Survival Analysis of Diabetic Patients: Investigating the Impact of Age on Long-Term Outcomes

- Latthika S
- 24MSP3033

This project focuses on conducting a survival analysis to examine the impact of age on the long-term survival of diabetic patients. Using a dataset of diabetic patients who underwent laser coagulation treatment for diabetic retinopathy, the study applies Kaplan-Meier estimation, log-rank tests, and Cox proportional hazards modeling to evaluate survival probabilities stratified by age groups (young vs. old). The Kaplan-Meier survival curves suggest a slight survival advantage for younger patients; however, statistical tests reveal no significant difference in survival between the two age groups. The findings emphasize the need for further exploration of additional variables to better understand survival outcomes for diabetic patients. This study provides insights that may guide clinical decision-making and prioritization of interventions for diabetic patients based on risk factors beyond age.

CODE:

```
# Load required libraries

library(survival)

library(survminer)

library(dplyr)


# Load the diabetic dataset

data('diabetic')

head(diabetic)


# Create age groups for analysis

diabetic <- diabetic %>% mutate(age_group = ifelse(age >= median(age), "old", "young"))

diabetic$age_group <- factor(diabetic$age_group)


# Kaplan-Meier method

surv_object <- Surv(time = diabetic$time, event = diabetic$status)
```

```
surv_object
fit1 <- survfit(surv_object ~ age_group, data = diabetic)
summary(fit1)

# Plot Kaplan-Meier survival curve
ggsurvplot(
  fit1,
  data = diabetic,
  conf.int = FALSE,
  pval = TRUE,
  risk.table = TRUE,
  legend.title = "Age Group",
  xlab = "Time",
  ylab = "Survival Probability",
  title = "Kaplan-Meier Survival Curve by Age Group",
  ggtheme = theme_minimal()
)

# Log-rank test
log_rank_test <- survdiff(Surv(time, status) ~ age_group, data = diabetic)
log_rank_test

# Cox Model
cox_model <- coxph(Surv(time, status) ~ age_group, data = diabetic)
cox_model
```

OUTPUTS:

```
> data('diabetic')
> head(diabetic)
  id laser age   eye trt risk  time status
1  5 argon  28  left   0    9 46.23      0
2  5 argon  28 right   1    9 46.23      0
3 14 xenon  12  left   1    8 42.50      0
4 14 xenon  12 right   0    6 31.30      1
5 16 xenon   9  left   1   11 42.27      0
6 16 xenon   9 right   0   11 42.27      0
```

Description: Results from a trial of laser coagulation for diabetic retinopathy.

Key Variables:

- time: Follow-up time (days).
- status: Event indicator (1 = failure, 0 = censored).
- laser: Treatment group (1 = treated, 0 = control).
- age: Age of the patient.
- eye: Which eye (1 = left, 2 = right).
- risk: Risk score (continuous variable).

```
> # Kaplan-Meier method
> surv_object <- Surv(time = diabetic$time, event = diabetic$status)
> surv_object
  [1] 46.23+ 46.23+ 42.50+ 31.30  42.27+ 42.27+ 20.60+ 20.60+  0.30  38.77+ 65.23+ 54.27  63.50+
 [14] 10.80  23.17+ 23.17+  1.47+  1.47+ 58.07+ 13.83  48.53+ 46.43  44.40+  7.90  39.57+ 39.57+
 [27] 30.83  38.57  66.27+ 14.10   6.90  20.17  41.40  58.43+ 58.20+ 58.20+ 57.43+ 57.43+ 56.03+
 [40] 56.03+ 67.53+ 67.53+ 61.40+  0.60  10.27   1.63  66.20+ 66.20+ 13.83   5.67  58.83+ 29.97
 [53] 60.27+ 26.37   1.33   5.77  35.53   5.90  21.90  25.63  14.80  33.90   6.20   1.73  22.00
 [66] 46.90+ 31.13+ 31.13+ 22.00  30.20  70.90+ 70.90+ 25.80  13.87  48.30   5.73  53.43+ 53.43+
 [79]  1.90  51.10+  9.90   9.90  34.20+ 34.20+  2.67  46.73+ 18.73+ 13.83  32.03+  4.27  13.90
 [92] 69.87+ 66.80+ 66.80+ 64.73+ 64.73+  1.70   1.70   1.77  43.03  29.03+ 29.03+ 56.57+ 56.57+
[105]  8.30   8.30  21.57+ 18.43  31.57+ 31.57+ 31.63  31.63+ 39.77+ 39.77+  6.53  18.70  18.90+
[118] 18.90+ 56.80+ 22.23  55.60+ 14.00  42.17  42.17   5.33  10.70+ 59.80  66.33+  5.83  52.33+
[131] 58.17+  2.17  48.43  14.30  25.83+ 25.83+ 45.40+ 45.40+ 47.60+ 47.60+  9.60  13.33  42.10+
[144] 42.10+ 39.93+ 39.93+  7.60  14.27   1.80  34.57   4.30  65.80+ 12.20   4.10  60.93+ 60.93+
[157] 57.20+ 57.20+ 38.07+ 12.73  54.10+ 54.10  59.27+  9.40   9.90  21.57  54.10+ 54.10+ 50.47+
[170] 50.47+ 46.17+ 46.17+ 46.30+ 46.30+ 38.83+ 38.83+ 44.60+ 44.60+ 43.07+ 43.07+ 40.03+ 26.23
[183] 41.60+ 18.03  38.07+ 38.07+ 65.23+ 65.23+  7.07  66.77+ 13.77  13.77   9.63   9.63+ 46.23+
[196] 46.23+  1.50  45.73+ 33.63  33.63  40.17+ 40.17+ 27.60  63.33  38.47   1.63  55.23+ 55.23+
[209] 25.30  52.77+ 46.20  57.17+  9.87+  1.70  57.90+ 57.90+  5.90+  5.90+ 32.20+ 32.20+ 10.33
[222]  0.83  50.90+  6.13  25.93  43.67+ 38.30+ 38.30+ 38.77+ 19.40  21.97  38.07+ 38.30+ 38.30+
[235] 70.03+ 26.20  18.03  62.57+  1.57  13.83  46.50+ 13.37   1.97  11.07  42.47+ 22.20  38.73+
[248] 38.73+ 51.13+ 51.13+ 46.50+  6.10  11.30   2.10  17.73  42.30+ 26.47+ 26.47+ 10.77+ 10.77+
[261] 55.33+ 55.33+ 58.67+ 58.67+  4.97  12.93  26.47  54.20+ 49.57+ 49.57+  9.87  24.43  50.23+
[274] 50.23+ 30.40  13.97  43.33+ 43.33  42.23+ 42.23+ 74.93+ 74.93+ 66.93+ 66.93+ 73.43+ 73.43+
[287] 67.47+ 38.57   3.67+   3.67  67.03+ 48.87  65.60+ 65.60+ 15.83  15.83+ 20.07+  8.83  67.43+
[300] 67.43+  1.47+  1.47+ 62.93+ 22.13   6.30  56.97+ 59.70+ 18.93  19.00  13.80  55.13+ 55.13+
[313]  5.43  13.57  42.20+ 42.20+ 38.27+ 38.27+  7.10+  7.10  26.17  63.63+ 24.73  59.00+ 54.37+
[326] 54.37+ 54.60+ 10.97  21.10  63.87+ 62.37+ 43.70  62.80+ 62.80+ 63.33+ 14.37  58.53+ 58.53+
[339] 58.07+ 58.07+ 58.50+ 58.50+ 14.37+  1.50  54.73+ 38.40  50.63+  2.83  51.10+ 51.10+ 49.93+
[352]  6.57  46.27+ 46.27  10.60+ 10.60+ 42.77+ 42.77+ 34.37  42.27+ 42.07+ 42.07+ 38.77+ 38.77+
[365] 61.83  74.97+ 66.97+  6.57  38.87  68.30+ 46.63  42.43  67.07+ 67.07+  2.70   2.70+ 63.80+
[378] 63.80+ 32.63+ 32.63+ 62.00+ 62.00+ 54.80+ 13.10   8.00+  8.00+ 42.33  51.60+ 49.97+  2.90
[391] 45.90+  1.43  41.93+ 41.93+
```

Survival Object Interpretation
The Surv object encapsulates time-to-event data and event status for the dataset. It represents:

1. **Survival Time**: Duration until the event occurs or is censored.
    - Example: 46.23 indicates an observation lasted 46.23 time units.
2. **Event Indicator**:
    - + denotes censored observations (event did not occur during the study period).
    - Absence of + indicates the event occurred.

This object is foundational for survival analysis, enabling Kaplan-Meier estimations, log-rank tests, and Cox proportional hazards models. It allows for comprehensive analysis of censored and uncensored survival data.

```
> fit1 <- survfit(surv_object ~ age_group, data = diabetic)
> summary(fit1)
Call: survfit(formula = surv_object ~ age_group, data = diabetic)

                age_group=old
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0.60    198       1    0.995 0.00504        0.985        1.000
  1.33    197       1    0.990 0.00711        0.976        1.000
  1.50    194       1    0.985 0.00871        0.968        1.000
  1.57    193       1    0.980 0.01005        0.960        1.000
  1.63    192       1    0.975 0.01122        0.953        0.997
  1.70    191       1    0.969 0.01227        0.946        0.994
  1.73    190       1    0.964 0.01322        0.939        0.991
  1.77    189       1    0.959 0.01410        0.932        0.987
  2.17    188       1    0.954 0.01492        0.925        0.984
  2.67    187       1    0.949 0.01569        0.919        0.980
  2.90    186       1    0.944 0.01641        0.912        0.977
  3.67    185       1    0.939 0.01710        0.906        0.973
  4.10    183       1    0.934 0.01776        0.900        0.969
  4.27    182       1    0.929 0.01839        0.893        0.965
  4.97    181       1    0.923 0.01899        0.887        0.961
  5.73    180       1    0.918 0.01956        0.881        0.958
  5.77    179       1    0.913 0.02012        0.875        0.954
  5.83    178       1    0.908 0.02065        0.869        0.949
  6.20    177       1    0.903 0.02116        0.862        0.945
  6.57    176       1    0.898 0.02165        0.856        0.941
```

age_group=young

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
0.30	196	1	0.995	0.00509	0.985	1.000
0.83	195	1	0.990	0.00718	0.976	1.000
1.43	194	1	0.985	0.00877	0.968	1.000
1.50	191	1	0.980	0.01013	0.960	1.000
1.63	190	1	0.974	0.01131	0.952	0.997
1.70	189	2	0.964	0.01333	0.938	0.991
1.80	187	1	0.959	0.01422	0.931	0.987
1.90	186	1	0.954	0.01505	0.925	0.984
1.97	185	1	0.949	0.01583	0.918	0.980
2.10	184	1	0.943	0.01656	0.912	0.976
2.70	183	1	0.938	0.01726	0.905	0.973
2.83	181	1	0.933	0.01792	0.899	0.969
4.30	180	1	0.928	0.01856	0.892	0.965
5.33	179	1	0.923	0.01916	0.886	0.961
5.43	178	1	0.918	0.01974	0.880	0.957
5.67	177	1	0.912	0.02030	0.873	0.953
5.90	176	1	0.907	0.02084	0.867	0.949
6.10	173	1	0.902	0.02137	0.861	0.945
6.13	172	1	0.897	0.02188	0.855	0.941
6.30	171	1	0.891	0.02237	0.849	0.936
6.53	170	1	0.886	0.02284	0.843	0.932
6.57	169	1	0.881	0.02330	0.836	0.928

The Kaplan-Meier survival analysis stratified by age groups (young and old) provides insight into the survival probabilities over time for diabetic patients.

Survival Trends:

- Both age groups demonstrate a gradual decline in survival probabilities over time.
- The "old" age group exhibits a slightly lower survival probability compared to the "young" group as time progresses.

Early Survival Rates:

- At the initial time points, survival rates are high for both groups (approximately 99.5%), indicating minimal risk of events in the early stages.

Divergence Over Time:

- By time 10, the survival probability for the "old" group drops to approximately 84.1%, while the "young" group maintains a slightly higher survival of 84.4%.

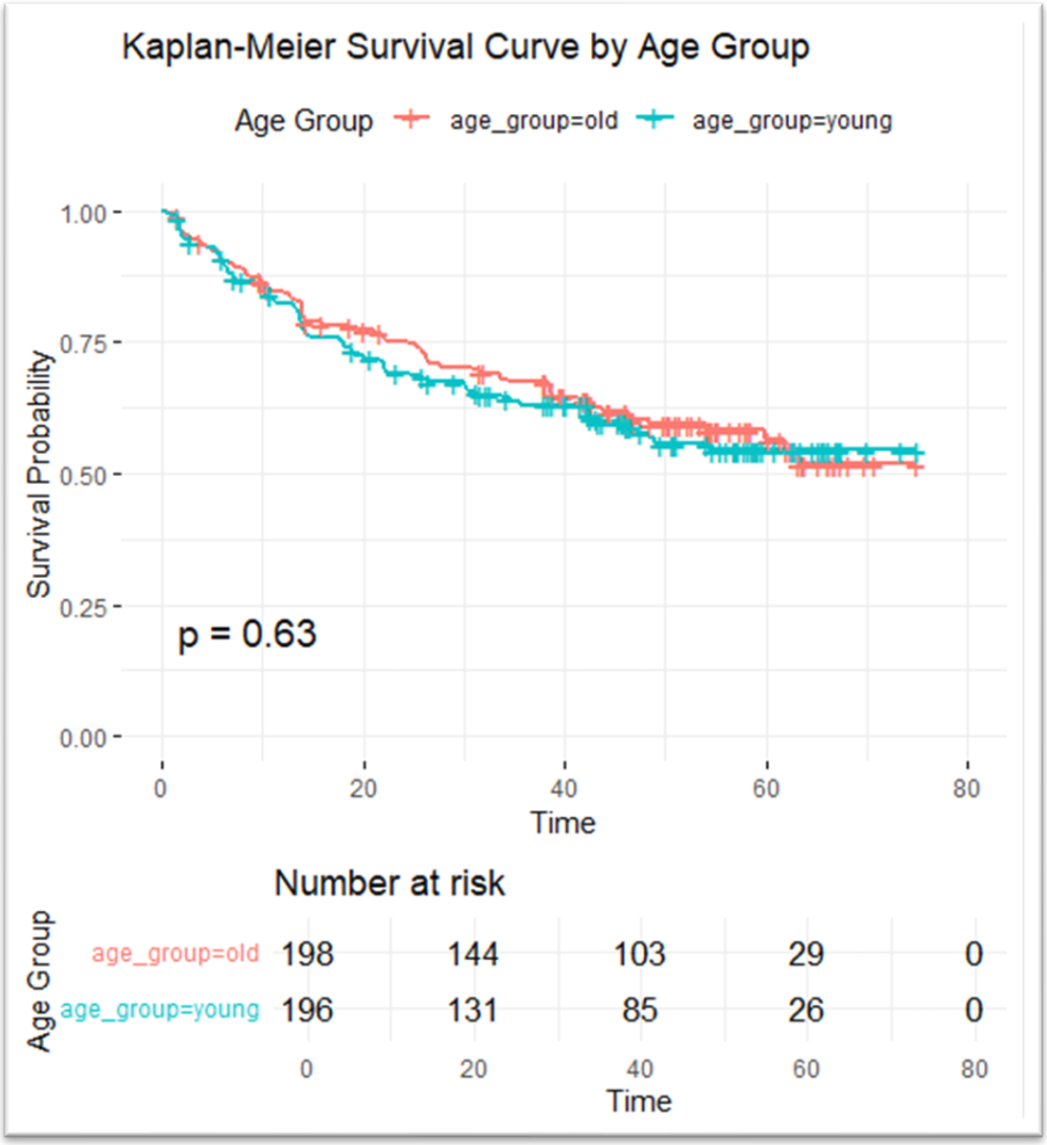
Steeper Decline in Older Group:

- As time progresses, the "old" group experiences a steeper decline, with survival probabilities reaching 54.3% by time 61.83, compared to 71.3% for the "young" group by time 21.57.

Confidence Intervals:

- The standard errors and confidence intervals indicate more uncertainty in survival estimates at later time points due to reduced sample sizes and higher event occurrences.

Older diabetic patients tend to have lower survival probabilities over time compared to their younger counterparts, emphasizing the influence of age as a significant factor in long-term survival. This trend can guide healthcare practitioners in prioritizing early interventions and monitoring for higher-risk groups.



The Kaplan-Meier survival curves for diabetic patients stratified by age group ("young" vs. "old") indicate the following:

1. **Survival Trends**:
    - The survival probabilities for both groups decrease over time, reflecting the occurrence of events (e.g., death, complications).
    - The curves are closely aligned, with minimal differences in survival between the two age groups.
2. **Statistical Comparison**:
    - The p-value of 0.63 indicates no statistically significant difference in survival between the "young" and "old" groups.
3. **Number at Risk**:
    - The number of individuals at risk decreases steadily over time in both groups, as expected in survival studies.

The analysis suggests that age does not significantly influence survival probabilities in this diabetic population. This finding highlights the potential importance of other factors in determining long-term survival outcomes.

```
> log_rank_test <- survdiff(Surv(time, status) ~ age_group, data = diabetic
> log_rank_test
Call:
survdiff(formula = Surv(time, status) ~ age_group, data = diabetic)

                  N Observed Expected (O-E)^2/E (O-E)^2/V
age_group=old   198       78       81     0.112     0.236
age_group=young 196       77       74     0.123     0.236

 Chisq= 0.2  on 1 degrees of freedom, p= 0.6
```

Log-Rank Test:

- The log-rank test compares survival distributions between the "old" and "young" age groups.
- Results:
    - Chi-squared statistic = 0.2, p-value = 0.6.
    - Conclusion: There is no statistically significant difference in survival between the two age groups.

```
> cox_model <- coxph(Surv(time, status) ~ age_group, data = diabetic)
> cox_model
Call:
coxph(formula = Surv(time, status) ~ age_group, data = diabetic)

                   coef exp(coef) se(coef)     z     p
age_groupyoung 0.07788   1.08100  0.16077 0.484 0.628

Likelihood ratio test=0.23  on 1 df, p=0.6281
n= 394, number of events= 155
```

Cox Proportional Hazards Model:

- This model estimates the hazard ratio for the "young" age group compared to the "old" group.
- Results:
  - Coefficient (coef) = 0.07788.
  - Hazard Ratio (exp(coef)) = 1.081 (indicating a slight, non-significant increase in risk for the young group).
  - p-value = 0.628.
- Likelihood Ratio Test: Chi-squared statistic = 0.23, p-value = 0.6281.
- Conclusion: Age group is not a significant predictor of survival.

Both the log-rank test and the Cox proportional hazards model confirm that age group ("young" vs. "old") does not have a significant impact on survival in this diabetic population. Other variables should be explored to understand survival outcomes better.