STAT 417 Final Project Report

New York Times Mini Crossword Completion

Alea Seifert, Alie Hall, Jett Palmer, Libby Brill

Table of contents

Introduction

Origin of Data

These data were student-collected by the members of this 417 project. We decided to collect data on completion of the New York Times Mini Crossword, referred to as the Mini henceforth. We recorded the following data on 51 Cal Poly students: time to Mini completion in seconds (up to 300 seconds), whether the time was complete or censored, number of playable squares, the participant's self-identified gender, and student reaction time in milliseconds (according to the Human Benchmark).

Action Plan

This study analyzed the time-to-event variable T, representing the time (in seconds) from when a subject presses "start" to when the mini is complete. The mini has a built-in timer to track participants time that was used for this study. The study was conducted over the course of five days, with each day featuring a different Mini puzzle. Participants were Cal Poly students, and no duplicate individuals were included in the study.

During each session, students attempted to solve the Mini. If they successfully completed the puzzle within 300 seconds (five minutes), we recorded their total completion time. If the five minutes expired before they won, their time was recorded as censored (300+ seconds). Additionally, if a student quit before 300 seconds without solving the puzzle, or turned on the autocheck feature, their time was marked as censored.

We also collected each participant's reaction time using the Human Benchmark reaction time test. This was to see if there is any association between reaction time and Mini performance.

Variables

Our final data set includes the Mini completion time, whether the time was censored or not, the number of squares to be filled in, the participant's self-identified gender, and the participant's reaction time.

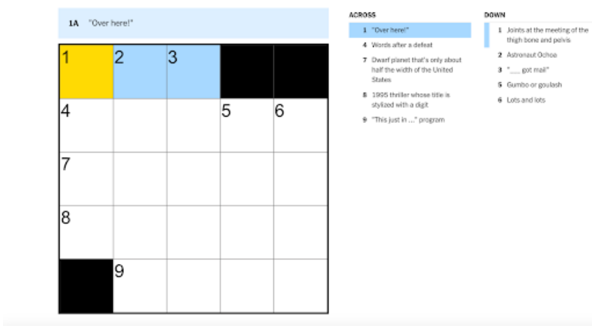
- **time** (quantitative)
  - Starts when the puzzle is opened, when the Mini clock begins counting
  - Measured in seconds, converted from the Mini clock time (which is given in minutes and seconds)
  - Censored at 300 seconds
- **censor** (indicator)

- 0: censored, 1: complete

- **gender** (categorical)
  - Participant self-identified (Female, Male, Non-Binary/Other)

- **reaction_time** (quantitative)
  - Measured in milliseconds, as given by the Human Benchmark reaction time test

- **squares** (quantitative)
  - The number of squares to be filled in
  - For example, the number of squares in the following puzzle would be recorded as 22:

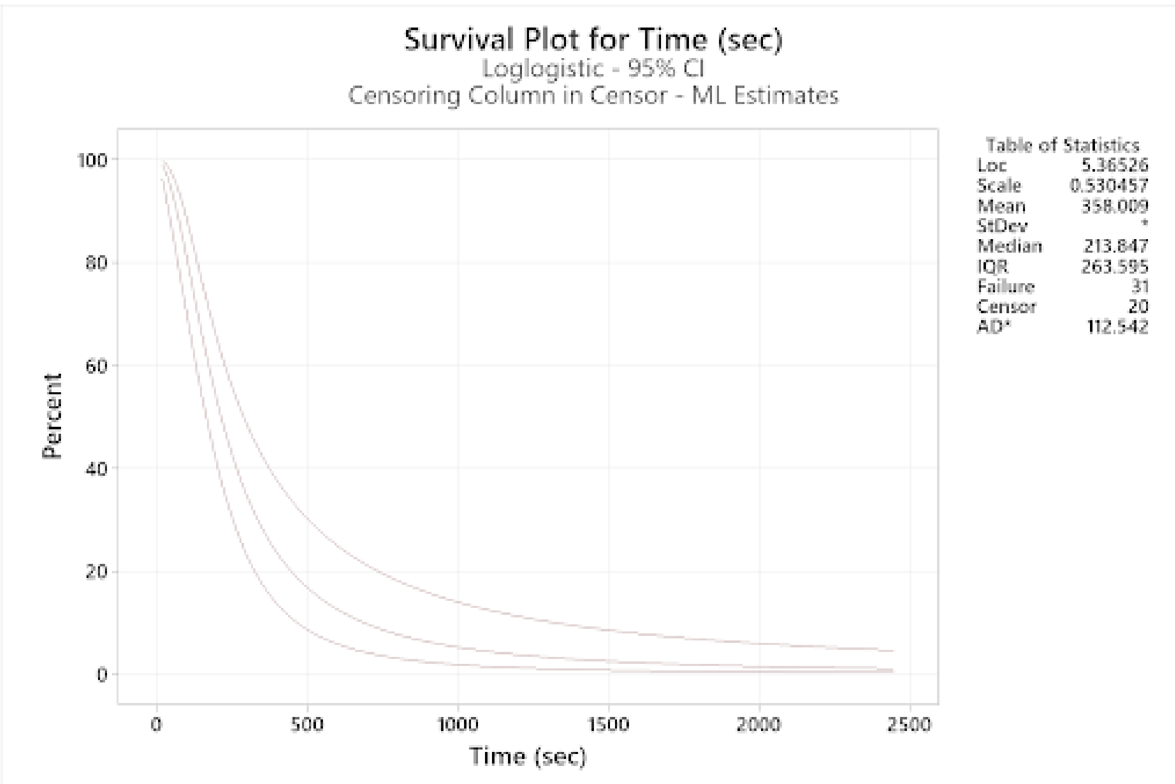


Main Report

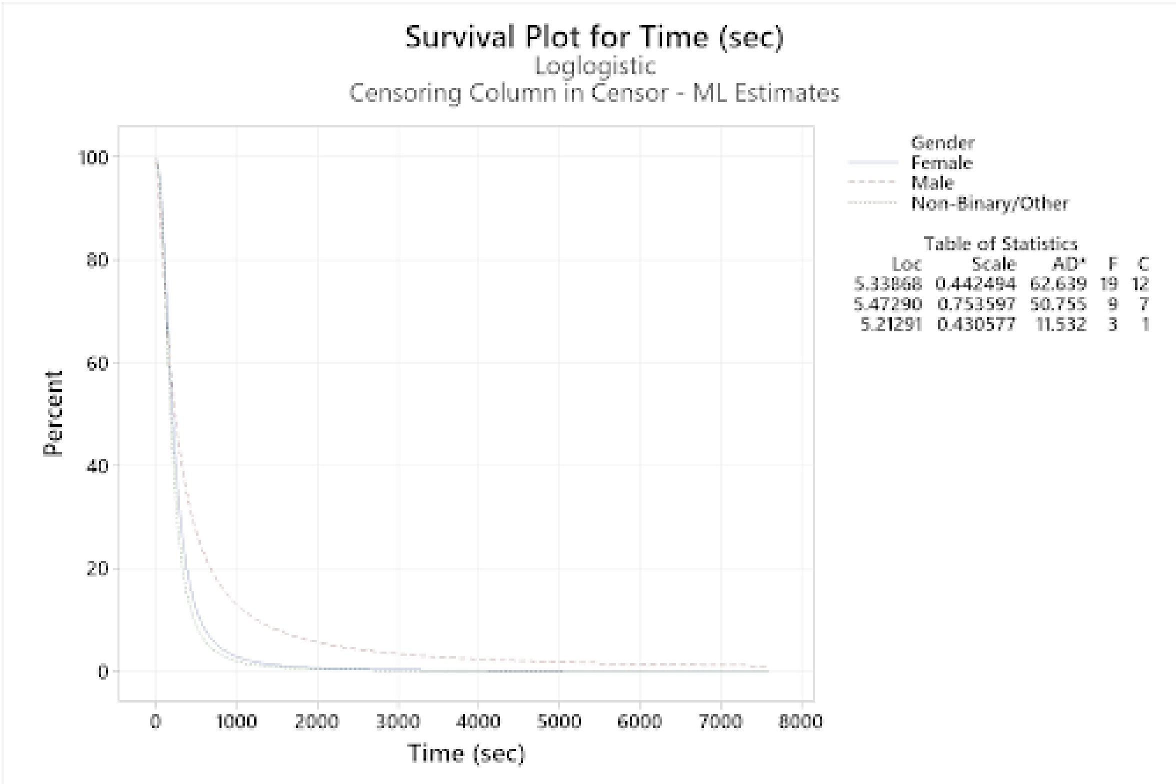
Parametric Survival Analysis

The loglogistic distribution (location = 5.365, scale = 0.530) appears to fit the data best, as supported by the lowest Anderson-Darling statistic (112.542) when compared to other distributions. Using a quantile plot (see appendix), we observed all the observations fell close to the line and between the 95% confidence error bars.

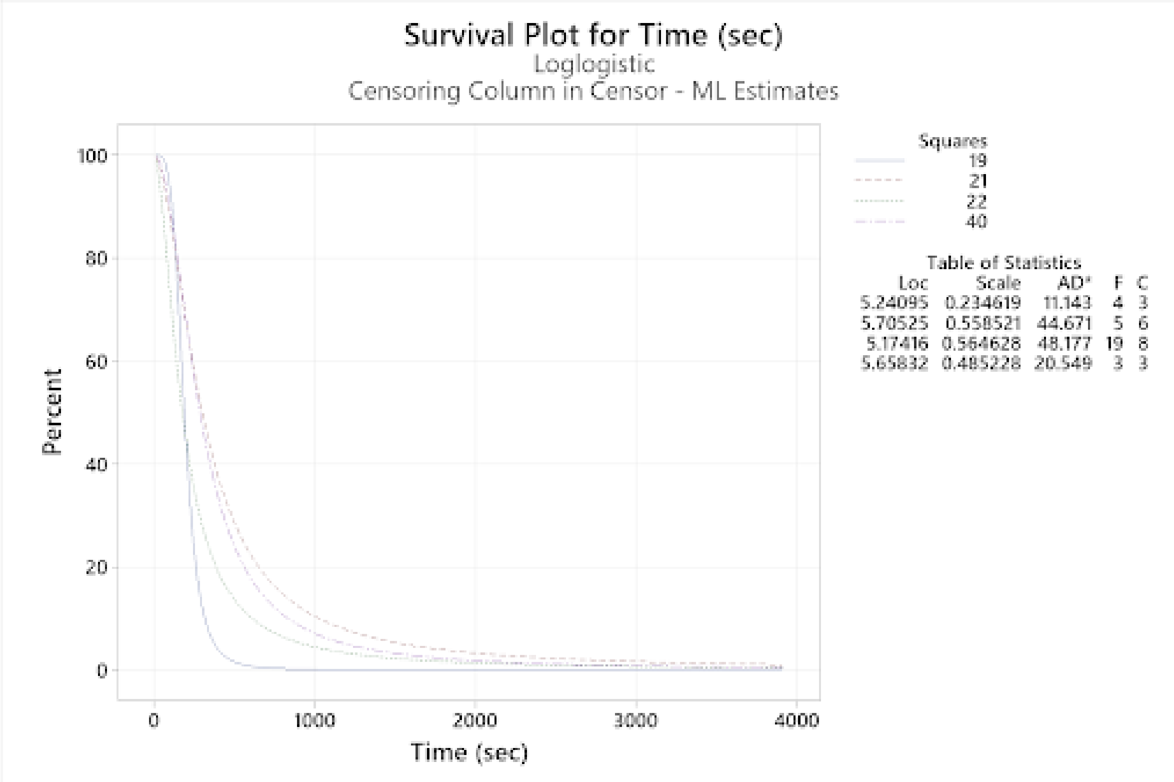
Survival Plots



The probability that a student has not completed the puzzle after time t, given they have not completed it by time t, decreases rapidly, then begins to plateau around 500 seconds.

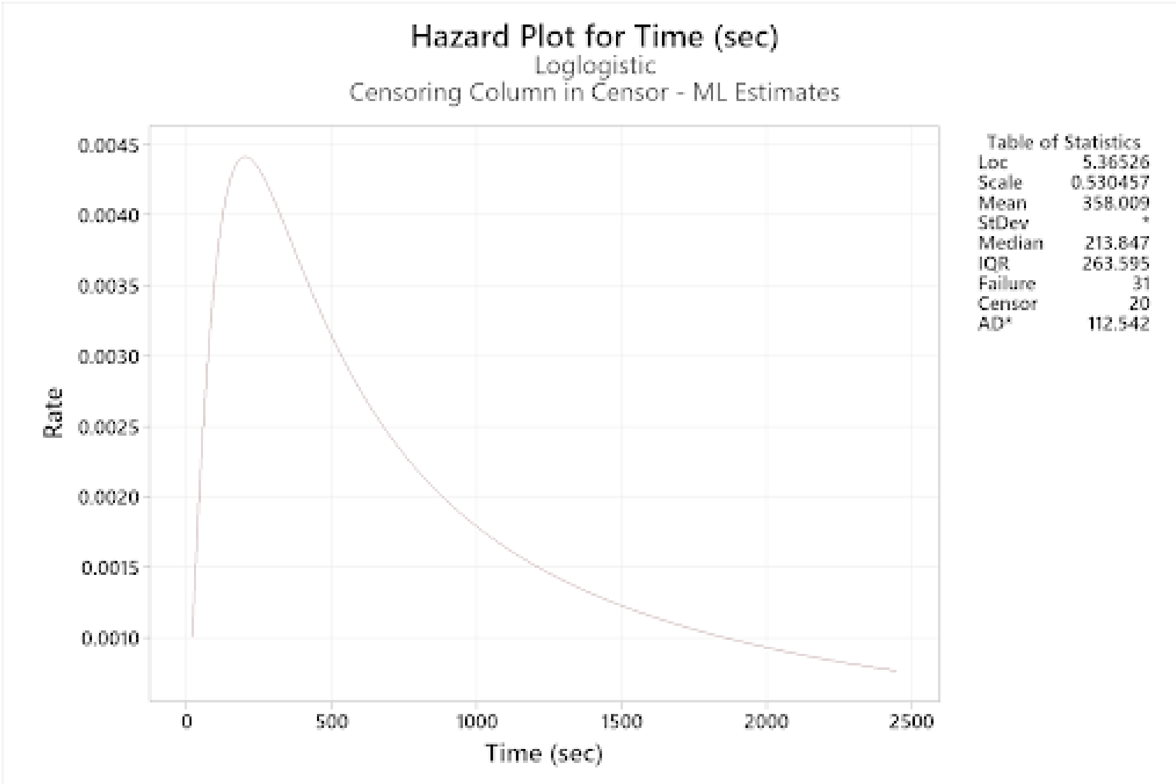


When we compare completion times based on the gender of the student, we observe similar survival curves for female and non-binary students. However, it appears that the probability that male students do not complete the puzzle after time t, given they have not completed it by t, is greater than the corresponding probability for female and non-binary students, across all time.

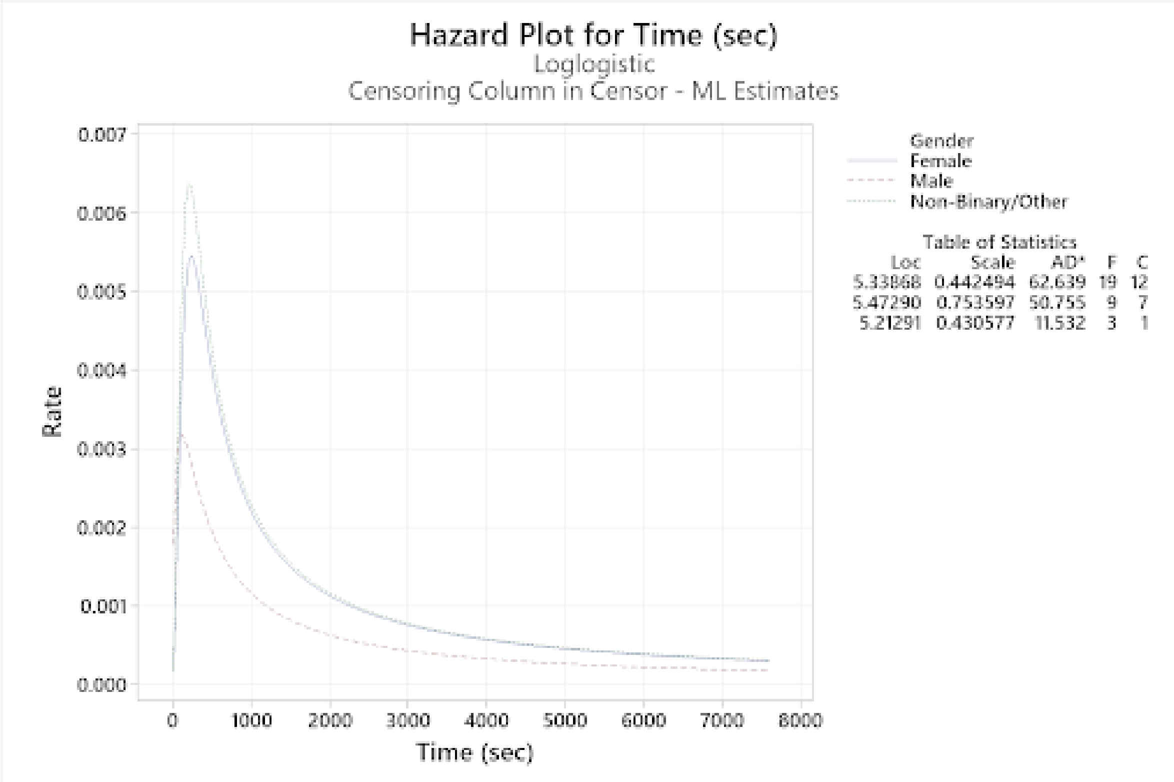


We also compared survival times based on the number of playable squares and found that, given subjects have not completed the puzzle by time t, the probability they do not complete puzzles with 19 squares beyond t is lowest, as compared to puzzles with more playable squares (21, 22, 40).

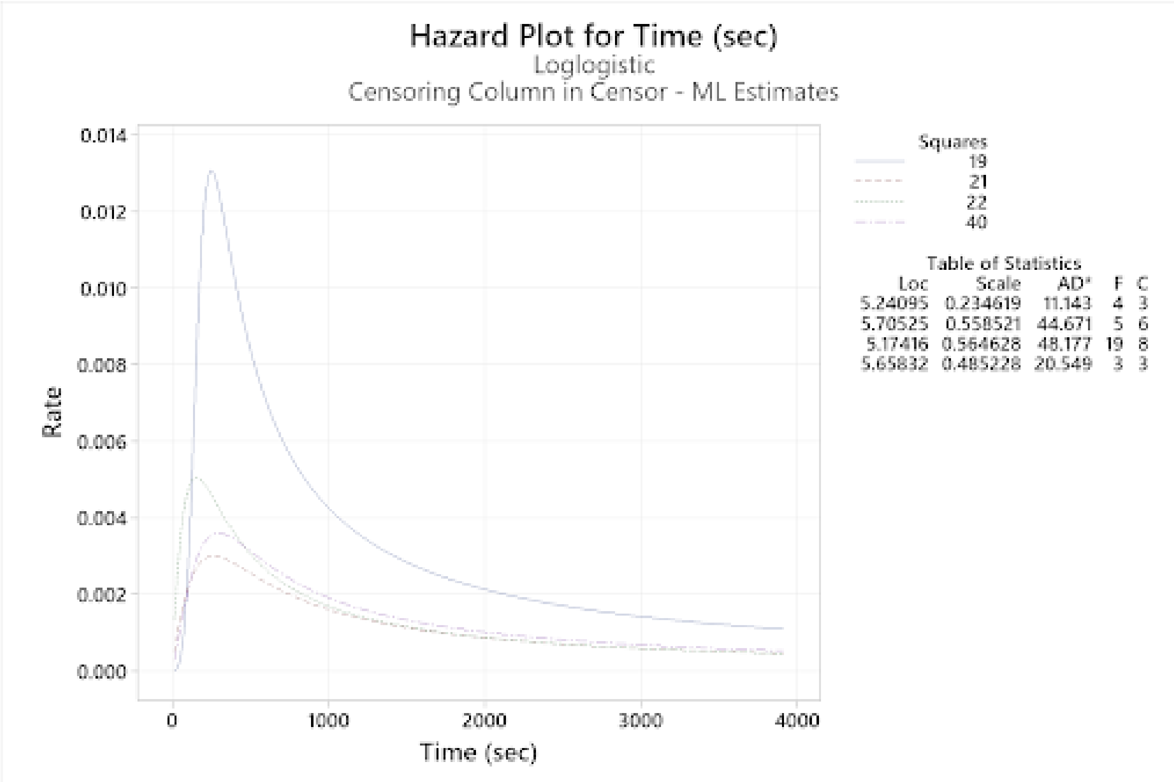
Hazard Plots



The potential of completing the puzzle increases rapidly until roughly 200 seconds. After 200 seconds, the longer students do not complete the puzzle, the lower their potential of completing the puzzle.



Before 200 seconds, the potential of completing the puzzle is lowest for male students and highest for non-binary students. After 200 seconds, the potential to complete the puzzle is roughly the same for female and non-binary students and remains lower for male students for the rest of time.



For all time t, the potential to complete the puzzle with 19 playable squares is notably higher than the potential to complete the puzzle for other size puzzles (21, 22, 40 squares).

Loglogistic Summary Data

Overall

Mean (s.)	Median (s.)
201.666	199

The estimated mean survival time for mini puzzle completion is 201.67 seconds. The estimated time at which 50% of the students have already completed the mini puzzle and 50% have not is 199 seconds.

Gender

	Mean (s.)	Median (s.)
Female	202.792	185
Male	200.313	229
Non-Binary/Other	183.5	120

Female students had a slightly higher mean survival time (202.79 seconds) than males (200.31 seconds), but their median survival time was lower (185 vs. 229 seconds). Non-binary students had the shortest mean and median completion times (183.50 seconds, 120 seconds).

Number of Squares

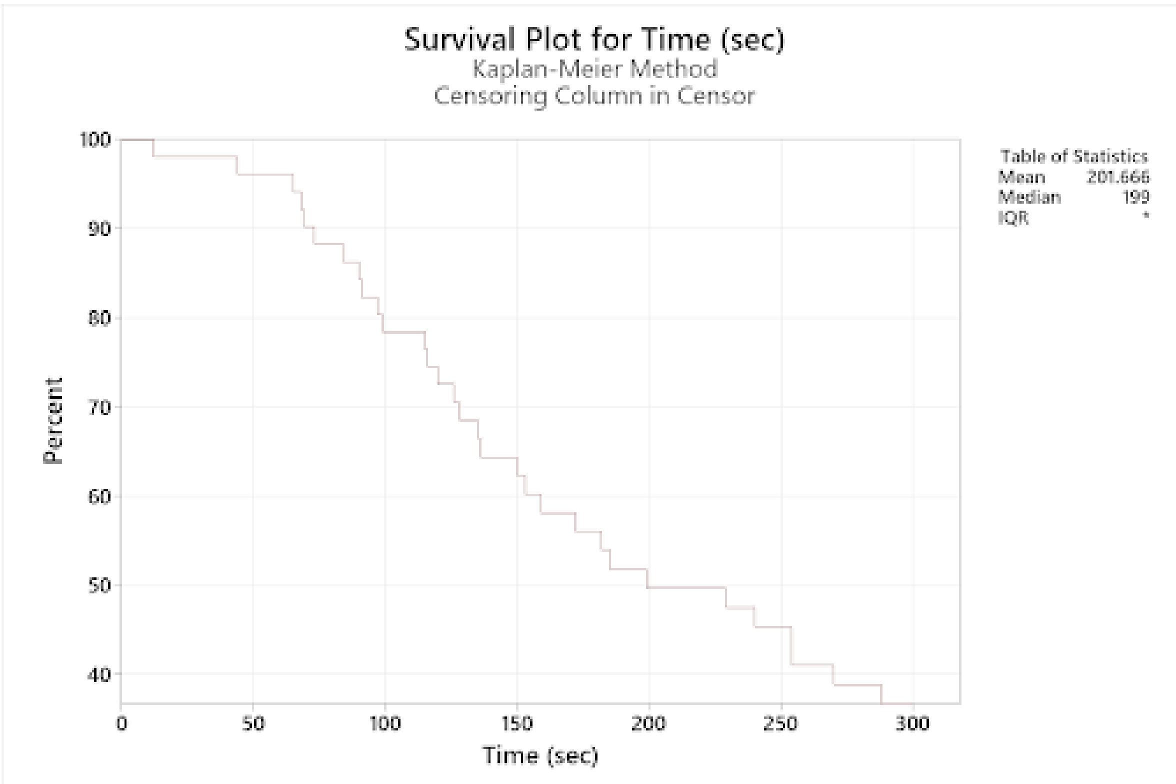
	Mean (s.)	Median (s.)
19	194.958	159
21	226.727	–
22	181.852	172
40	235.375	288

The 22-square puzzle had the shortest estimated mean completion time of 181.85 seconds and a median of 172 seconds, indicating it was generally solved more quickly. The 19-square puzzle had a slightly higher mean of 194.96 seconds and shortest median survival time of 159 seconds. The 21-square puzzle had a mean completion time of 226.73 seconds; however, no median could be determined, as 50% of participants did not finish the puzzle within the 300-second study limit. The 40-square puzzle took the longest to complete, with a mean of 235.38 seconds and a median of 288 seconds, reinforcing our hypothesis that puzzles with more squares usually require more time to finish.

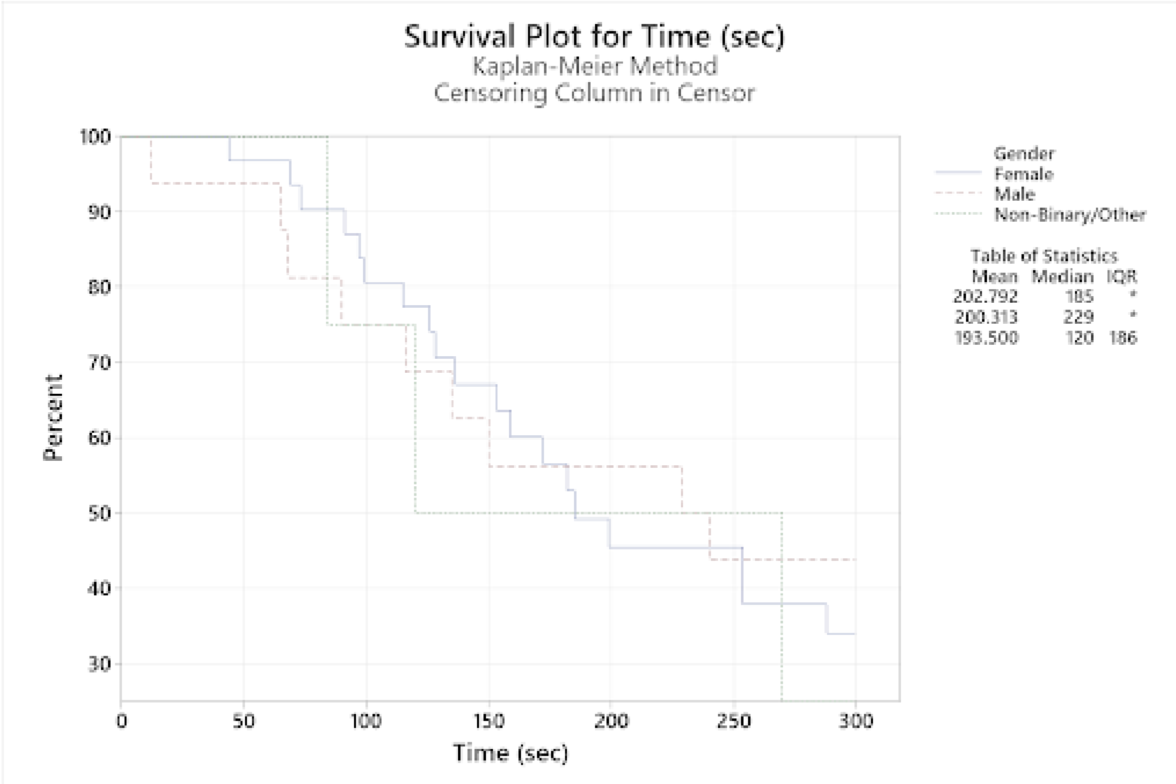
Non-Parametric Survival Analysis

Although the loglogistic distribution fit our data moderately well, since the actual distribution of the data is unknown, and we will proceed with a non-parametric analysis.

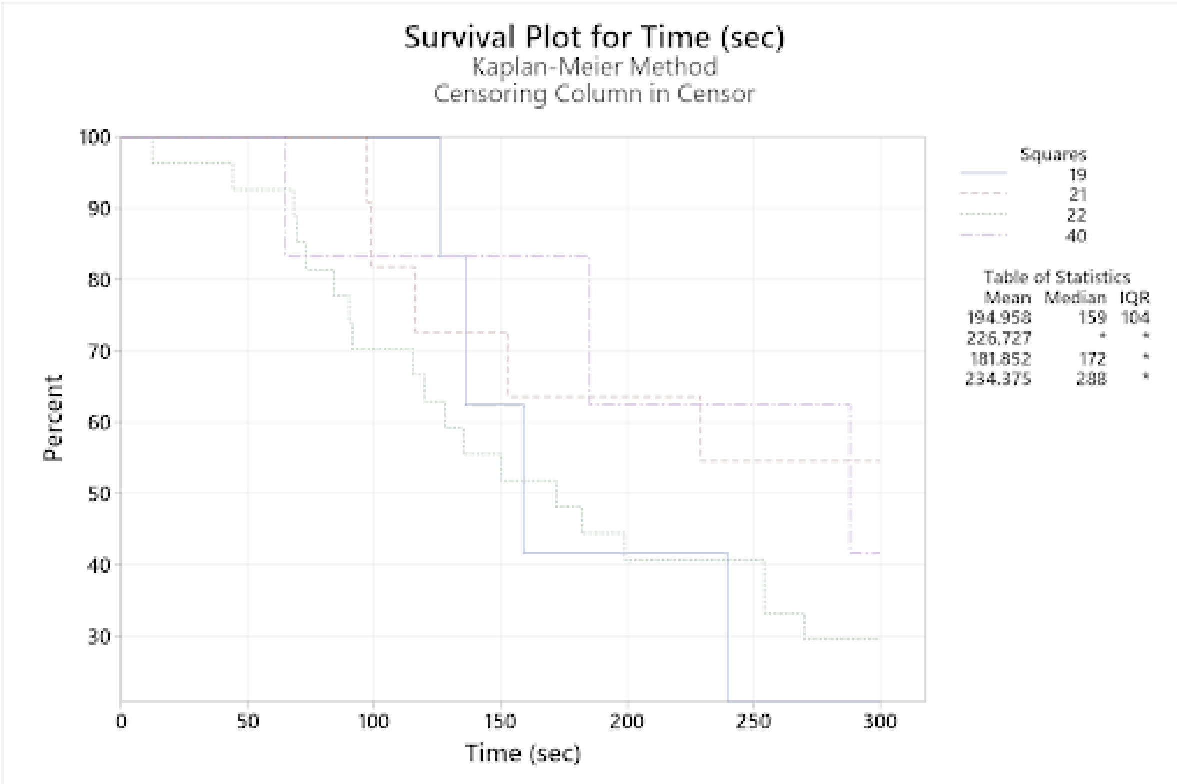
Survival Plots



Based on the Kaplan-Meier curve, we estimate about 35% of students do not complete the puzzle within the first five minutes. Given the puzzle has not been completed by time t in seconds, the estimated probability of not completing the puzzle beyond t decreases steadily for all time.

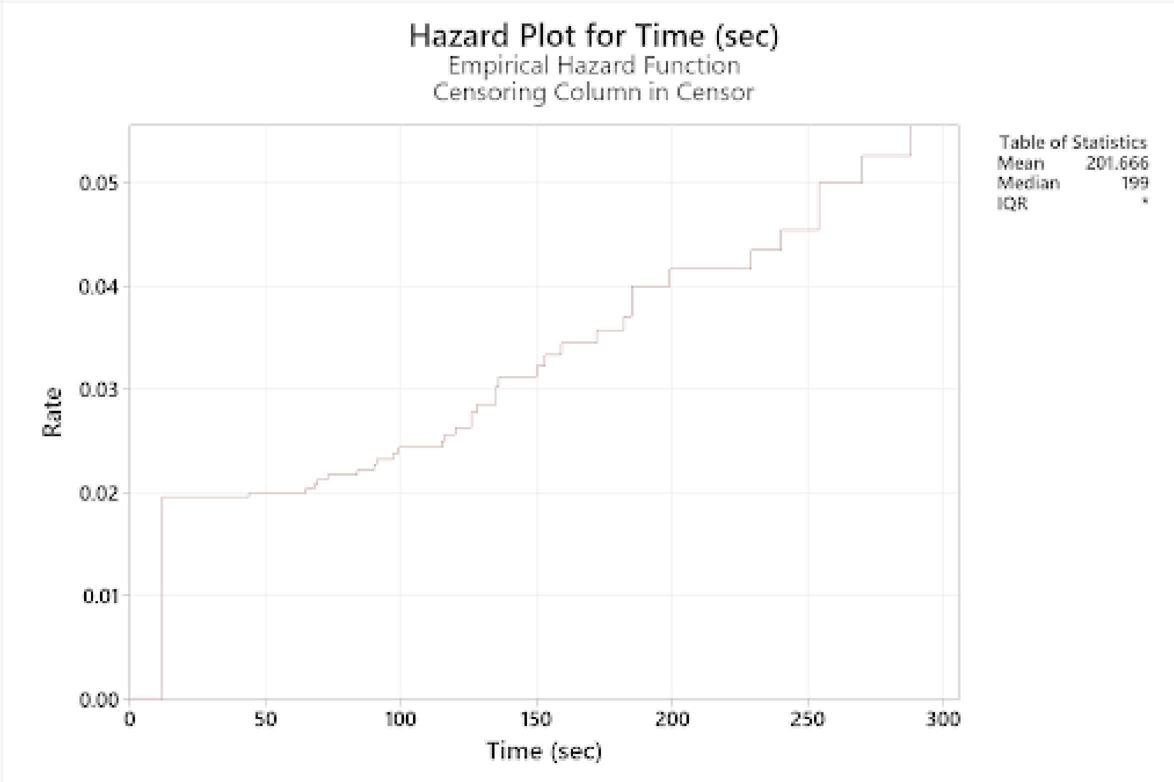


The estimated probability of not completing the puzzle by the end of the study appears to be similar across genders for all time. However, it appears that the probability of not completing the puzzle beyond t is higher for female students at most times prior to 175 seconds, then falls below male and non-binary students after 175 seconds, given they have not completed the puzzle by t.

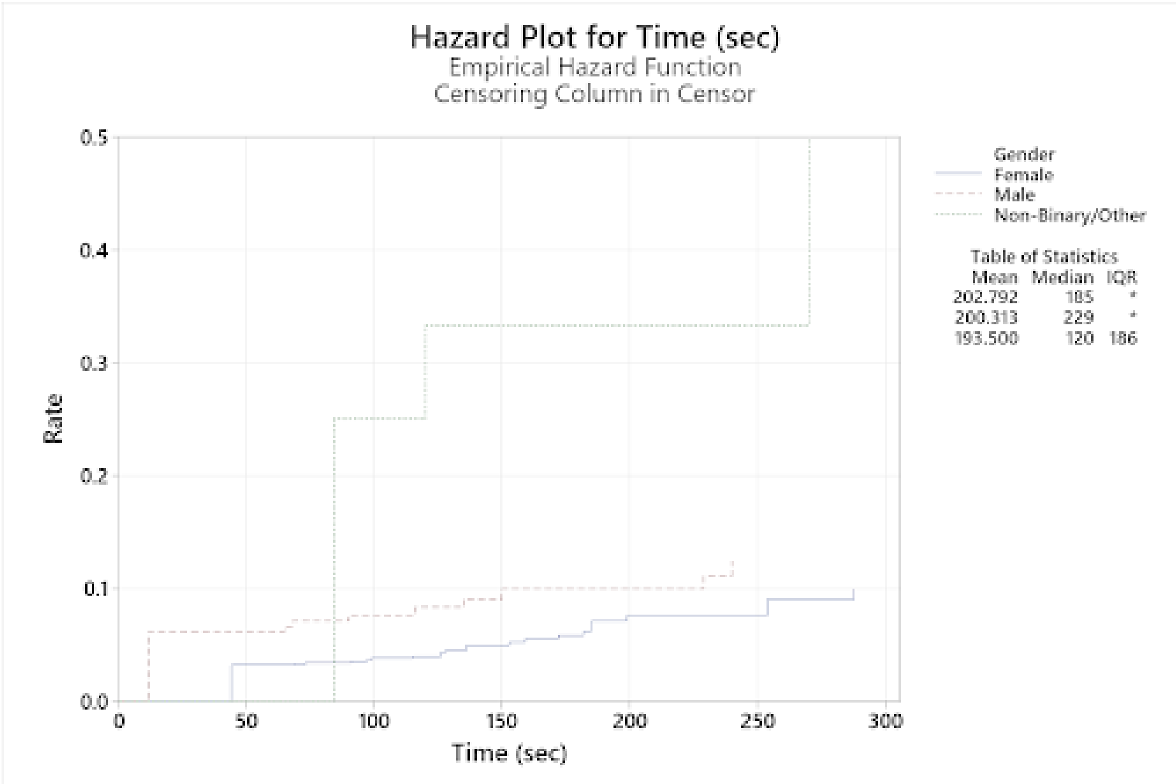


Given students have not completed puzzles with 22 squares by time t, the estimated probability that students do not solve 22 square puzzles beyond t is lowest for most of the study duration, with the exception of 175 to 200 seconds and briefly around 60 seconds. For most times, the estimated probability a student does not solve puzzles with 40 playable squares is highest, especially compared to puzzles with 22 squares.

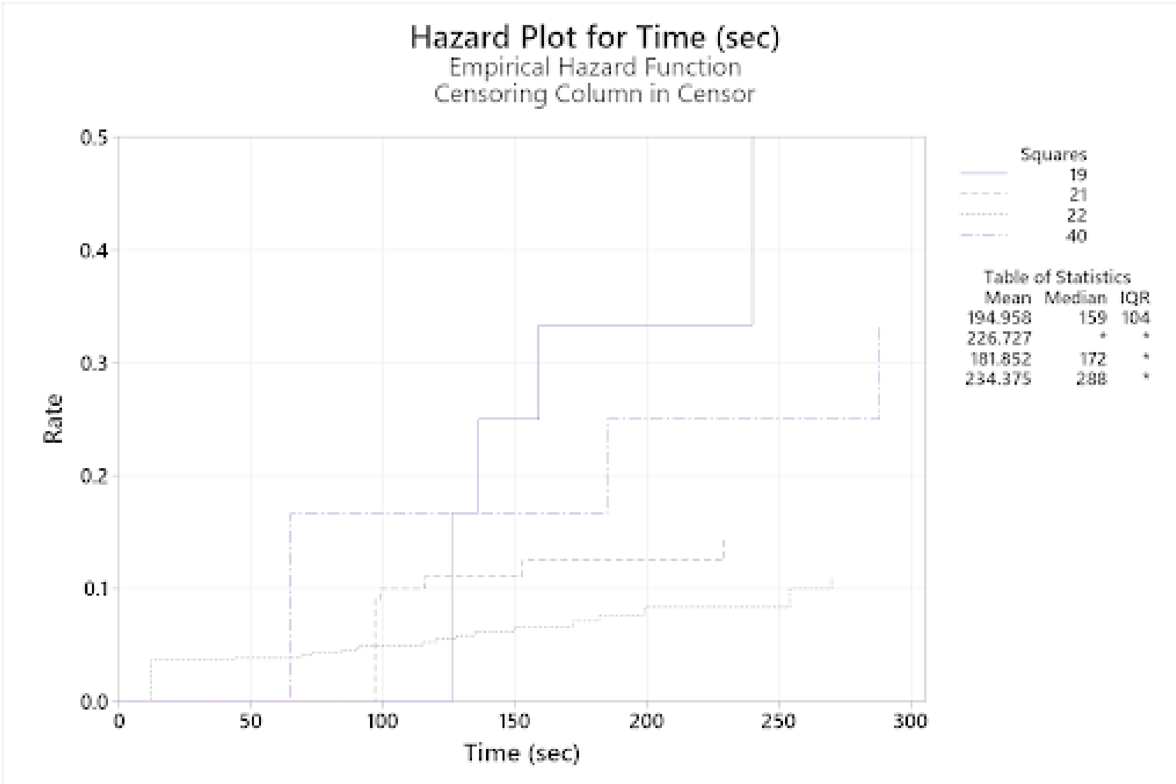
Hazard Plots



The estimated risk of completing the puzzle rapidly increases during the first 20 seconds. After 20 seconds, the estimated hazard is roughly constant until 60 seconds, then steadily increases for the remainder of the study.



The estimated risk of completing the puzzle is lowest for female students across all time, t. While male students exhibit slightly higher potential for puzzle completion than females, we generally observe a similar trend between these two groups. A much sharper increase in potential to complete the puzzle for non-binary students occurs around 80 seconds and continues for the rest of the study.



For most of the study, the estimated potential to complete puzzles with 19 playable squares is notably larger than that of the larger sized puzzles (21, 22, and 40 squares). After 150 seconds, we observe the estimated risk of completion for 21 and 22 square puzzles increase slightly, but 40 and 19 square puzzles increase more rapidly.

Log-Rank Test

Gender

Test Statistics

Method	Chi-Square	DF	P-Value
Log-Rank	0.249059	2	0.883
Wilcoxon	0.156630	2	0.925

Since the p-value 0.883 is larger than any reasonable level of significance, we have insufficient evidence to conclude that the probability of puzzle incompletion differs across male, female, and non-binary students for some time, t. This indicates a failure to reject the null hypothesis of the log-rank test.

Squares

Test Statistics

Method	Chi-Square	DF	P-Value
Log-Rank	2.46472	3	0.482
Wilcoxon	2.81468	3	0.421

Similarly, since the p-value 0.482 is larger than any reasonable level of significance, we have insufficient evidence to conclude that the probability of puzzle incompletion differs across puzzles with 19, 21, 22, or 40 playable squares. This indicates a failure to reject the null hypothesis of the log-rank test.

Comparing Parametric and Non-Parametric Survival Analyses

We found that both approaches lead to the same general conclusions when assessing mini completion times. While the nonparametric survival times were generally lower, they also did not include values beyond 300 seconds, as the study was limited to that time frame. Both methods provide some indication that the 21-square puzzle was more time consuming, with many participants failing to complete it within five minutes. The 19- and 22-square puzzles generally took less time and the 40-square puzzle was the most time consuming, with the highest overall mean and median completion times for both nonparametric and parametric survival analyses. Additionally, both methods agree that females and non-binary students tended to complete the puzzle faster than males, which aligns with trends observed in the survival and hazard plots, despite insignificant inferential tests.

Regression Analysis

Model Selection

Over the course of our analysis, we developed many Cox regression models. Through these analyses, we found insignificant evidence to conclude that one or more genders have significantly different Mini completion times. Similarly, there was insignificant evidence to conclude that different values of playable squares, reaction time, or any combination of these variables (both as main effects and interactions), have significantly different Mini completion times.

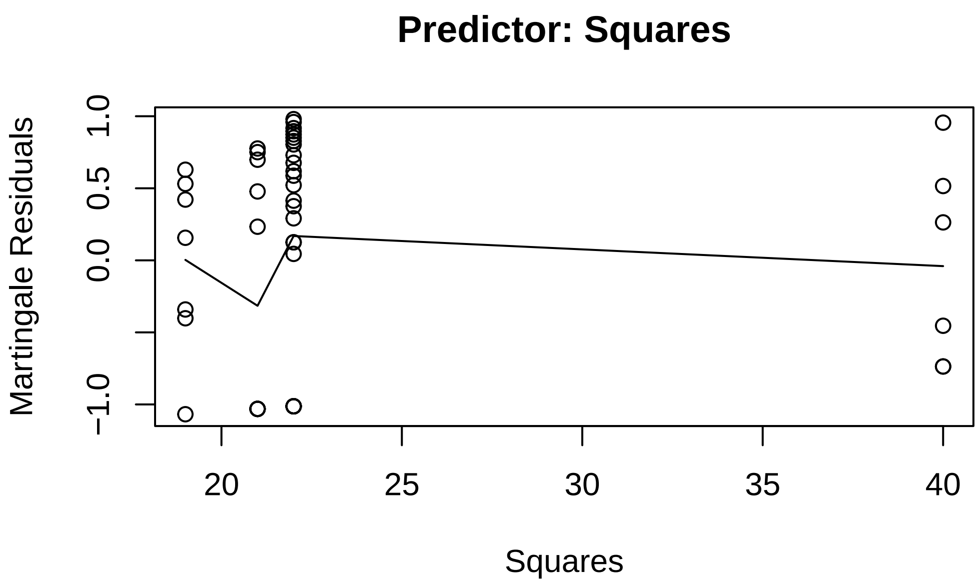
Our final model uses squares as the only predictor, as this model had the lowest AIC of any models with any explanatory variables, as determined with a step-wise model selection procedure.

```
Call:
coxph(formula = Surv(time, censor) ~ squares, data = mini_times)

  n= 51, number of events= 31

            coef exp(coef) se(coef)      z Pr(>|z|)
squares -0.01770   0.98246  0.03116 -0.568     0.57

        exp(coef) exp(-coef) lower .95 upper .95
squares    0.9825      1.018    0.9243     1.044

Concordance= 0.468  (se = 0.046 )
Likelihood ratio test= 0.35  on 1 df,   p=0.6
Wald test            = 0.32  on 1 df,   p=0.6
Score (logrank) test = 0.33  on 1 df,   p=0.6
```

For the purposes of this project we will use this model for further interpretation, however, we would not use this for professional statistical analysis due to insignificance of the likelihood ratio test.

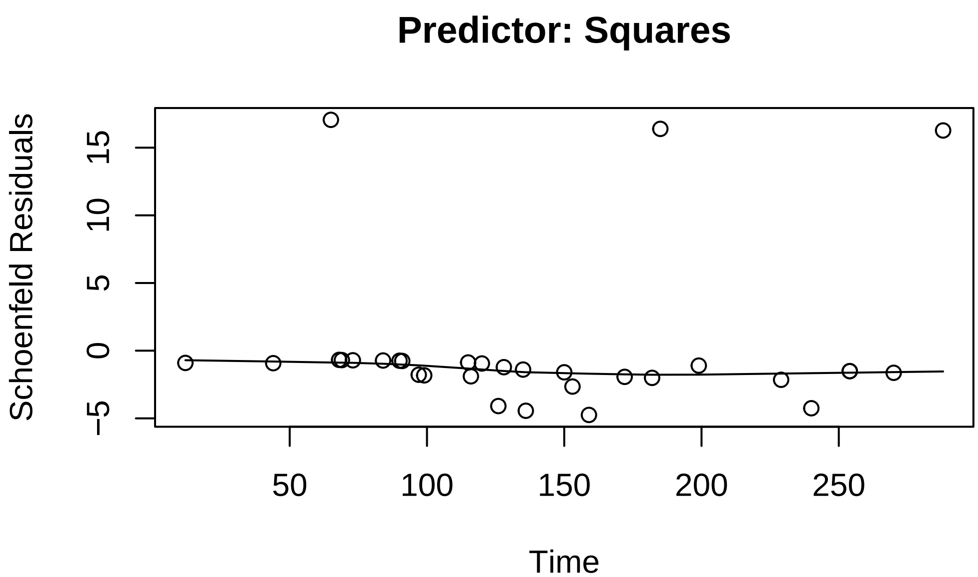
Model Assumptions

Martingale Residuals



The LOWESS is approximately horizontal, meaning the relationship between the predictor squares and log-hazard is approximately linear.

Schoenfeld Residuals



The LOWESS fit is approximately horizontal, meaning the proportional hazards assumption appears to not be violated. However, we will investigate this further with a formal test.

```
         chisq df    p
squares 0.0262  1 0.87
GLOBAL  0.0262  1 0.87
```

Testing Proportional Hazards with `cox.zph()` returns a p-value larger than any reasonable level of significance. This indicates that our P.H. assumption is not violated.

Interpretations

The estimated hazard of completing the mini crossword for each additional square present within the puzzle is 0.982

We are 95% confident that the true hazard of completing a mini crossword for each one-square increase in puzzle size is 7.57% lower and 4.4% higher.

Hazard Between 22 Squares and 19 Squares:

The estimated hazard of completing the mini crossword with 22 squares is 5.18% lower than the hazard of completing the mini crossword with 19 squares.

We are 95% confident that the true hazard of completing a mini crossword with 22 squares is between 21.03% lower and 13% higher when compared to the hazard of completing the mini crossword with 19 squares.

Hazard Between 40 Squares and 19 Squares:

The estimated hazard of completing the mini crossword with 40 squares is 31.04% lower than the hazard of completing the mini crossword with 19 squares.

We are 95% confident that the true hazard of completing a mini crossword with 40 squares is between 80.56% lower and 147% higher when compared to the hazard of completing the mini crossword with 19 squares.

These confidence intervals all capture 0%. This is consistent with the insignificance of our Likelihood Ratio Test, and is another indicator of the insignificance of squares as a predictor for mini completion time.

Conclusion

This study explored relationships between completion time of the NYT mini crossword and factors including gender, puzzle size, and reaction time. We should note that the true distribution of mini crossword completion times is unknown and our Cox regression models found the auxiliary variables insignificant predictors of completion time. Thus, in practice, non-parametric analysis with Kaplan-Meyer estimation is ideal for understanding mini crossword completion times. For the purposes of this project, we still conducted all three forms of analysis.

Parametric models (loglogistic) suggested a difference in mini completion times for male students as compared to female and non-binary students. The loglogistic model also identified 40 square puzzles as leading to the highest probabilities of puzzle incompletion for all time.

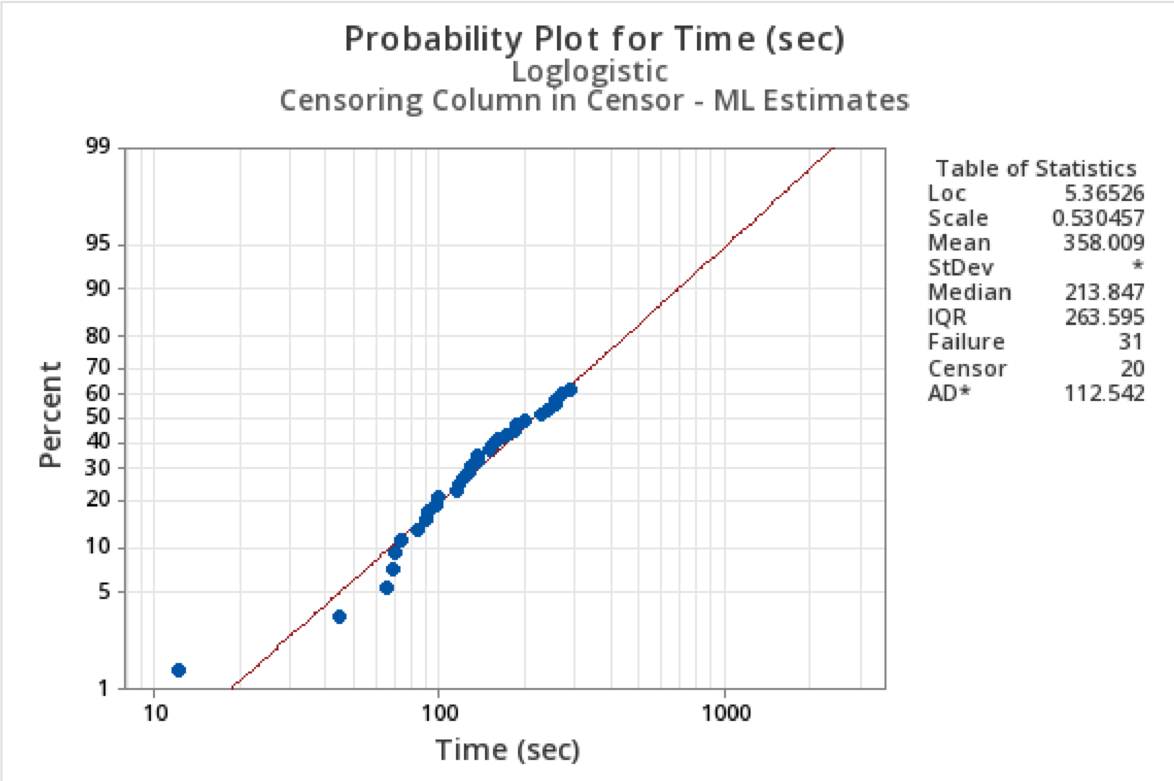
The Kaplan-Meier analysis revealed similar relationships between completion time and auxiliary variables. We observed a frequent intersection between the lines for each level which likely led to the insignificance of our non-parametric models (log-rank, $p > 0.05$).

Finally, our Cox regression models indicated that gender, puzzle size, and reaction time were insignificant predictors of puzzle completion time. We ran a stepwise variable selection function and found that a reduced model with only puzzle size (squares) returned the lowest AIC. Hence, we used this model to conduct further analyses that compared the hazard rates and survival probabilities between puzzles with various sizes, generally indicating that larger sized puzzles have lower hazard of completion, although our findings were not significant at the .05 level.

While these findings do not have immediate practical applications, they challenge our intuitive assumptions about puzzle completion times across variables. Future research could explore the high variability in completion times among individuals, possibly influenced by skill level or frequency of play. A larger sample size or tracking individuals over multiple days could provide more insight into these factors.

Appendix

Parametric and Non-Parametric Analyses



The graphs and tests for parametric and non-parametric analyses were conducted in Minitab.

Regression Analysis

```
model_squares <- coxph(formula = Surv(time, censor) ~ squares, data = mini_times)
summary(model_squares)
```

```
mart <- residuals(model_squares, type = "martingale")
plot(as.numeric(mini_times$squares), mart, type = "p", xlab = "Squares",
     ylab = "Martingale Residuals", main = "Predictor: Squares")
smooth.sres <- lowess(mini_times$squares, mart)
lines(smooth.sres$x, smooth.sres$y, lty = 1)
```

```r
schoen <- residuals(model_squares, type = "schoenfeld")

comp.times <- sort(mini_times[mini_times$censor!= 0,]$time)
plot(comp.times, schoen, xlab="Time", ylab = "Schoenfeld Residuals",
     main = "Predictor: Squares")
smooth.sres <- lowess(comp.times, schoen)
lines(smooth.sres$x, smooth.sres$y, lty = 1)
```

```r
cox.zph(model_squares, transform = "log")
```

Data

The data set, along with all of the code, can also be found at: github.com/AleaSeifert/stat417.git