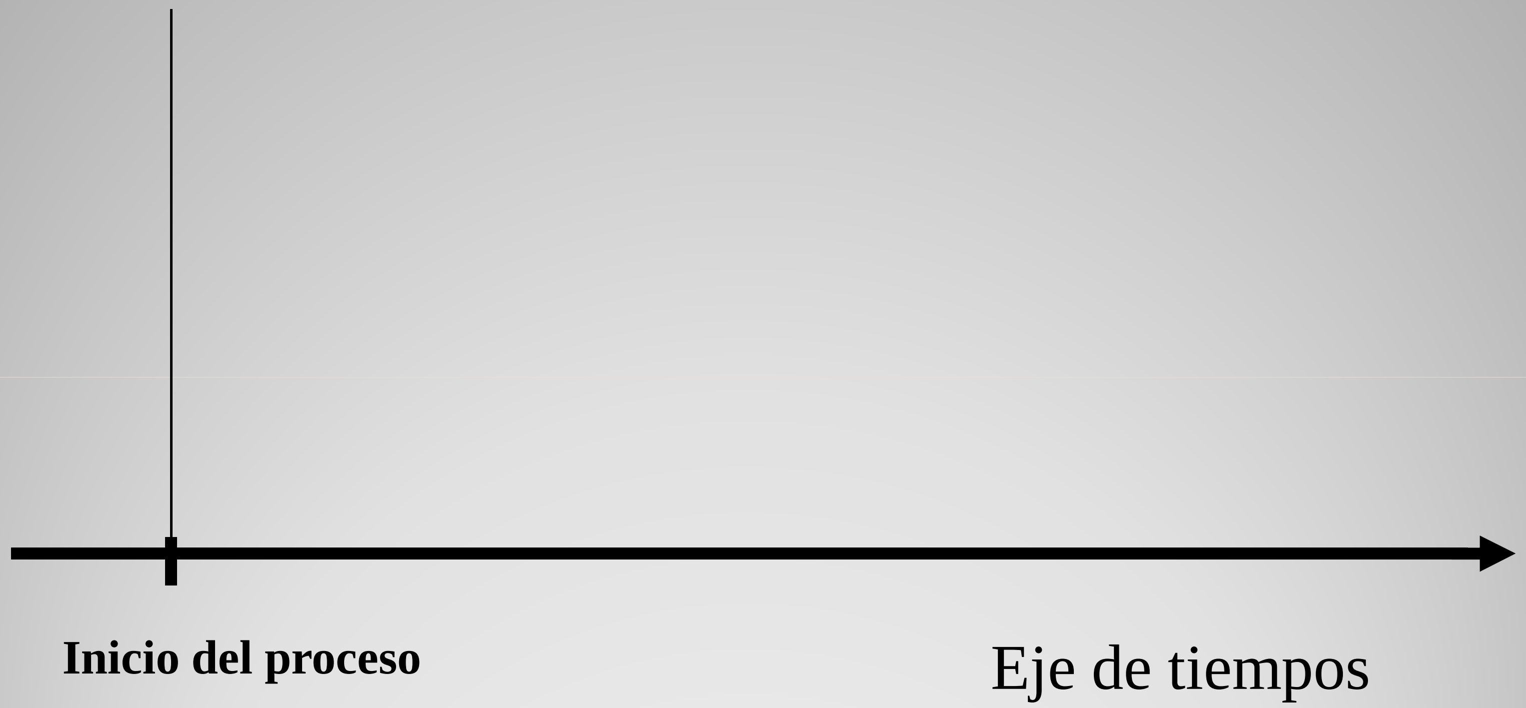
Introducción al ADS

- Vamos a estudiar *Duraciones de Procesos* que es algo muy común en muchas ciencias:
  - Duración de un componente (Fiabilidad)
  - Supervivencia de un paciente a un tratamiento (Medicina/farmacología)
  - Duración del desempleo (Economía)
  - Edad de las personas (Demografía y sociología)

Variables Aleatorias Positivas

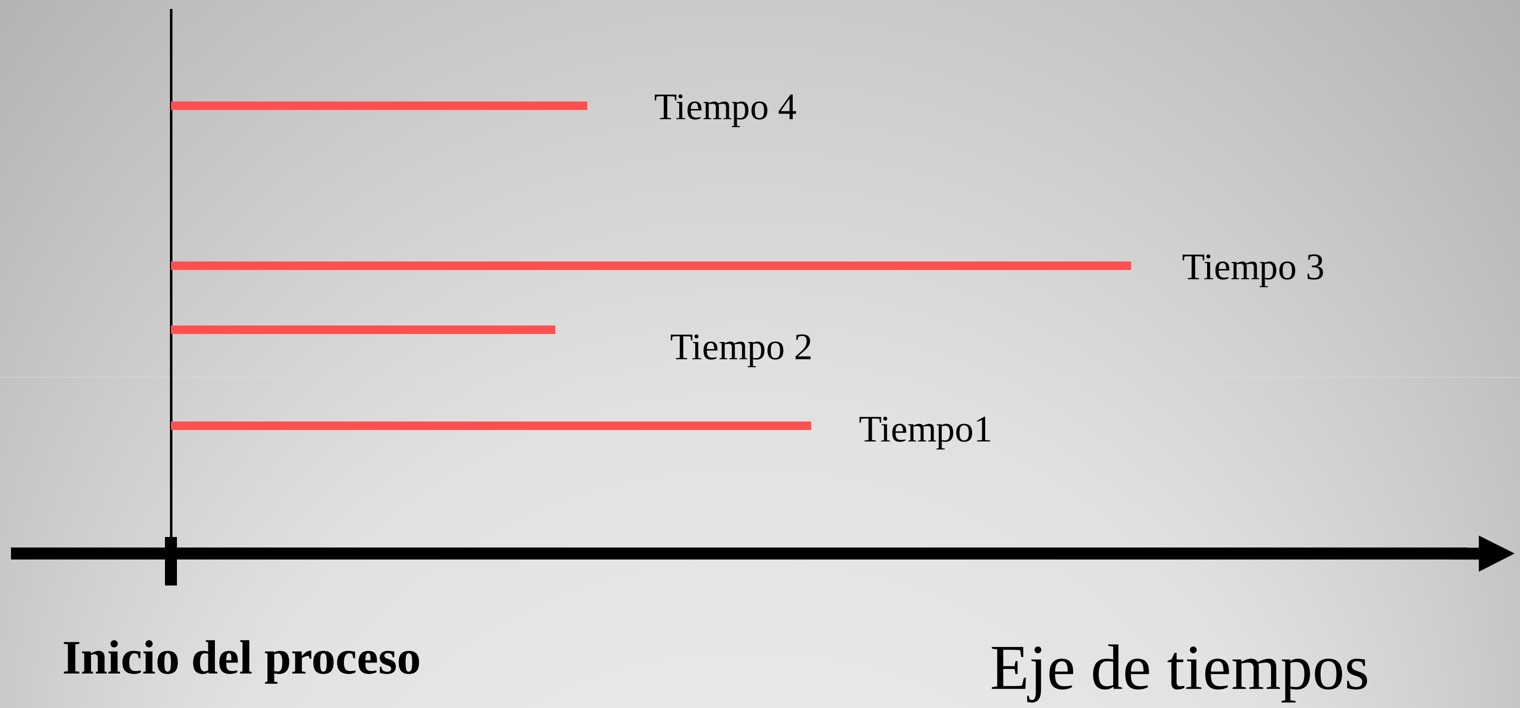
Conceptos básicos

Eje de tiempos



Conceptos básicos

Nuestros datos serán las duraciones de estos cuatro procesos: t_1, t_2, t_3 y t_4.



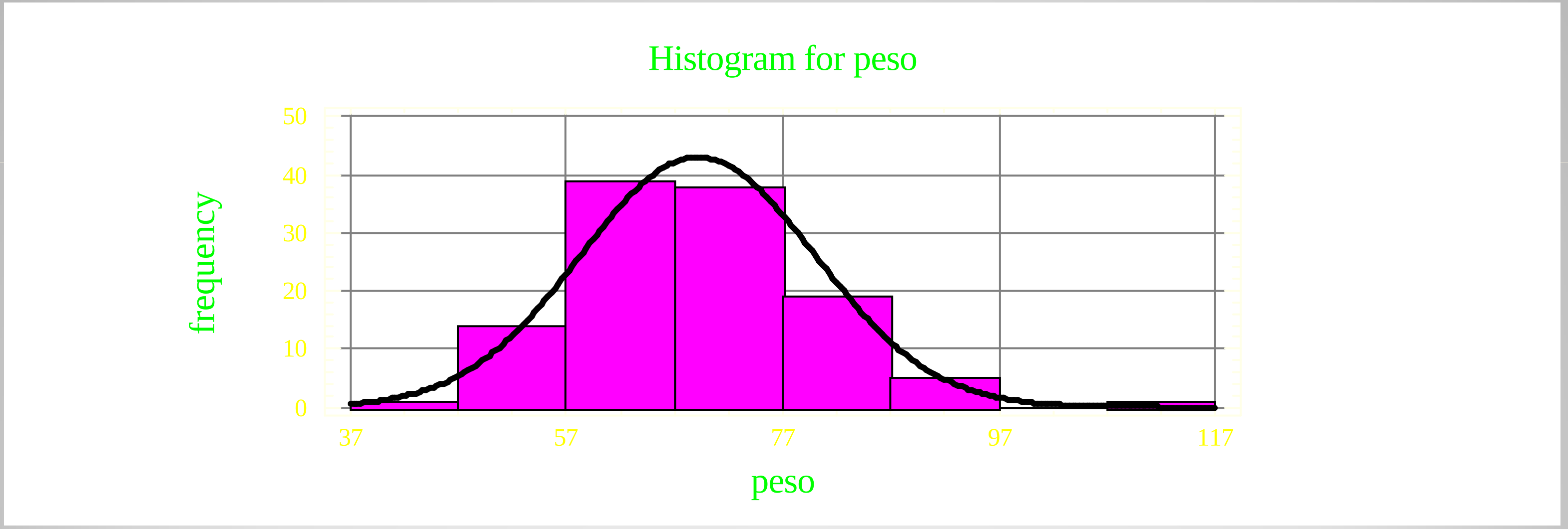
Conceptos básicos

- En otros análisis estadísticos hemos utilizado la Función de Densidad
- En ADS además usaremos:
  - Función de Supervivencia o Función de Fiabilidad (en ingeniería)

Funciones asociadas al ADS

Función de Densidad (repaso)

Se denomina f(t)



El área comprendida bajo la función de densidad es la probabilidad de encontrar observaciones en ese intervalo.

- Función de Supervivencia o de Fiabilidad

- $S(t)$=Prob(Durar más allá de t)=$P(X>t)$

Nuevas funciones

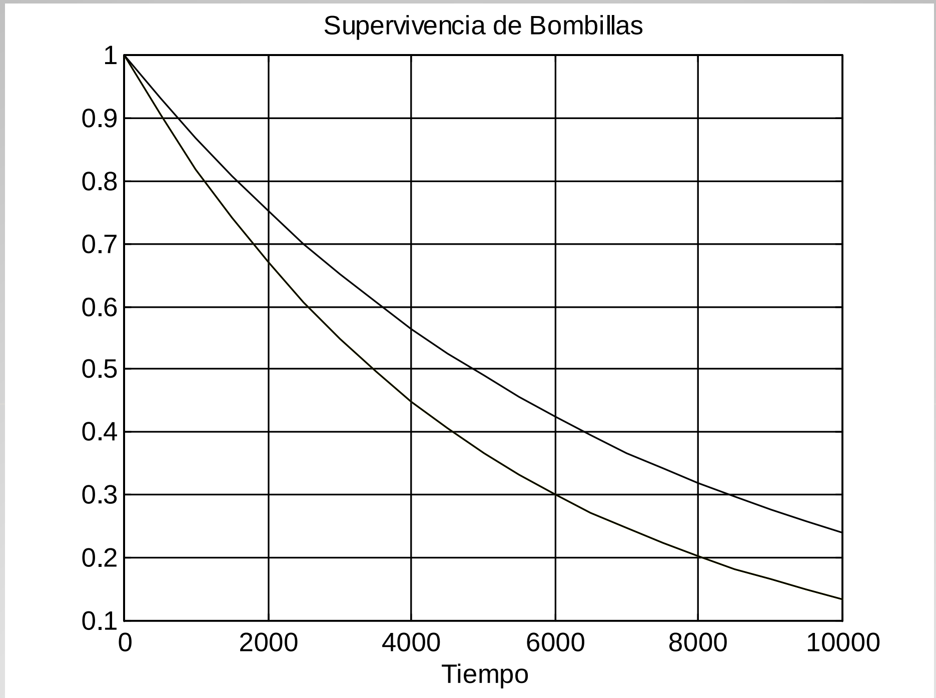
La función de supervivencia proporciona la probabilidad de que un componente/proceso/paciente esté funcionando/vivo al cabo de *t* horas.

Si un componente tiene una función de supervivencia:

$$S(1000)=0.89$$

quiere decir que la probabilidad de que el componente siga funcionando al cabo de 1000 horas es de 0.89.

Función de supervivencia



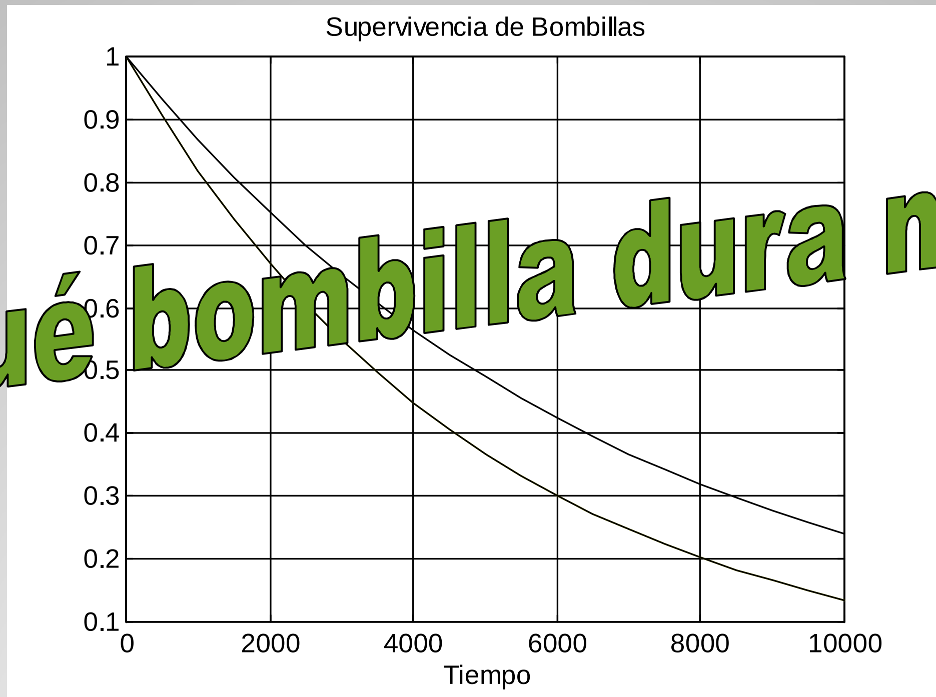
Evidentemente

S(0)=Pr(Durar mas de 0)=1

S(infinito)=Pr(Durar más de infinito)=0

Es decir que empieza en 1 llega hasta 0

Función de supervivencia



¿Qué bombilla dura más?

La probabilidad de que ambas estén funcionando al cabo de 6000 horas es de 0.3 y 0.42 respectivamente.

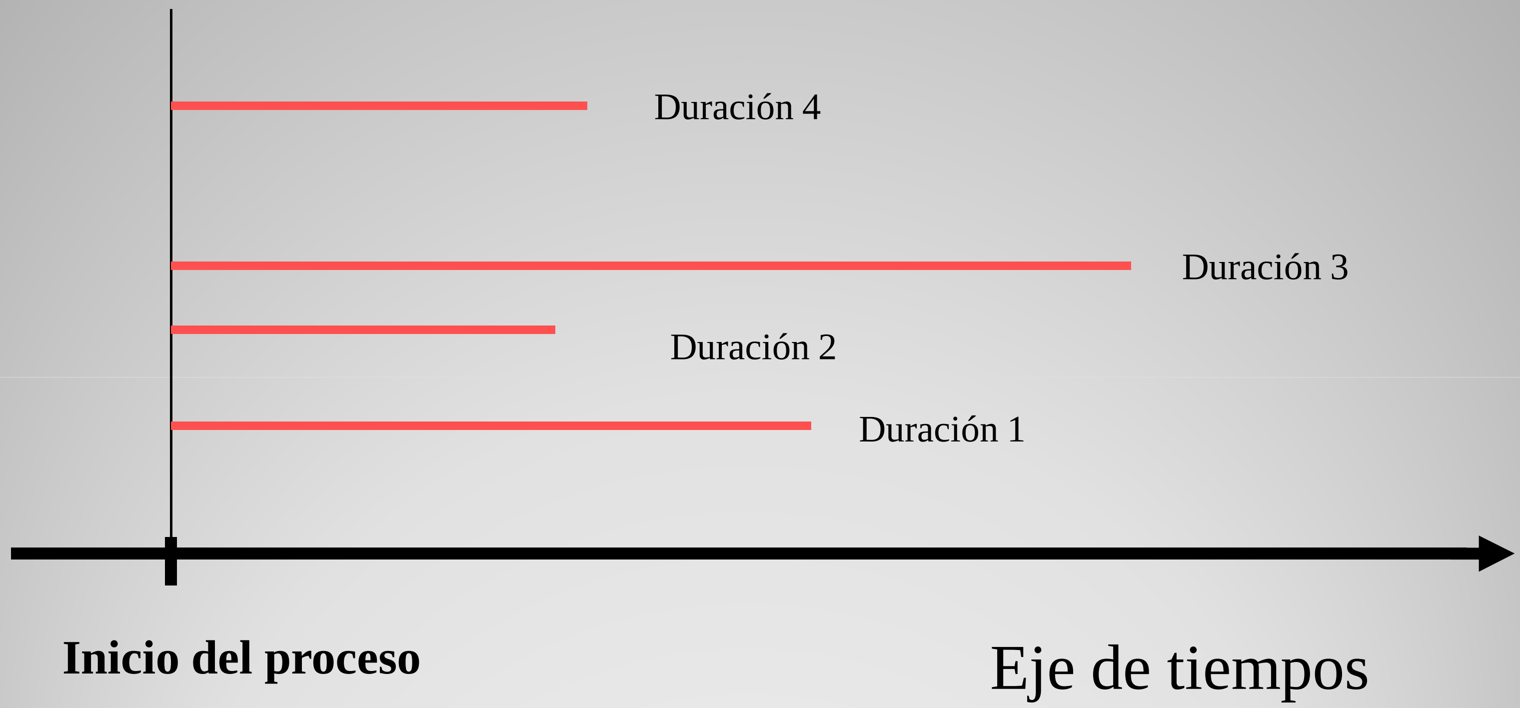
Datos incompletos

Censura

Censura

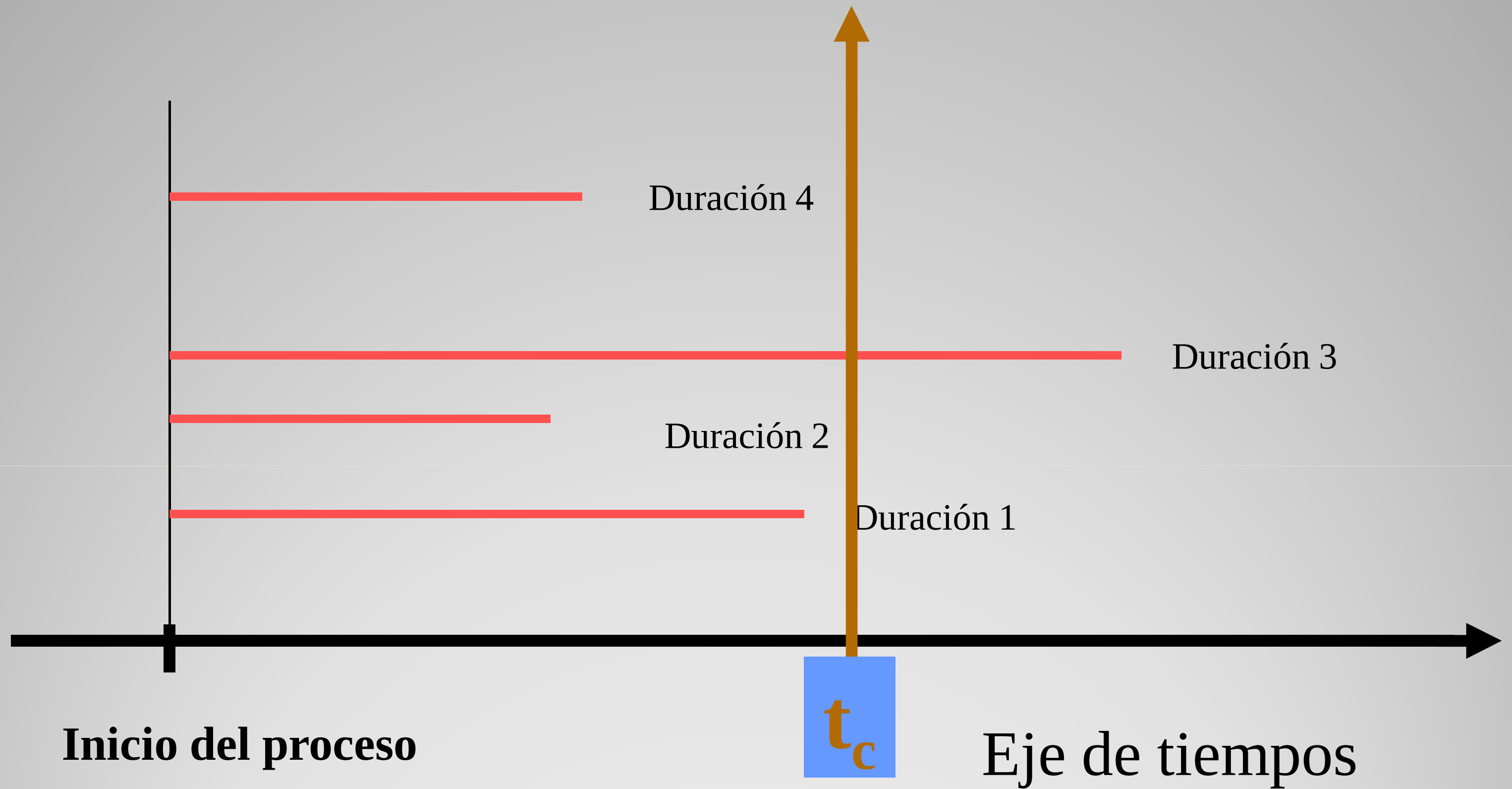
- Una observación esta censurada cuando solo contiene información parcial sobre la variable a estudiar.
- Esta situación es muy frecuente: la longitud del intervalo entre tránsitos impide muchas veces el seguimiento de la muestra hasta el transito final.
- Hay tres tipos de censura:
  - Censura por la derecha
  - Censura por la izquierda
  - Censura por intervalos

Sin censura



Nuestros datos serán las duraciones de estos cuatro objetos: t_1, t_2, t_3 y t_4.

Fin del estudio al cabo de un tiempo t_c

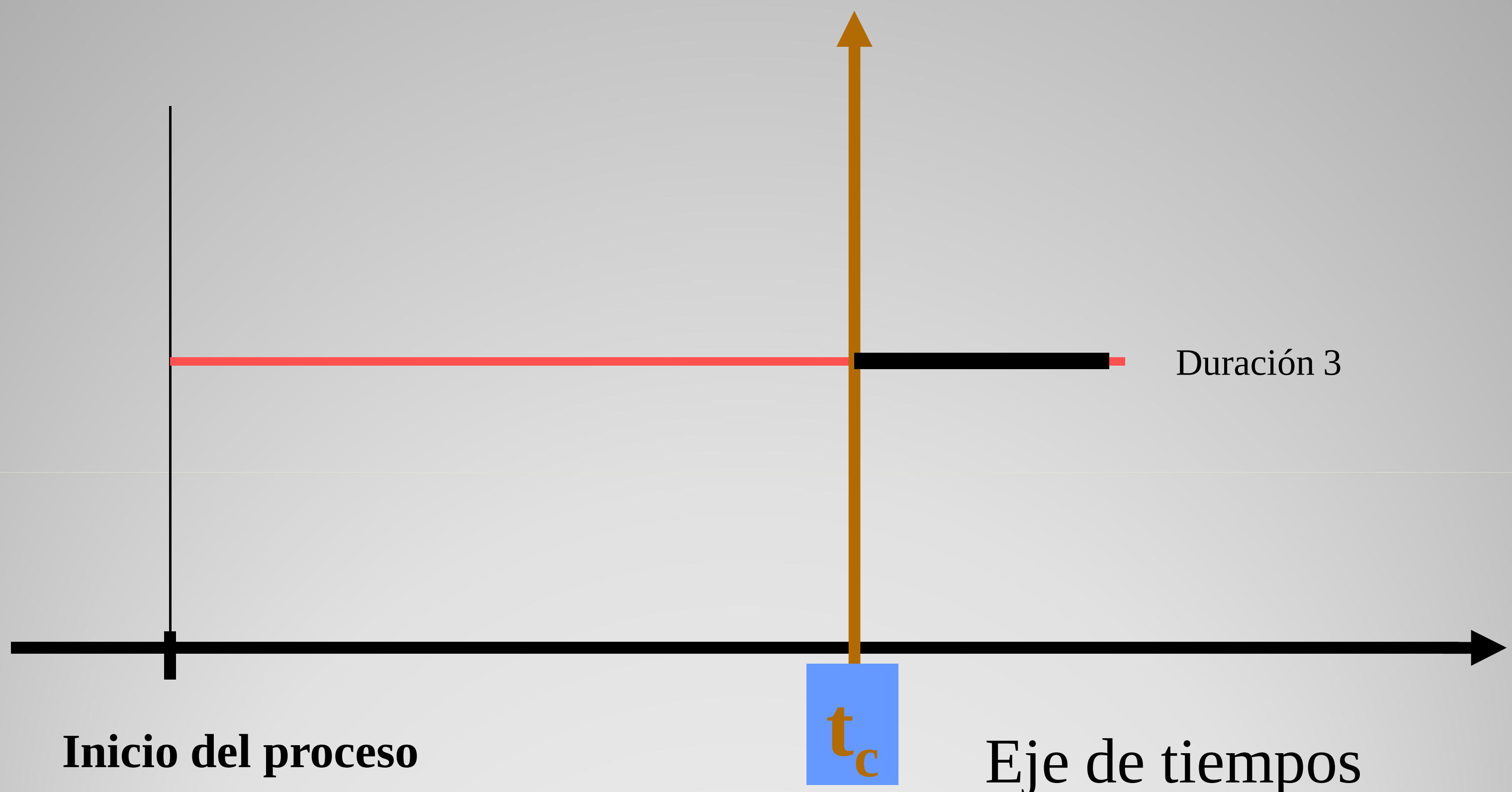


Nuestros datos serán las duraciones de los tres objetos:

t_1, t_2 y t_4

PERO NO OBSERVAMOS EL FINAL DE t_3. Sólo sabemos que $t_3 > t_c$

Fin del estudio al cabo de un tiempo t_c



Parte de la derecha no observada del componente 3
CENSURA POR LA DERECHA

- Economía: duración del desempleo suele obtenerse de encuestas que preguntan a los parados cuanto tiempo llevan en paro. Al no conocerse el tiempo adicional que van a permanecer sin trabajo, solo se sabe su duración censurada.
  - El paro es superior al que el entrevistado indica en la encuesta. Si una persona dice que lleva en paro 3 meses, su paro real será $t_i > 3$

Censura por la derecha

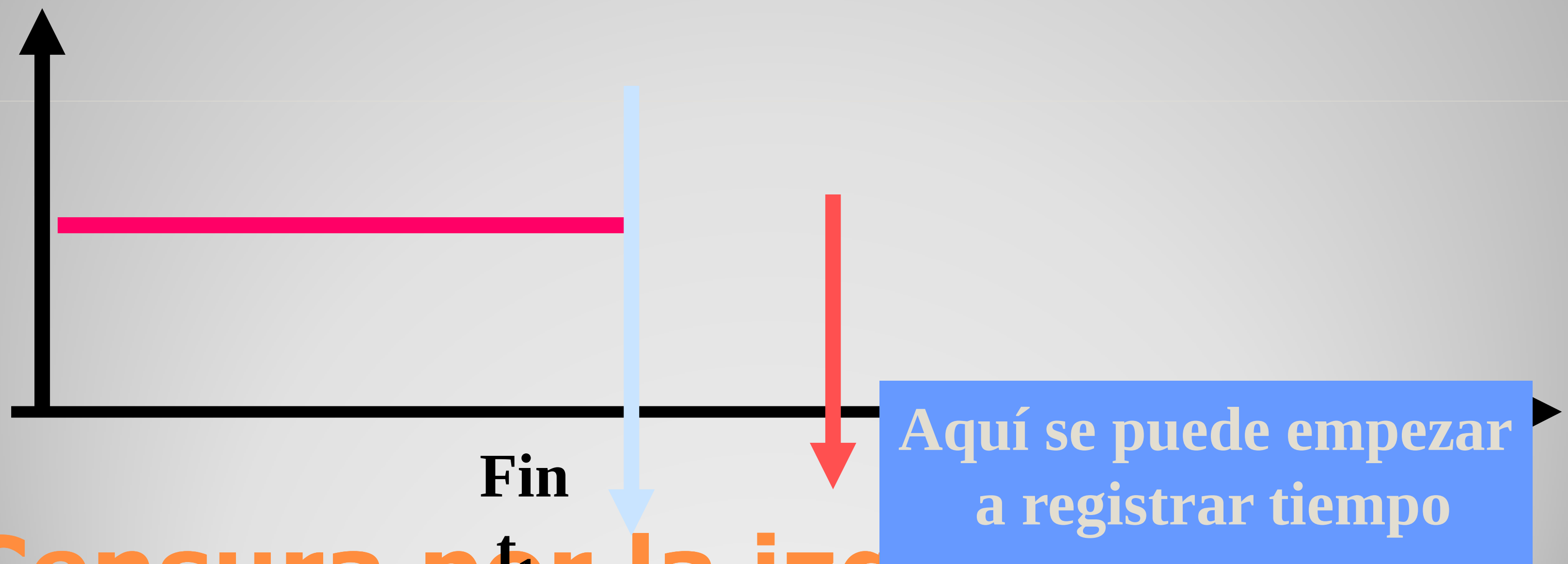
- **Fiabilidad**: es muy normal poner a prueba una partida de componentes y observar los fallos durante un periodo de tiempo determinado. Los elementos que fallen durante este periodo proporcionaran observaciones completas. Los que sigan en funcionamiento al final del periodo proporcionaran observaciones censuradas.
  - El tiempo que se recoge para los elementos censurados será $t_i > t_c$

Censura por la derecha

- Muy rara.
- Aparece en fisica nuclear

Censura por la izquierda

- Cuando no podemos observar un acontecimiento por ocurrir demasiado rápido (Vida de partículas subatómicas)



Aquí se puede empezar a registrar tiempo t_c

Censura por la izq

Registramos que la duración $t_1 < t_c$

- Mucho más compleja que con datos completos
- En general es imposible calcularlo a mano
- Lo haremos en ordenador

Estimación con datos censurados

- Si hay censura las técnicas descriptivas básicas no van a servir.
- No podremos realizar histogramas si no conocemos la longitud final de las observaciones.
- Usaremos el estimador de producto límite o estimador de Kaplan Meier

Estimación con datos censurados

- Estima la función de supervivencia cuando hay censura

- Vamos a estudiarlo con un ejemplo

Estimador de Kaplan Meier

Estimador de Kaplan Meier

- Se realiza un experimento para saber si una nueva droga es efectiva tratando una enfermedad mortal.
- Un grupo de pacientes es tratado con la nueva droga (6M) y el otro con placebo.
- El ensayo es doble ciego:
- Datos de primer grupo (6MP):
  - 6 6 6 6* 7* 9 10* 10* 11 13 16* 17* 19* 20 22 23* 25* 32* 32* 34* 35
- Datos del segundo grupo (Placebo)
  - 1 1 2 2 3 4 4 5 5 8 8 8 8 11 11 12 12 15 17 22 23

Estimador de Kaplan Meier

- Realiza las siguientes operaciones: (El ordenador)
  - Se ordenan los valores de menor a mayor
  - Para cada tiempo de fallo (Si hay varios fallos en el mismo momento, para el ultimo) se calcula el numero de individuos que quedan en riesgo.
  - El estimador para el primer tiempo de fallos será:
    - $S(t_1)=(n_1-d_1)/n_1$
  - n_1 representa el número de individuos que están en riesgo justo antes del primer tiempo de fallo.
  - d_1 es el número de fallos/muertes en el primer tiempo de fallo.
  - Para el segundo tiempo de fallo será
    - $S(t_2)=[(n_2-d_2)/n_2].S(t_1)$
    - $S(t_3)=[(n_3-d_3)/n_3].S(t_2)$

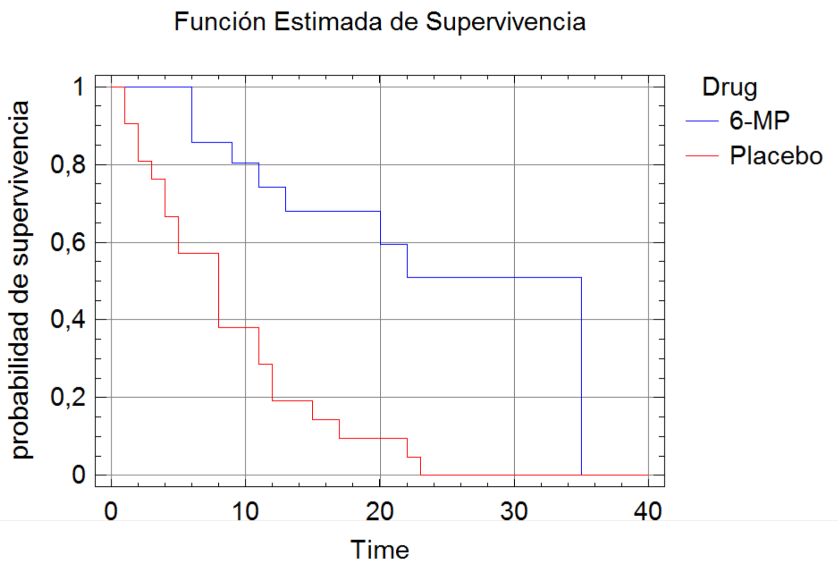
Se puede saltar

6MP): 6 6 6 6* 7* 9 10* 10* 11 13 16* 17* 19* 20 22 23* 25* 32* 32* 34* 35

Sólo se registran Tiempos de FALLO
No censuras. Por eso no aparece 7 o 9

Tiempo	Riesgo	Fallos	$\frac{n_i-d_i}{n_i}$	S(t)
6	21	3	$\frac{21-3}{21}$	$\frac{21-3}{21} = 0.8571$
9	16	1	$\frac{16-1}{16}$	$\frac{16-1}{16} 0.8571 = 0.8036$
11	13	1	$\frac{13-1}{13}$	$\frac{13-1}{13} .0.8036 = 0.7418$
13	12	1	$\frac{12-1}{12}$	$\frac{12-1}{12} 0.7418 = 0.6799$
20	8	1	$\frac{8-1}{8}$	$\frac{8-1}{8} .0.6799 = 0.5950$
22	7	1	$\frac{7-1}{7}$	$\frac{7-1}{7} .0.5950 = 0.51$
35	1	1	0	0

Se puede saltar



DROGA 6MP:Tiempo medio de supervivencia = 24,2411 Error estándar = 3,0095

PLACEBO:Tiempo medio de supervivencia = 8,66667 Error estándar = 1,4114