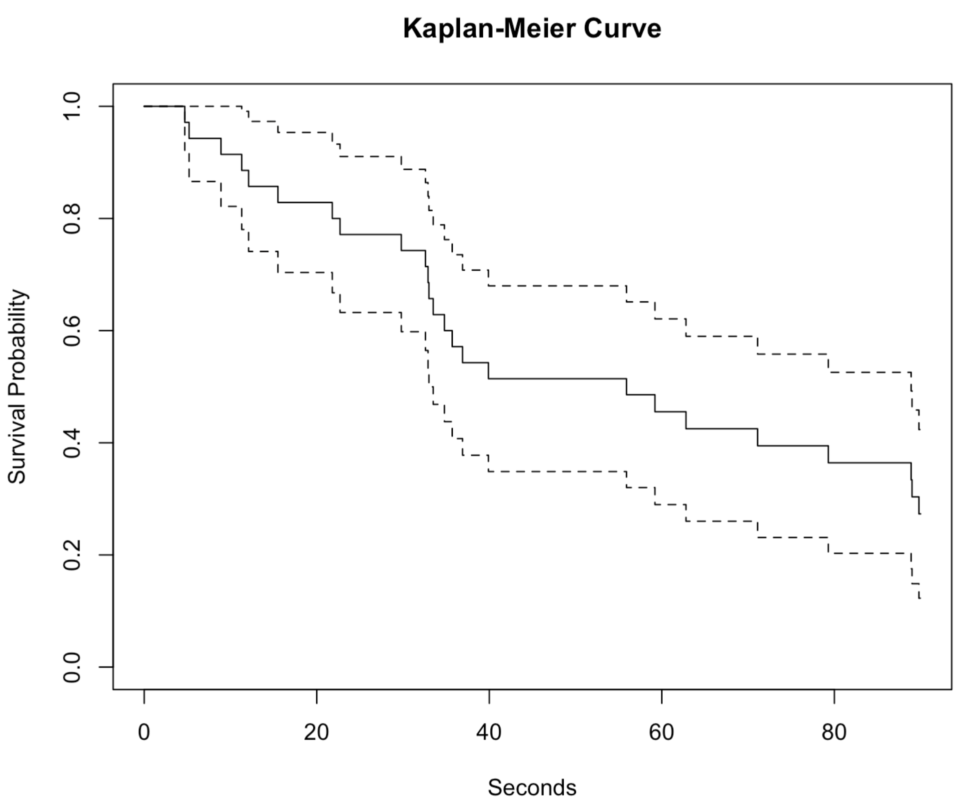
After grouping by familiarity, the estimated mean survival time for those unfamiliar with Bop-It is 45.94 seconds, and the estimated mean survival time for those familiar with Bop-It is 181.9 seconds. The estimated median survival time for those unfamiliar with Bop-It is 28.37 seconds, and the estimated median survival time for those familiar with Bop-It is 86.87 seconds. Looking at Figure 2.4, the survival plot of time for those familiar with Bop-It is consistently considerably higher at a given time than it is for those unfamiliar with Bop-It (much higher than the comparison between males and females discussed earlier). The hazard plot shows that both the hazard of failure for those familiar with Bop-It increases until it levels off at about 60 seconds, then barely decreases until 100 seconds. The hazard of failure for those unfamiliar with Bop-It increases until about a time of 30 seconds, then decreases for the rest of the time (until time is 100 seconds). The hazard of failure for those unfamiliar with Bop-It is always much higher than the hazard of failure for those who are familiar with Bop-It.

III. Non-parametric Survival Analysis:

To test whether our loglogistic distribution is fitting our time to failure Bop It data, we needed to compare our parametric methods used to explain our data to nonparametric methods of explaining our data. Nonparametric methods, such as Kaplan-Meier estimators for survival functions and hazard rates or Nelson-Aalen estimators for the cumulative hazard rate all help to explain our time to event data in ways parametric methods might not.

We started analyzing our data by creating a basic Kaplan-Meier graph of the estimated survival function to represent the probabilities of continuing to play the game for our population. We can see that the probability of continuing to play the game decreases at a constant rate up until about 30 seconds, to which the probability of continuing to play the game decreases rapidly until about 40 seconds. After 40 seconds the probability of survival doesn't start to decrease until about 58 seconds at which the survival probability decreases slowly to about 0.2 at our censored time of 90 seconds.

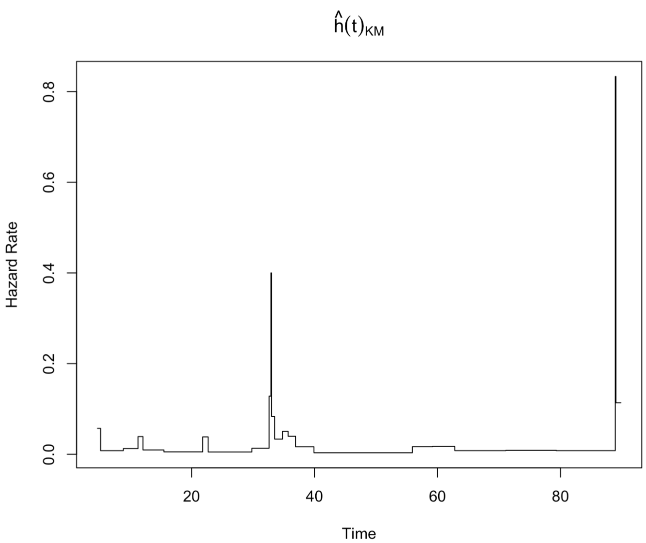
Figure 3.1: Kaplan-Meier Estimated Survival Function for All Individuals



Continuing our exploration of the time to the end of a Bop It game we were able to create a Kaplan-Meier graph of the hazard rate at a given time for individuals playing Bop It. Figure 3.2

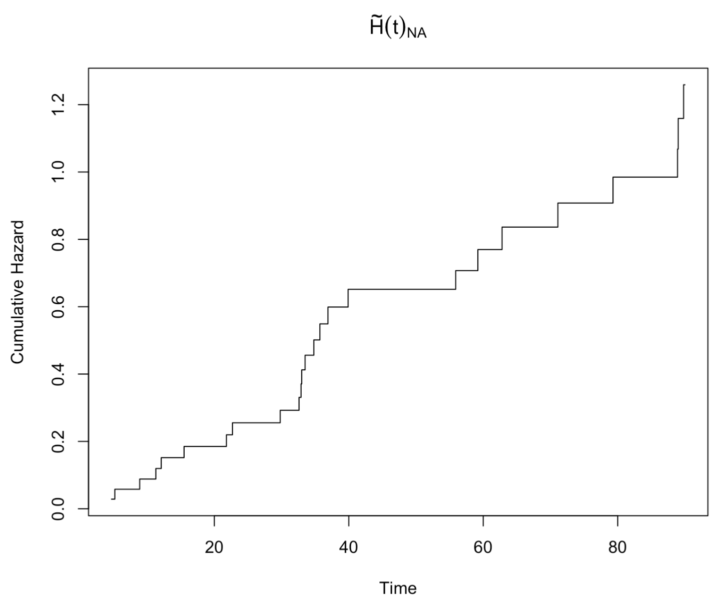
below represents the hazard or risk of losing the game at a time given an individual has survived till the time. Looking at our graph we can see small peaks of hazard or of losing at times 1 second, 10 seconds, and about 22 seconds. However, most of our hazard (excluding the hazard at 90 seconds) is between 30 and 40 seconds, meaning the hazard of not playing the game was highest at those times. This makes sense as we saw in our Kaplan-Meier graph of the survival function the probability of continuing to play decreased sharply between time 30 and 40 seconds. We can also see that the hazard stays very low for times past 40 seconds until 90 seconds which was our censored time. This was again represented by the small decrease in our Kaplan-Meier estimated survival function past 40 seconds. The large hazard at 90 seconds is indicative of most people finishing the game at time 90 seconds as we had 12 censored times. This point can be misleading as the hazard of not continuing to play the game is very high at 90 seconds however, not many players play for 90 seconds.

Figure 3.2: Kaplan-Meier Estimated Hazard Rate for All Individuals



To better represent our hazard over time we can use Nelson-Aalen's estimator for cumulative hazard to see how hazard is increasing over time. In many ways the cumulative hazard function resembles Kaplan-Meiers estimated survival function, just with a different interpretation. The next step in analyzing our data was to graph Nelson-Aalens estimated cumulative hazard function and see if our data continued to trend in the same pattern we have already observed. Looking at our cumulative hazard graph we can see hazard accumulates at a constant rate up until about 30 seconds, which is similar to how our survival probability decreased at a constant rate until about 30 seconds. We can then see that hazard begins to accumulate very rapidly from 30 to 40 seconds, which is consistent with both the hazard rate graph and survival function graph we have looked at before. As we begin to notice as a common trend with our data, hazard accumulates slowly after 40

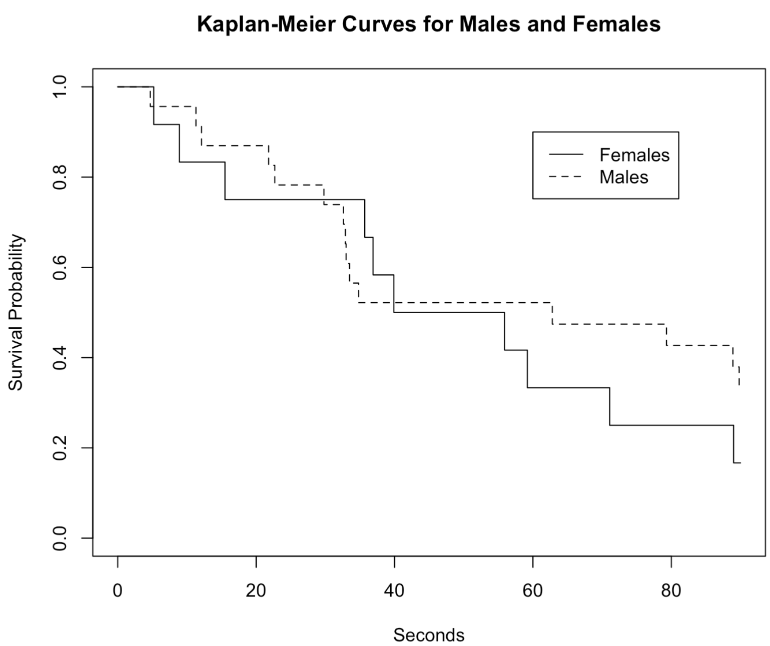
Figure 3.3: Kaplan-Meier Estimated Cumulative Hazard Rate for All Individuals



seconds meaning the hazard of not playing is not increasing by much until 90 seconds. This makes sense when compared to the hazard rate graph as the highest peak for hazard rate was at time 90 seconds.

After looking at the survival probabilities, hazard rates, and cumulative hazard of Bop It times for all individuals we decided to compare Bop It times between different groups of people in our study. By comparing different groups we could see whether the time to end of the game was different for these groups. The first group we decided to compare was males and females to try

Figure 3.4: Kaplan-Meier Estimated Survival Function for Males and Females



and test whether Bop It was a guy's toy or a girls toy. When looking at the survival probabilities of both groups we can see that the survival probabilities are about the same for both males and females. Their survival probabilities follow the same distribution as the overall survival probabilities with the exception of females survival decreasing more than males after 55 seconds. To test whether there was a significant difference between the survival probabilities we were able to use the log rank test to compare the groups. The output of the logrank test was done below indicating that we do not have enough evidence to say that the survival probabilities differ between the two groups.

Table 3.1: Summary Output of Log-Rank Test (Sex)

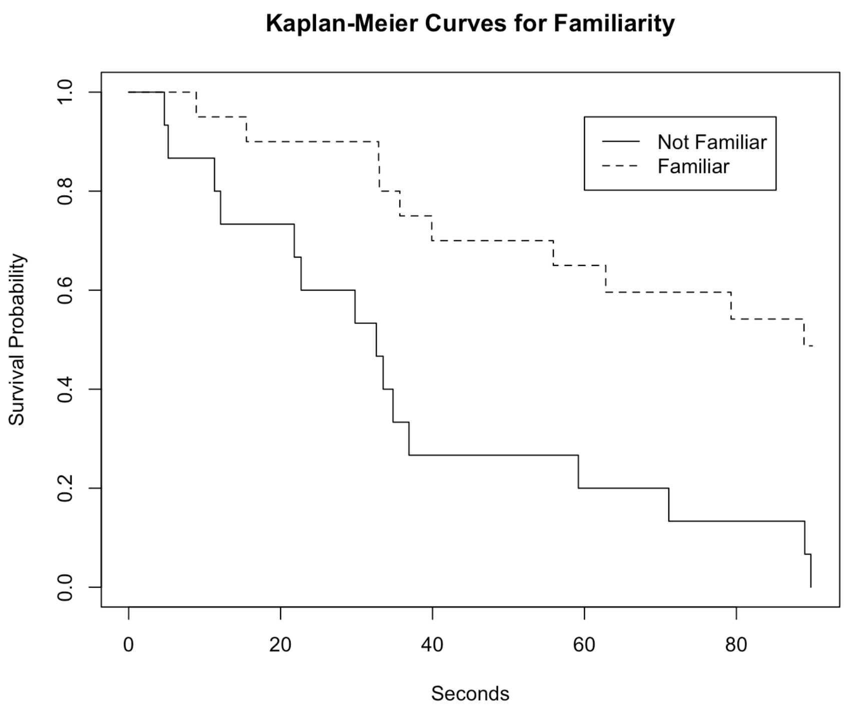
	N	Observed	Expected	Test Statistic
Female	12	10	8.19	0.601
Male	23	15	16.81	0.601

Table 3.2: Log-Rank Test Results (Sex)

Log-Rank Test Statistic	Test Statistic	Degrees of Freedom	p-value
	0.6	1	0.4

The next group we decided to compare were individuals who have played Bop It before against individuals who have not played Bop It before. Our prediction for the comparison was that individuals who have played Bop It before would have a greater survival probability than those who did not. When looking at the graph of the survival probabilities for the two groups we can see that the individuals who were familiar with Bop It had a much higher survival probability than those who were not familiar. We can see that for individuals who are familiar the probability of continuing to play past 90 seconds is almost 0.5 compared to the probability of those who aren't familiar being 0 at 90 seconds. The next step was to once again use the Log-Rank test to see whether the difference in their survival functions was significant.

Figure 3.5: Kaplan-Meier Estimated Survival Function for Individuals Based on Familiarity



Looking at our log-rank results for our two familiar groups we can see that there is a significant difference between the two survival functions. The output helps us to see that indeed individuals who were familiar with Bop It had a higher probability of continuing to play the game than those who are not familiar with Bop It.

Table 3.3: Summary Output of Log-Rank Test (Familiarity)

	N	Observed	Expected	Test Statistic
Not Familiar	15	15	7.11	12.7
Familiar	20	10	17.89	12.7

Table 3.4: Log-Rank Test Results (Familiarity)

Log-Rank Test Statistic	Test Statistic	Degrees of Freedom	p-value
	12.7	1	0.000

Finally, when comparing our non parametric methods of analyzing our time to end of game for different groups and for all individuals we can see that our non-parametric and parametric analysis tell us similar stories. Both overall survival probabilities decrease faster from times 30 to 40 seconds and start to level off after 40 seconds. Additionally, when we look at the graphs of survival curves for gender we see a similar story between the loglogistic model and the Kaplan-Meier estimated survival function. Both the male and female survival probabilities are similar until later times which is represented in both the parametric and non-parametric methods. Our last finding of a significant difference between familiar and unfamiliar individuals was also represented accurately in our parametric findings. When looking back at Figure 2.4 we can see the survival functions for the two groups looking similar to our non-parametric survival functions for individuals separated by familiarity.

IV. Regression Analysis:

In order to investigate the effects of the predictors on the hazard rate of losing a game of Bop It, we first tested many different Cox regression models. We tested different combinations of the predictors, while also testing the inclusion and exclusion on interaction effects between the different predictors. In the end, our final model includes the predictors: *Age*, *Sex*, and *Familiarity*. After deciding on our final model, we calculated the summary output of our model in R, which can be seen in Table 4.1.

Table 4.1: Summary Output for CR Model

Predictor	Coefficient	exp(Coefficient)	Lower 95% CI	Upper 95% CI	Standard Error	z-score	p-value
Age	0.0512	1.0525	1.0046	1.103	0.0238	2.153	0.0313
Sex (Male)	-0.11	0.896	0.372	2.159	0.449	-0.246	0.806
Familiarity (Yes)	-1.545	0.213	0.0872	0.522	0.457	-3.384	0.000715

For our first predictor, *Age*, we will interpret the parameter estimate, estimated hazard ratio, confidence interval for the hazard ratio, and the results of the Wald Test. The estimated coefficient for *Age* is 0.0512, which represents the estimated increase in natural log hazard of losing a game of Bop It associated with a one-year increase in *Age*, adjusting for *Sex* and *Familiarity*. Next, the estimated hazard ratio for *Age* is 1.0525, which means the hazard of losing a game of Bop It is estimated to be 5.25% higher for each one-year increase in *Age*, adjusting for *Sex* and *Familiarity*. The confidence interval for the hazard ratio tells us that we are 95% confident that the hazard of losing a game of Bop It is between 0.46% and 10.3% higher for each one-year increase in *Age*, adjusting for *Sex* and *Familiarity*. Since our confidence interval for the hazard ratio does not include the value 1, we can conclude that the hazard of losing a game of Bop It for *Age* is significantly different for different ages. Lastly, once conducting the Wald test to determine whether *Age* is a significant predictor of hazard of losing a game of Bop It, due to the test statistic of 2.153, and p-value of 0.0313, at the 0.05 level, we can conclude that *Age* is a significant predictor of hazard of losing a game of Bop It, after adjusting for *Sex* and *Familiarity*.

For our second predictor, *Sex*, we will interpret the parameter estimate, estimated hazard ratio, confidence interval for the hazard ratio, and the results of the Wald Test. The estimated coefficient for *Sex* is -0.11, which means the natural log hazard of losing a game of Bop It for males is estimated to be 0.11 lower than the natural log hazard of losing a game of Bop It for males, adjusting for *Age* and *Familiarity*. Next, the estimated hazard ratio for *Sex* is 0.896, which means the hazard of losing a game of Bop It for males is estimated to be 10.4% lower than the hazard of losing a game of Bop It for females, adjusting for *Age* and *Familiarity*. The confidence interval for the hazard ratio tells us that we are 95% confident that the hazard of losing a game of Bop It for males is estimated between 62.8% lower and 115.9% higher than the hazard of losing a game of Bop It for females, adjusting for *Age* and *Familiarity*. Since our confidence interval for the hazard ratio does include the value 1, this means that the hazards of losing a game of Bop It for males and females are not significantly different. Lastly, once conducting the Wald test to determine whether *Sex* is a significant predictor of hazard of losing a game of Bop It, due to the test statistic of -0.246, and p-value of 0.806, at the 0.05 level, we cannot conclude that *Sex* is a significant predictor of hazard of losing a game of Bop It, after adjusting for *Age* and *Familiarity*.

For our third predictor, *Familiarity*, we will interpret the parameter estimate, estimated hazard ratio, confidence interval for the hazard ratio, and the results of the Wald Test. The estimated coefficient for *Familiarity* is -1.545, which means the natural log hazard of losing a game of Bop It for those who have played with a Bop It before is estimated to be 1.545 lower than the natural log hazard of losing a game of Bop It for those who have not played with a Bop It before, adjusting for *Age* and *Sex*. Next, the estimated hazard ratio for *Familiarity* is 0.213, which means the hazard of losing a game of Bop It for those who have played with a Bop It before is estimated to be 78.7% lower than the hazard of losing a game of Bop It for those who have not played with a Bop It before, adjusting for *Age* and *Sex*. The confidence interval for the hazard ratio tells us that we are 95% confident that the hazard of losing a game of Bop It for those who have played with a Bop It before is estimated between 47.8% and 91.28% lower than the hazard of losing a game of Bop It for those who have not played with a Bop It before, adjusting for *Age* and *Sex*. Since our confidence interval for the hazard ratio does not include the value 1, this means that the hazards of losing a game of Bop It for those who have and have not played with a Bop It before are significantly different. Lastly, once conducting the Wald test to determine whether *Familiarity* is a significant predictor of hazard of losing a game of Bop It, due to the test statistic of -3.384, and p-value of 0.000715, at the 0.05 level, we can conclude that *Familiarity* is a significant predictor of hazard of losing a game of Bop It, after adjusting for *Age* and *Sex*.

In addition to investigating the effects of the predictors on the hazard rate, we also tested for the overall significance of the Cox regression model, which can be seen in Table 4.2.

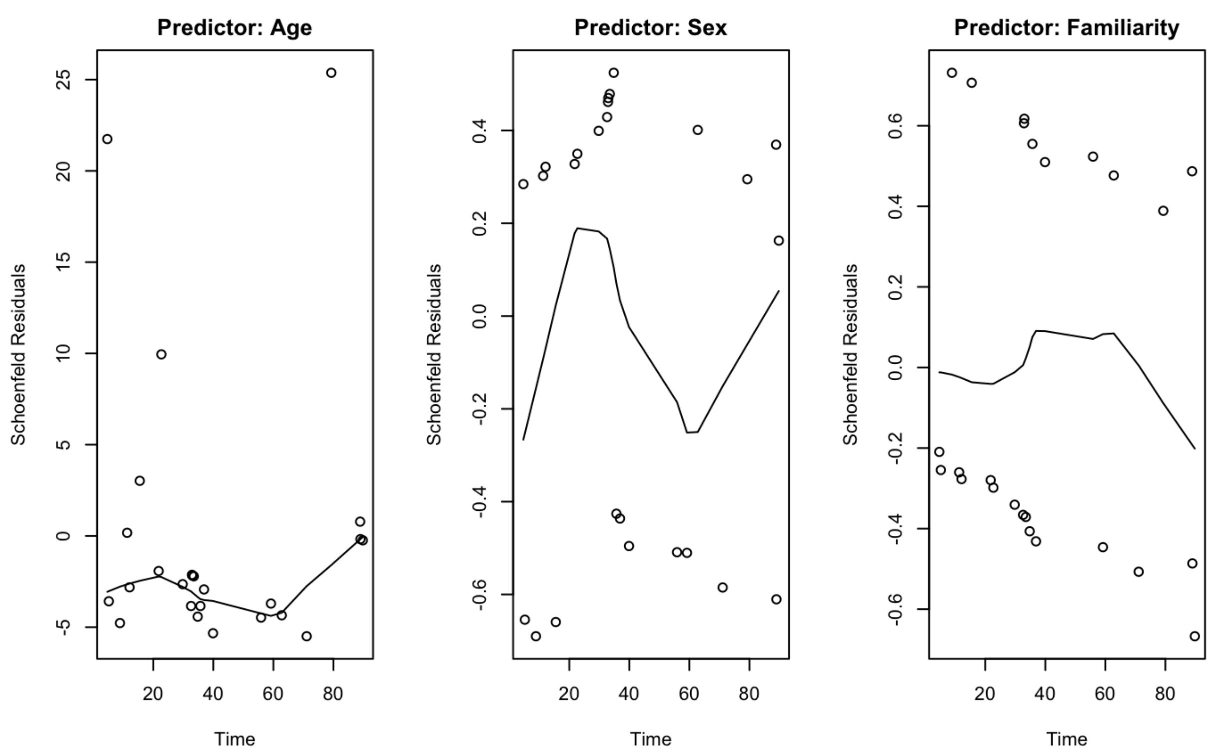
Table 4.2: Summary Output of Partial Likelihood Ratio Test

Partial Likelihood Ratio Test	Test statistic	Degrees of freedom	p-value
	14.83	3	0.002

By performing a Partial Likelihood Ratio Test to assess the overall significance of our Cox regression model, with a test statistic of 14.83 and a p-value of 0.002, we can conclude that *Age*, *Sex*, and *Familiarity* are collectively significantly useful for predicting hazard of losing a game of Bop It.

Now we will assess whether the model assumptions have been met. Before doing this, although we did have a wide range of times for the length of the game of Bop It, none of these times were too influential or outlying, because they the outlying cases were always coupled with other extreme times for really short or really long games, so we decided not to remove any cases and performed our analysis with all data points. To assess the model assumptions of proportional hazards which means that the association of the predictors on hazard does not depend on time, we will be using the Schoenfeld residuals which checks for violations of the proportional hazards assumption for each predictor. The Schoenfeld Residual Plots can be seen in Figure 4.1, and observing the results of a more formal test procedure based on scaled Schoenfeld residuals, which can be seen in Table 4.3.

Figure 4.1: Schoenfeld Residual Plots



By looking at the Schoenfeld residual plot for *Age*, the line appears fairly straight which would mean that the proportional hazard assumption for *Age* has not been violated, adjusting for *Sex* and *Familiarity*. By looking at the Schoenfeld residual plot for *Sex*, the line does not appear straight which would mean that the proportional hazard assumption for *Sex* has been violated, adjusting for *Age* and *Familiarity*. By looking at the Schoenfeld residual plot for *Familiarity*, the line does not appear straight which would mean that the proportional hazard assumption for *Familiarity* has been violated, adjusting for *Age* and *Sex*. However, since we did not have too many observations and looking at residual plots can be quite subjective, we will perform a formal test to assess the proportionality assumption.

Table 4.3: Proportional Hazards Assumption Test Results

	Chi-Square Test Statistic	Degrees of Freedom	p-value
Age	0.36468	1	0.55
Sex	0.0399	1	0.84
Familiarity	0.00424	1	0.95
Global	0.51403	3	0.92

After performing a formal test procedure based on scaled Schoenfeld residuals, due to a large p-value of 0.55 which is larger than 0.05, we can conclude that the proportional hazard assumption is not violated for *Age* after adjusting for *Sex* and *Familiarity*. Similarly, due to a large p-value of 0.84 which is larger than 0.05, we can conclude that the proportional hazard assumption is not violated for *Sex* after adjusting for *Age* and *Familiarity*. Also, due to a large p-value of 0.95 which is larger than 0.05, we can conclude that the proportional hazard assumption is not violated for *Sex* after adjusting for *Age* and *Sex*. Lastly, the global test has a large p-value of 0.92, meaning that the proportional hazard assumption is not violated overall.

Now that we know that we have investigated the effects of our predictors separately and determined that the proportional hazard assumption is not violated in our model, let's investigate one more particular comparison. Since we had students and professors participate in our data collection, we wanted to compute the hazard ratio that corresponds to students versus professors. To do this, we are comparing participants who are 20 years old who are female and have played with a Bop It before (which will represent students) to participants who are 50 years old who are male and have not played with a Bop it before (which will represent professors).

First, we will calculate the estimated hazard ratio using the summary output from Table 4.1.

$$\hat{HR} = \frac{h_0(t)exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3)}{h_0(t)exp(\hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3)}$$

$$\hat{HR} = \frac{exp(0.0512(20) - 0.11(0) - (1.545(1))}{exp(0.0512(50) - 0.11(1) - (1.545(0))}$$

$$\hat{HR} = 0.0513$$

Ultimately, the hazard of losing a game of Bop It for 20 year old females who have played with a Bop It before is estimated to be 94.87% lower than the hazard of losing a game of Bop It for 50 year old males who have not played with a Bop It before.

Next, we will calculate the confidence interval for the hazard ratio of 20 year old females who have played with a Bop It before to 50 year old males who have not played with a Bop It before. To calculate this confidence interval, we will need to first calculate the standard error using output from the Estimated Variance-Covariance Matrix (Appendix 4D).

$$SE(-30\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3) = \sqrt{(-30)^2\hat{V}(\hat{\beta}_1) + (-1)^2\hat{V}(\hat{\beta}_2) + (1)^2\hat{V}(\hat{\beta}_3) + 2(-30)(-1)\hat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + 2(-30)(1)\hat{Cov}(\hat{\beta}_1, \hat{\beta}_3) + 2(-1)(1)\hat{Cov}(\hat{\beta}_2, \hat{\beta}_3)}$$

$$SE(-30\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3) = \sqrt{(900)(0.0005651551) + 0.201491848 + 0.208519463 + (60)(-0.0024194869) + (-60)(-0.0028207106) + (-2)(-0.046520952)}$$

$$SE(-30\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3) = 1.017726$$

Now that we have calculated the standard error, we can now compute the confidence interval for the true hazard ratio for 20 year old females who have played with a Bop It before than 50 year old males who have not played with a Bop It before.

$$95\% CI for HR = exp[(-30\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3) \pm (1.96)(SE(-30\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3))]$$

$$95\% CI for HR = exp[((-30)(0.0512) - (-0.11) + (-1.545)) \pm (1.96)(1.017726)]$$

$$95\% CI for HR = (0.00697, 0.377)$$

This confidence interval means that we are 95% confidence that the hazard of losing a game of Bop It for 20 year old females who have played with a Bop It before is between 62.3% and 99.03% lower than the hazard of losing a game of Bop It for 50 year old males who have not played with a Bop It before. Additionally, because the value 1 is not included in this interval, we can conclude that there is a significant difference in hazard between these two groups.

V. Conclusion:

Across our parametric, nonparametric, and regression analysis, there have been consistent results for our grouping variables.

For sex, males tend to have a longer survival time (not failing Bop-It) at a given time compared to females. This can be observed in the parametric and nonparametric analysis by looking at the survival curves and Kaplan-Meyer curves respectively. This can also be confirmed in the regression analysis by observing the hazard of losing a game of Bop It for males is estimated to be 10.4% lower than the hazard of losing a game of Bop It for females (adjusting for *Age* and *Familiarity*). However, note that sex is not a significant predictor of time until failure of the Bop-It.

Similarly, for familiarity, those familiar with the Bop-It have a longer survival time at a given time than those who are not familiar with the Bop-It. This can be observed in the parametric and nonparametric analysis by looking at the survival curves and Kaplan-Meyer curves respectively. This can also be confirmed in the regression analysis by observing the hazard of losing a game of Bop It for those who have played with a Bop It before is estimated to be 78.7% lower than the hazard of losing a game of Bop It for those who have not played with a Bop It before, adjusting for *Age* and *Sex*. Importantly, familiarity IS a significant predictor of time until failure of the Bop-It.

Of note, at the .05 level, we can conclude that *Age* is a significant predictor of the hazard of losing a game of Bop It, after adjusting for *Sex* and *Familiarity*.

VI. Appendix

Appendix 3A: R Code for Loading in Packages, Data File, and Hazard Functions

```
library(survival)
BopIt <- read.csv("/Users/jacobubafu/Desktop/STAT417/FinalProject/BopIt.csv", header =
TRUE, sep = ",")
plot.haz <- function(KM.obj,plot="TRUE") {
  ti <- summary(KM.obj)$time
  di <- summary(KM.obj)$n.event
  ni <- summary(KM.obj)$n.risk

  #Est Hazard Function
  est.haz <- 1:(length(ti))
  for (i in 1:(length(ti)-1))
    est.haz[i] <- di[i]/(ni[i]*(ti[i+1]-ti[i]))
  est.haz[length(ti)] <- est.haz[length(ti)-1]

  if (plot=="TRUE") {
    plot(ti,est.haz,type="s",xlab="Time",ylab="Hazard
Rate",main=expression(paste(hat(h),(t)[KM])))
  }
  return(list(est.haz=est.haz,time=ti))
}
plot.chaz <- function(KM.obj,plot="TRUE") {
  ti <- summary(KM.obj)$time
  di <- summary(KM.obj)$n.event
  ni <- summary(KM.obj)$n.risk

  #Est Cumulative Hazard Function
  est.cum.haz <- 1:(length(ti))
  for (i in 1:(length(ti)))
    est.cum.haz[i] <- sum(di[1:i]/ni[1:i])

  plot.chaz <- 1:length(KM.obj$time)
  for (i in 1:length(plot.chaz))
    plot.chaz[i] <- sum((KM.obj)$n.event[1:i]/(KM.obj)$n.risk[1:i])

  if (plot=="TRUE") {
```

```
    plot((KM.obj)$time,plot.chaz,type="s",xlab="Time",ylab="Cumulative
Hazard",main=expression(paste(tilde(H),(t)["NA"])))
  }
  return(list(est.chaz=plot.chaz,time=(KM.obj)$time))
}
```

Appendix 3B: R Code for Figure 3.1

```
KM.obj <- survfit(Surv(Time, Censor)~1, conf.type = "plain", data = BopIt)
KM.obj.na <- survfit(Surv(Time, Censor)~1, type="fh", conf.type="none", data = BopIt)
KM.obj.km <- survfit(Surv(Time, Censor)~1, conf.type="none", data = BopIt)
summary(KM.obj)

print(KM.obj)

plot(KM.obj, xlab = "Seconds", ylab = "Survival Probability", main = "Kaplan-Meier Curve")
```

Appendix 3C: R Code for Figure 3.2

```
plot.haz(KM.obj)
```

Appendix 3D: R Code for Figure 3.3

```
plot.chaz(KM.obj)
```

Appendix 3E: R Code for Figure 3.4, Table 3.1, and Table 3.2

```
KM.obj <- survfit(Surv(Time, Censor)~Sex, data = BopIt)
plot(KM.obj, lty = 1:2, xlab = "Seconds", ylab = "Survival Probability", main = "Kaplan-Meier
Curves for Males and Females")
legend(60, .9, c("Females", "Males"), lty=1:2)
survdiff(Surv(Time, Censor)~Sex, data = BopIt)
```

Appendix 3F: R Code for Figure 3.5, Table 3.3, and Table 3.4

```
KM.obj <- survfit(Surv(Time, Censor)~Familiarity, data = BopIt)
plot(KM.obj, lty = 1:2, xlab = "Seconds", ylab = "Survival Probability", main = "Kaplan-Meier
Curves for Familiarity")
legend(60, .95, c("Not Familiar", "Familiar"), lty=1:2)
survdiff(Surv(Time, Censor)~Familiarity, data = BopIt)
```

Appendix 4A: R Code for Loading in Packages, Data File, Table 4.1, and Table 4.2

```
library(survival)
BopIt <- read.csv(file = 'BopIt.csv')
cr.object <- coxph(Surv(Time, Censor) ~ Age + as.factor(Sex) + as.factor(Familiarity),
```

```
                    data = BopIt)
summary(cr.object)
```

Appendix 4B: R Code for Figure 4.1

```
schoen <- residuals(cr.object, type = "schoenfeld")

comp.times <- sort(BopIt[BopIt$Censor != 0,]$Time)

par(mfrow=c(1,3))

plot(comp.times, schoen[,1], xlab = "Time", ylab = "Schoenfeld Residuals", main = "Predictor: Age")
smooth.sres <- lowess(comp.times, schoen[,1])
lines(smooth.sres$x, smooth.sres$y, lty=1)

plot(comp.times, schoen[,2], xlab = "Time", ylab = "Schoenfeld Residuals", main = "Predictor: Sex")
smooth.sres <- lowess(comp.times, schoen[,2])
lines(smooth.sres$x, smooth.sres$y, lty=1)

plot(comp.times, schoen[,3], xlab = "Time", ylab = "Schoenfeld Residuals", main = "Predictor: Familiarity")
smooth.sres <- lowess(comp.times, schoen[,3])
lines(smooth.sres$x, smooth.sres$y, lty=1)
```

Appendix 4C: R Code for Table 4.3

```
cox.zph(cr.object, transform = "log")
```

Appendix 4D: R Code for the Estimated Variance-Covariance Matrix

```
cr.object$var
```