**Survival Assignment**

Consider the college graduation data (**Graduation.csv**):  
**Years**: the number of years taken to complete the requirements for a bachelor’s  
**Censor** = 1: the degree is completed, Censor = 0: the degree is censoring  
**Gender**: Male and Female

**(a)** Create a Kaplan-Meier plot and compare the survival experiences to investigate whether there are differences in the time required to obtain a bachelor’s degree by gender.

**(b)** Conduct the log-rank test to compare the survival curves of the two genders. Write down the null and alternative hypotheses statements, calculate the test statistic value and p-value, give the conclusion.

**(c)** Estimate the mean and median time until the graduation for each gender. Construct 95% confidence intervals for the survival probabilities S(t) at Years = 4 for males and females respectively and interpret the results.   
     At what time would you estimate that at least 40% of males will graduate?  
     At what time would you estimate that at least 50% of females will graduate?

**(d)** Plot and examine the estimator of the cumulative hazard function for males. Discuss the changes in the rate of change of Hn(t)\_NA as each of the following time periods passes and discuss corresponding changes in the hazard rate of graduation as college students move from one time interval to another.   
    • 0–3.75 years   
    • 3.75–4.75 years   
    • > 4.75 years  
    • around 8 years

**=================================================================**

**(a)** **Kaplan-Meier plot.**

**Chart

Description automatically generated**

A slight difference in time can be seen in the graph while comparing the graduation time of male and female. Male students take more time for graduation compared to female students. No students graduated till 3.7 years including male and female.

**b. Log Rank Test to compare the survival curves of the two genders**

H0: Survival curves of male and female are equal for all times t

Ha: Survival curves of male and female are different for some values of t

Test Statistic Value – 6.1

P value – 0.01

P value < 0.05

Decision – Reject Ho

Conclusion – We have sufficient evidence to conclude that survival experiences of male and female are different at 5% level of significance.

**c. Mean time until the graduation for each gender**

Mean time for female to graduate is 4.98 years

Mean time for male to graduate is 5.40 years

**Median time until the graduation for each gender**

Median time for female until graduation is 4.58 years

Mean time for male until graduation is 4.75years

**95% confidence intervals for the survival probabilities S(t) at Years = 4 for males**

Confidence Interval for male = (0.643, 0.738)

Interpretation - With 95% confidence we can conclude that the probability that the male students will take longer than 4 years to graduate is between 0.643 and 0.738.

**95% confidence intervals for the survival probabilities S(t) at Years = 4 for females**

Confidence Interval for male = (0.5463,0.640)

Interpretation – With 95% confidence we can conclude that the probability that the female students will take longer than 4 years to graduate is between 0.54 and 0.64.

**At what time would you estimate that at least 40% of males will graduate?**

At least 40% of males will graduate after 4.41 years

**At what time would you estimate that at least 50% of females will graduate?**

At least 40% of males will graduate after 4.41 years

d.

Chart, line chart, scatter chart

Description automatically generated

**0–3.75 years**

Till 3.75 years the slope is 0. Rate of change in H(t) is 0, then hazard function, h(t) is 0. Majority of students failed to graduate in this interval except for a very few students graduated towards the end of 3.75 years.

**3.75–4.75 years**

From 3.75 to 4.75 years the slope of change in H(t) is increasing over time then, hazard function h(t) is increasing over the same interval of time which means the students graduated between this interval is increasing.

**> 4.75 years**

Here the slope of change in H(t) is slowly increasing over time same as before. So, h(t) also increases, and the students who got graduated in this period is noticeable from the graph.

**Around 8 years**

Around 8 years, slope is 0. Rate of change in H(t) is 0, then hazard function, h(t) is 0. No students graduated in this interval.

Recipes (recipe\_name, **recipie\_id**, recipe\_type, recipeauthor, recipe\_description)

Restuarant (restaurant\_name, **restaurant\_id, recipie\_id**)

Category (cate\_name, **cate\_id, recipe\_id**)

Feedback ( **rating\_id**, ratings, **recipe\_id**)

Ingredients (**ingredient\_id**, name)

cook\_steps (**step\_id**, description, sequence, **recipe\_id**)

Recipe\_Ingredients (**recipie\_id**, name, **ingredient\_id, step\_id**)