

Análisis de Supervivencia

Giancarlo Sal y Rosas

Marzo, 2017

Contents

1	Introducción	5
2	Kaplan-Meier	9
3	Prueba de hipótesis	13
4	Modelos paramétricos	17
5	Modelo de Cox	19
5.1	Interpretación	19
5.2	Construcción	20
6	Censura por intervalo	21
	Bibliography	23
A	Bases de datos	25
A.1	Encuesta de satisfacción: SUSALUD	25
A.2	Cancer	25
A.3	Estudio de notificación de parejas	26
A.4	Estudio de recaídas en drogas	26

Chapter 1

Introducción

1. Leer el artículo de Bray et al. [2012], que puede ser obtenido gratis en pubmed (<http://www.ncbi.nlm.nih.gov/pubmed/22930645>) y conteste las siguientes preguntas:
 - (a) ¿Cuál es la variable principal de estudio?
 - (b) ¿Cuál fue el periodo de estudio?
 - (c) ¿Qué tipo de censura se observó y cuáles fueron las razones principales de censura?
 - (d) ¿Cuál era el objetivo del estudio?
2. Leer el artículo de Giovannetti et al. [2011] y responder las siguientes preguntas:
 - (a) ¿Cuál es la variable principal de estudio?. Defina el evento de inicio y el evento de termino de la variable de interes.
 - (b) ¿Cuál fue el periodo de estudio?
 - (c) ¿Cuáles son las variables independientes?
 - (d) Si hubo datos censurados, ¿qué tipo de censura se observó y cuáles fueron las razones principales de censura?

Las siguientes preguntas estan en base a el cuestionario 1 correspondiente a la Encuesta Nacional de Satisfacción de Usuarios en Salud - ENSUSALUD (vea los anexos por orientación en esto).

3. Sea T el tiempo que le toma a un paciente llegar al centro de salud (en minutos). Esta variable se puede construir, usando codigo de **R**, de la siguiente manera
`as.numeric(salud$C1P8_H)*60 + as.numeric(salud$C1P8_M)`
 - (a) Asumiendo que T sigue un modelo Exponencial, encuentre la función de supervivencia T

- (b) Asumiendo que T sigue un modelo Weibull, encuentre la función de supervivencia de T
 - (c) Construya el estimador no paramétrico de la función de supervivencia de T
4. Repita el problema anterior con la variable T = tiempo de duración de la atención (C1P14)
5. Supongamos que el tiempo de duración de la atención tiene una distribución **Log logística** con parámetros α (escala) y β (forma). Este modelo tiene una función de supervivencia dada por

$$S(t) = 1 - \frac{1}{1 + \left(\frac{t}{\alpha}\right)^{-\beta}}, \quad \alpha > 0, \quad \beta > 0$$

- (a) Encuentre la función de supervivencia estimada del tiempo de duración de la atención
 - (b) Estime la probabilidad que la atención de una persona dure mas de 10 minutos
 - (c) Estime la probabilidad que la atención de una persona dure entre 10 y 13 minutos
6. Compare los modelos construidos en las preguntas anteriores (5 y 6) y discuta cual de ellos explica mejor los datos.
7. Se desea estudiar si tiempo de duración de la atención depende del sistema de salud en el que una persona se atiende. La variable INSTITUCION de la base de datos tiene la siguiente estructura

```
> attr( "salud$INSTITUCION", "labels" )
      MINSA – GR          ESSALUD FF.AA. Y PNP          CLINICAS
           1              2              3              4
```

- (a) Construya la función de supervivencia, del tiempo de duración de la atención, asumiendo un modelo Weibull para los pacientes que se atienden en el MINSA, ESSALUD, FF.AA y PNP y Clinicas
 - (b) En base a el resultado anterior, ¿cree usted que existe evidencia para afirmar que la duración de la atención de un paciente depende del sistema de salud que usa ? Justifique
8. Consideremos el tiempo de duración de la atención, T se comporta como un modelo log logístico. Estamos interesados en saber si T esta asociado al idioma que habla el paciente:

```
> attr( "salud$C1P80", "labels" )
      Espanol  Quechua      Otro
           1         2         3
```

Desarrolle lo siguiente:

- (a) Encuentre la función de riesgo instantáneo de que termine la atención para cada grupo en base al idioma.
- (b) Interprete el comportamiento de estas curvas
- (c) Discuta en base a la pregunta anterior si existe evidencia (o no) que el tiempo de duración de la atención depende o está asociado a el idioma que habla el paciente.

Chapter 2

Kaplan-Meier

1. La base de datos **jasa** en la librería **survival** en R, contiene los datos de 103 pacientes que formaron parte del programa de trasplante de corazón de la Universidad de Stanford. El estudio trató de responder si el tiempo de vida de un paciente se extendía al recibir un trasplante de estos pacientes y las siguientes variables serán consideradas para su análisis.

- **futime**: Tiempo hasta muerte o censura
- **fustat**: Muerte (1) o no (0)
- **transplant**: Indicador de trasplante (1) o no

Responda las siguientes preguntas

- (a) Estime la probabilidad sobrevivir al menos 500 para estos pacientes. Calcule su intervalo de confianza al 95% e interprete sus resultados.
 - (b) Estime e interprete el riesgo acumulado de fallecer en los primeros 51 días.
 - (c) Estime e interprete el percentil 25, 50 (mediana) y 75 del tiempo de vida y sus intervalos de confianza asociados.
 - (d) Estime e interprete la media del tiempo de vida.
 - (e) ¿Existe evidencia que el tiempo de vida es más prolongado para pacientes que recibieron trasplante de corazón?. Sí/No, ¿por qué?
2. La base de datos **cancer** en la library **survival** en R describe la información de 228 pacientes con cáncer de pulmón avanzado. Las siguientes variables fueron medidas y fueron consideradas para el análisis

- **time**: Tiempo hasta muerte o censura
- **status**: Indicador de muerte o censura
- **sex**: Sexo del paciente (1=hombre, 2=mujer)

- (a) Estime la probabilidad sobrevivir al menos 500 para estos pacientes. Calcule su intervalo de confianza al 95% e interprete sus resultados.
 - (b) Estime e interprete el riesgo acumulado de fallecer en los primeros 51 días.
 - (c) Estime e interprete el percentil 25, 50 (mediana) y 75 del tiempo de vida y sus intervalos de confianza asociados.
 - (d) Estime e interprete la media del tiempo de vida y su intervalo de confianza asociado.
 - (e) ¿Existe evidencia que el tiempo de vida es más prolongado para pacientes que recibieron transplante de corazón?. Sí/No, ¿por qué?
3. Probar que, en el caso que no se observan datos censurados, el estimador de Kaplan-Meier, \hat{S}_{KM} , se reduce al estimador de la función de supervivencia empírico

$$\hat{S}_n(t) = 1 - \hat{F}_n(t) = 1 - \frac{1}{n} \sum_{i=1}^n I(T_i \leq t)$$

4. **Media esperada residual:** La función del valor esperado residual está definida por

$$r(t) = E[T - t \mid T \geq t], \quad t > 0$$

Su estimador, en el caso de datos censurados por la derecha, está dado por

$$r_n(t) = \frac{\sum_{i=1}^n (t_i - t) I(t_i > t) w_i}{\sum_{i=1}^n I(t_i > t) w_i}$$

donde $w_i = \hat{F}_n^{KM}(t_i) - \hat{F}_n^{KM}(t_i^-)$ y \hat{F}_n es el estimador de Kaplan-Meier.

Usando la base **jasa** (librería **survival** en R), responda lo siguiente

- (a) Grafique la función de valor esperado residual para aquellos pacientes que recibieron transplante
- (b) Grafique la función de valor esperado residual para aquellos pacientes que no recibieron transplante
- (c) De una breve explicación las diferencias y/o similitudes entre ambos grupos de pacientes.

5. Simulación

- (a) Variable T : Genere 300 datos de una distribución Weibull con parámetros $(\alpha, \beta) = (0.5, 0.5)$ (**Hint:** `rweibull`)
- (b) Variable Y : Genere 300 tiempos de una distribución uniforme en $(0, 5)$ (**Hint:** `runif`)

- (c) Genere $X = \min(T, Y)$, es decir tiempos sujetos a censura por la derecha (**Hint:** pmin)
 - (d) Genere $\Delta = I(T \leq Y)$, es decir el indicador de censura.
 - (e) Para los datos $(X_1, \Delta_1), \dots, (X_n, \Delta_n)$ generados, calcule el estimador de Kaplan y Meier (\hat{S}^{KM})
 - (f) Para los datos $(X_1, \Delta_1), \dots, (X_n, \Delta_n)$ generados, calcule la función de supervivencia asumiendo que la variable T tiene una distribución exponencial.
 - (g) De los dos estimadores calculados, ¿Cuál es el mejor estimador? ¿Por qué?
6. **Inferencia bayesiana***: Considere las variables **time** y **status** de la base de datos **cancer** de la librería survival en R. Supongamos que el tiempo de vida, T , sigue un modelo exponencial (es decir $T \sim Exponencial(\lambda)$).
- (a) Enfoque frecuentista: Construya la función de verosimilitud y calcule el estimador de λ y su intervalo de confianza al 95%.
 - (b) Enfoque bayesiano: Asuma una distribución a priori para λ : $\lambda \sim Gamma(\gamma, \eta)$. Compruebe que la distribución posterior esta dada por

$$(\lambda \mid Data) \sim Gamma\left(\sum_i^n \delta_i + \gamma, \sum_{i=1}^n w_i + 1/\eta\right)$$
 - (c) Asuma que $(\gamma, \eta) = (1, 1)$ y use las variables **time** y **status** para calcular $E[\lambda \mid Data]$ y un intervalo de credibilidad para λ del 95%.
 - (d) Describa brevemente las similitudes y diferencias de sus resultados en (a) y (c).
7. **Censura por intervalo tipo I***: Sean T y Y variables aleatorias no negativas con funciones de distribución F (función de densidad f) y G (función de densidad g), respectivamente. Si T y Y son independientes, construya la función de densidad $(Y, 1_{[T \leq Y]})$.
8. **Censura por intervalo tipo I***: Supongamos que obtenemos una muestra aleatoria $(Y_1, \Delta_1), \dots, (Y_n, \Delta_n)$ de censura por intervalo tipo I (es decir $\Delta = I(T \leq Y)$). Suponga que T sigue un model exponencial con parametro λ
- (a) Construya la función de verosimilitud.
 - (b) ¿Cuál es la función de score para λ y su distribución?
 - (c) ¿Cuál es el estimador de máxima verosimilitud de λ y presente su distribución asintótica?.
 - (d) ¿Cuál es el estadístico del cociente de funciones de verosimilitud para el problema de prueba de hipótesis $H_0 : \lambda = \lambda_0$ vs. $H_0 : \lambda = \lambda_0$?

- (e) La base de datos **pns.csv** corresponde al "Estudio de notificación de parejas" descrito por Sal y Rosas and Hughes [2011] y Golden et al. [2005]. La variable tiempo es tiempo a reinfección de una gonorrhea y/o clamidia y esta base de datos tiene una estructura de censura por intervalo tipo I. Asuma que T sigue un model exponencial con parametro λ e implemente (a)-(d) para este conjunto de datos
9. Los tiempos de supervivencia para dos tratamientos A y B estan dados por 4, 5+, 6 y 10+ (tratamiento A) y 6, 2 y 6 (tratamiento B)
- (a) Estime la función de Kaplan-Meier para el tratamiento A y grafíquelo
 - (b) Construya el intervalod de confianza puntual asociado a la función de supervivencia contruida en a)
 - (c) Estime la mediana del tiempo de vida para cada grupo

Chapter 3

Prueba de hipotesis

1. Considere la base de datos **cancer** (de la libreria survival) y responda las siguientes preguntas.

- (a) Grafique la función de supervivencia para hombres y mujeres
- (b) ¿Existe evidencia que el tiempo de vida en pacientes con cancer esta asociado con el sexo del paciente? Justifique usando la prueba de log rank

2. Usando la base de datos **cancer** (de la libreria survival) la siguiente es una salida en R para estudiar la relación entre edad (en el momento del diagnostico) y cancer:

```
> cancer$cage <- cut(cancer$age, breaks=quantile(cancer$age),
+ include.lowest = TRUE)
> survdiff(Surv(time, status)~cage, data=cancer)
Call:
survdiff(formula = Surv(time, status) ~ cage, data = cancer)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
cage=[39,56]	58	39	44.3	0.6353	0.873
cage=(56,63]	59	41	44.5	0.2689	0.370
cage=(63,69]	55	39	40.8	0.0795	0.106
cage=(69,82]	56	46	35.4	3.1492	4.049

Chisq= 4.2 on 3 degrees of freedom, p= 0.244

- (a) Interprete el valor de p
 - (b) ¿Existe evidencia que el tiempo de vida en pacientes con cancer esta asociado con la edad del paciente?
3. En la base de datos **cancer** se midio la variable **ph.ecog** que es una escala (0=bueno a 5=malo) para medir de manera práctica la calidad de vida de un paciente. La salida siguiente es parte de la prueba de log rank implementada en R:

```
> survdiff(Surv(time, status) ~ ph.ecog, data=cancer)
Call:
survdiff(formula = Surv(time, status) ~ ph.ecog, data = cancer)
```

n=227, 1 observation deleted due to missingness.

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
ph.ecog=0	63	37	54.153	5.4331	8.2119
ph.ecog=1	113	82	83.528	0.0279	0.0573
ph.ecog=2	50	44	26.147	12.1893	14.6491
ph.ecog=3	1	1	0.172	3.9733	4.0040

Chisq= 22 on 3 degrees of freedom, p= 6.64e-05

- Interprete el número 26.147 de la salida del software
 - Interprete el número 14.6491 de la salida del software
 - ¿Existe evidencia que el tiempo de vida en pacientes con cancer esta asociado con el score ECOG de performance? Justifique
4. Possolli et al. [2015] publicaron recientemente un artículo (<http://jped.elsevier.es/en/hiv-testing-in-maternity-ward/articulo/S0021755715000613/>) en el cual evaluaron que factores podrian estar asociados con el inicio (o no) de la lactancia maternal. Mediante una lectura de las secciones de métodos y resultados, conteste las siguientes preguntas
- ¿Cuál fue el evento de interes?
 - ¿Cómo se define la variable tiempo?
 - Describa los datos censurados (si los hubiera)
 - La prueba estadística de Peto es una prueba de log rank ponderada (usando $\rho = 1$). ¿Considera que este artículo hace una adecuada interpretaci3n de esta prueba estadística ? Sí/No, por qué.
 - En el parrafo 12 de la secci3n de métodos los autores proponen dividir la data en dos grupos. ¿Está usted de acuerdo con esta idea? Sí/No, ¿por qué?
5. Se desea estudiar si el tiempo de vida de pacientes con gliomas malignos recurrentes y si este depende del tipo de tumor que el paciente tiene (Astrocitoma o Glioblastoma multiforme). Los tiempos de vida para cada grupo son los siguientes fueron publicados por Bland and Altman [2004] y se presentan a continuaci3n

Grupo A: 6, 13, 21, 30, 31⁺, 37, 38, 47⁺, 49, 50, 63, 79, 80⁺, 82⁺, 82⁺, 86, 96, 149⁺, 202, 219

Grupo G: 0, 10, 12, 13, 14, 15, 16, 17, 18, 20, 24, 24, 25, 28, 30, 33, 34⁺, 35, 37, 40, 40, 40⁺, 46, 48, 70⁺, 76, 81, 82, 91, 112, 181.

Responda las siguientes preguntas, sin usar software

- (a) Construya el estimador de Kaplan-Meier de la función de supervivencia para cada grupo
- (b) Responda si el tiempo de vida es diferente (o no) entre estos dos grupos de pacientes.

Chapter 4

Modelos paramétricos

1. Considere la base *cancer* (ver A.2) describe la información de 228 pacientes con cáncer de pulmón avanzado. Asumiendo un modelo de regresión exponencial donde las variables de explicación son sexo y edad, responda las siguientes preguntas
 - (a) Para los datos del estudio, podríamos afirmar que mujeres están en mayor riesgo de muerte ? Justifique
 - (b) Grafique la función de supervivencia para hombres y para mujeres controlando por la variable edad (Ayuda: Use la media de edades como referencia)
 - (c) Para los datos del estudio, podríamos afirmar que la edad (al momento de diagnóstico) está asociada con un mayor riesgo de muerte ? Justifique
 - (d) Grafique la relación entre la variable edad y el tiempo de vida de los pacientes, controlando por el sexo de estos (Ayuda: Estratifique por sexo del paciente, si es posible en un solo gráfico).
 - (e) Estime e interprete el riesgo acumulado de fallecer en los primeros 51 días para un hombre y para una mujer, ambos de 63 años
2. Repita la pregunta 1 considerando el modelo Weibull
3. Repita la pregunta 1, considerando el modelo Log logístico
4. En base a las tres preguntas anteriores
 - (a) Use el estadístico AIC para evaluar cual de los tres modelos ajusta mejor los datos
 - (b) ¿Podría usar el cociente de funciones de verosimilitud para comparar los modelos ?
 - (c) Gráficamente evalúe cual de los tres modelos ajusta mejor los datos.
5. La base de datos *uis* describe la información de pacientes con problemas de adicción a drogas (ver A.4). Asumiendo un modelo de regresión Weibull donde las variables de explicación son tratamiento, raza y edad, responda las siguientes preguntas

- (a) Contronlando por raza y edad, calcule el cociente de riesgos entre personas que recibieron el tratamiento largo vs. aquellas que recibieron el tratamiento corto
- (b) ¿Podríamos afirmar que mujeres estan en mayor riesgo de muerte ? Si su respuesta es si, podemos cuantificar este efecto ?
- (c) Grafique la función de supervivencia para ambos tratamientos controlando por edad y raza de la persona
- (d) Para los datos del estudio, podríamos afirmar que la raza esta asociada con un mayor riesgo de recaída en drogas ? Justifique
- (e) Para los datos del estudio, podríamos afirmar que la edad esta asociada con un mayor riesgo de recaída en drogas ? Justifique
- (f) Considera que un modelo Exponencial describiria suficientemente bien los datos, en comparación con el modelo Weibull. Justifique

6. **Censura por intervalo tipo I***: La base de datos **pns.csv** corresponde al "Estudio de notificación de parejas" descrito por Sal y Rosas and Hughes [2011] y Golden et al. [2005]. La variable tiempo, T , es el tiempo a reinfección de una gonorrea y/o clamidia y la variable arm, Z , es una variable dicotomica que registra si el paciente se asigno al grupo placebo o al grupo de la intervención. Asumiendo que

$$T \mid Z \sim Exponencial(e^{\beta_0 + \beta_1 Z})$$

- (a) Construya la función de verosimilitud.
- (b) Estime los parámetros del modelo
- (c) ¿Cuál es la función de supervivencia estimada para el grupo control e intervención?.
- (d) ¿Podríamos afirma que el riesgo de re-infección es mas alto en el grupo control que en el de la intervención? Justifique

Chapter 5

Modelo de Cox

5.1 Interpretación

1. Considere la base de datos *cancer* (see A.2) que describe la información de 228 pacientes con cáncer de pulmón avanzado. Asumiendo un modelo de riesgos proporcionales considerando las variables *sex*, *age* and *ph.ecog* en el modelo
 - (a) Los datos del estudio ofrecen evidencia que las mujeres estan en mayor riesgo de muerte ? Justifique
 - (b) Grafique la función de supervivencia para hombres y para mujeres controlando por la variable edad
 - (c) Los datos del estudio ofrecen evidencia para afirmar que la edad (al momento de diagnostico) esta asociada con un mayor riesgo de muerte ? Justifique
 - (d) Estime e interprete el riesgo acumulado de fallecer en los primeros 51 días para un hombre y para una mujer, ambos de 63 años
2. Considerando la base de datos *uis* (ver A.4) que describe un estudio aleatorizado de pacientes con riesgo de recaer en drogas. Asumiendo un modelo de riesgos proporcionales y considerando las covariables *race*, *age*, *treat* y *ndrugtx*
 - (a) ¿Existe evidencia para afirmar que el tratamiento largo redujo el riesgo de recaer en drogas?
 - (b) Grafique la función de supervivencia para ambos tratamientos controlando por edad, raza y número de tratamientos previos
 - (c) Compare el riesgo de recaer en drogas para una persona que recibo un tratamiento largo con 30 años de edad vs. una persona que recibio un tratamiento corto de 25 años de edad?. Calcule el intervalo de confianza de su estimación
 - (d) ¿ Podemos excluir la variable *race* del modelo? Justifique

5.2 Construcción

1. Considere que el tiempo a la ocurrencia de un evento, T , sigue el modelo de Cox con función de riesgo basal igual a 1. Es decir

$$\lambda(t) = \exp(0.1X_1 - 0.5X_2 + 0.4X_3 + 0X_4 + \dots 0X_{10})$$

donde $X_1 \sim \text{Bernoulli}(0.5)$, $X_2 \sim \text{Normal}(1.7, 0.1^2)$, $X_3 \sim \text{Uniforme}(0, 1)$ y $X_k \sim \text{Bernoulli}(0.5)$ para $k = 4, \dots, 10$ son covariables que se miden de la persona y que afectan el tiempo T .

Sea $Y \sim U(0, 5)$ el tiempo a censura, entonces la data observada en el caso de censura por la derecha es de la forma

$$(T \wedge Y, \Delta = I(T \leq Y), X_1, X_2, X_3, X_4)$$

- (a) Simule 1000 bases de datos de tamaño 2000 ($n = 1000$) con características descritas arriba.
 - (b) Aplique el método de stepwise para encontrar mejor modelo que describe los datos y calcule la proporción de veces que cada variable es seleccionada
 - (c) Aplique el método de lasso para encontrar mejor modelo que describe los datos y calcule la proporción de veces que cada variable es seleccionada
 - (d) Discuta sus resultados de b) y c)
2. Revise el artículo de Possolli et al. [2015] que estudia los factores asociados con el tiempo al inicio de lactancia en mujeres brasileñas. Discuta las limitaciones del análisis.

Chapter 6

Censura por intervalo

1. Doubly censored Sea T_1, T_2, \dots, T_n una muestra aleatoria con función de distribución F_T y sea $(V_1, \Delta_1), \dots, (V_n, \Delta_n)$ donde

$$V_i = \begin{cases} T_i & D_i < T_i \leq C_i & \Delta_i = 1 \\ C_i & T_i > C_i & \Delta_i = 2 \\ D_i & T_i \leq D_i & \Delta_i = 3 \end{cases}$$

donde (C_i, D_i) es independiente de T_i . Deriva la función de verosimilitud para F_T

Bibliography

- J. Bland and D. Altman. The logrank test. *BMJ: British Medical Journal*, 328(7447):1073, 2004.
- S. Bray, J. Gedeon, A. Hadi, and et. al. Predictive value of cd4 cell count nadir on long-term mortality in HIV-positive patients in Uganda. *HIV AIDS (Auckl)*, 4:135–140, 2012.
- G. Giovannetti, G. Ricchiuti, and M. Velucchi. Size, innovation and internationalization: a survival analysis of Italian firms. *Applied Economics*, 43(12):1511–20, 2011.
- M. Golden, W. Whittington, H. Handsfield, J. Hughes, W. Stamm, M. Hogben, A. Clark, C. Malinski, J. Helmers, K. Thomas, and K. Holmes. Effect of expedited treatment of sex partners on recurrent or persistent gonorrhea or chlamydial infection. *New England Journal of Medicine*, 352(7):676–685, 2005.
- G. Possolli, M. De Carvalho, and M. De Oliveira. HIV testing in the maternity ward and the start of breastfeeding a survival analysis. *J Pediatr (Rio J)*, 91(1):397–404, 2015.
- V. Sal y Rosas and J. Hughes. Nonparametric and semiparametric analysis of current status data subject to outcome misclassification. *Statistical Communications in Infectious Diseases*, 1(3), 2011.

Appendix A

Bases de datos

A.1 Encuesta de satisfacción: SUSALUD

Las encuestas nacionales de satisfacción en el sistema de salud se encuentran disponible en esta pagina web <http://iinei.inei.gob.pe/microdatos/> en formato .sav (archivo SPSS)

Nuestro enfoque sera en el capítulo 1, en la parte dirigida a usuarios de consulta externa (Código Módulo 550)

A.2 Cancer

La base de datos **cancer** en la library **survival** describe la información de 228 pacientes con cáncer de pulmón avanzado. Las siguientes variables fueron medidas y fueron consideradas para el análisis

- *inst*: Código de la institución donde de enrolo al paciente
- *time*: Tiempo hasta muerte o censura
- *status*: Indicador de muerte o censura
- *age*: Edad del paciente
- *sex*: Sexo del paciente (1=hombre, 2=mujer)
- *ph.ecog*: Score ECOG que mide performance del paciente. Es una escala que mide el progreso de la enfermedad y mide como esta afecta las actividades diarias de la persona (0=bueno,5=muerte)
- *ph.karno*: Score de Karnofsky. Es una escala de calidad de vida medida por el medico (0 = malo y 100 = bueno)
- *pat.karno*: Score de Karnofsky. Escala de calidad de vida medida por el paciente (0= malo y 100 = bueno)

- *meal.cal*: Calorias consumidas en la comida
- *wt.loss*: Peso perdido en los ultimos seis meses.

A.3 Estudio de notificación de parejas

A.4 Estudio de recaídas en drogas

La base de datos **uisurv** disponible en ftp://ftp.wiley.com/public/sci_tech_med/survival/ describe la información pacientes con problemas de adicción a drogas. Las siguientes variables fueron medidas y fueron consideradas para el análisis

- *time*: Tiempo hasta recaída en drogas (o no)
- *sensor*: Indicador de recaída en drogas (1) o censura (0)
- *race*: Raza del paciente (0=blanco, 1=otro)
- *age*: Edad del paciente en el momento de diagnostico
- *treat*: Tratamiento corto (0) o largo (1)