|  |  |  |
| --- | --- | --- |
| Assignment 2 – Part 2 | August 13  15338673 | |
| Paul-Willem Janse van Rensburg | | Survival Analysis |

# Table of contents

**Question 4** ………………………………………………………………………………………………………………………………………………**2**

**Question 5** ………………………………………………………………………………………………………………………………………………**3**

**Question 6** ………………………………………………………………………………………………………………………………………………**5**

**Question 7** ………………………………………………………………………………………………………………………………………………**6**

**Question 8** ………………………………………………………………………………………………………………………………………………**6**

**Question 9** ………………………………………………………………………………………………………………………………………………**7**

**Question 10** ……………………………………………………………………………………………………………………………………………**7**

**Question 11** ……………………………………………………………………………………………………………………………………………**8**

**Question 12** ……………………………………………………………………………………………………………………………………………**8**

**Appendix A** …………….………………………………………………………………………………………………………………………………**9**

Question 4

As we will be estimating survival functions, the Kaplan-Meier is more appropriate in that it estimates the survival function directly, as opposed to Nelson-Aalen which would require a conversion from cumulative hazard function to the survival function for an estimate.

We estimate the mean of the survival function for surgical placement, using the aforementioned Kaplan-Meier method as follows:

We set up a 95% confidence interval, implementing the below equation:

Around the estimated mean, , with the following result (α = 0.05):

We repeat the above for the percutaneous placement, with the following results:

Question 5

Continuing from before, we estimate the survival function using the Kaplan-Meier method and retrieve the below summary (for surgical placement, percutaneous placement to follow with same methodology applied).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| time | n.risk | n.event | n.censor | Estimated S(t) | std.err of  S(t) estimate | upper | lower |
| 1.5 | 43 | 1 | 0 | 0.976744 | 0.023531 | 1 | 0.93272 |
| 2.5 | 42 | 0 | 2 | 0.976744 | 0.023531 | 1 | 0.93272 |
| 3.5 | 40 | 1 | 3 | 0.952326 | 0.034565 | 1 | 0.889946 |
| 4.5 | 36 | 2 | 1 | 0.899419 | 0.053186 | 0.998237 | 0.810383 |
| 5.5 | 33 | 1 | 1 | 0.872163 | 0.061447 | 0.983788 | 0.773205 |
| 6.5 | 31 | 0 | 2 | 0.872163 | 0.061447 | 0.983788 | 0.773205 |
| 7.5 | 29 | 0 | 4 | 0.872163 | 0.061447 | 0.983788 | 0.773205 |
| 8.5 | 25 | 2 | 1 | 0.80239 | 0.08517 | 0.948162 | 0.67903 |
| 9.5 | 22 | 1 | 1 | 0.765918 | 0.097049 | 0.926383 | 0.633249 |
| 10.5 | 20 | 1 | 1 | 0.727622 | 0.109773 | 0.902287 | 0.586769 |
| 11.5 | 18 | 1 | 1 | 0.687199 | 0.123766 | 0.875855 | 0.539179 |
| 12.5 | 16 | 0 | 2 | 0.687199 | 0.123766 | 0.875855 | 0.539179 |
| 13.5 | 14 | 0 | 1 | 0.687199 | 0.123766 | 0.875855 | 0.539179 |
| 14.5 | 13 | 0 | 2 | 0.687199 | 0.123766 | 0.875855 | 0.539179 |
| 15.5 | 11 | 1 | 0 | 0.624726 | 0.156234 | 0.848547 | 0.459942 |
| 16.5 | 10 | 1 | 0 | 0.562254 | 0.188468 | 0.813497 | 0.388605 |
| 18.5 | 9 | 1 | 0 | 0.499781 | 0.222281 | 0.772655 | 0.323276 |
| 21.5 | 8 | 0 | 2 | 0.499781 | 0.222281 | 0.772655 | 0.323276 |
| 22.5 | 6 | 0 | 2 | 0.499781 | 0.222281 | 0.772655 | 0.323276 |
| 23.5 | 4 | 1 | 0 | 0.374836 | 0.364338 | 0.765537 | 0.183534 |
| 25.5 | 3 | 0 | 1 | 0.374836 | 0.364338 | 0.765537 | 0.183534 |
| 26.5 | 2 | 1 | 0 | 0.187418 | 0.795451 | 0.891046 | 0.03942 |
| 27.5 | 1 | 0 | 1 | 0.187418 | 0.795451 | 0.891046 | 0.03942 |

Estimating the median, as retrieved from the above table as the largest smaller than 0.5. It is estimated as below:

We calculate a 95% confidence interval using the below equation:

Where is the median for the treatment, around the aforementioned mean, with the standard error as retrieved from the table and with a , with the following result (α = 0.05):

We highlight the lower bound in the above table, but we can only estimate a lower bound of 4.5, but no upper bound as the study does not contain data beyond time=27.5.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| time | n.risk | n.event | n.censor | Estimated S(t) | std.err of  S(t) estimate | upper | lower |
| 0.5 | 76 | 6 | 10 | 0.921053 | 0.033583 | 0.983717 | 0.86238 |
| 1.5 | 60 | 0 | 4 | 0.921053 | 0.033583 | 0.983717 | 0.86238 |
| 2.5 | 56 | 2 | 5 | 0.888158 | 0.042299 | 0.964928 | 0.817495 |
| 3.5 | 49 | 1 | 5 | 0.870032 | 0.047057 | 0.954092 | 0.793378 |
| 4.5 | 43 | 0 | 3 | 0.870032 | 0.047057 | 0.954092 | 0.793378 |
| 5.5 | 40 | 0 | 5 | 0.870032 | 0.047057 | 0.954092 | 0.793378 |
| 6.5 | 35 | 1 | 1 | 0.845174 | 0.055269 | 0.941871 | 0.758404 |
| 7.5 | 33 | 0 | 3 | 0.845174 | 0.055269 | 0.941871 | 0.758404 |
| 8.5 | 30 | 0 | 3 | 0.845174 | 0.055269 | 0.941871 | 0.758404 |
| 9.5 | 27 | 0 | 2 | 0.845174 | 0.055269 | 0.941871 | 0.758404 |
| 10.5 | 25 | 0 | 3 | 0.845174 | 0.055269 | 0.941871 | 0.758404 |
| 11.5 | 22 | 0 | 2 | 0.845174 | 0.055269 | 0.941871 | 0.758404 |
| 12.5 | 20 | 0 | 4 | 0.845174 | 0.055269 | 0.941871 | 0.758404 |
| 14.5 | 16 | 0 | 2 | 0.845174 | 0.055269 | 0.941871 | 0.758404 |
| 15.5 | 14 | 1 | 0 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |
| 16.5 | 13 | 0 | 2 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |
| 18.5 | 11 | 0 | 1 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |
| 19.5 | 10 | 0 | 3 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |
| 20.5 | 7 | 0 | 1 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |
| 22.5 | 6 | 0 | 1 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |
| 24.5 | 5 | 0 | 1 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |
| 25.5 | 4 | 0 | 1 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |
| 26.5 | 3 | 0 | 2 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |
| 28.5 | 1 | 0 | 1 | 0.784805 | 0.092462 | 0.940731 | 0.654723 |

As at no point does the survival estimate reach 0.5 (as can be seen in the above table). With an undefined median, we cannot estimate a confidence interval, i.e.:

Question 6

Considering only the surgically placed catheter, we again implement the Kaplan-Meier method to estimate a new survival function, with the results as below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| time | n.risk | n.event | n.censor | Estimated S(t) | std.err of  S(t) estimate | upper | lower |
| 1.5 | 43 | 1 | 0 | 0.976744 | 0.023531 | 0.995468 | 0.885238 |
| 2.5 | 42 | 0 | 2 | 0.976744 | 0.023531 | 0.995468 | 0.885238 |
| 3.5 | 40 | 1 | 3 | 0.952326 | 0.034565 | 0.984862 | 0.855191 |
| 4.5 | 36 | 2 | 1 | 0.899419 | 0.053186 | 0.954618 | 0.78509 |
| 5.5 | 33 | 1 | 1 | 0.872163 | 0.061447 | 0.93676 | 0.750981 |
| 6.5 | 31 | 0 | 2 | 0.872163 | 0.061447 | 0.93676 | 0.750981 |
| 7.5 | 29 | 0 | 4 | 0.872163 | 0.061447 | 0.93676 | 0.750981 |
| 8.5 | 25 | 2 | 1 | 0.80239 | 0.08517 | 0.890015 | 0.659683 |
| 9.5 | 22 | 1 | 1 | 0.765918 | 0.097049 | 0.86367 | 0.615552 |
| 10.5 | 20 | 1 | 1 | 0.727622 | 0.109773 | 0.835095 | 0.570608 |
| 11.5 | 18 | 1 | 1 | 0.687199 | 0.123766 | 0.804108 | 0.524423 |
| 12.5 | 16 | 0 | 2 | 0.687199 | 0.123766 | 0.804108 | 0.524423 |
| 13.5 | 14 | 0 | 1 | 0.687199 | 0.123766 | 0.804108 | 0.524423 |
| 14.5 | 13 | 0 | 2 | 0.687199 | 0.123766 | 0.804108 | 0.524423 |
| 15.5 | 11 | 1 | 0 | 0.624726 | 0.156234 | 0.761519 | 0.443815 |
| 16.5 | 10 | 1 | 0 | 0.562254 | 0.188468 | 0.714558 | 0.372886 |
| 18.5 | 9 | 1 | 0 | 0.499781 | 0.222281 | 0.664039 | 0.30882 |
| 21.5 | 8 | 0 | 2 | 0.499781 | 0.222281 | 0.664039 | 0.30882 |
| 22.5 | 6 | 0 | 2 | 0.499781 | 0.222281 | 0.664039 | 0.30882 |
| 23.5 | 4 | 1 | 0 | 0.374836 | 0.364338 | 0.586967 | 0.164103 |
| 25.5 | 3 | 0 | 1 | 0.374836 | 0.364338 | 0.586967 | 0.164103 |
| 26.5 | 2 | 1 | 0 | 0.187418 | 0.795451 | 0.464645 | 0.025788 |
| 27.5 | 1 | 0 | 1 | 0.187418 | 0.795451 | 0.464645 | 0.025788 |

Considering the probability that a patient will experience no infection for at least 6 months, extracting the probability from the above table, we get to:

We estimate a 90% confidence interval using the below equation:

Arriving at the following result and highlighting in the table as well:

Question 7

The table below shows the patients at risk, Y, as a function of age.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | 62 | 63 | 65 | 66 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 76 | 77 |
| di | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| Yi | 6 | 9 | 10 | 8 | 12 | 11 | 10 | 11 | 10 | 9 | 9 | 7 | 5 |

With the initial patient at entering the study at age 58 and dying before age 60, they were included in the persons exposed at age 63, but not at 62. The reason being the person dying at age 62 only entered the study once the other person had already died.

Question 8

From the table in Q7, we set up the above plot of the survival estimates. The last person to die was aged 77. The function is a strictly decreasing function as expected and as can be seen from the plot. We can also discern a person with diabetes surviving to age 60 has just over 0.1 chance of surviving past 77 years of age.

Question 9

The survival function was estimated anew taking into account the survival rate of someone surviving to age 70. The conditional survival rates (conditional on surviving to age 70) appear to paint a slightly better picture than the conditional survival rates (conditional on surviving to age 60) of Q8, indicating you will have a better chance of surviving to age 77, once you have reached aged 70, than the long term survival rate of surviving from 60 to 77 (with diabetes).

Question 10

Considering all patients, without truncation, changes the table to:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | 60 | 62 | 63 | 65 | 66 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 76 | 77 |
| di | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 |
| Yi | 3 | 6 | 8 | 10 | 8 | 12 | 11 | 10 | 11 | 10 | 9 | 9 | 7 | 5 |

Question 11

Estimating the survival function without taking the truncation into consideration has a slight impact on the survival estimates. We see it tapering off slightly quicker than before, as the person dying at age 60 is no longer considered for the survival estimate at age 63.

Question 12

The same as with Q11, we can see the survival estimates tapering off more rapidly than before, as the number of patients exposed to the event is less than with truncation.

Appendix A

library(here)

library(dplyr)

library(data.table)

library(survival)

section1\_4 <- fread(here('Section1\_4.dat'),

sep = ' ', col.names = c('Time\_To\_Infection','Censored','Treatment'),

colClasses = c('numeric', 'numeric', 'numeric'))

kmp<-survfit(Surv(section1\_4$Time\_To\_Infection,section1\_4$Censored)~section1\_4$Treatment,type="kaplan-meier")

df\_summary <- fortify(kmp)

print(df\_summary)

print(kmp, print.rmean=TRUE)

print(18.6 + (qnorm(0.975,0,1)\*1.69))

print(18.6 - (qnorm(0.975,0,1)\*1.69))

print(23.3 + (qnorm(0.975,0,1)\*1.35))

print(23.3 - (qnorm(0.975,0,1)\*1.35))

surg\_placed = section1\_4[section1\_4$Treatment == 1]

kmp<-survfit(Surv(surg\_placed$Time\_To\_Infection,surg\_placed$Censored)~surg\_placed$Treatment,type="kaplan-meier", conf.int=0.9, conf.type='log-log')

surg\_placed\_summary <- fortify(kmp)

print(surg\_placed\_summary)

print(0.8721635 + (qnorm(0.95,0,1)\*0.06144666))

print(0.8721635 - (qnorm(0.95,0,1)\*0.06144666))