# INTRODUCCION AS
# TAREA 1

5.0

Por:

Daniela Pico

Juan Sebastián Falcón

Alexis Nieto

Entregado a:

Juan Carlos Salazar

**Universidad Nacional de Colombia**
**Sede Medellín**

FACULTAD DE CIENCIAS

**1) En las siguientes situaciones identifique: el evento de interes y la variable respuesta.**

**a)** Se realiza un seguimiento durante 13 años a un grupo de adultos de al menos 60 años de edad para ver que tanto tiempo permanecen vivos.

Evento de interés: La muerte

Variable respuesta: Tiempo de supervivencia en años para adultos de al menos 60 años de edad

**b)** Un grupo de pacientes con un trasplante de corazon son monitoreados para ver cuanto tiempo sobreviven.

Evento de interés: La muerte

Variable respuesta: Tiempo de supervivencia para pacientes que han sido sometidos a transplante de corazon

**2) Resuelva a MANO, usando los datos sobre Hepatitis Crónica Activa (CAH). Estos datos son de un ensayo clinico descrito en Kirk (1980). 44 pacientes con CAH se aleatorizaron a una droga llamada prednisolona o a un grupo de control sin tratamiento. El tiempo de supervivenvia en meses, despues de ser admitido al ensayo es la variables respuesta. Con estos datos obtenga una estimación de KM y el de NA. Note que debe evidenciar tablas de vida. No olvide el analisis descriptivo. Interprete**

A continuación se presenta los datos del ensayo clinico para 44 pacientes, los cuales están clasificados en dos grupos, un grupo de control con tratamiento y un grupo de control sin tratamiento, además se asigna el número 1 para las fallas y 0 para las censuras como se ilustra en la siguiente tabla

| id | group | time | status | id | group | time | status |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 1 | 1 | 2 | 2 | 1 |
| 2 | 1 | 2 | 1 | 2 | 2 | 3 | 1 |
| 3 | 1 | 12 | 1 | 3 | 2 | 4 | 1 |
| 4 | 1 | 54 | 1 | 4 | 2 | 7 | 1 |
| 5 | 1 | 56 | 0 | 5 | 2 | 10 | 1 |
| 6 | 1 | 68 | 1 | 6 | 2 | 22 | 1 |
| 7 | 1 | 89 | 1 | 7 | 2 | 28 | 1 |
| 8 | 1 | 96 | 1 | 8 | 2 | 29 | 1 |
| 9 | 1 | 96 | 1 | 9 | 2 | 32 | 1 |
| 10 | 1 | 125 | 0 | 10 | 2 | 37 | 1 |
| 11 | 1 | 128 | 0 | 11 | 2 | 40 | 1 |
| 12 | 1 | 131 | 0 | 12 | 2 | 41 | 1 |
| 13 | 1 | 140 | 0 | 13 | 2 | 54 | 1 |
| 14 | 1 | 141 | 0 | 14 | 2 | 61 | 1 |
| 15 | 1 | 143 | 1 | 15 | 2 | 63 | 1 |
| 16 | 1 | 145 | 0 | 16 | 2 | 71 | 1 |
| 17 | 1 | 146 | 1 | 17 | 2 | 127 | 0 |
| 18 | 1 | 148 | 0 | 18 | 2 | 140 | 0 |
| 19 | 1 | 162 | 0 | 19 | 2 | 146 | 0 |
| 20 | 1 | 168 | 1 | 20 | 2 | 158 | 0 |
| 21 | 1 | 173 | 0 | 21 | 2 | 167 | 0 |
| 22 | 1 | 181 | 0 | 22 | 2 | 182 | 0 |

Para resumir la información anterior de manera que se pueda obtener una aproximación de la curva de supervivencia se presentara las tablas de vida para cada grupo esto con el fin de facilitar la obteción de dicha curva

| $t_{(j)}$ | $d_j$ | $q_j$ | $n_j$ |
|---|---|---|---|
| $t_0$=0 | 0 | 0 | 22 (22 unidades sobreviven $\geq$ 0 meses) |
| $t_1$=2 | 2 | 0 | 22 (22 unidades sobreviven $\geq$ 2 meses) |
| $t_2$=12 | 1 | 0 | 20 (20 unidades sobreviven $\geq$ 12 meses) |
| $t_3$=54 | 1 | 1 | 19 (19 unidades sobreviven $\geq$ 54 meses) |
| $t_4$=68 | 1 | 0 | 17 (17 unidades sobreviven $\geq$ 68 meses) |
| $t_5$=89 | 1 | 0 | 16 (16 unidades sobreviven $\geq$ 89 meses) |
| $t_6$=96 | 2 | 5 | 15 (15 unidades sobreviven $\geq$ 96 meses) |
| $t_7$=143 | 1 | 1 | 8 (8 unidades sobreviven $\geq$ 143 meses) |
| $t_8$=146 | 1 | 2 | 6 (6 unidades sobreviven $\geq$ 146 meses) |
| $t_9$=168 | 1 | 2 | 3 (3 unidades sobreviven $\geq$ 168 meses) |
| TOTAL | 11 | 11 | 11+11=22 |

Table 1: Tabla de vida Grupo 1

Se procede a Analizar el tiempo promedio $\bar{T}$ y el Hazard promedio $\bar{h}$ para el Grupo 1 que se definen respectivamente como:

$$\bar{T} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

$$\bar{T} = \frac{1}{22}\sum_{i=1}^{22} t_i = 109.3636$$

$$\bar{h} = \frac{fallas}{n\bar{T}}$$

$$\bar{h} = \frac{11}{22(109.3636)} = 0.004571905$$

| $t_{(j)}$ | $d_j$ | $q_j$ | $n_j$ |
|---|---|---|---|
| $t_0$=0 | 0 | 0 | 22 (22 unidades sobreviven $\geq$ 0 meses) |
| $t_1$=2 | 1 | 0 | 22 (22 unidades sobreviven $\geq$ 2 meses) |
| $t_2$=3 | 1 | 0 | 21 (21 unidades sobreviven $\geq$ 3 meses) |
| $t_3$=4 | 1 | 0 | 20 (20 unidades sobreviven $\geq$ 4 meses) |
| $t_4$=7 | 1 | 0 | 19 (19 unidades sobreviven $\geq$ 7 meses) |
| $t_5$=10 | 1 | 0 | 18 (18 unidades sobreviven $\geq$ 10 meses) |
| $t_6$=22 | 1 | 0 | 17 (17 unidades sobreviven $\geq$ 22 meses) |
| $t_7$=28 | 1 | 0 | 16 (16 unidades sobreviven $\geq$ 28 meses) |
| $t_8$=29 | 1 | 0 | 15 (15 unidades sobreviven $\geq$ 29 meses) |
| $t_9$=32 | 1 | 0 | 14 (14 unidades sobreviven $\geq$ 32 meses) |
| $t_{10}$=37 | 1 | 0 | 13 (13 unidades sobreviven $\geq$ 37 meses) |
| $t_{11}$=40 | 1 | 0 | 12 (12 unidades sobreviven $\geq$ 40 meses) |
| $t_{12}$=41 | 1 | 0 | 11 (11 unidades sobreviven $\geq$ 41 meses) |
| $t_{13}$=54 | 1 | 0 | 10 (10 unidades sobreviven $\geq$ 54 meses) |
| $t_{14}$=61 | 1 | 0 | 9 (9 unidades sobreviven $\geq$ 61 meses) |
| $t_{15}$=63 | 1 | 0 | 8 (8 unidades sobreviven $\geq$ 63 meses) |
| $t_{16}$=71 | 1 | 6 | 7 (7 unidades sobreviven $\geq$ 71 meses) |
| TOTAL | 16 | 6 | 16+6=22 |

Table 2: Tabla de vida Grupo 2

El tiempo promedio $\bar{T}$ y el Hazard promedio $\bar{h}$ para el Grupo 2 que se definen respectivamente como:

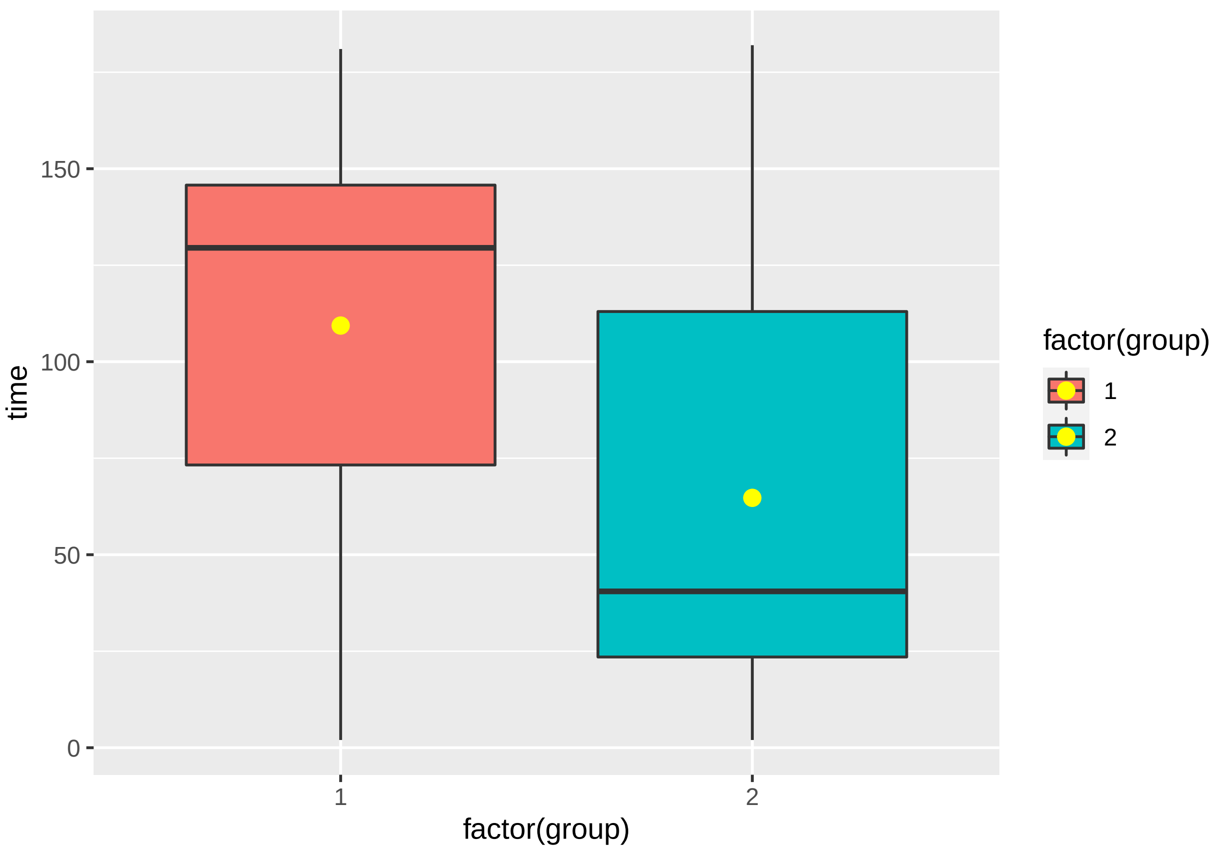
$$\bar{T} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

$$\bar{T} = \frac{1}{22}\sum_{i=1}^{22} t_i = 64.72727$$

$$\bar{h} = \frac{fallas}{n\bar{T}}$$

$$\bar{h} = \frac{11}{22(64.72727)} = 0.007724719$$

A mayor Hazard menor probabilidad de supervivencia, en este caso se oberva que $\bar{h_1} < \bar{h_2}$ y $\bar{T_1} > \bar{T_2}$ por lo que de una manera muy general se puede argumentar que la experiencia de supervivencia del grupo 2 es menor que en el grupo 1




En el Gráfico se observa el Boxplot correspondiente a cada grupo, en general se evidencia que los individuos del grupo 1 en comparación a los individuos del grupo 2 presentan un tiempo de supervivencia promedio mayor, además se observa que la media representada por el punto amarillo es mayor en el grupo 1 con respecto al grupo 2, finalmente en el grupo 1 el 50% de los datos esta aproximadamente por encima de 120 meses y el 50% por debajo, mientras que en el grupo 2 el 50% esta aproximadamente por encima de 40 meses y el 50% por debajo, es decir, que se presenta una diferencia significativa en el promedio de los tiempos hasta que el individuo falla entre el grupo 1 y el grupo 2.

## Estimación con KM - Grupo 1

| $t_{(j)}$ | $d_j$ | $q_j$ | $n_j$ | $\hat{S}_{KM}(t_j)$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 22 | 1-$\frac{0}{22}=1$ |
| 2 | 2 | 0 | 22 | $1\text{x}(1-\frac{2}{22})=\frac{10}{11}\approx 0.9091$ |
| 12 | 1 | 0 | 20 | $\frac{10}{11}\text{x}(1-\frac{1}{20})=\frac{19}{22}\approx 0.8636$ |
| 54 | 1 | 1 | 19 | $\frac{19}{22}\text{x}(1-\frac{1}{19})=\frac{9}{11}\approx 0.8182$ |
| 68 | 1 | 0 | 17 | $\frac{9}{11}\text{x}(1-\frac{1}{17})=\frac{144}{187}\approx 0.7701$ |
| 89 | 1 | 0 | 16 | $\frac{144}{187}\text{x}(1-\frac{1}{16})=\frac{135}{187}\approx 0.7221$ |
| 96 | 2 | 5 | 15 | $\frac{135}{187}\text{x}(1-\frac{2}{15})=\frac{117}{187}\approx 0.6257$ |
| 143 | 1 | 1 | 8 | $\frac{117}{187}\text{x}(1-\frac{1}{8})=\frac{819}{1496}\approx 0.5475$ |
| 146 | 1 | 2 | 6 | $\frac{819}{1496}\text{x}(1-\frac{1}{6})=\frac{1365}{2992}\approx 0.4562$ |
| 168 | 1 | 2 | 3 | $\frac{1365}{2992}\text{x}(1-\frac{1}{3})=\frac{455}{1496}\approx 0.3041$ |

## Estimación con KM - Grupo 2

| $t_{(j)}$ | $d_j$ | $q_j$ | $n_j$ | $\hat{S}_{KM}(t_j)$ |
|---|---|---|---|---|
| 0 | 0 | 0 | 22 | 1-$\frac{0}{22}=1$ |
| 2 | 1 | 0 | 22 | $1\text{x}(1-\frac{1}{22})=\frac{21}{22}\approx 0.9545$ |
| 3 | 1 | 0 | 21 | $\frac{21}{22}\text{x}(1-\frac{1}{21})=\frac{10}{11}\approx 0.9091$ |
| 4 | 1 | 0 | 20 | $\frac{10}{11}\text{x}(1-\frac{1}{20})=\frac{19}{22}\approx 0.8636$ |
| 7 | 1 | 0 | 19 | $\frac{19}{22}\text{x}(1-\frac{1}{19})=\frac{9}{11}\approx 0.8182$ |
| 10 | 1 | 0 | 18 | $\frac{9}{11}\text{x}(1-\frac{1}{18})=\frac{17}{22}\approx 0.7727$ |
| 22 | 1 | 0 | 17 | $\frac{17}{22}\text{x}(1-\frac{1}{17})=\frac{8}{11}\approx 0.7273$ |
| 28 | 1 | 0 | 16 | $\frac{8}{11}\text{x}(1-\frac{1}{16})=\frac{15}{22}\approx 0.6818$ |
| 29 | 1 | 0 | 15 | $\frac{15}{22}\text{x}(1-\frac{1}{15})=\frac{7}{11}\approx 0.6364$ |
| 32 | 1 | 0 | 14 | $\frac{7}{11}\text{x}(1-\frac{1}{14})=\frac{13}{22}\approx 0.5909$ |
| 37 | 1 | 0 | 13 | $\frac{13}{22}\text{x}(1-\frac{1}{13})=\frac{6}{11}\approx 0.5455$ |
| 40 | 1 | 0 | 12 | $\frac{6}{11}\text{x}(1-\frac{1}{12})=\frac{1}{2}\approx 0.5$ |
| 41 | 1 | 0 | 11 | $\frac{1}{2}\text{x}(1-\frac{1}{11})=\frac{5}{11}\approx 0.4545$ |
| 54 | 1 | 0 | 10 | $\frac{5}{11}\text{x}(1-\frac{1}{10})=\frac{9}{22}\approx 0.4091$ |
| 61 | 1 | 0 | 9 | $\frac{9}{22}\text{x}(1-\frac{1}{9})=\frac{4}{11}\approx 0.3636$ |
| 63 | 1 | 0 | 8 | $\frac{4}{11}\text{x}(1-\frac{1}{8})=\frac{7}{22}\approx 0.3182$ |
| 71 | 1 | 6 | 7 | $\frac{7}{22}\text{x}(1-\frac{1}{7})=\frac{3}{11}\approx 0.2727$ |

## Estimación con NA - Grupo 1

| Falla Número | meses | $\hat{h}_t$ | $\hat{H}_t$ | $S(t)$ |
|---|---|---|---|---|
| 1-2 | 2 | $\frac{2}{22}$ | $\frac{2}{22}\approx 0.0909$ | 0.9131 |
| 3 | 12 | $\frac{1}{20}$ | $\frac{2}{22}+\frac{1}{20}=\frac{31}{220}\approx 0.1409$ | 0.8686 |
| 4 | 54 | $\frac{1}{19}$ | $\frac{31}{220}+\frac{1}{19}=\frac{809}{4180}\approx 0.1935$ | 0.8240 |
| 5 | 68 | $\frac{1}{17}$ | $\frac{809}{4180}+\frac{1}{17}=\frac{17933}{771060}\approx 0.2524$ | 0.7769 |
| 6 | 89 | $\frac{1}{16}$ | $\frac{17933}{771060}+\frac{1}{16}=\frac{89497}{284240}\approx 0.3149$ | 0.7299 |
| 7-8 | 96 | $\frac{2}{15}$ | $\frac{89497}{284240}+\frac{2}{15}=\frac{382187}{852720}\approx 0.4482$ | 0.6388 |
| 9 | 143 | $\frac{1}{8}$ | $\frac{382187}{852720}+\frac{1}{8}\approx 0.5732$ | 0.5637 |
| 10 | 146 | $\frac{1}{6}$ | $0.5732+\frac{1}{6}\approx 0.7399$ | 0.4772 |
| 11 | 168 | $\frac{1}{3}$ | $0.7399+\frac{1}{3}\approx 1.0732$ | 0.3419 |

**Estimación con NA - Grupo 2**

| Falla Número | meses | $\hat{h}_t$ | $\hat{H}_t$ | $S(t)$ |
|---|---|---|---|---|
| 1 | 2 | $\frac{1}{22}$ | $\frac{1}{22} \approx 0.0455$ | 0.9556 |
| 2 | 3 | $\frac{1}{21}$ | $\frac{1}{22} + \frac{1}{21} = \frac{43}{462} \approx 0.0931$ | 0.9111 |
| 3 | 4 | $\frac{1}{20}$ | $\frac{43}{462} + \frac{1}{20} = \frac{661}{4620} \approx 0.1431$ | 0.8667 |
| 4 | 7 | $\frac{1}{19}$ | $\frac{661}{4620} + \frac{1}{19} = \frac{17179}{87780} \approx 0.1957$ | 0.8223 |
| 5 | 10 | $\frac{1}{18}$ | $\frac{17179}{87780} + \frac{1}{18} = \frac{66167}{263340} \approx 0.2513$ | 0.7778 |
| 6 | 22 | $\frac{1}{17}$ | $\approx 0.3101$ | 0.7334 |
| 7 | 28 | $\frac{1}{16}$ | $\approx 0.3716$ | 0.6896 |
| 8 | 29 | $\frac{1}{15}$ | $\approx 0.4393$ | 0.6445 |
| 9 | 32 | $\frac{1}{14}$ | $\approx 0.5107$ | 0.6001 |
| 10 | 37 | $\frac{1}{13}$ | $\approx 0.5876$ | 0.5557 |
| 11 | 40 | $\frac{1}{12}$ | $\approx 0.6709$ | 0.5112 |
| 12 | 41 | $\frac{1}{11}$ | $\approx 0.7618$ | 0.4668 |
| 13 | 54 | $\frac{1}{10}$ | $\approx 0.8618$ | 0.4224 |
| 14 | 61 | $\frac{1}{9}$ | $\approx 0.9730$ | 0.3779 |
| 15 | 63 | $\frac{1}{8}$ | $\approx 1.0980$ | 0.3335 |
| 16 | 71 | $\frac{1}{7}$ | $\approx 1.2408$ | 0.2891 |

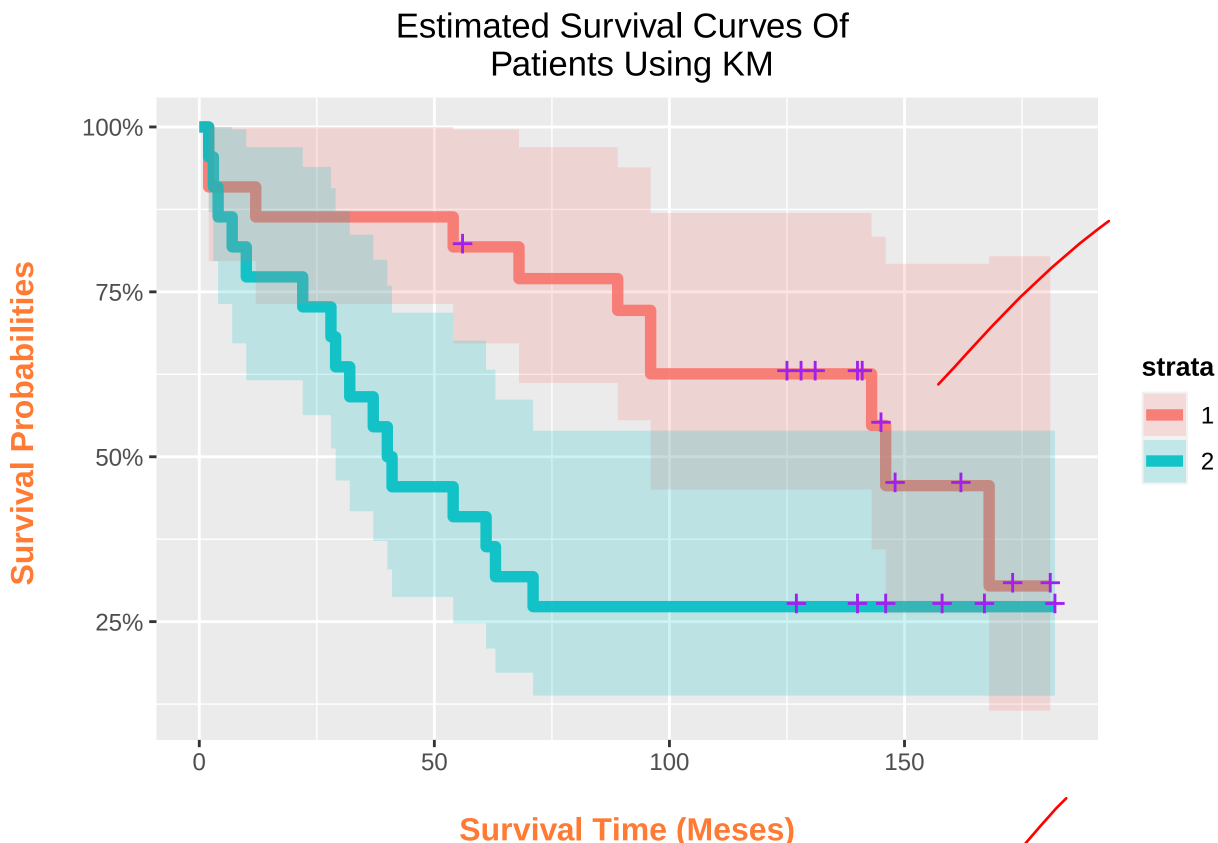
**3) Usando SAS o R y los datos del numeral anterior, obtenga estimaciones para la curva de supervivencia usando KM y NA**

```
CAH$group<-as.factor(CAH$group)
df=data.frame(CAH)
KM_fit <- survfit(Surv(time, status) ~ group, data = df)
summary(KM_fit)
```

```
## Call: survfit(formula = Surv(time, status) ~ group, data = df)
##
##                 group=1
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     2     22       2    0.909  0.0613        0.797        1.000
##    12     20       1    0.864  0.0732        0.732        1.000
##    54     19       1    0.818  0.0822        0.672        0.996
##    68     17       1    0.770  0.0904        0.612        0.969
##    89     16       1    0.722  0.0967        0.555        0.939
##    96     15       2    0.626  0.1051        0.450        0.870
##   143      8       1    0.547  0.1175        0.359        0.834
##   146      6       1    0.456  0.1285        0.263        0.793
##   168      3       1    0.304  0.1509        0.115        0.804
##
##                 group=2
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     2     22       1    0.955  0.0444        0.871        1.000
##     3     21       1    0.909  0.0613        0.797        1.000
##     4     20       1    0.864  0.0732        0.732        1.000
##     7     19       1    0.818  0.0822        0.672        0.996
##    10     18       1    0.773  0.0893        0.616        0.969
##    22     17       1    0.727  0.0950        0.563        0.939
##    28     16       1    0.682  0.0993        0.513        0.907
```

```
## 29 15 1 0.636 0.1026 0.464 0.873
## 32 14 1 0.591 0.1048 0.417 0.837
## 37 13 1 0.545 0.1062 0.372 0.799
## 40 12 1 0.500 0.1066 0.329 0.759
## 41 11 1 0.455 0.1062 0.288 0.718
## 54 10 1 0.409 0.1048 0.248 0.676
## 61  9 1 0.364 0.1026 0.209 0.632
## 63  8 1 0.318 0.0993 0.173 0.587
## 71  7 1 0.273 0.0950 0.138 0.540
```

Se observan las estimaciones calculadas implementando R en la columna "survival", note que dichas estimaciones son aproximadamente iguales a las estimaciones calculadas a mano, finalmente se procede a Graficar la estimación de la curva de supervivencia a través de KM

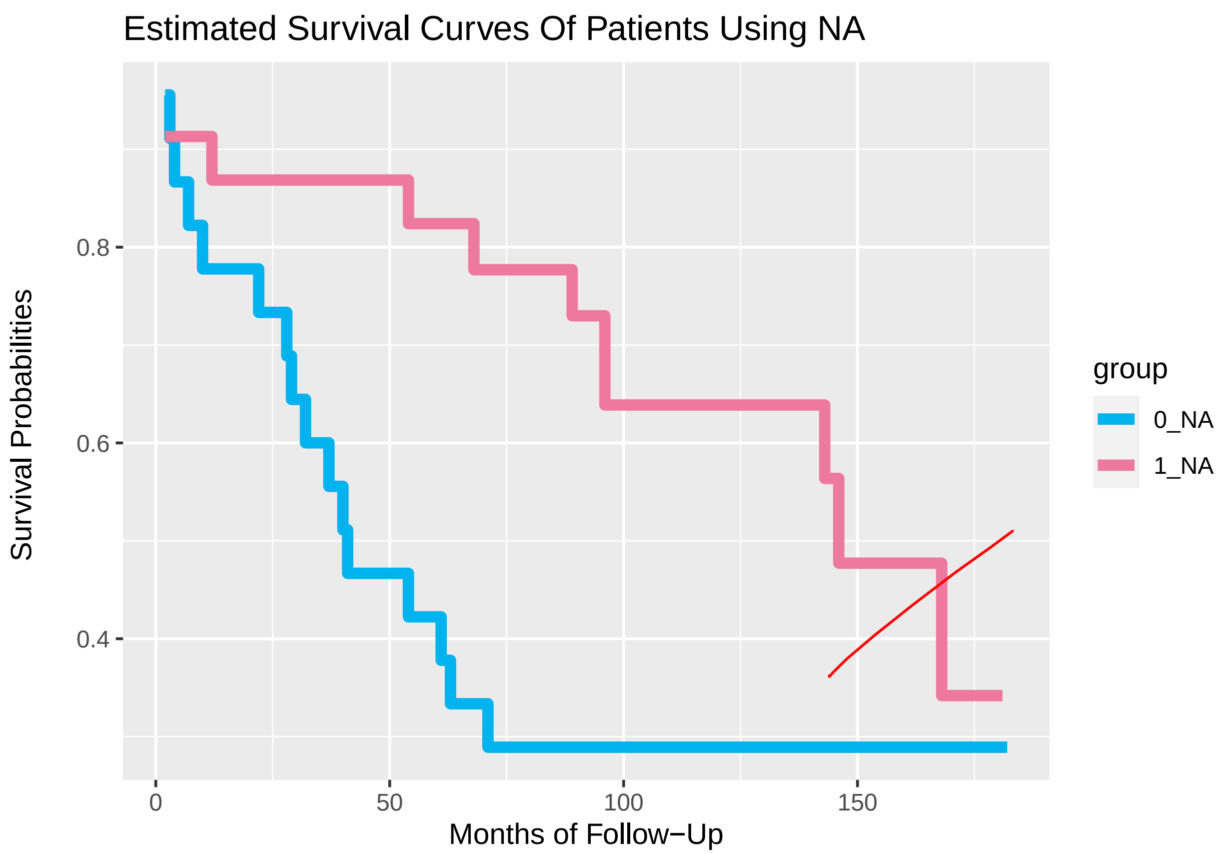



En la Gráfica anterior se observa la función de supervivencia que proporciona la probabilidad de que una persona sobreviva más allá de algún tiempo especifico, por ejemplo en el tiempo 0 (al inicio del estudio), como nadie ha experimentado el evento de interes la probabilidad de supervivencia es del 100% y a medida que transcurren los meses está probabilidad disminuye para cada grupo, sin embargo, se observa que la probabilidad decrece más rapidamente para el grupo 2, por ejemplo, cuando han pasado aproximadamente 8 años la probabilidad de supervivencia para el grupo sin control de tratamiento es aproximadamente del 27%, mientras que el grupo 1 que está bajo el efecto de la droga prednisilona tiene una probabilidad de supervivencia del 62.5%, es decir, que la medicina aporta una supervivencia significativamente mejorada con respecto al grupo de control sin tratamiento.

```
##    time        N_A group method        ST
## 1     2 0.09090909  1_NA    N-A 0.9131007
## 2    12 0.14090909  1_NA    N-A 0.8685683
## 3    54 0.19354067  1_NA    N-A 0.8240363
## 4    56 0.19354067  1_NA    N-A 0.8240363
## 5    68 0.25236420  1_NA    N-A 0.7769617
## 6    89 0.31486420  1_NA    N-A 0.7298880
## 7    96 0.44819753  1_NA    N-A 0.6387785
## 8   125 0.44819753  1_NA    N-A 0.6387785
## 9   128 0.44819753  1_NA    N-A 0.6387785
## 10  131 0.44819753  1_NA    N-A 0.6387785
## 11  140 0.44819753  1_NA    N-A 0.6387785
## 12  141 0.44819753  1_NA    N-A 0.6387785
## 13  143 0.57319753  1_NA    N-A 0.5637200
## 14  145 0.57319753  1_NA    N-A 0.5637200
## 15  146 0.73986420  1_NA    N-A 0.4771787
## 16  148 0.73986420  1_NA    N-A 0.4771787
## 17  162 0.73986420  1_NA    N-A 0.4771787
## 18  168 1.07319753  1_NA    N-A 0.3419135
## 19  173 1.07319753  1_NA    N-A 0.3419135
## 20  181 1.07319753  1_NA    N-A 0.3419135
```

```
##    time        N_A group method        ST
## 1     2 0.04545455  0_NA    N-A 0.9555630
## 2     3 0.09307359  0_NA    N-A 0.9111264
## 3     4 0.14307359  0_NA    N-A 0.8666903
## 4     7 0.19570517  0_NA    N-A 0.8222546
## 5    10 0.25126073  0_NA    N-A 0.7778195
## 6    22 0.31008426  0_NA    N-A 0.7333852
## 7    28 0.37258426  0_NA    N-A 0.6889516
## 8    29 0.43925092  0_NA    N-A 0.6445190
## 9    32 0.51067950  0_NA    N-A 0.6000877
## 10   37 0.58760257  0_NA    N-A 0.5556578
## 11   40 0.67093591  0_NA    N-A 0.5112299
## 12   41 0.76184500  0_NA    N-A 0.4668044
## 13   54 0.86184500  0_NA    N-A 0.4223821
## 14   61 0.97295611  0_NA    N-A 0.3779641
## 15   63 1.09795611  0_NA    N-A 0.3335521
## 16   71 1.24081325  0_NA    N-A 0.2891490
## 17  127 1.24081325  0_NA    N-A 0.2891490
## 18  140 1.24081325  0_NA    N-A 0.2891490
## 19  146 1.24081325  0_NA    N-A 0.2891490
## 20  158 1.24081325  0_NA    N-A 0.2891490
## 21  167 1.24081325  0_NA    N-A 0.2891490
## 22  182 1.24081325  0_NA    N-A 0.2891490
```

Se observan las estimaciones calculadas para los grupos 1 y 2 respectivamente implementando R en la columna "ST" que fue hallada mediante la relación entre el Hazard acumulado y S(t) de la siguiente manera: $S(t) = e^{-H(t)}$. Note que dichas estimaciones son aproximadamente iguales a las estimaciones calculadas a mano, finalmente se procede a Graficar la estimación de la curva de supervivencia a través de la relación entre NA y S(t).




En esta Gráfica se observa que la funcion de supervivencia tiene un comportamiento que es aproximadamente igual a la calculada con KM, por tanto, tienen la misma interpretación.

**4) Usando de nuevo los datos de CAH y A MANO obtenga estimaciones para H(t) usando -Log(S(t)) y NA. Grafique las estimaciones a mano. Note que se deben evidenciar las tablas usadas en el proceso. Interprete.**

Teniendo en cuenta la relacion entre el Hazard acumulado y $S(t)$, donde $\hat{H}_{NA}(t) = -Log(\hat{S}_{KM}(tj))$. Se calculan las estimaciones para ambos grupos de la siguiente manera

**Estimación con -Log(S(t)) - Grupo 1**

| $t_{(j)}$ | $d_j$ | $q_j$ | $n_j$ | $\hat{S}_{KM}(t_j)$ | $\hat{H}_{NA}(t_j)$ |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 22 | $1-\frac{0}{22}=1$ | 0 |
| 2 | 2 | 0 | 22 | $1x(1-\frac{2}{22})=\frac{10}{11}\approx 0.9091$ | 0.09530 |
| 12 | 1 | 0 | 20 | $\frac{10}{11}x(1-\frac{1}{20})=\frac{19}{22}\approx 0.8636$ | 0.14664 |
| 54 | 1 | 1 | 19 | $\frac{19}{22}x(1-\frac{1}{19})=\frac{9}{11}\approx 0.8182$ | 0.20065 |
| 68 | 1 | 0 | 17 | $\frac{9}{11}x(1-\frac{1}{17})=\frac{144}{187}\approx 0.7701$ | 0.26123 |
| 89 | 1 | 0 | 16 | $\frac{144}{187}x(1-\frac{1}{16})=\frac{135}{187}\approx 0.7221$ | 0.32559 |
| 96 | 2 | 5 | 15 | $\frac{135}{187}x(1-\frac{2}{15})=\frac{117}{187}\approx 0.6257$ | 0.46888 |
| 143 | 1 | 1 | 8 | $\frac{117}{187}x(1-\frac{1}{8})=\frac{819}{1496}\approx 0.5475$ | 0.6024 |
| 146 | 1 | 2 | 6 | $\frac{819}{1496}x(1-\frac{1}{6})=\frac{1365}{2992}\approx 0.4562$ | 0.7848 |
| 168 | 1 | 2 | 3 | $\frac{1365}{2992}x(1-\frac{1}{3})=\frac{455}{1496}\approx 0.3041$ | 1.1904 |

**Estimación con -Log(S(t)) - Grupo 2**

| $t_{(j)}$ | $d_j$ | $q_j$ | $n_j$ | $\hat{S}_{KM}(t_j)$ | $\hat{H}_{NA}(t_j)$ |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 22 | 1-$\frac{0}{22} = 1$ | 0 |
| 2 | 1 | 0 | 22 | 1x$(1 - \frac{1}{22}) = \frac{21}{22} \approx 0.9545$ | 0.04657 |
| 3 | 1 | 0 | 21 | $\frac{21}{22}$x$(1 - \frac{1}{21})$=$\frac{10}{11} \approx 0.9091$ | 0.0953 |
| 4 | 1 | 0 | 20 | $\frac{10}{11}$x$(1 - \frac{1}{20})$=$\frac{19}{22} \approx 0.8636$ | 0.14665 |
| 7 | 1 | 0 | 19 | $\frac{19}{22}$x$(1 - \frac{1}{19})$=$\frac{9}{11} \approx 0.8182$ | 0.20065 |
| 10 | 1 | 0 | 18 | $\frac{9}{11}$x$(1 - \frac{1}{18})$=$\frac{17}{22} \approx 0.7727$ | 0.25786 |
| 22 | 1 | 0 | 17 | $\frac{17}{22}$x$(1 - \frac{1}{17})$=$\frac{8}{11} \approx 0.7273$ | 0.31842 |
| 28 | 1 | 0 | 16 | $\frac{8}{11}$x$(1 - \frac{1}{16})$=$\frac{15}{22} \approx 0.6818$ | 0.38302 |
| 29 | 1 | 0 | 15 | $\frac{15}{22}$x$(1 - \frac{1}{15})$=$\frac{7}{11} \approx 0.6364$ | 0.45193 |
| 32 | 1 | 0 | 14 | $\frac{7}{11}$x$(1 - \frac{1}{14})$=$\frac{13}{22} \approx 0.5909$ | 0.52611 |
| 37 | 1 | 0 | 13 | $\frac{13}{22}$x$(1 - \frac{1}{13})$=$\frac{6}{11} \approx 0.5455$ | 0.60605 |
| 40 | 1 | 0 | 12 | $\frac{6}{11}$x$(1 - \frac{1}{12})$=$\frac{1}{2} \approx 0.5$ | 0.69315 |
| 41 | 1 | 0 | 11 | $\frac{1}{2}$x$(1 - \frac{1}{11})$=$\frac{5}{11} \approx 0.4545$ | 0.78856 |
| 54 | 1 | 0 | 10 | $\frac{5}{11}$x$(1 - \frac{1}{10})$=$\frac{9}{22} \approx 0.4091$ | 0.8938 |
| 61 | 1 | 0 | 9 | $\frac{9}{22}$x$(1 - \frac{1}{9})$=$\frac{4}{11} \approx 0.3636$ | 1.0117 |
| 63 | 1 | 0 | 8 | $\frac{4}{11}$x$(1 - \frac{1}{8})$=$\frac{7}{22} \approx 0.3182$ | 1.14508 |
| 71 | 1 | 6 | 7 | $\frac{7}{22}$x$(1 - \frac{1}{7})$=$\frac{3}{11} \approx 0.2727$ | 1.29938 |

Posteriormente se estima H(t) mediante Nelson-Aalen de la siguiente manera

**Estimación con NA - Grupo 1**

| Falla Número | meses | $\hat{h}_t$ | $\hat{H}_t$ |
|---|---|---|---|
| 1-2 | 2 | $\frac{2}{22}$ | $\frac{2}{22} \approx 0.0909$ |
| 3 | 12 | $\frac{1}{20}$ | $\frac{2}{22} + \frac{1}{20} = \frac{31}{220} \approx 0.1409$ |
| 4 | 54 | $\frac{1}{19}$ | $\frac{31}{220} + \frac{1}{19} = \frac{809}{4180} \approx 0.1935$ |
| 5 | 68 | $\frac{1}{17}$ | $\frac{809}{4180} + \frac{1}{17} = \frac{17933}{771060} \approx 0.2524$ |
| 6 | 89 | $\frac{1}{16}$ | $\frac{17933}{771060} + \frac{1}{16} = \frac{89497}{284240} \approx 0.3149$ |
| 7-8 | 96 | $\frac{2}{15}$ | $\frac{89497}{284240} + \frac{2}{15} = \frac{382187}{852720} \approx 0.4482$ |
| 9 | 143 | $\frac{1}{8}$ | $\frac{382187}{852720} + \frac{1}{8} \approx 0.5732$ |
| 10 | 146 | $\frac{1}{6}$ | $0.5732 + \frac{1}{6} \approx 0.7399$ |
| 11 | 168 | $\frac{1}{3}$ | $0.7399 + \frac{1}{3} \approx 1.0732$ |

**Estimación con NA - Grupo 2**

| Falla Número | meses | $\hat{h}_t$ | $\hat{H}_t$ |
|---|---|---|---|
| 1 | 2 | $\frac{1}{22}$ | $\frac{1}{22} \approx 0.0455$ |
| 2 | 3 | $\frac{1}{21}$ | $\frac{1}{22} + \frac{1}{21} = \frac{43}{462} \approx 0.0931$ |
| 3 | 4 | $\frac{1}{20}$ | $\frac{43}{462} + \frac{1}{20} = \frac{661}{4620} \approx 0.1431$ |
| 4 | 7 | $\frac{1}{19}$ | $\frac{661}{4620} + \frac{1}{19} = \frac{17179}{87780} \approx 0.1957$ |
| 5 | 10 | $\frac{1}{18}$ | $\frac{17179}{87780} + \frac{1}{18} = \frac{66167}{263340} \approx 0.2513$ |
| 6 | 22 | $\frac{1}{17}$ | $\frac{66167}{263340} + \frac{1}{17} \approx 0.3101$ |
| 7 | 28 | $\frac{1}{16}$ | $0.3101 + \frac{1}{16} \approx 0.3716$ |
| 8 | 29 | $\frac{1}{15}$ | $0.3716 + \frac{1}{15} \approx 0.4393$ |
| 9 | 32 | $\frac{1}{14}$ | $0.4393 + \frac{1}{14} \approx 0.5107$ |
| 10 | 37 | $\frac{1}{13}$ | $0.5107 + \frac{1}{13} \approx 0.5876$ |
| 11 | 40 | $\frac{1}{12}$ | $0.5876 + \frac{1}{12} \approx 0.6709$ |
| 12 | 41 | $\frac{1}{11}$ | $0.6709 + \frac{1}{11} \approx 0.7618$ |
| 13 | 54 | $\frac{1}{10}$ | $0.7618 + \frac{1}{10} \approx 0.8618$ |
| 14 | 61 | $\frac{1}{9}$ | $0.8618 + \frac{1}{9} \approx 0.9730$ |
| 15 | 63 | $\frac{1}{8}$ | $0.9730 + \frac{1}{8} \approx 1.0980$ |
| 16 | 71 | $\frac{1}{7}$ | $1.0980 + \frac{1}{7} \approx 1.2408$ |

**5) Haga los mismo del numeral anterior pero usando SAS o R**

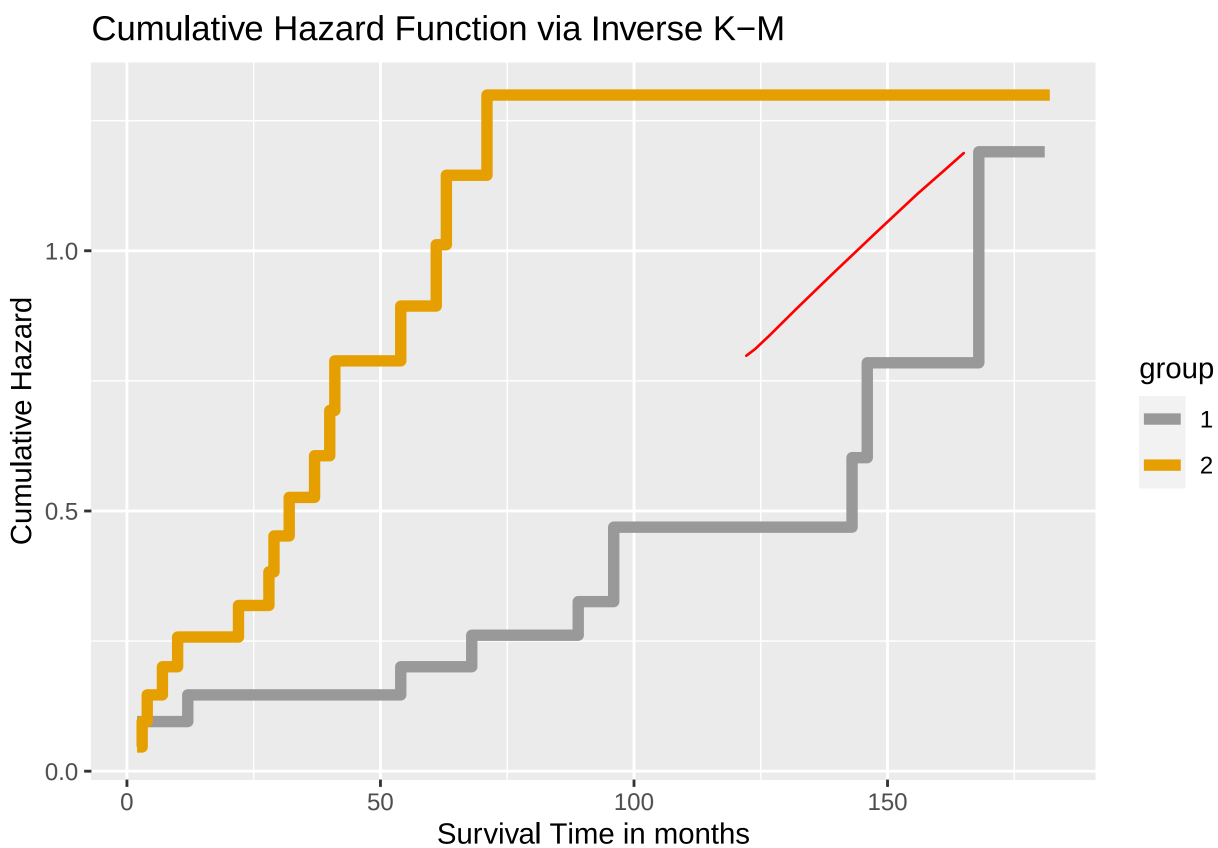
con -Log(S(t))

```
##    time        N_A group     method
## 1     2 0.09531018     1 Inverse_KM
## 2    12 0.14660347     1 Inverse_KM
## 3    54 0.20067070     1 Inverse_KM
## 4    56 0.20067070     1 Inverse_KM
## 5    68 0.26129532     1 Inverse_KM
## 6    89 0.32583384     1 Inverse_KM
## 7    96 0.46893468     1 Inverse_KM
## 8   125 0.46893468     1 Inverse_KM
## 9   128 0.46893468     1 Inverse_KM
## 10  131 0.46893468     1 Inverse_KM
## 11  140 0.46893468     1 Inverse_KM
## 12  141 0.46893468     1 Inverse_KM
## 13  143 0.60246607     1 Inverse_KM
## 14  145 0.60246607     1 Inverse_KM
## 15  146 0.78478763     1 Inverse_KM
## 16  148 0.78478763     1 Inverse_KM
## 17  162 0.78478763     1 Inverse_KM
## 18  168 1.19025274     1 Inverse_KM
## 19  173 1.19025274     1 Inverse_KM
## 20  181 1.19025274     1 Inverse_KM
```

```
##    time        N_A group     method
## 1     2 0.04652002     2 Inverse_KM
## 2     3 0.09531018     2 Inverse_KM
## 3     4 0.14660347     2 Inverse_KM
## 4     7 0.20067070     2 Inverse_KM
## 5    10 0.25782911     2 Inverse_KM
```

```
## 6     22 0.31845373      2 Inverse_KM
## 7     28 0.38299225      2 Inverse_KM
## 8     29 0.45198512      2 Inverse_KM
## 9     32 0.52609310      2 Inverse_KM
## 10    37 0.60613580      2 Inverse_KM
## 11    40 0.69314718      2 Inverse_KM
## 12    41 0.78845736      2 Inverse_KM
## 13    54 0.89381788      2 Inverse_KM
## 14    61 1.01160091      2 Inverse_KM
## 15    63 1.14513230      2 Inverse_KM
## 16    71 1.29928298      2 Inverse_KM
## 17   127 1.29928298      2 Inverse_KM
## 18   140 1.29928298      2 Inverse_KM
## 19   146 1.29928298      2 Inverse_KM
## 20   158 1.29928298      2 Inverse_KM
## 21   167 1.29928298      2 Inverse_KM
## 22   182 1.29928298      2 Inverse_KM
```

Se estiman el H(t) mediante R y se observa que los valores son muy similares a los estimados en la tabla a mano. Se procede a graficar ambos resultados.

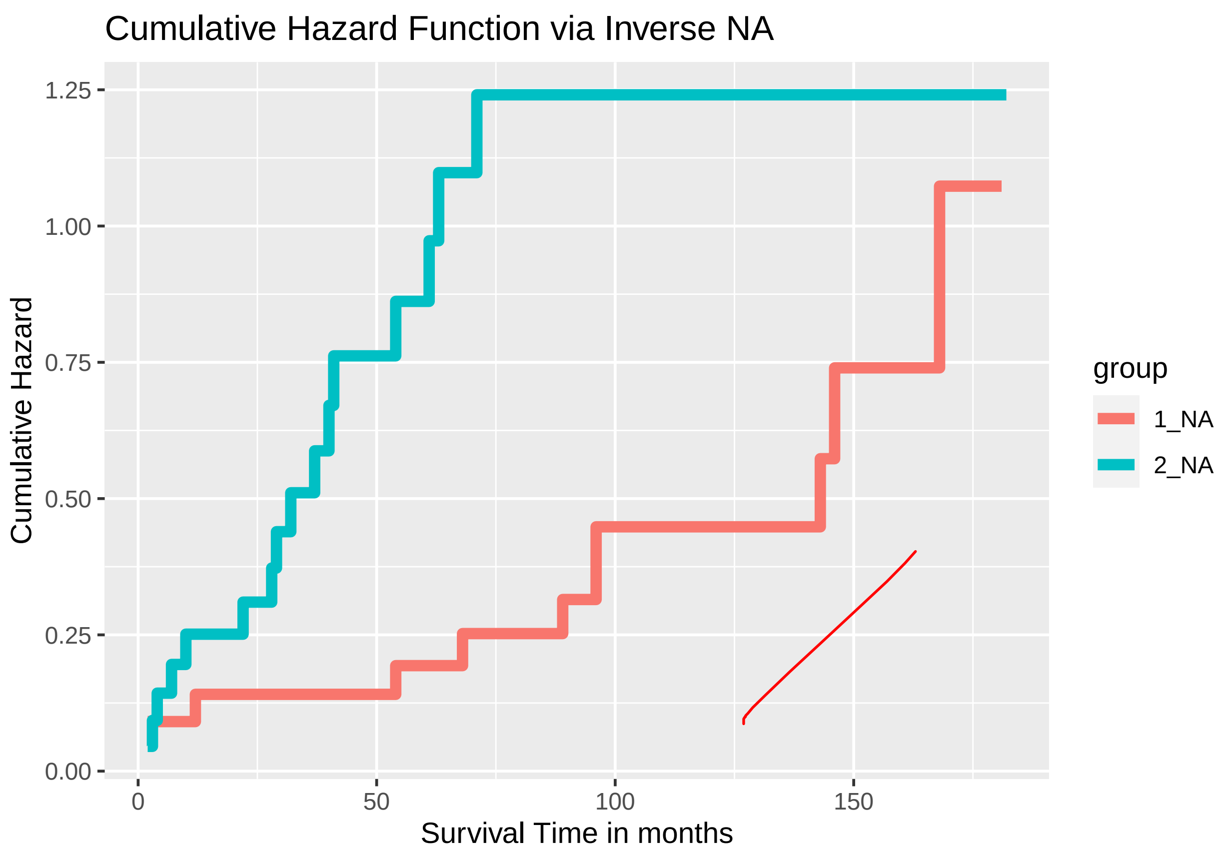



En la grafica del Hazard acumulado se observa el potencial de fallo instantaneo acumulado a medida que pasa el tiempo, por ello en el primer tiempo este valor es muy bajo, pues hay muy poco potencial de que alguno de los pacientes se muera. Sin embargo se observa desde los primeros meses, este potencial instantaneo crece rapidamente para el grupo 2 en comparacion con el grupo 1. Por ejemplo se observa que para el grupo 1, cuando el Hazard acumulado es de aproximadamente 0.8, han pasado casi 150 meses, mientras que para el grupo 2 solo tuvieron que pasar unos 40 meses.

Se ralizan las estimaciones y las graficas mediante NA

```
##    time        N_A group method
## 1     2 0.09090909  1_NA    N-A
## 2    12 0.14090909  1_NA    N-A
## 3    54 0.19354067  1_NA    N-A
## 4    56 0.19354067  1_NA    N-A
## 5    68 0.25236420  1_NA    N-A
## 6    89 0.31486420  1_NA    N-A
## 7    96 0.44819753  1_NA    N-A
## 8   125 0.44819753  1_NA    N-A
## 9   128 0.44819753  1_NA    N-A
## 10  131 0.44819753  1_NA    N-A
## 11  140 0.44819753  1_NA    N-A
## 12  141 0.44819753  1_NA    N-A
## 13  143 0.57319753  1_NA    N-A
## 14  145 0.57319753  1_NA    N-A
## 15  146 0.73986420  1_NA    N-A
## 16  148 0.73986420  1_NA    N-A
## 17  162 0.73986420  1_NA    N-A
## 18  168 1.07319753  1_NA    N-A
## 19  173 1.07319753  1_NA    N-A
## 20  181 1.07319753  1_NA    N-A
```

```
##    time        N_A group method
## 1     2 0.04545455  2_NA    N-A
## 2     3 0.09307359  2_NA    N-A
## 3     4 0.14307359  2_NA    N-A
## 4     7 0.19570517  2_NA    N-A
## 5    10 0.25126073  2_NA    N-A
## 6    22 0.31008426  2_NA    N-A
## 7    28 0.37258426  2_NA    N-A
## 8    29 0.43925092  2_NA    N-A
## 9    32 0.51067950  2_NA    N-A
## 10   37 0.58760257  2_NA    N-A
## 11   40 0.67093591  2_NA    N-A
## 12   41 0.76184500  2_NA    N-A
## 13   54 0.86184500  2_NA    N-A
## 14   61 0.97295611  2_NA    N-A
## 15   63 1.09795611  2_NA    N-A
## 16   71 1.24081325  2_NA    N-A
## 17  127 1.24081325  2_NA    N-A
## 18  140 1.24081325  2_NA    N-A
## 19  146 1.24081325  2_NA    N-A
## 20  158 1.24081325  2_NA    N-A
## 21  167 1.24081325  2_NA    N-A
## 22  182 1.24081325  2_NA    N-A
```




De igual manera se observa que el potencial instantaneo de fallo acumulado es mucho mayor en el grupo 2, y que mediante este metodo las estimaciones son muy similares.

**6) Usando de nuevo los datos de CAH y A MANO obtenga el logrank test. Interprete**

La finalidad del logrank test es estudiar si existe evidencia estadística a favor o en contra de que las 2 curvas de supervivencias correspondientes a cada grupo son iguales o no. Todo esto a partir de la siguiente prueba estadística:

$$H_0 : S_1(t) = S_2(t) \text{ vs } H_1 : S_1(t) \neq S_2(t)$$

$S_1(t)$ función de sepervivencia para el Grupo 1: Tratamiento Prednisolona; $S_2(t)$ función de sepervivencia para el Grupo 2: Control. El estadístico de prueba asociado distribuye como $\chi^2_{k-1}$, con $k-1$ grados de libertad y $k$ como el número de grupos bajo comparación. Se define como:

$$Log - Rank = \frac{(O_i - E_i)^2}{var(O_i - E_i)}, \ i = 1, 2.$$

En esta expresión $O_i$: valor observado en el grupo $i$, $E_i$: valor esperado en el grupo $i$ $(i = 1, 2)$.Para obtener su valor se parte de los datos de supervivencia de cada grupo y con ellos se obtienen una serie de términos que componen la siguiente tabla resumen y que con ellos se obtiene el valor numérico del $Log - Rank$:

Tratamiento con Prednisolona: 2, 2, 12, 54, 56+, 68, 89, 96, 96, 125+, 128+, 131+, 140+, 141+, 143, 145+, 146, 148+, 162+, 168, 173+ 181+.

Control: 2, 3, 4, 7, 10, 22, 28, 29, 32, 37, 40, 41, 54, 61, 63, 71, 127+, 140+, 146+, 158+, 167+, 182+.

| $t_{(j)}$ | $m_{1j}$ | $m_{2j}$ | $n_{1j}$ | $n_{2j}$ | $e_{1j}$ | $e_{2j}$ | $m_{1j}-e_{1j}$ | $m_{2j}-e_{2j}$ | $Var(O_j-E_j)$ |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 2 | 1 | 22 | 22 | 1.5000 | 1.5000 | 0.5000 | -0.5000 | 0.7151 |
| 3 | 0 | 1 | 20 | 21 | 0.4878 | 0.5122 | -0.4878 | 0.4878 | 0.2499 |
| 4 | 0 | 1 | 20 | 20 | 0.5000 | 0.5000 | -0.5000 | 0.5000 | 0.2500 |
| 7 | 0 | 1 | 20 | 19 | 0.5128 | 0.4872 | -0.5128 | 0.5128 | 0.2498 |
| 10 | 0 | 1 | 20 | 18 | 0.5263 | 0.4737 | -0.5263 | 0.5263 | 0.2493 |
| 12 | 1 | 0 | 20 | 17 | 0.5405 | 0.4595 | 0.4595 | -0.4595 | 0.2484 |
| 22 | 0 | 1 | 19 | 17 | 0.5278 | 0.4722 | -0.5278 | 0.5278 | 0.2492 |
| 28 | 0 | 1 | 19 | 16 | 0.5429 | 0.4571 | -0.5429 | 0.5429 | 0.2482 |
| 29 | 0 | 1 | 19 | 15 | 0.5588 | 0.4412 | -0.5588 | 0.5588 | 0.2465 |
| 32 | 0 | 1 | 19 | 14 | 0.5758 | 0.4242 | -0.5758 | 0.5758 | 0.2443 |
| 37 | 0 | 1 | 19 | 13 | 0.5938 | 0.4062 | -0.5938 | 0.5938 | 0.2412 |
| 40 | 0 | 1 | 19 | 12 | 0.6129 | 0.3871 | -0.6129 | 0.6129 | 0.2373 |
| 41 | 0 | 1 | 19 | 11 | 0.6333 | 0.3667 | -0.6333 | 0.6333 | 0.2322 |
| 54 | 1 | 1 | 19 | 10 | 1.3103 | 0.6897 | -0.3103 | 0.3103 | 0.4357 |
| 61 | 0 | 1 | 18 | 9 | 0.6667 | 0.3333 | -0.6667 | 0.6667 | 0.2222 |
| 63 | 0 | 1 | 18 | 8 | 0.6923 | 0.3077 | -0.6923 | 0.6923 | 0.2130 |
| 68 | 1 | 0 | 18 | 7 | 0.7200 | 0.2800 | 0.2800 | -0.2800 | 0.2016 |
| 71 | 0 | 1 | 17 | 7 | 0.7083 | 0.2917 | -0.7083 | 0.7083 | 0.2066 |
| 89 | 1 | 0 | 17 | 6 | 0.7391 | 0.2609 | 0.2609 | -0.2609 | 0.1928 |
| 96 | 2 | 0 | 16 | 6 | 1.4545 | 0.5455 | 0.5455 | -0.5455 | 0.3778 |
| 143 | 1 | 0 | 14 | 6 | 0.7000 | 0.3000 | 0.3000 | -0.3000 | 0.2100 |
| 146 | 1 | 0 | 13 | 6 | 0.6842 | 0.3158 | 0.3158 | -0.3158 | 0.2161 |
| 168 | 1 | 0 | 12 | 6 | 0.6667 | 0.3333 | 0.3333 | -0.3333 | 0.2222 |
| **Total** | $\mathbf{O_1}$ | $\mathbf{O_2}$ | | | $\mathbf{E_1}$ | $\mathbf{E_2}$ | $\mathbf{O_1-E_1}$ | $\mathbf{O_2-E_2}$ | **Var(O-E)** |
| | 11 | 16 | | | 16.4548 | 10.5452 | -5.4548 | 5.4548 | 6.1594 |

- $m_{ij}$ : Número de fallos observados en en grupo $i$ en el tiempo $t_{(j)}$; $i = 1,2 \ \wedge \ j = 1,2,...,23$.
- $n_{ij} = n_{i(j-1)} - m_{i(j-1)}$, total de sujetos vivos y a riesgo al tiempo $t_j$ en el $i-ésimo$ grupo; $i = 1,2 \ \wedge \ j = 1,2,...,23$.
- $e_{ij} = \left(\frac{n_{ij}}{n_{1j}+n_{2j}}\right)(m_{1j}+m_{2j})$, conteo esperado en cada categoría; $i = 1,2 \ \wedge \ j = 1,2,...,23$.
- $E_i = \sum_{j=1}^{\#fallas} e_{ij}$, valor esperado en el $i-ésimo$ grupo; $i = 1,2$.
- $O_i - E_i = \sum_{j=1}^{\#fallas}(m_{ij} - e_{ij})$.

En esta tabla, a partir de la suma total de la penúltima columna $(m_{2j}-e_{2j})$ se obtiene el valor correspondiente a $(O_i-E_i)$ y junto a la suma total de la última columna se puede encontrar el valor del estadístico de prueba. Así,

$$Log-Rank = \frac{(O_2-E_2)^2}{var(O_2-E_2)} = \frac{(16-10.5452)^2}{6.1594} = \frac{(5.4548)^2}{6.1594} \approx 4.8308$$

De esta forma, el paso a seguir es encontrar el valor p asociado a esta prueba de la siguiente forma,

$$P(\chi_1^2 > 4.8308) \approx 0.028$$

Comparando este resultado con un nivel de significancia $\alpha = 0.05$ se concluye que existe evidencia estadística bajo una confianza del 95% para sugerir que las curvas de supervivencia asociadas al grupo de personas bajo

Tratamiento con Prednisolona (Grupo 1) y al grupo de personas en Control (Grupo 2) son distintas, pues el valor p es menor al nivel de significancia $\alpha$. De esta forma podemos ver, bajo análisis estadístico, que sí existe efectividad del tratamiento a la hora de prolongar la supervivencia de los individuos del grupo 1, dado lo mostrado en el gráfico comparativo de las curvas de supervivencia en los puntos anteriores.

**7) Haga los mismo del numeral anterior pero usando SAS o R.**

Usando R, con ayuda de la librería **Survival**, se aplica este test del siguiente modo,

```
CAH <- read.table(file= "CAH.txt",header=T) # Lectura de datos de supervivencia
CAH$group <- as.factor(CAH$group)
df=data.frame(CAH)

library(survival) # Librería necesaria para aplicar Log-Rank Test
log_rank <- survdiff(Surv(time, status) ~ group, data = df, rho = 0) # Log-Rank test
log_rank # Resultados
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ group, data = df, rho = 0)
##
##          N Observed Expected (O-E)^2/E (O-E)^2/V
## group=1 22       11     16.3      1.74      4.59
## group=2 22       16     10.7      2.67      4.59
##
##  Chisq= 4.6  on 1 degrees of freedom, p= 0.03
```

```
round(log_rank$chisq,4) # Valor del Estadístico de prueba Log-Rank
```

```
## [1] 4.5915
```

```
round(pchisq(log_rank$chisq,1,lower.tail = F),4) # Valor p asociado a este test
```

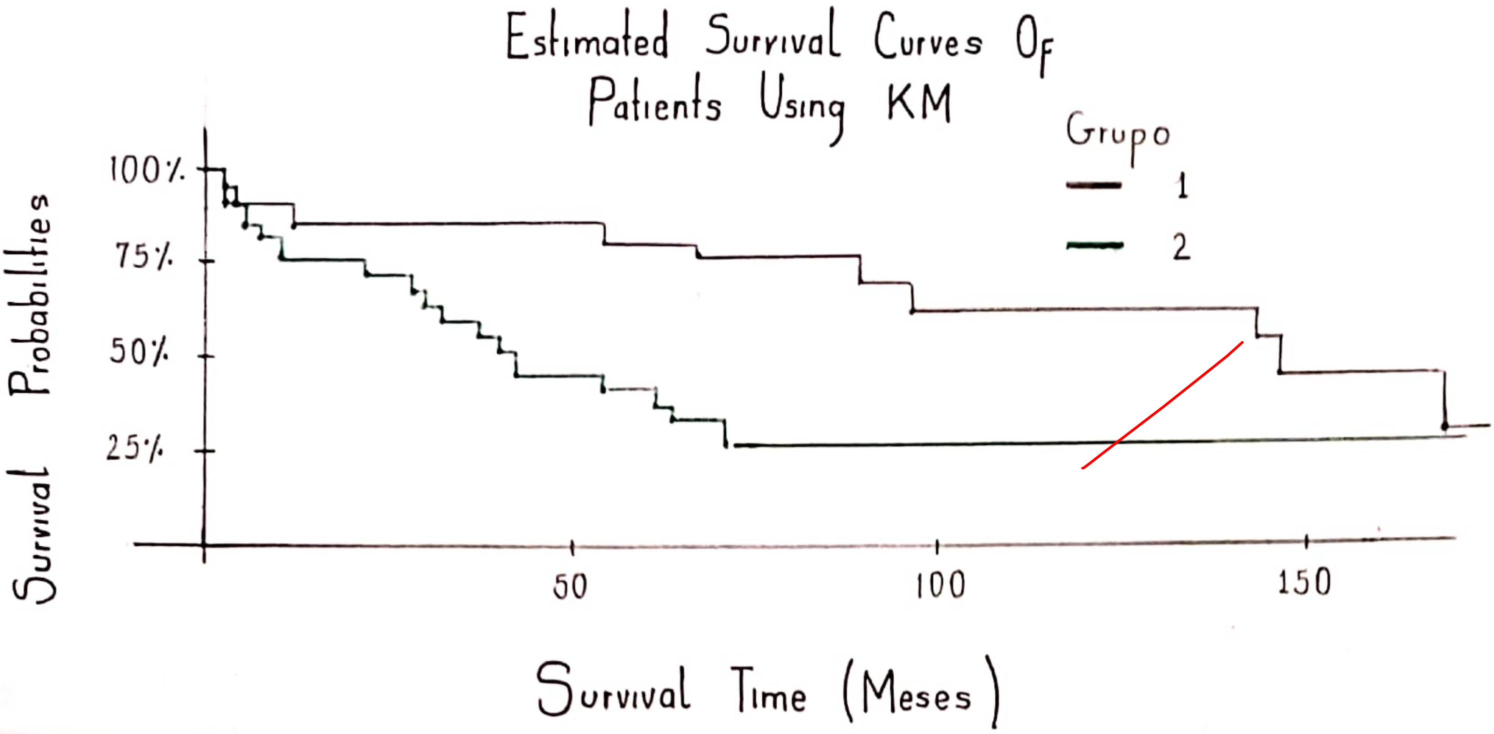
```
## [1] 0.0321
```

El valor del estadístico de prueba obtenido en R difiere con el obtenido "a mano" en el punto anterior; como consecuencia de esto, también el p valor obtenido respecto a cada cuantil es algo distinto. Pero si se redondean todos estos valores a 2 cifras decimales vemos como el valor p coincide en ambas maneras de calculo:

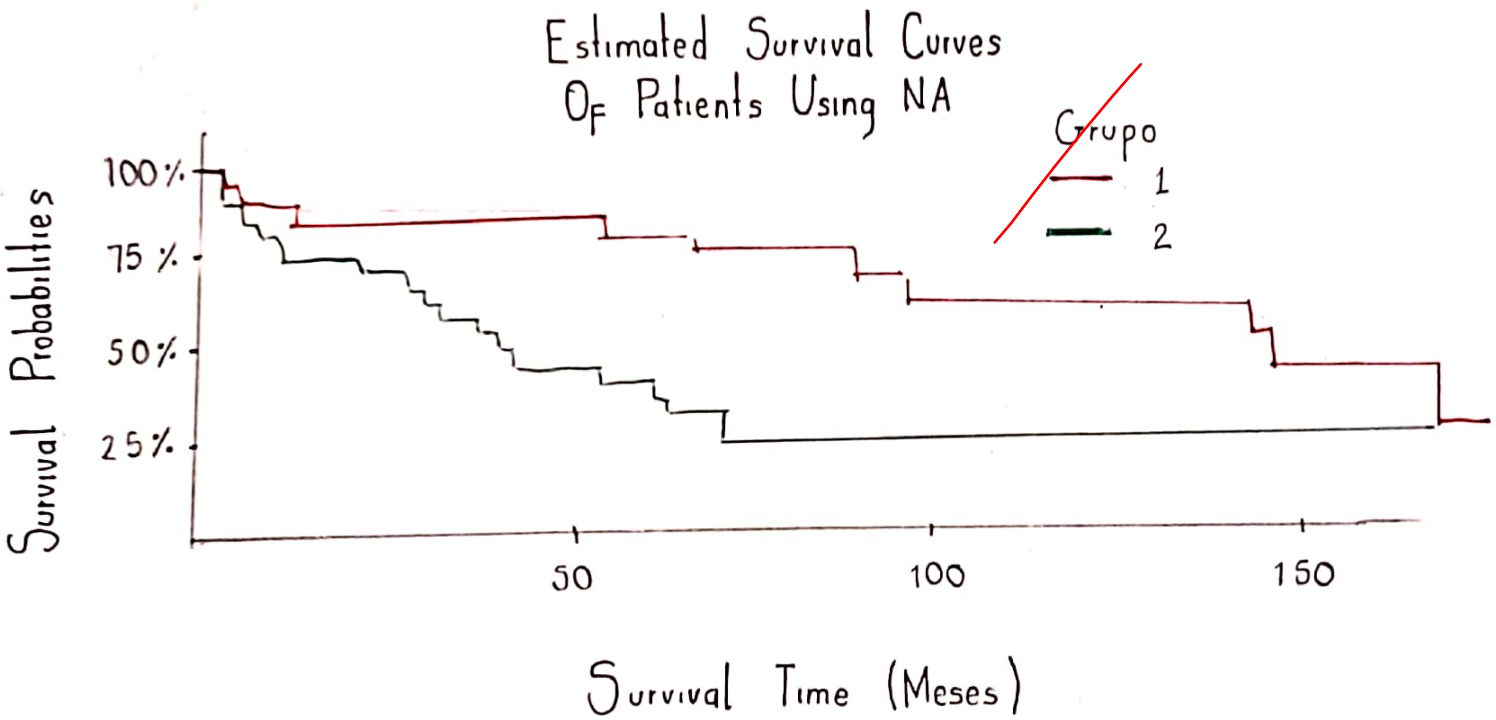
| | A mano | R |
|---|---|---|
| Cuantil | 4.83 | 4.59 |
| Valor P | 0.03 | 0.03 |

Por consiguiente, por ambos caminos se llegan a resultados similares. Finalmente, de este ejercicio comparativo queda es reafirmar el análisis hecho y lo concluido en el punto anterior sobre las curvas de supervivencia y decir que sí existe diferencia en las experiencias de supervivencias de los grupos comparados, el primero bajo Tratamiento con Prednisolona y, el segundo, personas en Control.

## Anexos

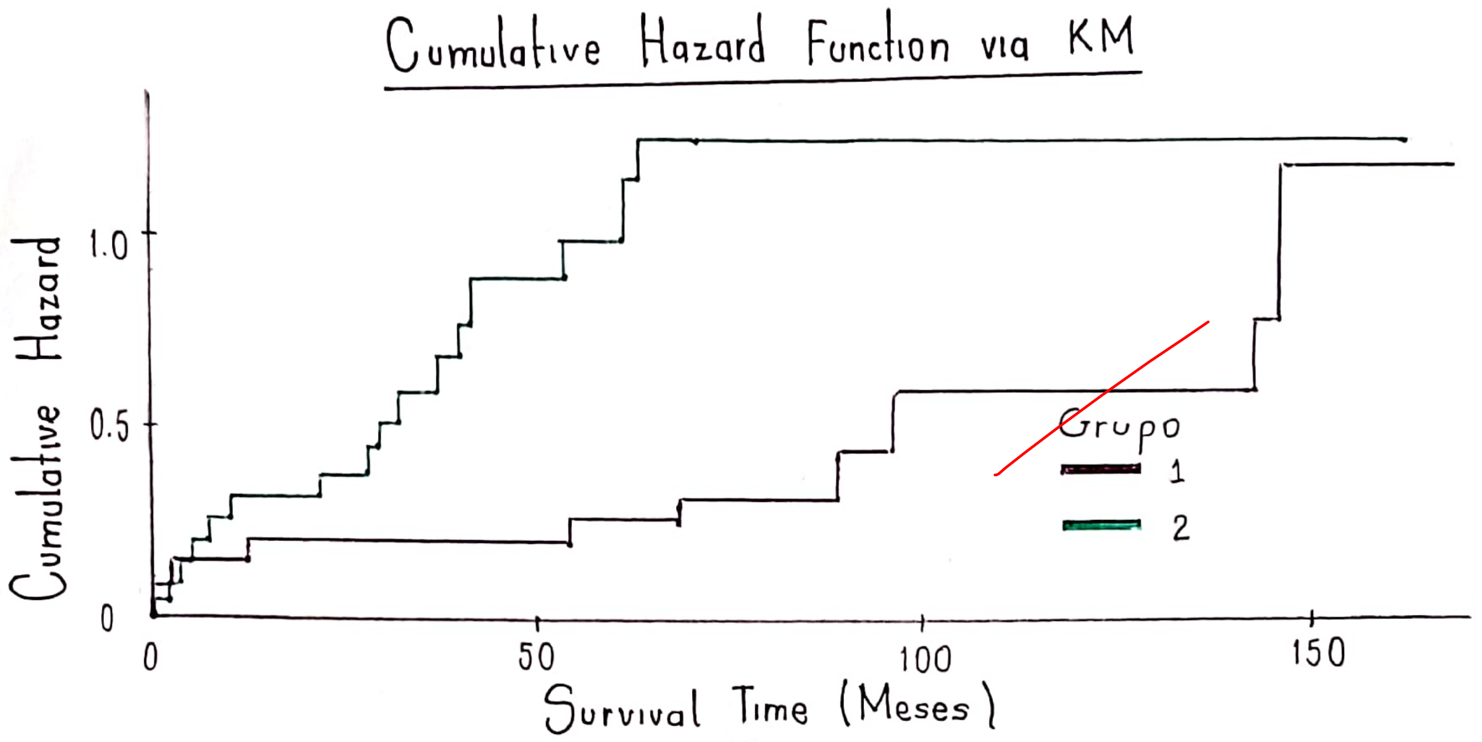


Estimated Survival Curves Of Patients Using KM
Grupo
1
2
100%
75%
50%
25%
Survival Probabilities
50
100
150
Survival Time (Meses)

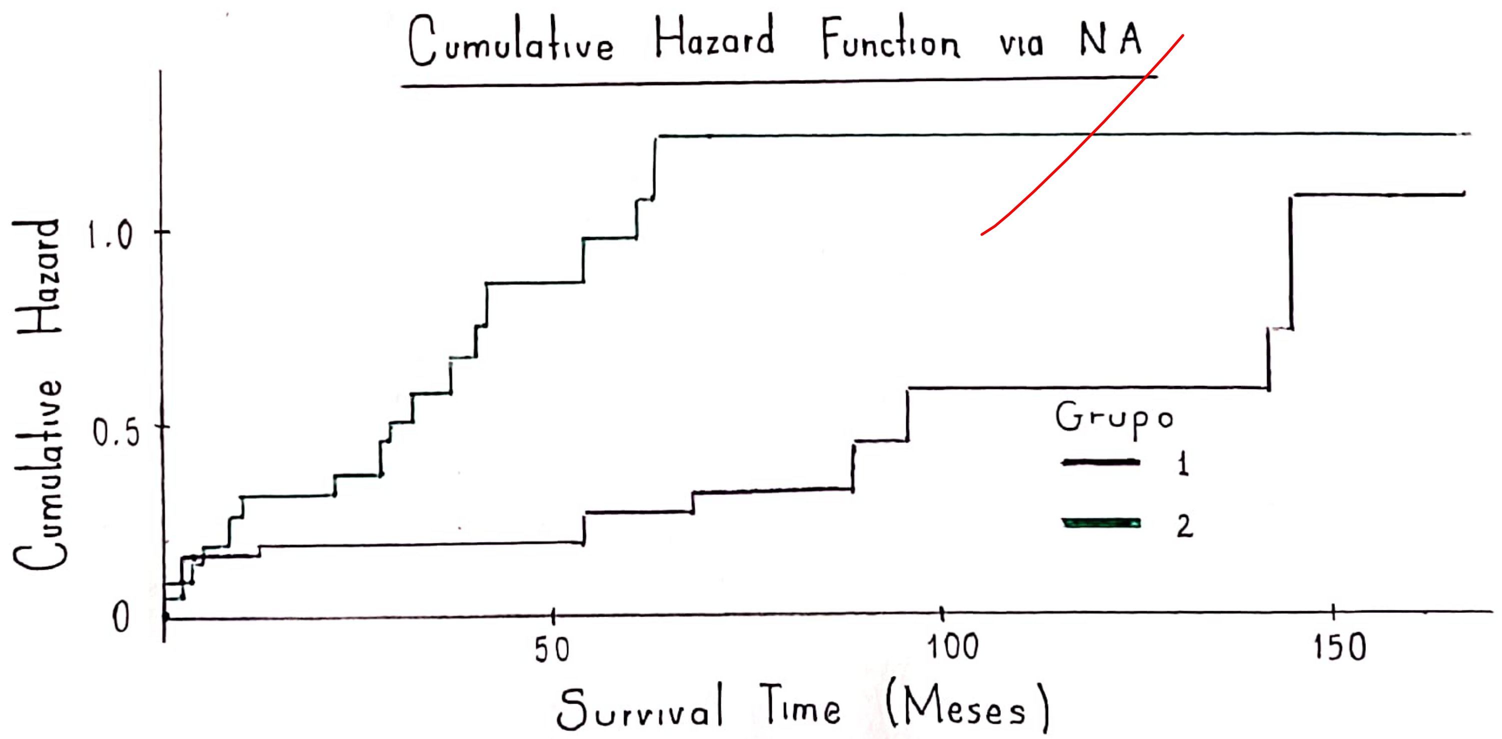



Estimated Survival Curves Of Patients Using NA
Grupo
1
2
100%
75%
50%
25%
Survival Probabilities
50
100
150
Survival Time (Meses)

Anexos



Cumulative Hazard Function via KM
Cumulative Hazard
1.0
0.5
0
0
50
100
150
Survival Time (Meses)
Grupo
1
2

Cumulative Hazard Function via NA
Cumulative Hazard
1.0
0.5
0
50
100
150
Survival Time (Meses)
Grupo
1
2