Estratificación y covariables que cambian en el tiempo

Giancarlo Sal y Rosas

Departmento de Ciencias
Pontificia Universidad Católica del Perú

May 31, 2017

Outline

Caso 1: Estudio HPTN 039

- **Hipótesis**: El tratamiento contra VHS-2 (virus del herpes typo 2) reduce el riesgo de infección por VIH-1
- **Población**: mujeres y hombres que tienen sexo con otros hombres VIH-1 negativas y VHS-2 positivas en alto riesgo de adquirir VIH
- **Grupos**: Participantes fueron aleatorizados a:
  - **Intervención**: Aciclovir un medicamente muy efectivo para tratar VHS-2 y sin efectos adversos conocidos
  - **Control**: Un placebo que lucia fisicamente similar al aciclovir

Caso 1: Estudio HPTN 039

- **Localización**: El estudio incluyo 4 ciudades de EEUU, 3 de Peru y 6 de Africa.
- **Visitas**: El periodo de seguimiento se dio cada 3 meses y duro entre 12 a 18 meses.
- **Respuesta**: Tiempo hasta la acquisición de VIH.
- **Objetivo científico**: El tratamiento con aciclovir reduce el riesgo de infección de VIH.
- **Objetivo estadístico**: La función de supervivencia es diferente (mayor) para el grupo que tomo aciclovir que para el que no lo tomo.

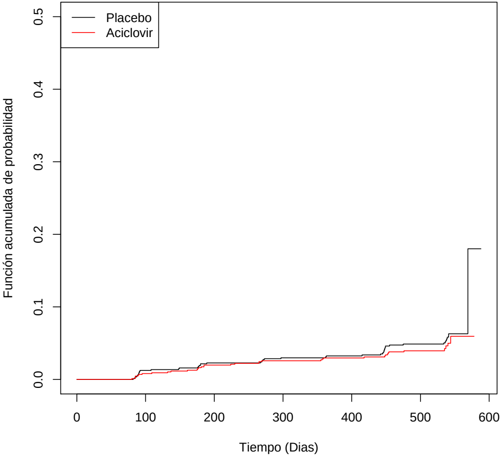
Estructura de la data

```
> hptn[1:5,c(2:3,5:13)]
       ptid  t1 age circum qsmp1isx qsmp1rsx qsmp2isx qsmp2rsx qsmp3isx qsmp3rsx site
2 203000045   0  60      1        1        0        1        0        1        0   NY
3 203000045  90  60      1        2        0        1        0       10        0   NY
4 203000045 175  60      1        8        0        0        0        0        0   NY
5 203000045 266  60      1        8        0        1        0        2        0   NY
7 203000053   0  39      0        1        0       35        0        1        0   NY

> table(hptn$site)

    Lima        NY Pucallpa  Seattle        SF
    5545       166     1558     1252       818
```

donde

- *qsmp1isx*: Número de actos insertivos con la primera pareja descrita
- *qsmp3rsx*: Número de actos receptivos con la tercera pareja descrita
- *site*: Ciudad que participo en el estudio

Estudio HPTN 039



Modelo

- Se propone un análisis estratificado por una variable S si queremos ajustar por esta, pero es de menor importancia que el resto de covariables Z
  - S ya se estudio suficiente en investigaciones previas
  - S puede ser fija por diseño (ciudad o país).
- El modelo de riesgos proporcionales para el estrato $S = s$ es
$$\lambda_s(t \mid X, \beta) = \lambda_{s0}(t) \exp\left[Z^t \beta\right]$$
- Note que se asume que β es constante a lo largo de todos los estratos

Modelo

Para el estrato $S = s$, la función parcial de verosimilitud es

$$L_{n_s}(\beta) = \prod_{i=1}^{n_s} \left[\frac{\exp\left[z_{si}^t \beta\right]}{\sum_{j \in R(t_{si})} \exp\left[z_{sj}^t \beta\right]} \right]^{\delta_{si}}$$

donde

- n_s es el número de observaciones en el estrato $S = s$
- t_{si} es el tiempo de falla *i-ésimo* en el estrato $S = s$
- $R(t_{si})$ es el conjunto de observaciones en riesgo hasta antes del tiempo t_{si}

Modelo

- La función parcial de verosimilitud para todos los datos esta dada por

$$L_n(\beta) = L_{n_1}(\beta) \times \cdots \times L_{n_S}(\beta) = \prod_{s=1}^{S} L_s(\beta)$$

- Note que cada estrato construye sus propios conjuntos de riesgo ($R(t_{si})$) y esa es la forma como ajustamos por la variable S.
- $\hat{\beta}$ es el estimador de máxima verosimilitud y se calculan de la misma forma como en el caso sin estratificación.

HPTN 039

- No estratificación

```
> model1 <- coxph(Surv(t1,t2,pos)~ arm,data=hptn)
> summary(model1)

  n= 9339, number of events= 85

       coef exp(coef) se(coef)      z Pr(>|z|)
arm -0.1835    0.8323   0.2182 -0.841      0.4

    exp(coef) exp(-coef) lower .95 upper .95
arm    0.8323      1.201    0.5427     1.277
```

- Estratificación

```
> model2 <- coxph(Surv(t1,t2,pos)~ arm + strata(site),data=hptn)
> summary(model2)

  n= 9339, number of events= 85

       coef exp(coef) se(coef)      z Pr(>|z|)
arm -0.1829    0.8329   0.2184 -0.837    0.402

    exp(coef) exp(-coef) lower .95 upper .95
arm    0.8329      1.201    0.5428     1.278
```

HPTN 039

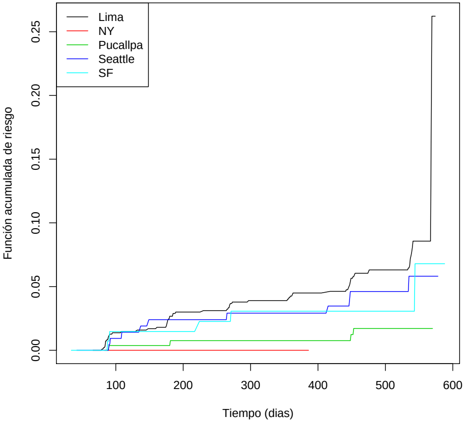


Figure : Funciones basales de acuerdo al site

Estructura de la data

```
> hptn[1:10,c(2:3,5:13)]
        ptid  t1 age circum qsmp1isx qsmp1rsx qsmp2isx qsmp2rsx qsmp3isx qsmp3rsx site
2  203000045   0  60      1        1        0        1        0        1        0   NY
3  203000045  90  60      1        2        0        1        0       10        0   NY
4  203000045 175  60      1        8        0        0        0        0        0   NY
5  203000045 266  60      1        8        0        1        0        2        0   NY
7  203000053   0  39      0        1        0       35        0        1        0   NY
8  203000053  86  39      0       20        1        4        1        1        1   NY
9  203000053 175  39      0       25        0        2        0        0        0   NY
10 203000053 265  39      0       10        0        1        0        2        0   NY
12 203000061   0  38      1        0        2        0        2        1        3   NY
13 203000061  88  38      1        0        2        0        0        0        0   NY
```

donde

- *qsmp1isx*: Número de actos insertivos con la primera pareja descrita
- *qsmp3rsx*: Número de actos receptivos con la tercera pareja descrita

son variables que cambian en el tiempo

Motivación

- Al estudiar un modelo de regresión, algunos actores podrian aumentar su correlación con el evento a medida que estos sean medidos en un tiempo cercano a su ocurrencia.
- Desde un punto de vista conceptual, implementar un modelo de este tipo parece muy dificil.
- Debemos tener cuidado que el valor de la covariable en mención dependaria del tiempo definido en el estudio y no del tiempo cronológico

Modelo

- $x(t)$ es el valor de la covariable medido en el tiempo t
- $x_l(t_i)$ es el valor de la covariable para el sujeto l-ésimo medido en el tiempo t_i
- El vector de covariables para el sujeto l-ésimo en el tiempo t_i es

$$x_l(t_i) = (x_{l1}(t_i), x_{l2}(t_i), \ldots, x_{lp}(t_i))$$

- Si una covariable es fija en el tiempo

$$x_{lk}(t_i) = x_{lk}(t = 0) = x_{lk}$$

Modelo

- El modelo de Cox se define (de manera general)

$$\lambda(t, X^H(t), \beta) = \lambda_0(t) \exp\left[Z(t)\beta\right]$$

  donde:

  - $X^H(t)$ es la historia de la variable X hasta el tiempo t y su efecto se captura con $Z(t)$
  - Asumimos que el efecto de las covariables es constante en el tiempo

- La generalización de la función de verosimilitud parcial es

$$l_p(\beta) = \prod_{i=1}^{n} \left\{ \frac{e^{Z_i(t_i)\beta}}{\sum\limits_{l \in R(t_{(i)})} e^{Z_l(t_i)\beta}} \right\}^{\delta_i}$$

HPTN 039: Circumcisión y VIH

- El modelo considerado es

$$\lambda(t \mid X) = \lambda_0(t) e^{\beta_1 Z_1 + \beta_2 Z_2(t) + \beta_3 Z_1 \times Z_2(t)}$$

donde

- Z_1 es una variable indicador si la persona es circuncidada o no

$$Z_1 = \begin{cases} 1 & , \quad \textit{circuncidado} \\ 0 & , \quad \textit{no circuncidado} \end{cases}$$

- $Z_2(t)$ es el rol que el participante tuvo con sus parejas en el tiempo t

$$Z_2(t) = \frac{\#\textit{actos insertivos en el tiempo } t}{\#\textit{total de actos en el tiempo } t}$$

HPTN 039: Estructura de los datos

```
> hptn[1:25,c("ptid","t1","t2","tii","ta","pia","circum")]
        ptid  t1  t2 tii ta       pia circum
2  203000045   0  90   3  3 1.0000000      1
3  203000045  90 175  13 13 1.0000000      1
4  203000045 175 266   8  8 1.0000000      1
5  203000045 266 357  11 11 1.0000000      1
7  203000053   0  86  37 37 1.0000000      0
8  203000053  86 175  25 28 0.8928571      0
9  203000053 175 265  27 27 1.0000000      0
10 203000053 265 358  13 13 1.0000000      0
12 203000061   0  88   1  8 0.1250000      1
13 203000061  88 182   0  2 0.0000000      1
14 203000061 182 265   1  5 0.2000000      1
15 203000061 265 358   1  4 0.2500000      1
17 203000086   0  86  11 11 1.0000000      1
18 203000086  86 182   6  6 1.0000000      1
19 203000086 182 266  12 12 1.0000000      1
20 203000086 266 363  11 11 1.0000000      1
22 203000127   0  86   7 13 0.5384615      1
23 203000127  86 175   5 12 0.4166667      1
24 203000127 175 266   0  3 0.0000000      1
25 203000127 266 359   0  2 0.0000000      1
27 203000130   0  86   0  3 0.0000000      0
28 203000130  86 175   2  5 0.4000000      0
29 203000130 175 273   2  2 1.0000000      0
30 203000130 273 358   7  7 1.0000000      0
32 203000143   0  90   2  9 0.2222222      1
```

HPTN 039: Circuncisión y VIH

- El cociente de riesgos instantaneos de acquirir VIH de una persona *circuncidada* vs. una *no circuncidada* es
  - Si tiene el "rol" insertivo todo el tiempo ($Z_2(t) = 1$)

$$HR = \frac{e^{\beta_1+\beta_2+\beta_3}}{e^{\beta_2}} = e^{\beta_1+\beta_3}$$

  - Si tiene el "rol" insertivo casi todo el tiempo ($Z_2(t) = 0.8$)

$$HR = \frac{e^{\beta_1+0.8\beta_2+0.8\beta_3}}{e^{0.8\beta_2}} = e^{\beta_1+0.8\beta_3}$$

  - Si tiene el "rol" insertivo la mayoria de veces ($Z_2(t) = 0.6$)

$$HR = \frac{e^{\beta_1+0.6\beta_2+0.6\beta_3}}{e^{0.6\beta_2}} = e^{\beta_1+0.6\beta_3}$$

HPTN 039: Efecto de circuncisión

- Modelo

```
> model3 <- coxph(Surv(t1,t2,pos)~ circum*pia + strata(site),data=hptn)
> summary(model3)
Call:
coxph(formula = Surv(t1, t2, pos) ~ circum * pia + strata(site),
    data = hptn)

  n= 9327, number of events= 85
   (12 observations deleted due to missingness)

              coef exp(coef) se(coef)      z Pr(>|z|)
circum      0.3802    1.4625   0.4290  0.886    0.376
pia         0.3866    1.4719   0.3136  1.233    0.218
circum:pia -0.9095    0.4027   0.6817 -1.334    0.182

           exp(coef) exp(-coef) lower .95 upper .95
circum        1.4625     0.6837    0.6308     3.391
pia           1.4719     0.6794    0.7961     2.722
circum:pia    0.4027     2.4830    0.1059     1.532
```

- Interpretación: No existe evidencia que sugiera que circuncisión tiene un rol protector.

HPTN 039: Circuncisión y VIH

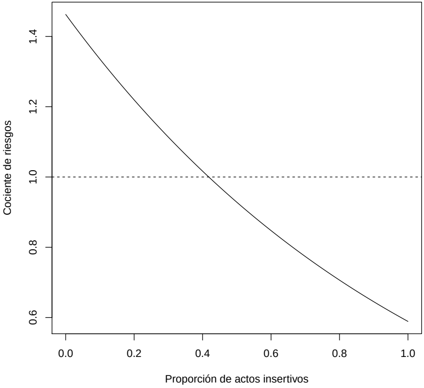


Figure : Efecto de circumcisión para diferentes niveles de comportamiento insertivo