

Análisis de Sobrevida

Análisis de Sobrevida

Qué es?

➤ Un conjunto de metodologías estadísticas utilizadas para estimar la probabilidad de ocurrencia de un determinado evento en diferentes momentos a lo largo de un periodo bajo estudio; utilizando para ello distintas técnicas que pueden ser de tipo no paramétrico, semiparamétrico o paramétrico.

Otros nombres:

- Survival Analysis (Medicina) Análisis de Sobrevida
- Failure-Time Analysis (procesos industriales)
- Analysis of Time to an Event
- Analyses of Time at Risk
- Analyses of Time to an End-Point

Aplicaciones

General: Estudios de seguimiento (observacionales o experimentales)

➤ **Medicina Humana (Veterinaria)**

Evaluación de sobrevida después de una intervención.

Evaluación de tiempo a presentación de una afección

Comparación de curvas de vida (expectativas de vida)

(En general para los mismos casos en que aplicaría una regresión logística, con la diferencia de que en AS los periodos bajo riesgo pueden ser de distinta longitud para cada individuo)

➤ **Producción Animal:**

Evaluación de longevidad

Análisis de intervalos reproductivos (parto-concepción)

Presentación de Mastitis a través de la lactancia

Comparación de curvas de vida (expectativas de vida)

➤ **Sociología** (Durabilidad de matrimonios, Deserción escolar)

➤ **Economía** (Durabilidad de empresas)

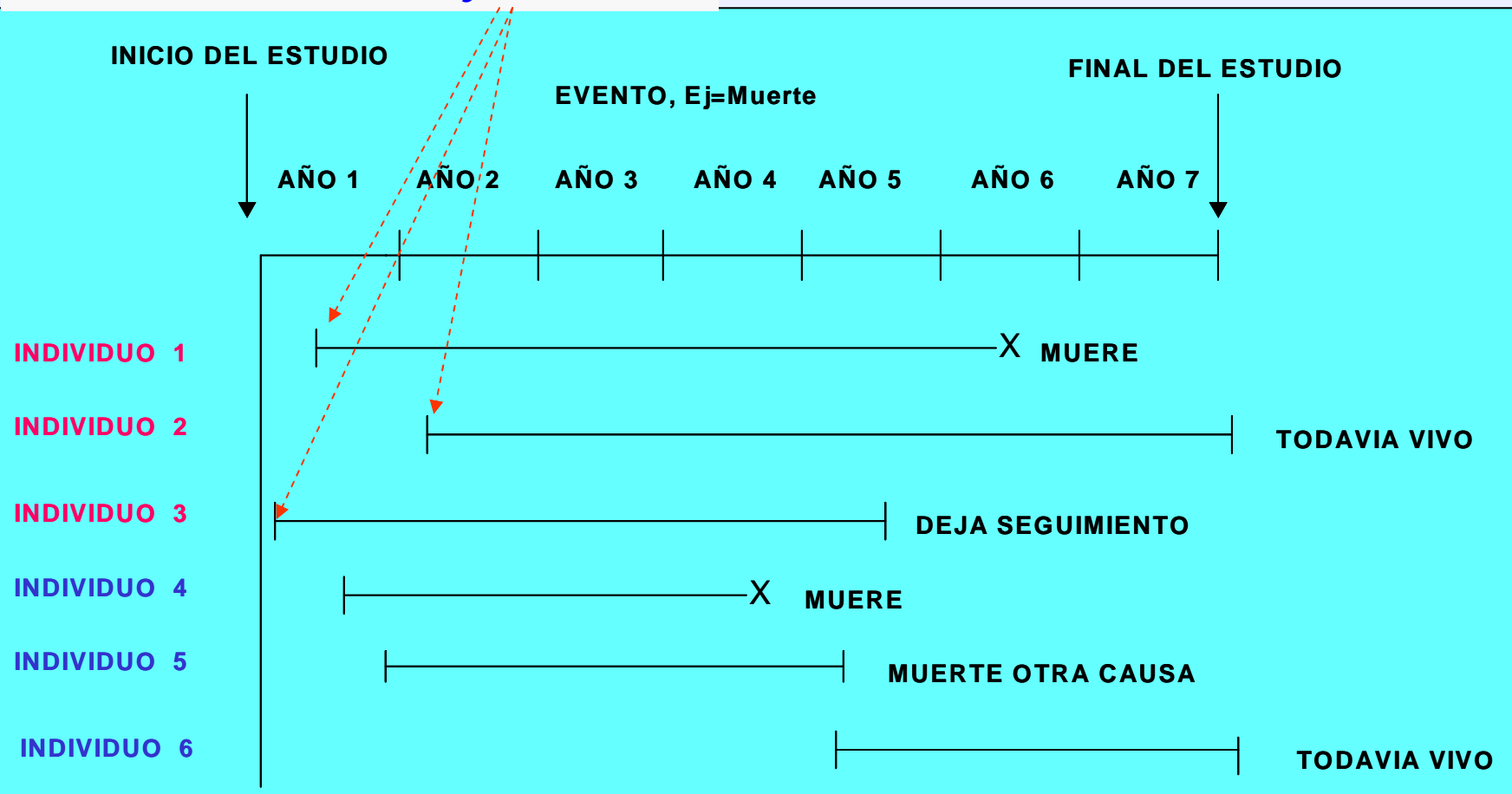
➤ **Ingeniería** (Análisis de vida útil)

Características comunes a cualquier metodología de Anál. de sobrevida

- Un evento final de interés de carácter dicotómico (ej muerte)
- Un periodo de análisis (ej 7 años)
- Un punto de referencia inicial (ej intervención)
- Un grupo de individuos que presentan el mismo estatus al momento de entrar al estudio (ej. todos enfermos)
- Entrada y Salida de Individuos en el estudio puede ser en distintos momentos a lo largo del periodo. Una vez que entran al estudio los individuos se encuentran bajo "riesgo" de sufrir el evento
- Algunos individuos presentan el evento durante el periodo de estudio (No censurados) mientras que otros no lo presentan (censurados)

Análisis de sobrevida/ Representación gráfica

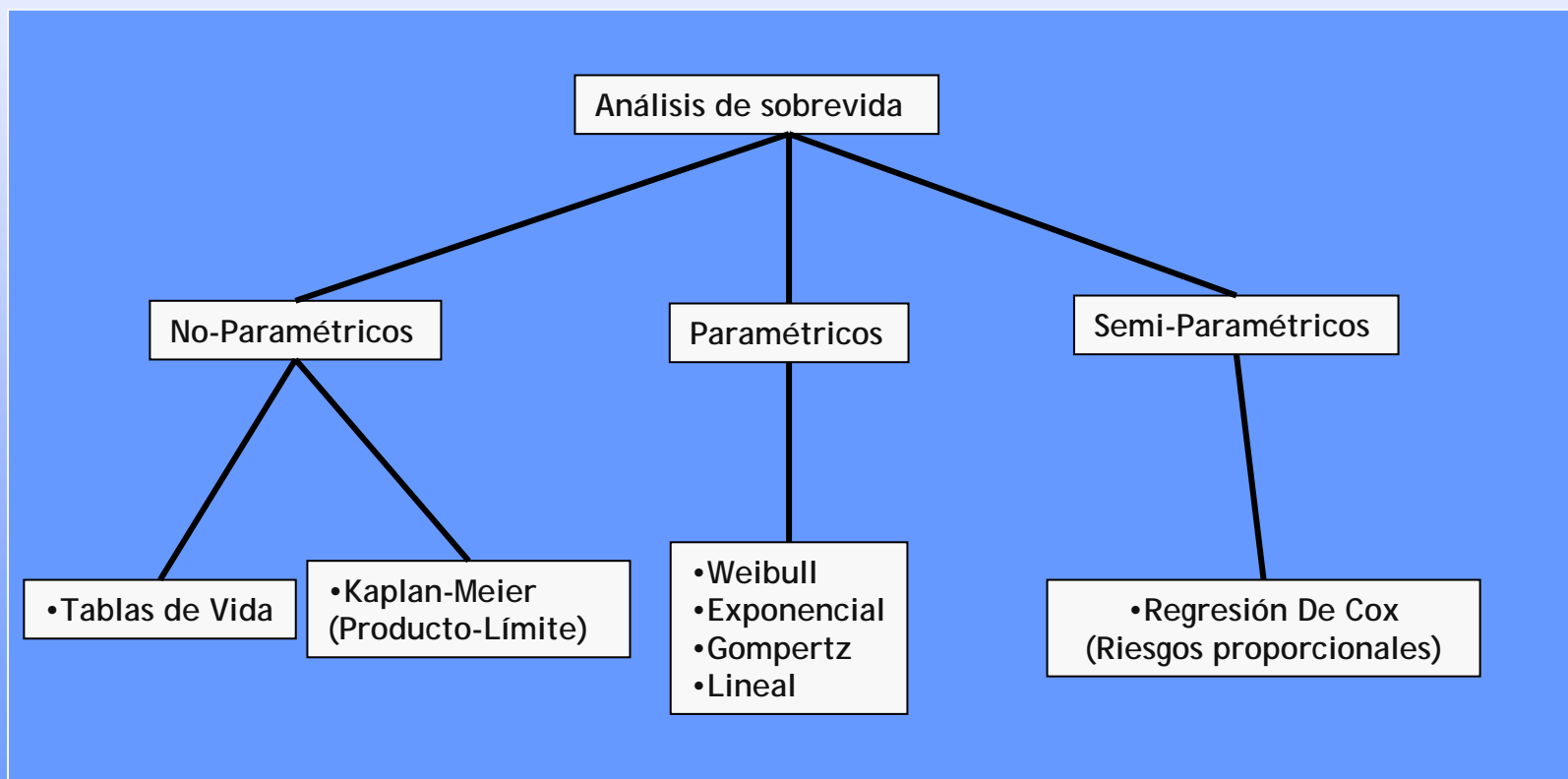
Punto de referencia inicial, ej. Intervención



Terminología utilizada en Análisis de sobrevida

- **Tiempo de Falla (Failure-Time):** Intervalo (horas, días, semanas, etc) transcurrido desde el **punto de referencia inicial** hasta el momento de ocurrencia del **evento** analizado.
- **Tiempo de Censura:** Intervalo transcurrido entre el punto de referencia inicial y el **momento de censura**
- **Aclaración:**
El término de Tiempo de Falla (Failure-Time) se utiliza debido al empleo original del análisis (análisis de mortalidad). Sin embargo, el evento representado por la falla puede ser también de carácter positivo (ej. recuperación)

Análisis de sobrevida /Métodos utilizados



Cuando usar estos métodos?

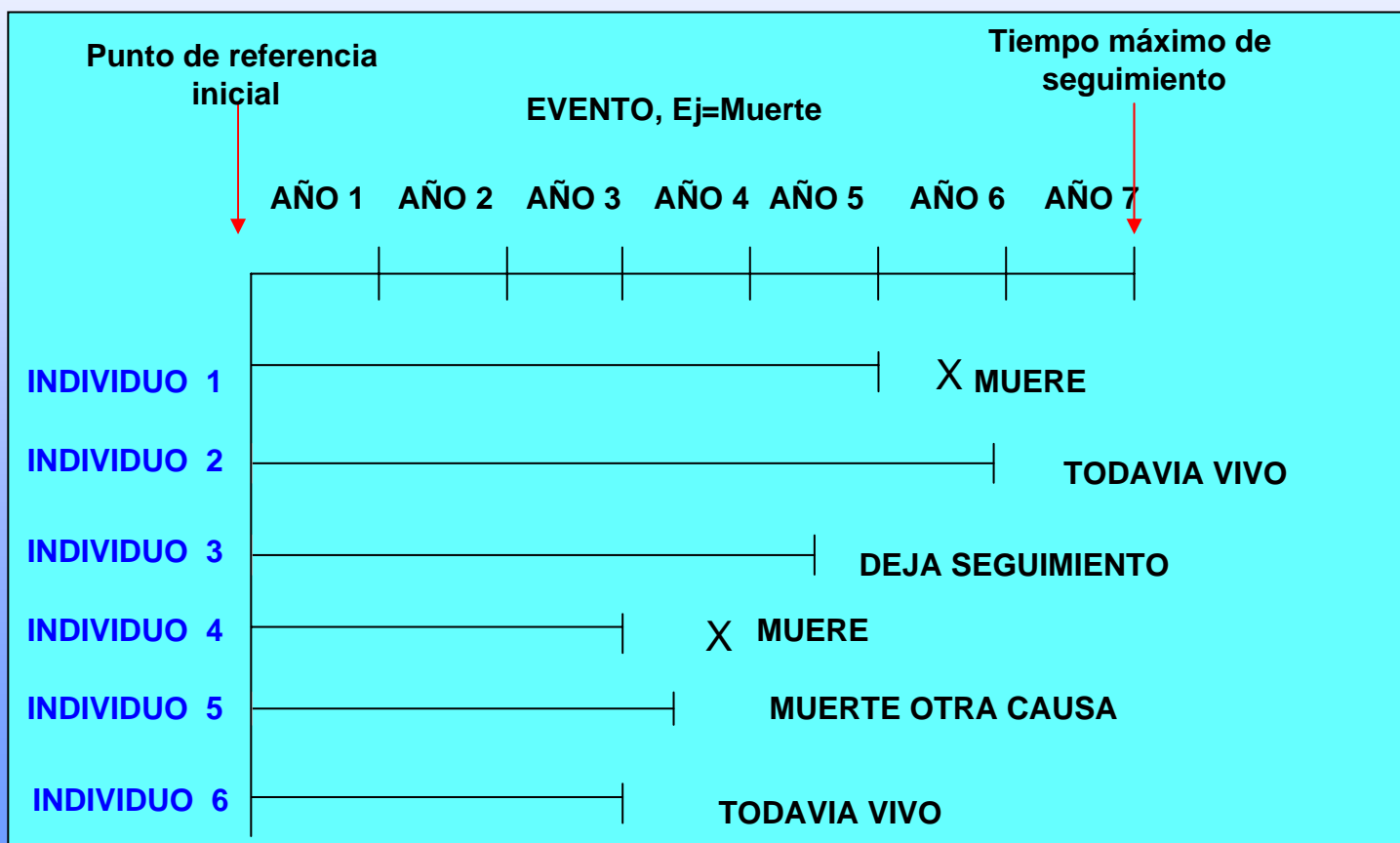
- Los métodos **No-Paramétricos** (Tablas de vida, Kaplan Meier) se utilizan generalmente en estudios de carácter **exploratorio o descriptivo**. La ventaja es que no requieren previo conocimiento de la *distribución subyacente* de la variable por analizar.
- Los métodos **Paramétricos** (regresión Weibull, Exponencial, Gompertz, LogNormal) se utilizan cuando se desea realizar estudios de carácter **predictivo** y cuando se tiene certeza sobre la verdadera *distribución subyacente* de la variable por analizar
- Los métodos Semi Paramétricos (Regresión de **Cox**) se utiliza cuando no se tiene certeza sobre la *distribución subyacente* de la variable por analizar y cuando sea razonable asumir que los **riesgos** de grupos expuestos a distintos niveles de exposición son **proporcionales** a través del tiempo.

Los métodos paramétricos aportan más información pero requieren que la distribución asumida sea en realidad la que corresponde a la variable por analizar

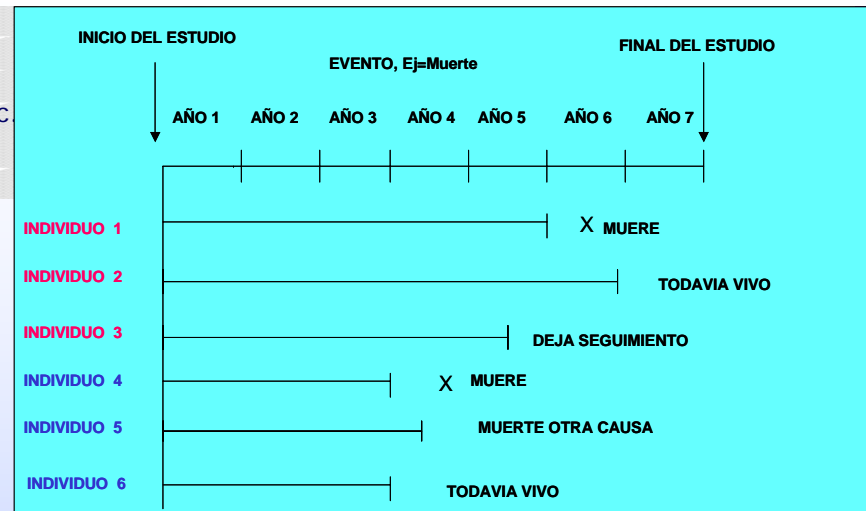
Cronograma

- **Vie. 17-11** Métodos No Paramétricos
Construcción de Tablas de vida
Curvas Kaplan Meier(Producto-Límite)
Comparación no paramétrica de sobrevida en 2 muestras
- **Vie. 24-11** Modelos de sobrevida Paramétricos
Modelos Paramétricos (Weibull, Exponencial, Gompertz)
Comparación de muestras múltiples
- **Vie. 01-12** Modelos de sobrevida Semi-Paramétricos
Regresión de Cox (Riesgos proporcionales)

*Punto de Referencia:
Momento de diagnóstico*



Codificación de datos requerida para un Análisis de sobrevida



<u>Individuo</u>	<u>Censura</u> (0=Censurado, Incompleto 1=No censurado, Completo)	<u>Tiempo a Evento</u> (o censura) (años)
Individuo 1	1	5.0
Individuo 2	0	5.8
Individuo 3	0	4.3
Individuo 4	1	3.0
Individuo 5	0	3.3
Individuo 6	0	3.0

➤ Aclaración:

El código utilizado para falla/censura es indiferente, aunque generalmente se utiliza 1 para No censurado y 0 para censurados)

Diferencia con respecto a otros procedimientos de Regresión:

➤ **Regresión Lineal:**

A diferencia de la Reg. Lineal el Anál. de Sobrevida incorpora información 'incompleta' (censurada). Esta información es valiosa puesto que muchas veces los individuos que no fallan (ej. sobreviven) son también de interés para el investigador.

➤ **Regresión Logística:**

A diferencia de la Reg. Logística en el A. de Sobrev. no solo interesa la ocurrencia del evento en sí, sino también el **momento** en el que se presenta y los patrones de variación a través del tiempo.

Métodos No Paramétricos/ Tablas de Vida

Qué es?

Representación tabular probabilística del patrón de ocurrencia de un determinado evento a través del tiempo

Uso

Principalmente de carácter exploratorio y descriptivo

Construcción

Ver Ejemplo Excel y Statistica

Ejemplo/ Datos de Sobrevida al cáncer de pulmón

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	censoring	dayssurv	perfstatus	months	ageyrs	priortherp	treatment	cell	PriorCat						
2	1	1	50	7	35	0	1	1	0						
3	1	1	20	21	65	10	1	1	1		Fuente:				
4	1	2	40	36	44	10	1	2	1		Se describe la sobrevida de individuos que han sido diagnosticados con distintos tipos de tumor en pulmones				
5	1	3	30	3	43	0	2	3	0						
6	1	4	40	2	35	0	2	2	0						
7	1	7	50	7	70	0	2	2	0		censoring	0=observación censurada, 1=fallo observado (muerte)			
8	1	7									dayssurv	Tiempo transcurrido entre el diagnóstico y la muerte o tiempo hasta el último momento visto con vida			
9	variable de censura			4	58	0	1	3	0						
10	1	8	80	2	68	0	1	2	0		perfstatus	El estado del paciente al inicio de la prueba según el tipo de tumor			
11	1	8	20	19	61	10	2	3	1			10-30 hospitalizado,			
12	1	8	50	5	66	0	1	3	0			40-60 parcialmente confinado			
13	1	8	40	58	63	10	2	1	1			70-90 capaz de atenderse solo.			
14	1	10	40	23	67	10	2	2	1						
15	1	10	20	5	49	0	2	1	0		months	Número de meses transcurridos entre el diagnóstico y la muerte			
16	1	11	70	11	48	10	2	1	1		ageyrs	Edad del paciente en años.			
17	1	12	50	4	63	10	2	3	1		Priortherp?	(0 o 10)			
18	1	12	40	12	68	10	2	0	1		treatment	1=estándar, 2=quimioterapia.			
19	1	13	60	4	56	0	2	2	0		cell	Tipo de tumor, clasificado según tipo de célula:			
20	1	13	30	2	62	0	1	2	0			0=squamous?,			
21	1	15	50	13	40	10	1	1	1			1=small cell?			
22	1	15	30	5	63	0	1	0	0			2=adenoma,			
23	1	16	30	4	53	10	2	2	1			3=big cell?			
24	1	18	30	4	60	0	2	2	0		PriorCat	(0 o 1)			
25	1	18	40	5	69	10	1	3	1						

Ejemplo/ Análisis de Tablas de Vida en Statistica

Variable de tiempo a evento

Variable de censura

datos pueden ser "crudos" (ejemplo) o pre-tabulados (interv, fallas, censuras)

Código para respuestas completas (1)
y respuestas censuradas (0)

Seleccionar:

No. de intervalos (min 3, max 30)

o

Ancho de intervalo

(el ancho escogido no debe producir
menos de 3 o más de 30 intervalos!!)

opción para corregir
intervalos sin eventos
(solo si se desea ajustar
distribuciones)

Statística/ Tablas de Vida

Life Table (9ansobnp) Log-Likelihood for data: -175.350																	
Interval	Interval Start	Mid Point	Interval Width	Number Entering	Number Withdrwn	Number Exposed	Number Dying	Proportn Dead	Proportn Survivng	Cum.Prop Survivng	Problty Density	Hazard Rate	Std.Err. Cum.Surv	Std.Err. Prob.Den	Std.Err. Haz.Rate	Median Life Exp	Std.Err. Life Exp
Intno.1	0.0000	45.4091	90.81818	137	3	135.0000	73	0.538745	0.461255	1.000000	0.005932	0.008119	0.000000	0.000472	0.000883	84.2867	7.2409
Intno.2	90.8182	136.2273	90.81818	61	4	59.0000	30	0.508475	0.491525	0.461255	0.002582	0.007508	0.042825	0.000408	0.001289	89.3045	11.6265
Intno.3	181.6364	227.0455	90.81818	27	2	26.0000	9	0.346154	0.653846	0.226718	0.000864	0.004609	0.036665	0.000272	0.001502	147.8021	36.3203
Intno.4	272.4546	317.8636	90.81818	16	0	16.0000	6	0.375000	0.625000	0.148239	0.000612	0.005082	0.031971	0.000238	0.002019	127.1455	36.3273
Intno.5	363.2727	408.6818	90.81818	10	0	10.0000	5	0.500000	0.500000	0.092649	0.000510	0.007341	0.026855	0.000219	0.003095	90.8182	143.5962
Intno.6	454.0909	499.5000	90.81818	5	0	5.0000	1	0.200000	0.800000	0.046325	0.000102	0.002447	0.019872	0.000101	0.002432	158.9318	50.7689
Intno.7	544.9091	590.3182	90.81818	4	0	4.0000	2	0.500000	0.500000	0.037060	0.000204	0.007341	0.017928	0.000142	0.004894	363.2727	90.8182
Intno.8	635.7273	681.1364	90.81818	2	0	2.0000	0	0.000000	1.000000	0.018530	0.000000	0.000000	0.012891	0.000000	0.000000	317.8636	64.2181
Intno.9	726.5455	771.9545	90.81818	2	0	2.0000	0	0.000000	1.000000	0.018530	0.000000	0.000000	0.012891	0.000000	0.000000	227.0455	64.2181
Intno.10	817.3636	862.7727	90.81818	2	0	2.0000	0	0.000000	1.000000	0.018530	0.000000	0.000000	0.012891	0.000000	0.000000	136.2273	64.2181
Intno.11	908.1818	953.5909	90.81818	2	0	2.0000	1	0.500000	0.500000	0.018530	0.000102	0.007341	0.012891	0.000101	0.006921	45.4091	64.2181
Intno.12	999.0000			1	0	1.0000	1	1.000000	0.000000	0.009265			0.009191				

Desventaja de Tabla de Vida: Los valores de los estimados dependen del número de intervalos que se seleccionan

**ver cálculos e interpretación en Excel*

Representación gráfica / Función de Sobrevida



Función de Sobrevida (Survival Function)

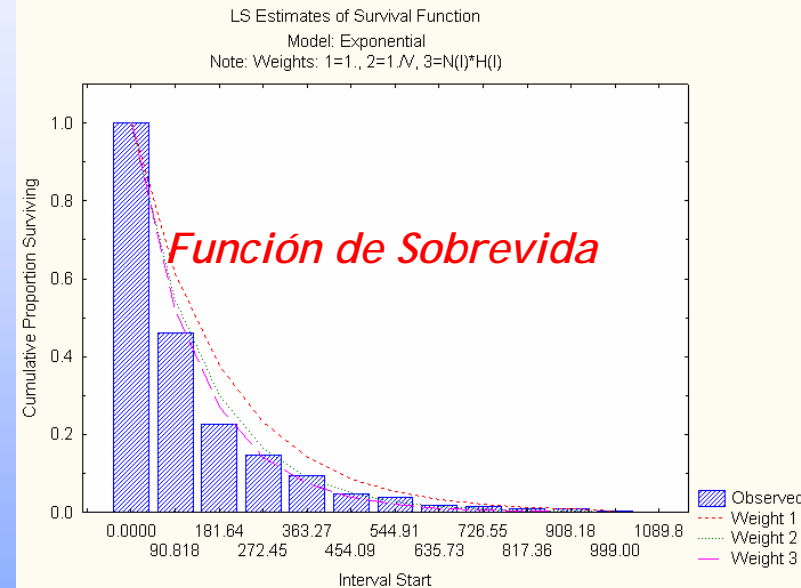
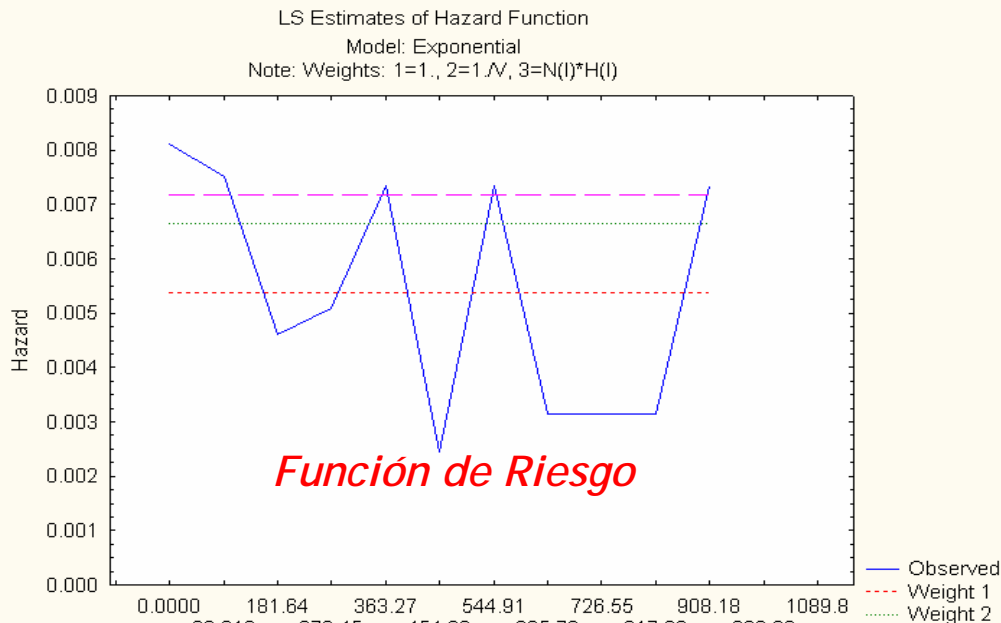
Prob.Acumulada de Sobrevivencia

Es la probabilidad de que un individuo sobreviva al menos **hasta el inicio del intervalo respectivo**

Ej. Aprox. 23% de los individuos sobreviven hasta 181 d.

Generalmente los estimados de sobrevida se comparan en periodos específicos (p.e 1 año, 5 años, etc) dependiendo del evento que se analiza

Representación gráfica/ Función de Riesgo



Función de riesgo (Hazard Function)

Estima el riesgo 'inmediato' de falla

Es una tasa (mortalidad x unidad de tiempo)

Se relaciona con la pendiente de la curva de sobrevivencia

p.e, en momentos más críticos el HR tiende a subir, en momentos menos críticos tiende a bajar

En el ejemplo no hay una tendencia clara (sube y baja)

OJOs!!

Tamaño de muestra para Tablas de Vida:

- Para obtener estimados confiables de las funciones de supervivencia, densidad de probabilidad y de riesgo el tamaño mínimo de muestra (eventos y censuras) recomendado dentro de cada intervalo es 30.

Periodo de estudio:

- Debe seleccionarse un periodo de longitud “razonable” (ni muy corto ni muy largo) según el evento analizado, de manera que se obtenga una proporción razonable de eventos-censuras

Proporción de individuos censurados:

- Debido a que no existe certeza sobre el verdadero tiempo de supervivencia de los individuos censurados, algunos autores recomiendan que el % de censurados no exceda del 40%

Supuestos del análisis de sobrevida No Paramétrico:

- No hay cambio en métodos de diagnóstico, tratamientos o procedimientos de seguimiento durante el estudio
- Los pacientes censurados experimentarían la misma distribución de consecuencias (eventos) que los que permanecen en el estudio
- La censura es independiente de los tratamientos comparados (supuesto de censura NO informativa!)

Estimados Kaplan-Meier (Product-Limit)

Qué es?

Este método estima la función de sobrevivencia directamente para **cada uno** de los distintos tiempos de falla.

Ventaja sobre Tabla de vida:

Los estimados resultantes **no dependen del número de intervalos** que se definan

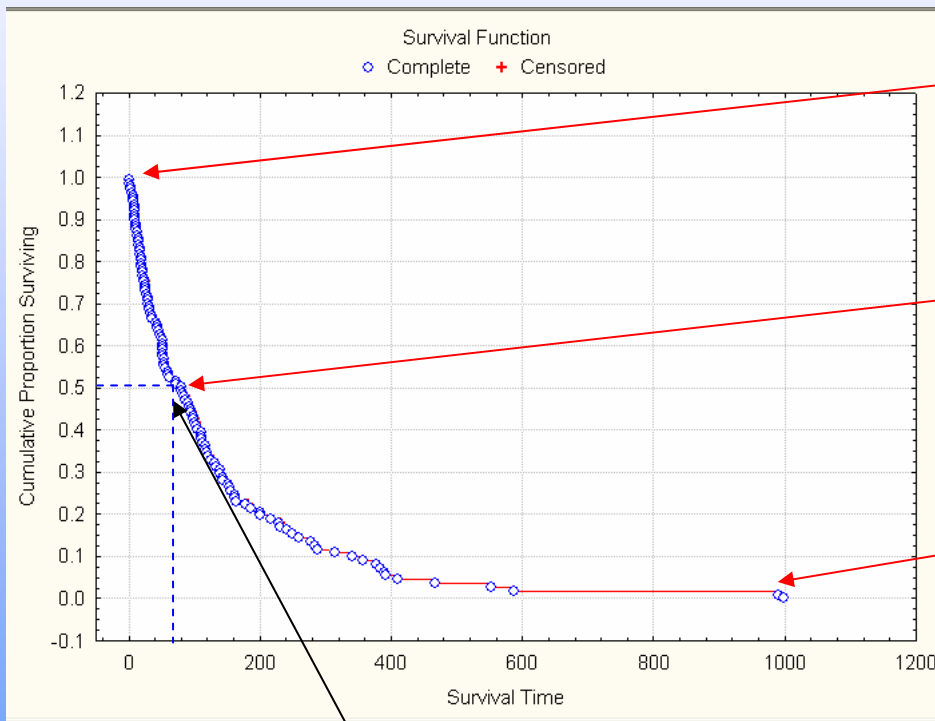
Nota:

En realidad el método de Producto-Límite y el de Tabla de Vida son idénticos si los intervalos de la Tabla se ajustaran para contener exactamente una observación.

Cálculo:

Ver Ejemplo Excel y Statistica

Ejemplo Cáncer de Pulmón/ Curva Kaplan-Meier (Producto-Límite)



Al principio el 100% de los individuos están vivos

El 50% de los individuos sobrevive hasta aprox. los 80 días (ver tabla percentiles abajo)

Momento en que todos los individuos de la población han muerto o han sido censurados (aprox. 1000 días)

Percentiles	Survival Time
25 th percentile (lower quartile)	24,2500
50 th percentile (median)	80,0000
75 th percentile (upper quartile)	159,9050

Percentiles:

Ej El 50% de los individuos sobrevive hasta los 80 días (Mediana de expectativa de vida)
ojo! Era 84 mediante Tablas de Vida

Comparaciones No Paramétricas de Funciones de Supervivencia en 2 muestras

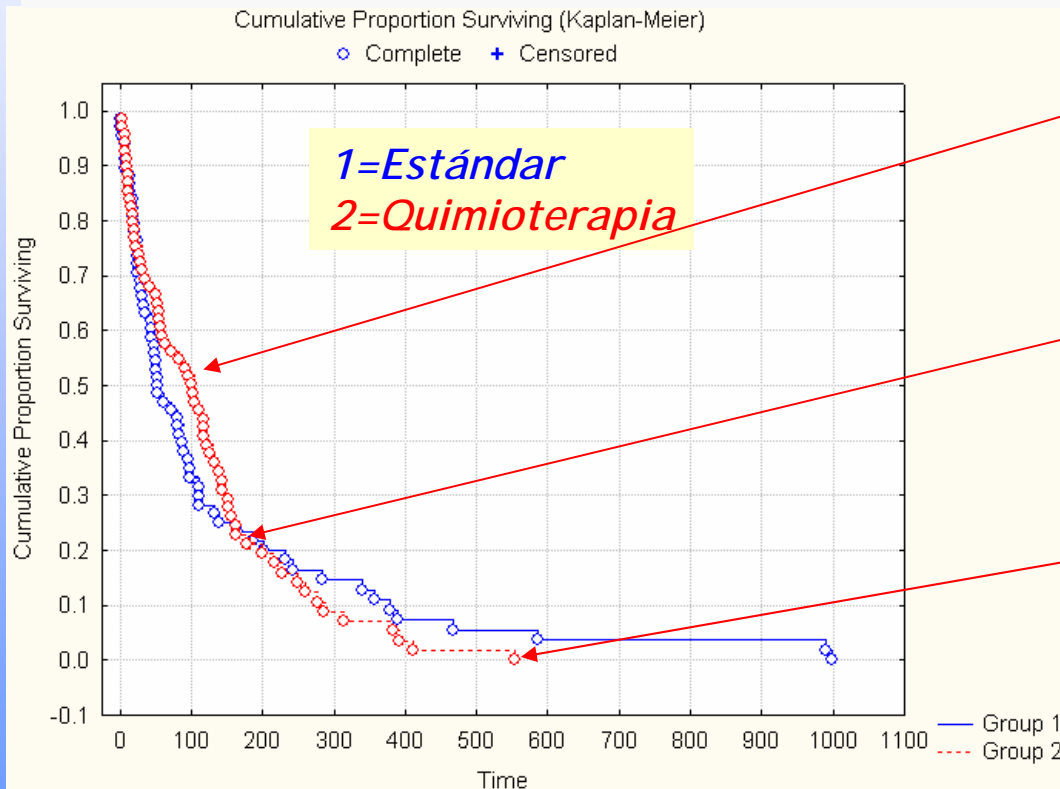
- La comparación de curvas de supervivencia en 2 grupos se puede hacer por métodos gráficos (Kaplan Meier) o por pruebas estadísticas
- La mayoría de las pruebas estadísticas para comparación de supervivencia en 2 grupos son de carácter NO-PARAMETRICO

Métodos No Paramétricos utilizados en STATISTICA:

- Test de Wilcoxon test (Gehan)
- Prueba F de Cox
- Prueba Cox-Mantel
- Log-rank
- Peto & Peto

Todos estos métodos prueban la significancia estadística de las diferencias en supervivencia en 2 grupos.

Ejemplo Cáncer de Pulmón/ Curvas Kaplan Meier para trat. estándar vs quimioterapia



Al principio la sobrevivencia es mayor en el grupo bajo quimioterapia

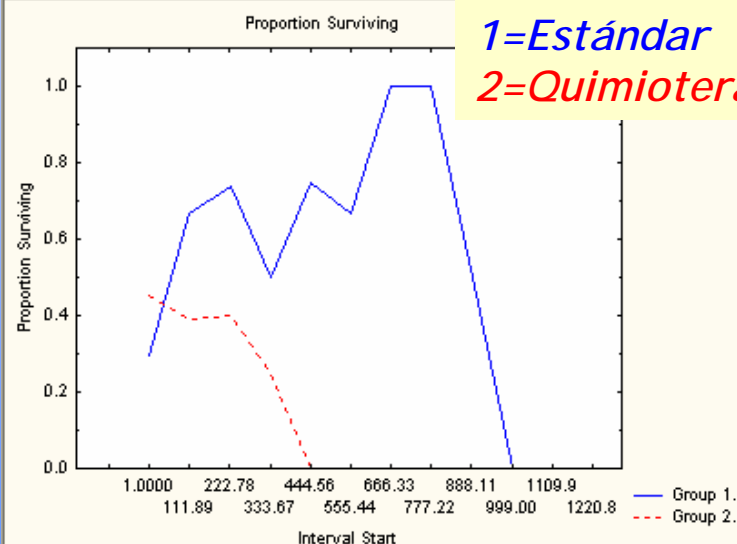
A los 200 días la proporción de individuos sobrevivientes es aprox. igual en ambos tratamientos

Al final la proporción que sobrevive es similar individuos bajo quimioterapia mueren más rápido (sobrev = 0 en día 550)

**** En este ejemplo los censurados (+) no se aprecian porque son solo 9**

Ejemplo Cáncer de Pulmón/ Comparación de 2 grupos; trat est(1) vs quimiot(2) Tablas de sobrevivencia por grupo y gráfico de prop. sobrev x grupo e intervalo

Life Table for Group 1 and Group 2 (An_Sob1b)										
Group 1: Code 1.0000 Group 2: Code 2.0000										
Lower Limit	Group 1: No.Enter	Group 2: No.Enter	Group 1: No.Cnsrd	Group 2: No.Cnsrd	Group 1: No.Dying	Group 2: No.Dying	Group 1: % Srwng	Group 2: % Srwng	Group 1: Cum.%Sr	Group 2: Cum.%Sr
1.00000	68	69	3	3	47	37	29.3233	45.18518	100.0000	100.0000
111.8889	18	29	0	2	6	17	66.6667	39.28571	29.3233	45.1852
222.7778	12	10	1	0	3	6	73.9130	40.00000	19.5489	17.7513
333.6667	8	4	0	0	4	3	50.0000	25.00000	14.4492	7.1005
444.5555	4	1	0	0	1	1	75.0000	0.00000	7.2246	1.7751
555.4445	3	0	0	0	1	0	66.6667	0.00000	5.4184	0.0000
666.3333	2	0	0	0	0	0	100.0000	0.00000	3.6123	0.0000
777.2222	2	0	0	0	0	0	100.0000	0.00000	3.6123	0.0000
888.1111	2	0	0	0	1	0	50.0000	0.00000	3.6123	0.0000
999.0000	1	0	0	0	1	0	0.0000	0.00000	1.8061	0.0000



Distribución porcentual de los individuos
que sobreviven en cada grupo para cada
intervalo

Diferencias entre distintos TEST de comparación de 2 muestras

No existe un test apropiado para todas las situaciones

Algunas recomendaciones:

➤ Prueba F tiende a ser más potente (menor error tipo II) que Gehan's Wilcoxon cuando:

- (1) Tamaños de muestra son pequeños (p.e., n por grupo <50)
- (2) Cuando las muestras siguen una distribución exponencial o Weibull
- (3) Cuando no hay observaciones censuradas.

➤ Cox Mantel y Log- rank son más poderosas que Gehan's Wilcoxon cuando las muestras provienen de una población que sigue una distribución exponencial o Weibull. En estos casos hay poca diferencia entre CoxMantel y Log Rank.

➤ Log-Rank es más poderoso que Wilcoxon para detectar diferencias **tardías** en la probabilidades de sobrevivencia, mientras que Gehan's Wilcoxon y Peto son más eficiente para detectar diferencias **tempranas**.

➤ En el ejemplo los resultados de todas las pruebas llevan a la misma conclusión (no hay diferencias significativas entre grupos)

Cox's F-Test (An_Sob1b)
T1 = 60.87314 T2 = 68.12685
F(128, 128) = 1.119161 p = .26257

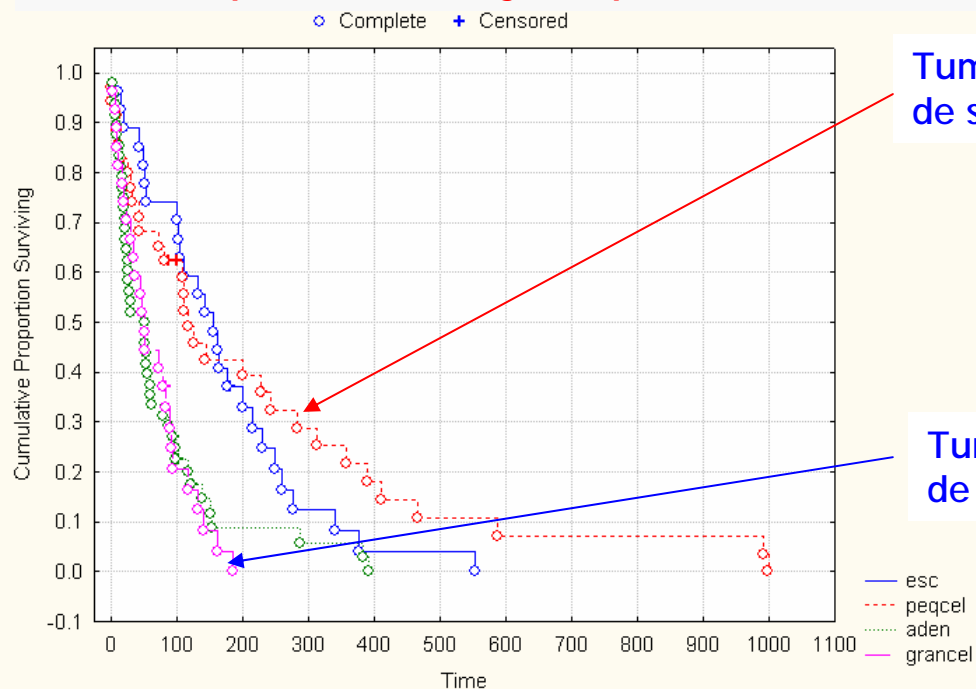
Gehan's Wilcoxon Test (An_Sob1b)
WWW = -447.0 Sum = 8273E2 Var = 2083E2
Test statistic = -.977160 p = .32849

Cox-Mantel Test (An_Sob1b)
I = 30.41039 U = -.500197
Test statistic = -.090705 p = .92773

Log-Rank Test (An_Sob1b)
WWW = .50020 Sum = 121.85 Var = 30.685
Test statistic = .0902984 p = .92805

Peto & Peto Wilcoxon Test (An_Sob1b)
WWW = 3.1422 Sum = 45.085 Var = 11.353
Test statistic = .9325330 p = .35106

Ejemplo Cáncer de Pulmón/ Comparación gráficas de múltiples grupos Curvas Kaplan Meier según tipo de tumor



Tumores de células pequeñas tienen una curva de sobrevida más alta

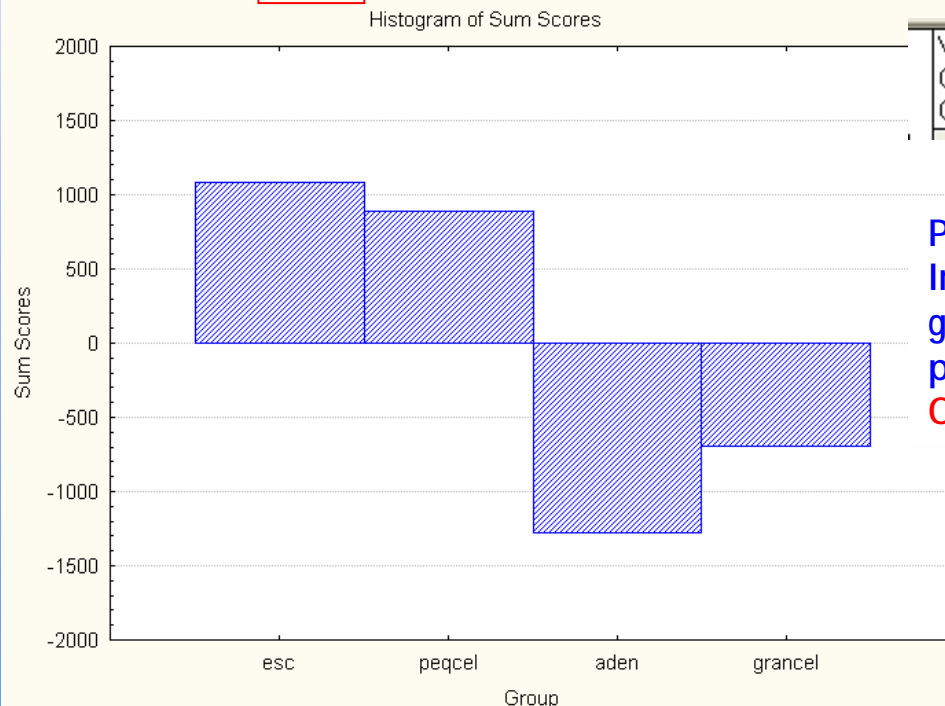
Tumores de células grandes tienen una curva de sobrevida más baja

Ejemplo Cáncer de Pulmón/ Scores de sobrevida para comparación de múltiples grupos

Summary statistics for each group (An_Sob1b)							
Group	Code for Group	Sum of Scores	N.uncsrd	Percent uncensrd	N.censrd	Percent censored	Total N
esc	0	1085.00	26	96.29630	1	3.70370	27
peqcel	1	890.00	31	88.57143	4	11.42857	35
aden	2	1278.00	45	93.75000	3	6.25000	48
grancel	3	-697.00	26	96.29630	1	3.70370	27

Scores de sobrevida:

Los grupos con mayores scores son los que tienden a tener mayor sobrevivencia



Variables: dayssurv by cell (4 groups) (An_Sob1b)
Censoring var.: censoring (Censored cases are marked with +)
Chi² = 19.4405 df = 4 p = .00022

Prueba chi-cuadrado

Indica que las diferencias en sobrevida entre grupos (tipos de tumor) son significativas.. pero... Cuáles grupos difieren?

Opción: Hacer comparación por pares

Referencias

➤ Para este curso:

Daniel, W. W. 2002. Bioestadística. Base para el análisis de las ciencias de la salud. Traduc. Por F. León Hernández. Edit. Limusa Wiley, México D.F. 908 p. [Capítulo 12.8](#)

Bull, K; Spiegelhalter, D. 1997. Tutorial in biostatistics. Survival analysis in observational studies. Statistics in Medicine, vol 16 1041-1074

***Este artículo condensa los 3 tipos de métodos

Noordhuizen et al. 1997. Application of quantitative methods in veterinary epidemiology. Cap. 7.

Assesing time-to-and event as an endpoint. Annotated guidelines for reporting statistical information. Reporting Survival Analysis. Chpater 9. (Solo No Paramétrico)

Altman, D. 1991. Practical statistics for medical research. Cap.13

➤ Información adicional:

J.D.Kalbfleisch and R.L. Prentice (1980). The Statistical Analysis of Failure Time Data. John Wiley and Sons.

D.R. Cox and D. Oaks (1984). Analysis of Survival Data. Chapman and Hall, London.