

# Supervivencia

Using python.

Iván Andrés Trujillo Abella

Facultad de Ingeniería  
Pontificia Universidad Javeriana

[trujilloiv@javeriana.edu.co](mailto:trujilloiv@javeriana.edu.co)



Prediction or explication? what matter?



Weightman, W. M., Gibbs, N. M., Pavey, W. A., LARBalestier, R. I., Newman, M. A., Sheminant, M., Matzelle, S. (2021). The Influence of Choice of Surgical Procedure on Long-Term Survival After Cardiac Surgery. Heart, Lung and Circulation.

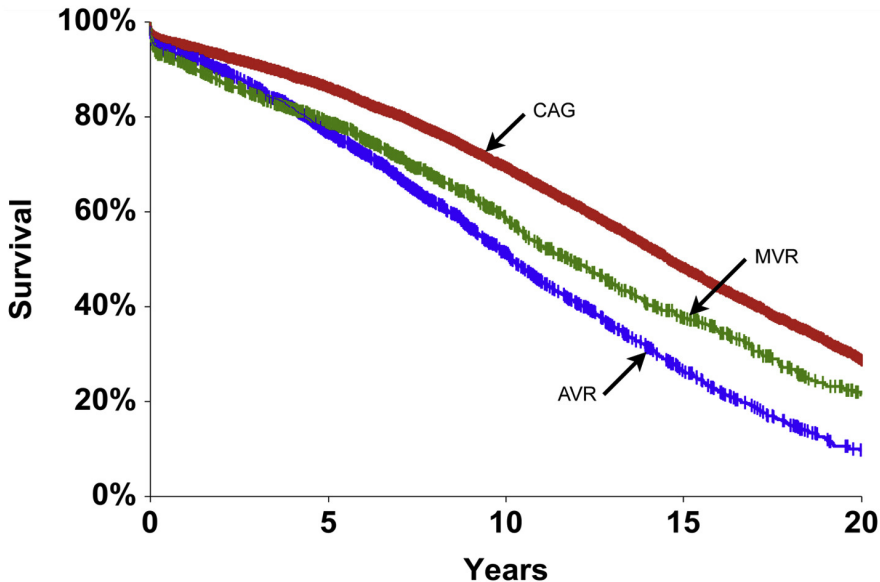
This article was published in September i a British journal Q2, with 46 index.



# The Influence of Choice of Surgical Procedure on Long-Term Survival After Cardiac Surgery (2021)

*"There is some interest in long-term survival after various cardiac surgical strategies, including off-pump versus on-pump coronary artery surgery (CAG), mitral valve (MV) repair versus replacement, and aortic valve (AV) bioprosthetic versus mechanical replacement."*





Number at risk

CAG	8456	5930	3954	2123	701
MVR	1949	1223	596	203	40
AVR	1949	1223	596	203	40

# Que es una función?

En términos prácticos es una regla que asocia un valor de entrada con un valor de salida, por ejemplo los valores:

$$f(x) = 2x + 1 \quad (1)$$



# Que es una variable aleatoria?

Determinadas sus condiciones su resultado no está predeterminado, expectativa de vida, el resultado al lazar una moneda. **Distribución muestral de la media e inferencia, valor P.**



# introducción

En el Análisis de supervivencia se pretende establecer una función de supervivencia:

$$S(t) = \text{prob}(T > t)$$

una formula que nos indique con que probabilidad ocurrirá un evento en un periodo determinado  $t$ . Aunque históricamente se asocia con la determinación del evento de defunción no es necesario que este sea el evento de interés, se extiende a todo evento que se presente en un lapso determinado de tiempo.





# Estudios de supervivencia

- Tiempo transcurrido desde el diagnóstico de cáncer de estómago hasta el fallecimiento del individuo.
- Tiempo de aparición de un tumor después de su resección.
- Tiempo transcurrido desde el inicio de tratamiento hasta la recuperación del movimiento de individuos afectados con Guillain-Barré.
- el tiempo transcurrido hasta que un prematuro obtenga su peso ideal.



# variables de estudio

particularmente bajo ciertos supuestos (detallaremos más adelante) y en un análisis básico solo necesitamos la variable de ocurrencia del evento ( si se presenta o no ) y el tiempo de observación, este ultimo depende del inicio de intervención o tratamiento hasta el cierre del control.



# Ejemplo de una base de datos pero y el ¿Estado 2?

caso	estado	tiempo en meses
1	1	24
2	1	14
3	0	60
4	1	17
5	2	38
6	1	20
7	0	48
8	1	9
9	1	1
10	0	10

Table: Fuente: Rebase (2005)



Aunque para este tipo de estimación básica solamente se puede presentar el evento una vez, es decir si un individuo presenta el evento, termina su seguimiento; no puede volver a presentar el evento. Aunque todas las observaciones se tomaron de esta forma existe un concepto que hay que tener en cuenta el de **censura**.



# La censura

**la censura** se presenta por varias razones; entre ellas está por que al terminar el periodo de estudio el individuo no presentó el evento de interés o no se pudo establecer cual fue su resultado por otras razones que desconocemos, por lo tanto son observaciones censuradas:

- al finar los 3 años de seguimiento por diagnostico de cáncer de estomago el individuo sigue vivo.
- se censuró porque presentó un evento que limitó su seguimiento.
- se trato de cáncer de mama pero murió de un accidente de tránsito.
- el individuo cambió su lugar de residencia y perdimos su control.

el tipo de censura que nos interesa se denomina **censura por la derecha** lo que implica que hay otros tipos de censura.



# La censura aporta información

Aunque no conozcamos el resultado final del individuo ( o el evento de interés no se presentó ) nos aporta información debido a que contribuye a la población susceptible en el periodo antes de su censura.



# Censura en la base de Rebasa(2005)

Rebasa(2005) codificó la variable *estado* de modo que 0 fallecido por cáncer (objeto de estudio) ; 1 vivo ; 2 perdido.

caso	estado	tiempo en meses
1	1	24
2	1	14
3	0	60
4	1	17
5	2	38
6	1	20
7	0	48
8	1	9
9	1	1
10	0	10

Table: Fuente: Rebasa (2005)



# ¿Como analizamos esta información?

Se debe tener en cuenta que algunas variables que implican tiempo, tienen una distribución particular; Pensemos en el tiempo de estancia hospitalaria ( Usualmente sesgada a la derecha) ¿por qué?

ejemplo: Tiempo para que un individuo infecto por VIH desarrolle el Síndrome de Inmunodeficiencia Adquirida (SIDA)

La censura limita el conocimiento preciso de cuando se presenta el evento de interés.





# La función de supervivencia $S(t)$

debe tener unas particularidades intuitivas:

- que al empezar el estudio la supervivencia global sea del 100%
- si seguimos al individuo por "la infinidad" el evento se presentará.



# Kaplan - Meier

Una forma tradicional de estimar  $S(t)$  es por medio del estimador no paramétrico **Kaplan-Meier**.



# Probabilidad

## Bag model

Supongamos que en el momento del tiempo  $t_0$  o en la bolsa  $t_0$  hay  $n$  canicas (individuos), de los cuales  $k$  son blancas(vivos) y  $j$  son rojas (muertes) y  $k + j = n$ . La probabilidad de obtener en ese momento una bola roja (una muerte) no es mas que;

$$P(roja) = \frac{j}{k+j} = \frac{j}{n}$$



para el caso contrario la probabilidad de no obtener ninguna bola roja(defunción) es;

$$1 - P(roja) = 1 - \frac{j}{n}$$

Ahora supongamos que para el periodo siguiente esas  $j$  bolas ya no están debido a que se han perdido o no se regresan a la bolsa y que  $i$  bolas de las  $k$  blancas se vuelven rojas con el lapso de tiempo por lo que hay  $k - i$  blancas:

$$P(roja) = \frac{i}{k}$$



# kaplan Meier

En el análisis de los fenómenos que hemos analizado en cada momento del tiempo existen diferentes individuos en riesgo, así para el periodo  $t_i$  hay  $n_i$  individuos en riesgo de presentar el evento y en dicha muestra se presentan  $d_i$  eventos por lo cual para ese periodo la probabilidad de ocurrencia es  $\frac{d_i}{n_i}$ , por el contrario de no ocurrencia es  $1 - \frac{d_i}{n_i}$ . así el estimador de **K.M.** se define como

$$\hat{S}(t_i) = \prod_{t \leq t_i} \left(1 - \frac{d_i}{n_i}\right)$$



id	meses	evento
1	3	1
2	4	0
3	5	1
4	4	1
5	3	0
6	5	1
7	6	0
8	7	1

Table: Fuente: Elaboración propia

tiempo	Sujetos riesgo	excluidos	$P(a)$	$1-P(a)$	K.M(t)
3	8	2	$1/8$	$7/8$	$(7/8)$
4	6	2	$1/6$	$5/6$	$(7/8) (5/6)$
5	4	2	$1/2$	1	$(7/8) (5/6) (2/2)$
6	2	1	0	1	1
7	1	1	1	0	0

Table: Estimador de kaplan Meier

# La base organizada de manera adecuada

caso	estado	tiempo en meses
1	0	24
2	0	14
3	1	60
4	0	17
5	0	38
6	0	20
7	1	48
8	0	9
9	0	1
10	1	10

Table: Fuente: Rebase (2005)





# ¿ el artículo está mal?

TABLA 2. Cálculo de la supervivencia

Tiempo	Estatus	Supervivencia acumulada	¿Cuántos fallecen en el intervalo?	¿Cuántos estaban vivos al inicio del intervalo?	Cómo se calcula
1	t(1)	1	0,9000	1	9
9	t(9)	1	0,8000	1	8
10		0			
14	t(14)	1	0,6857	1	7
17	t(17)	1	0,5714	1	6
20	t(20)	1	0,4571	1	5
24	t(24)	1	0,3429	1	4
38		3			
48		0			
60		0			

Figure: Fuente: extraído de Rebasa(2005)



Utilicemos el paquete estadístico para calcular la función de supervivencia por medio del estimador **K.M.**

# Estimación en Stata

Para llevar a cabo la estimación de  $S(t)$  por el estimador no paramétrico **K.M.** debemos organizar la variable de evento de tal forma que los datos **censurados** se codifiquen con el valor cero(**0**) y los no censurados con la unidad(**1**)



Para poder llevar el análisis, previamente a la obtención de  $\hat{S}(t)$  debemos establecer o indicarle al programa cual es la variable de censura-No censura y el tiempo de ocurrencia.

```
stset vartime, failure(evento)
```

posteriormente la función de supervivencia  $\hat{S}(t)$  estimada por **K.M.** se obtiene por el comando:

```
sts list
```



# ¿Que utilidad tiene esto?

La comparación de dos curvas de supervivencia en dos poblaciones homogéneas con una diferencia de exposición nos permite establecer comparaciones entre dos curvas.

- Existe diferencias en la supervivencia entre un grupo A y un grupo B.
- tienen mas supervivencia los individuos diagnosticado cáncer A o B.
- Es mas efectivo un tratamiento A o B para recuperar el movimiento.



# logrank test en stata

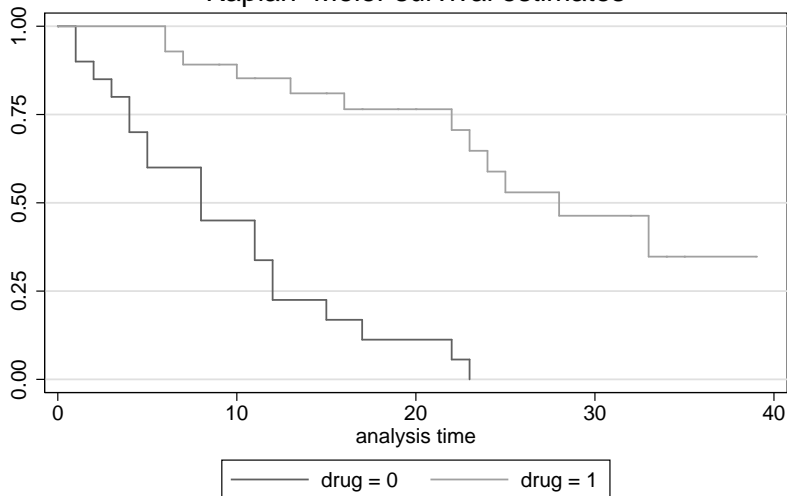
Para probar que las curvas son iguales o que no existen diferencias significativas entre ellas partimos de la hipótesis nula  $H_0$  de que ambas curvas son similares y no que tienen diferencias significativas entre ellas y utilizamos el comando:

```
sts test group, logrank
```



# Presentación - artículos

Kaplan–Meier survival estimates



Fuente: Base drugtr Stata