

Análisis sobre la duración en clínica de adictos a la heroína

Alejandro Barón García

Abril 2019

1 Introducción

En este artículo se expone un **análisis sobre el tiempo de permanencia en una clínica para el tratamiento de heroinómanos**. Los datos se obtuvieron en un estudio australiano de 1991, época en la que junto a los 80 la drogadicción golpeó la sociedad occidental con una virulencia brutal. La motivación de este estudio es doble, por un lado los intereses clínicos para ajustar tratamientos y por otro lado el interés humano, poder dar una estimación a los cercanos al paciente de su recuperación.

2 Marco del estudio

Esta recogida de datos se efectuó en Sydney en 1991 en dos clínicas con tratamientos distintos a lo largo de 1076 días, con 238 pacientes en el estudio. Las dos clínicas aplicaban políticas de tratamiento distintas, por lo que será interesante considerar las dos poblaciones como disociadas ya que lo que quiero plasmar en este estudio son los factores externos. Así mismo se recogió información sobre si los pacientes estuvieron alguna vez en prisión y la dosis de metadona, un opiáceo sintético utilizado como sustitutivo en el tratamiento de desintoxicación de drogadictos.

2.1 Censuras y truncamientos

El conjunto de datos no estaba muy bien descrito (ver fuentes del documento), pero he considerado que no había truncamientos. Por lo que he conseguido buscar (el conjunto de datos aparece en algunos cursos pero analizado y explicado de forma escueta), sí que existen censuras las cuales son de tipo fallecimiento o

abandono prematuro de la clínica. En el conjunto de datos tenemos muchas censuras, quizás demasiadas, un 63.025% de los datos son censuras.

3 Método

Para realizar este estudio, se utilizará la metodología estadística conocida como **Análisis de Supervivencia**, en el cual se estudian variables de tipo "Tiempo hasta que ocurre un evento". Dentro de este estudio, se analizarán las funciones de supervivencia y de riesgo, y finalmente se ajustará un modelo de regresión de Cox de riesgos proporcionales.

3.1 Variables

3.1.1 Tiempo en clínica

Variable respuesta del estudio, es el tiempo (en días) que un paciente permaneció en la clínica hasta su salida.

Clínica

Toma los valores 1 y 2, siguiendo cada clínica un tratamiento distinto. Utilizaremos esta variable como principal variable del estudio y además para estratificar y poder ver cómo difieren ambos tratamientos,

Dosis de Metadona

Se recogió la dosis máxima de metadona que recibió cada paciente al día, en unidades de mg/día. La consideraremos como covariable dentro del estudio.

Historial penal

Se tiene también un registro sobre si el paciente pasó en algún momento de su vida por prisión, lo cual se sospecha que puede ser un factor influyente en la drogadicción y capacidad de desintoxicación

4 Resultados

4.1 Estudio de las variables

Previamente a la estimación de un modelo de Cox, tenemos que ver si se cumple o no la hipótesis de riesgos proporcionales a lo largo del tiempo. Para ello, veremos las funciones de impacto/riesgo de las variables según el tiempo

Clínica

En este caso tenemos un 68.48% de pacientes de la Clínica 1 frente a un 31.52% de la Clínica 2. Lo primero que vemos en la gráfica (Figura 1) es que no se cumple la hipótesis de riesgos proporcionales, se alejan rápidamente las curvas de ambas clínicas. Debido a esto, he decidido realizar un modelo de regresión de Cox **estratificado**, ya que, teniendo una muestra suficientemente grande de las dos, creo que son dos poblaciones distintas y es necesario estudiarlas por separado. Si quisiésemos comparar los tratamientos, lo que también puede resultar muy interesante, un modelo estratificado no permite ver el efecto del tratamiento en la respuesta, pero en este estudio he preferido ver para cada clínica por separado la duración de los pacientes en cada una.

Prisión

Para esta variable tenemos un 46.63% de pacientes que han pasado por el presidio y un 53.57% que no. En este caso (Figura 2) sí que tenemos claro que se cumple la hipótesis de riesgos proporcionales, además de forma bastante clara aunque se crucen a veces las curvas, siguen prácticamente el mismo trazado. Debido a esto, se podría pensar que no hay diferencias significativas en el impacto para la salida de la clínica. Sin embargo, esa diferencia en los primeros 500 días

me hace pensar que sí que puede influir, luego entrará en el modelo.

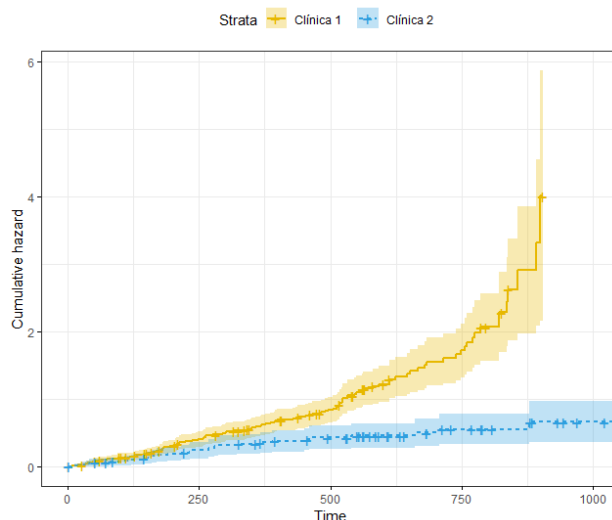


Figura 1: Riesgo según la Clínica

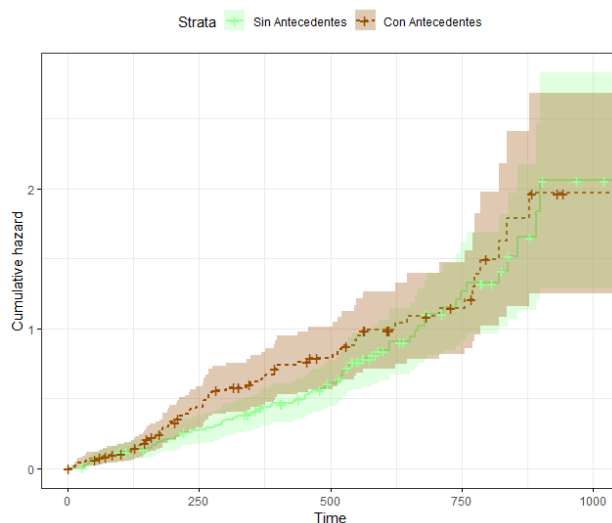


Figura 2: Riesgo según los antecedentes penales

Dosis

La dosis ha requerido de una transformación de la variable original. Dado que es una variable que podemos suponer continua, se ha discretizado para facilitar el análisis. En primera instancia, se consideró la división en lo que algunos expertos consideran como dosis baja (hasta 60 mg/día), media (de 60 a 100 mg/día) o alta (superior a los 100 mg). Sin embargo, si vemos el siguiente histograma (Figura 3), vemos que para la dosis alta habría escasas observaciones, lo cual nos hará incurrir en una mala estimación de los resultados. Es por esto que finalmente se van a considerar 2 niveles, baja (menor a 60mg/día) y alta (mayor de 60mg/día). En el sentido estadístico esta división tiene sentido para estos datos; la media y la mediana coinciden con este punto de corte para discretizar.

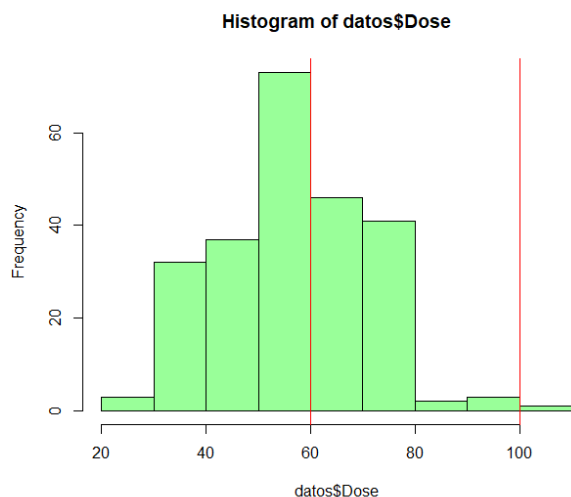


Figura 3: Riesgo según la dosis

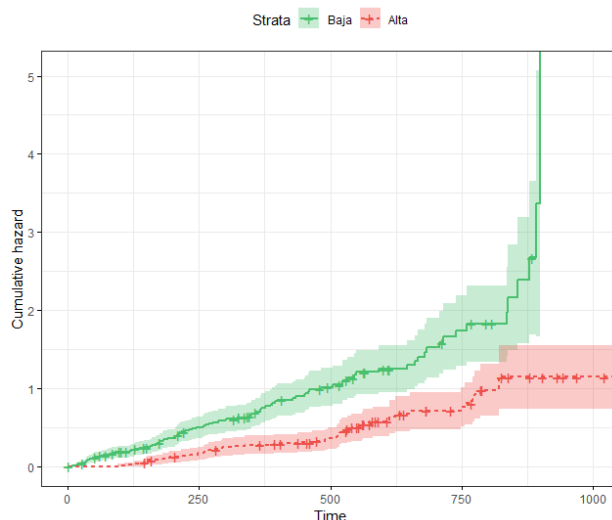


Figura 4: Riesgo según la dosis

Una vez se discretiza la variable, podemos ver como aquí también se mantienen proporcionales los riesgos según avanza el tiempo, de hecho son prácticamente paralelos hasta el final. Ese despunte brusco de la función de riesgo no nos debería preocupar en principio, pues son las observaciones finales (pocas), lo que sumado a las censuras nos puede dar una mala estimación. No obstante, tendremos cuidado en el análisis con esta variable.

4.2 Modelo

Como se comentó anteriormente, vamos a implementar un modelo de regresión de Cox estratificado por Clínica. La formulación del modelo será:

$$h(t|Z, S) = h_0^s(t) \exp(B'Z)$$

Con $S = \{Clínica\ 1, Clínica\ 2\}$ como estratos.

Como solo tenemos dos variables, ajustaremos dos modelos que serán, en notación de R:

Modelo sin interacción:

```
Días ~ Strata(Clínica) + Prisión + Dosis
```

Modelo con interacción:

Días~
Strata(Clinica)+Prisión+Dosis+Prisión*Dosis

Antes de empezar con el estratificado, vamos a ver el modelo sin estratificar y ver cómo se rechaza con el logrank test que los riesgos sean equitativos.

4.3 Modelo base (sin estratos)

Variable	β	Z-value	Pvalue
<i>Clinic</i>	-0.9676	-4.468	7.88-e06
<i>Prision</i>	0.2684	1.615	0.106
<i>Dosis</i>	0.7269	-4.059	4.93e-05

Modelo:

R ²	Likelihood	Wald	Score(Logrank)
0.194	4.085e-11	1.351e-09	2.378e-10

AIC=1365.711 . Vemos que las variables son significativas salvo Prision, pero lo interesante de este modelo es el logrank test, para ver que como sospechábamos en el apartado anterior, no se cumple la hipótesis de riesgos proporcionales. Los test globales del modelo (Wald Likelihood y Score(Logrank)) Rechazan que el modelo sea nulo.

Test de riesgos proporcionales

	rho	chisq	p
Clinic	-0.2681	12.4885	0.000409
Prison	-0.0111	0.0186	0.891658
DoseLevel	0.1362	2.8617	0.090713
GLOBAL	NA	14.1461	0.002713

Se rechaza fuertemente (pvalor casi 0) que la variable Clinic cumpla la hipótesis de riesgos proporcionales, confirmando así nuestras sospechas iniciales. El test para el modelo en global también se rechaza. Si vemos los residuos del modelo, podemos observar patrones temporales (curva hacia abajo para β_{clinic} o crecientes en $\beta_{Prision}$), lo que indica problemas en el modelo.

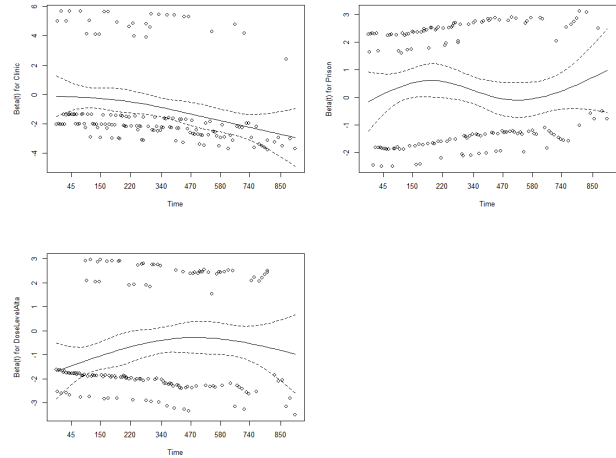


Figura 5: Residuos del modelo sin estratos

4.4 Modelo Estratificado sin interacción

Variable	β	Z-value	Pvalue
<i>Prision</i>	0.3328	1.980	0.0477
<i>Dosis</i>	-0.6930	-3.898	9.7e-05

Modelo:

R ²	Likelihood	Wald	Score(Logrank)
0.081	4.198e-05	6.814e-05	5.035e-05

Para este modelo, AIC=1212.791. Como vemos, el R^2 ha bajado (variabilidad explicada por el modelo, malo) pero esto puede deberse a que hemos quitado una variable (menor sobreajuste). Sin embargo, el AIC se ha minimizado, con lo que el modelo según este criterio es mejor. Además, ahora prisión es significativo a nivel 0.05 (antes no) Veamos los test y los residuos. Los test globales del modelo (Wald Likelihood y Score(Logrank)) Rechazan que el modelo sea nulo.

	rho	chisq	p
Prison	0.00834	0.0106	0.9181
DoseLevel	0.15372	3.5042	0.0612
GLOBAL	NA	3.5130	0.1726

Los pvalores suben y el global no se rechaza a un nivel 0.05, buena señal de que el modelo estratificado va por mejor camino que el sin estratificar. Los residuos no son malos, parece que sí que siguen una línea horizontal (la foto en el documento está achatada).

4.5 Modelo Estratificado con interacción

Variable	β	Z-value	Pvalue
<i>Prision</i>	0.2971	1.456	0.14541
<i>Dosis</i>	-0.7436	-3.061	0.00221
<i>Dosis*Prision</i>	0.1088	0.309	0.75763

Modelo:

R^2	Likelihood	Wald	Logrk
0.082	0.0001505	0.0002894	0.0001776

AIC=1214.695. Como vemos, el pvalor de *Dosis*Prision* es muy alto, la interacción de ambas variables que pensábamos que podía ser interesante parece que no que cuenta para el modelo. *Prisión* pasa a no ser significativa aunque *Dosis* sí, sin embargo la dejaría en el modelo porque nos interesa estimar su efecto. Los test globales del modelo (Wald Likelihood y Score(Logrank)) Rechazan que el modelo sea nulo, aunque han subido los pvalores, lo cual ya da pistas de que esta interacción podría no ser estadísticamente útil.

Si miramos los tests de riesgos proporcionales, podemos ver que tampoco se rechaza a nivel $\alpha = 0.05$ la hipótesis de riesgos proporcionales. Los residuos también parecen buenos, no se ven patrones temporales (parece que siguen para los tres β una línea horizontal).

	rho	chisq	p
Prison	0.0159	0.0383	0.845
DoseLevel	0.1302	2.5719	0.109
Prison:DoseLevel	-0.0253	0.0952	0.758
GLOBAL	NA	3.7220	0.293

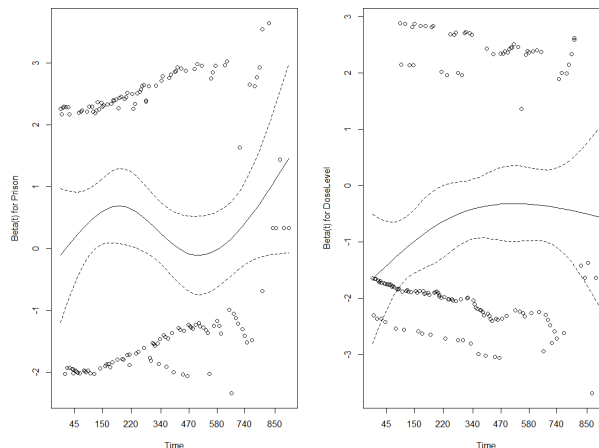


Figura 6: Residuos del modelo estratificado

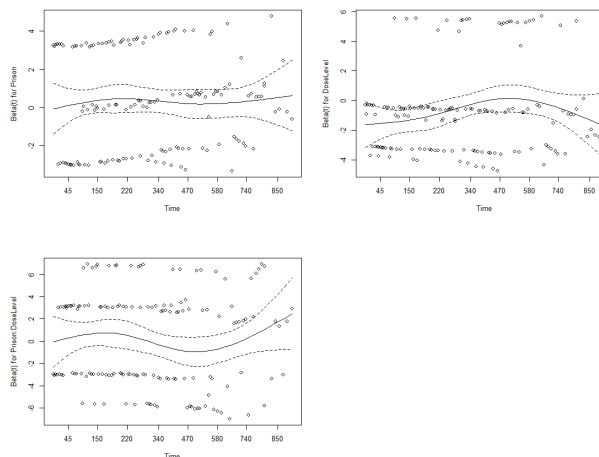


Figura 7: Residuos del modelo con interacción

Sin embargo, habría que valorar si el añadir una variable nos interesa, posiblemente estemos cometiendo un mayor sobreajuste, además el AIC aumenta y los pvalores son peores. Además, la interacción no parece interesante, quizás si se tiene especial interés clínico se podría considerar para los modelos, pero el pvalor me parece demasiado alto.

5 Conclusiones

5.1 Modelo final

En la elección del modelo final, aunque los tres propuestos no describen muy bien la realidad de este conjunto de datos, he decidido quedarme con el segundo, ya que es el que mejores propiedades estadísticas tiene (cumple RP y se rechaza que los β sean 0 a nivel 0.05).

5.2 Supervivencia

Hay que tener en cuenta que para este análisis, la supervivencia es algo "malo". El evento que estamos midiendo es "Tiempo hasta la salida de la clínica", luego tener una supervivencia alta implica alta probabilidad de que no suceda el evento, lo cual es algo que es benigno en este caso (al contrario de otros estudios clínicos como "Tiempo hasta la muerte" o estudios de durabilidad de "Tiempo hasta que falla un componente").

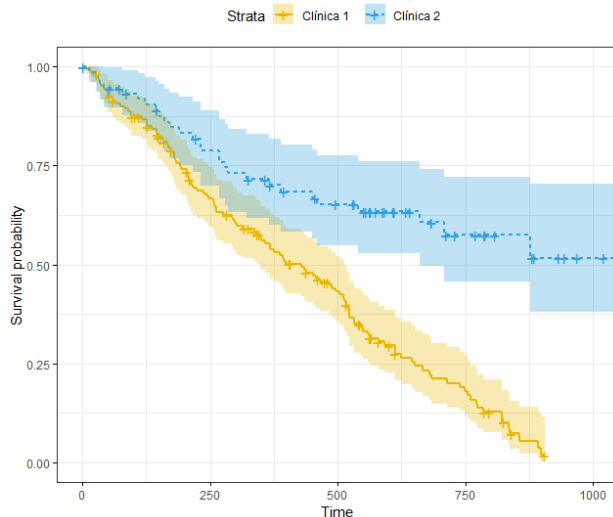


Figura 8: Supervivencia por clínica

Clínica

Como hemos visto, la clínica no cumple la hipótesis de riesgos proporcionales. Lo que sí que podemos

decir, dado que la curva para la Clínica 2 está siempre por encima, podemos decir que los pacientes de la clínica 2 por lo general están más tiempo hasta su recuperación.

Dosis

Para la dosis, su valor β era de -0.7436, lo que nos indica que el aumentar en una unidad (Alta=1 vs Baja=0) disminuye el riesgo de sufrir el evento, en este caso, salir de la clínica. Posiblemente esté relacionado con que al tener que suministrar una dosis alta, el paciente esté más "enganchado" a la droga y sea más difícil su recuperación.

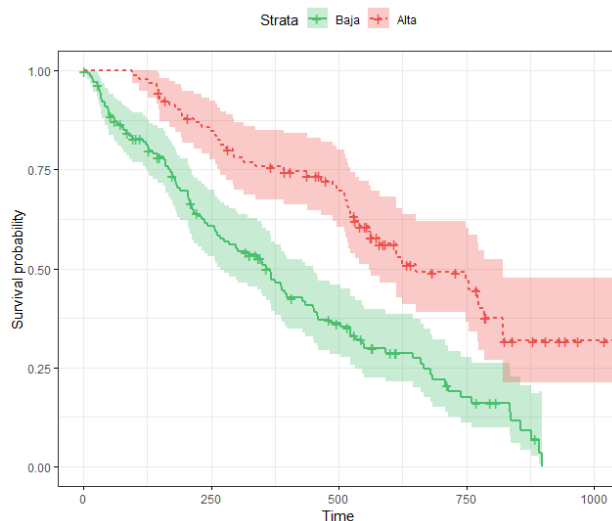


Figura 9: Supervivencia según la dosis

Prisión

Para esta variable, podemos ver que la supervivencia de los pacientes con antecedentes es menor que la de los que no han pasado por prisión. Esto es sorprendente, ya que la idea preconcebida es que los pacientes con historial conflictivo posiblemente hayan tenido mayor contacto con drogas y por tanto una mayor adicción. El parámetro para esta variable es $\beta = 0.2971$, lo cual nos indica que los pacientes con presencia penal anterior (Prision=1) tienen mayor riesgo

de sufrir el evento(salida de la clínica). Tratando de buscar alguna posible causa, quizás sea gente acostumbrada a una vida dura y tengan mayor fuerza de voluntad ante la adversidad, pero no me atrevo a asegurar nada.

Interaccion

La interacción de las variables podría arrojar algo de luz sobre lo anterior. Cuando toma el valor 1, implica que es un paciente que ha estado en prisión y que ha necesitado de una dosis alta de metadona. Quizás podríamos decir que los pacientes con un pasado conflictivo y con una necesidad de droga sustitutiva alta les cueste salir más porque su adicción sea mayor. Este grupo de pacientes tiene una supervivencia más alta todo el tiempo, lo que implica menos probabilidad de salir de la clínica. No he sido capaz de encajar esto en el análisis anterior.

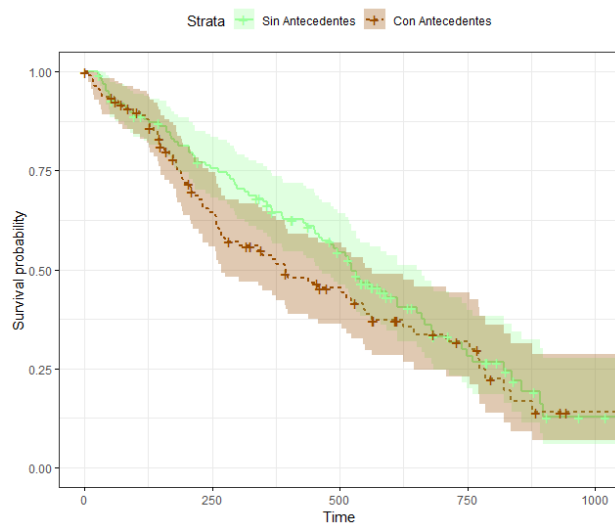


Figura 10: Supervivencia segun la estancia previa en prisión

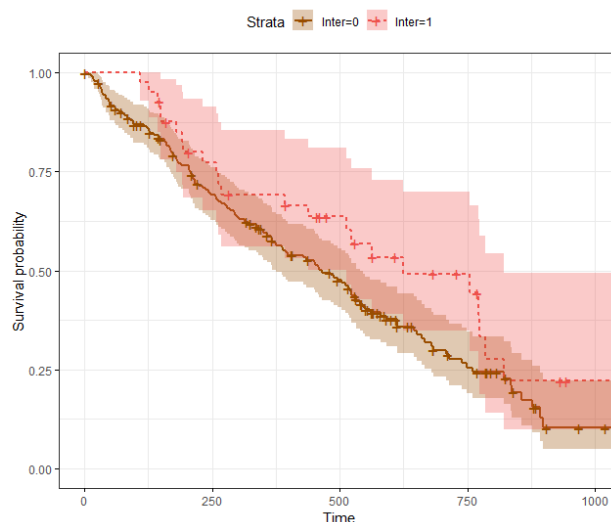


Figura 11: Supervivencia según la interacción

5.3 Posibles vía alternativa de modelado

Otra opción no desarrollada en este estudio por falta de tiempo es la de un modelo de Cox extendido, sin estratificar, de tal forma que se introduzca una variable $g(t)$ que multiplique a la variable clínica. Este enfoque tendría dos motivaciones; por un lado, se podría estimar el efecto del tratamiento y por otro, sería un posible baipás a la violación de la hipótesis de riesgos proporcionales. También se podría clasificar a los pacientes en 4 grupos, según las combinaciones de Prisión y nivel de dosis para hacer cuatro perfiles de paciente.

References

- [1] Handbook of Small Data Sets. Conjunto de datos de la universidad de Stanford-¿ Conjunto heroin.dat <https://www2.stat.duke.edu/courses/Spring03/sta113/Data/Hand/Hand.html>
- [2] *Divergent survivor curves Australian study of heroin addicts*, Caplehorn, et al.

(1991). Estudio original al que no he conseguido tener acceso. Dataset explicado en <http://web1.sph.emory.edu/dkleinb/allDatasets/surv2datasets/addicts.dat>

- [3] Material docente de la asignatura *Métodos Estadísticos de Computación Intensiva* - Grado en Estadística - Cristina Rueda Sabater.
- [4] <http://www.sthda.com/english/wiki/cox-model-assumptions>
- [5] <http://www.sthda.com/english/wiki/cox-proportional-hazards-model>
- [6] <https://www.vademecum.es/principios-activos-metadona-n07bc02>