Prueba de hipótesis

Giancarlo Sal y Rosas, Ph.D.

Departmento de Ciencias
Pontificia Universidad Católica del Perú
vsalyrosas@pucp.edu.pe

8 de abril de 2017

Outline

Outline

1. Motivación
   - HPTN 039
   - Cáncer cervical

2. Prueba de Log Rank
   - Prueba de Log Rank estratificado
   - Prueba de log rank ponderado

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 VIH-1 negativas y VHS-2 positivas en alto riesgo de adquirir VIH-1.
- **Grupos**: Participantes fueron aleatorizados a:
  - **Intervención**: Aciclovir un medicamente muy efectivo para tratar VHS-2 y sin efectos adversos conocidos
  - **Control**: Un placebo que lucía fisicamente similar al aciclovir

Estudio: HPTN 039

- **Localización**: El estudio incluyó 4 ciudades de EEUU, 3 de Perú y 6 de África (solo analizaremos data del Perú).
- **Visitas**: El periodo de seguimiento se dio cada 3 meses y duró entre 12 a 18 meses.
- **Respuesta**: Tiempo hasta la adquisición 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 tomó aciclovir que para el que no lo tomó.
- Celum et al. [1] publicaron los resultados del estudio.

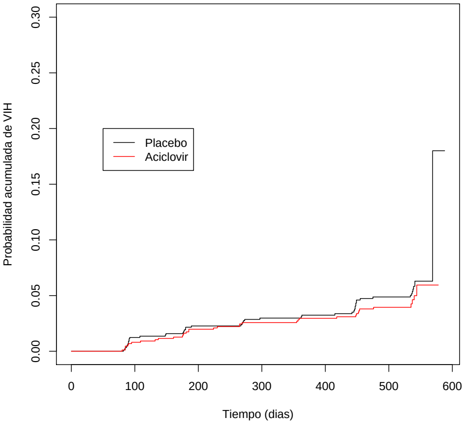

Figura 1: Estudio: HPTN 039

Outline

1. Motivación
   - HPTN 039
   - Cáncer cervical
2. Prueba de Log Rank
   - Prueba de Log Rank estratificado
   - Prueba de log rank ponderado

Estudio: Cáncer cervical

- El cáncer cervical afecta a un número importante de mujeres a nivel mundial.
- Exiten dos formas de tratamiento: Radioterapia y histerectomía (extracción total de útero)
- Aún cuando un paciente se le realiza histerectomía, esta podría necesitar una ronda de radioterapia.
- Deseamos estudiar la recurrencia o mortalidad en pacientes que se realizaron la histerectomía y aquellas que la abandonaron
- Se realizó un estudio retrospectivo en 268 mujeres en estadíos tempranos de cáncer [2]

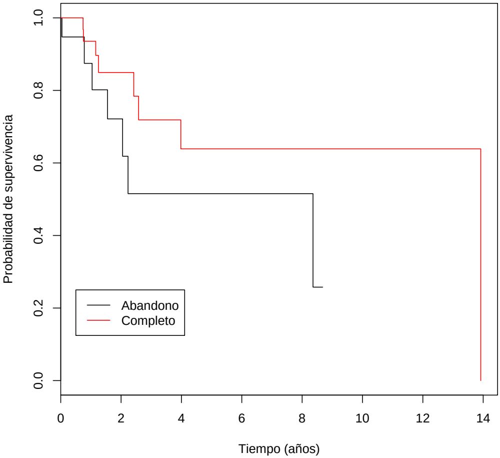

Figura 2: Probabilidad estimada de supervivencia para pacientes que abandonaron la histerectomia (negro) y aquellos que se la hicieron (rojo)

Cáncer cervical

```
> summary(survfit(Surv(time=time,event=rec)~Group,data=cancer))
Call: survfit(formula = Surv(time = time, event = rec) ~ Group, data = cancer)

                Group=Abandoned
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
 0.04     19       1    0.947  0.0512       0.8521        1.000
 0.78     13       1    0.874  0.0845       0.7236        1.000
 1.04     12       1    0.802  0.1042       0.6213        1.000
 1.55     10       1    0.721  0.1208       0.5197        1.000
 2.05      7       1    0.618  0.1408       0.3958        0.966
 2.23      6       1    0.515  0.1504       0.2909        0.913
 8.36      2       1    0.258  0.1971       0.0575        1.000

                Group=Completed
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  0.74     31       1    0.968  0.0317        0.908        1.000
  0.75     30       1    0.935  0.0441        0.853        1.000
  1.16     24       1    0.897  0.0570        0.792        1.000
  1.25     19       1    0.849  0.0709        0.721        1.000
  2.42     13       1    0.784  0.0907        0.625        0.983
  2.58     12       1    0.719  0.1040        0.541        0.954
  3.98      9       1    0.639  0.1192        0.443        0.921
 13.92      1       1    0.000     NaN           NA           NA
```

Prueba de Log rank

- Nuestor objetivo es comparar la función de supervivencia entre dos poblaciones.
- Matemáticamente deseamos estudiar la hipótesis

$$H_0 : S_0(t) = S_1(t) \ , \ \forall \ t > 0$$

que es equivalente a

$$H_0 : \lambda_0(t) = \lambda_1(t) \ , \ \forall \ t > 0$$

- Supongamos que para estudiar esta hipótesis tomamos una muestra de ambas poblaciones.

Prueba de Log rank

- Los datos observados tienen la forma

$$\begin{array}{lcl} \textit{Grupo } 0 & : & (X_1^0, \Delta_1^0), (X_2^0, \Delta_2^0), \ldots, (X_{n_0}^0, \Delta_{n_0}^0) \\ \textit{Grupo } 1 & : & (X_1^1, \Delta_1^1), (X_2^1, \Delta_2^1), \ldots, (X_{n_1}^1, \Delta_{n_1}^1) \end{array}$$

  donde n_0 y n_1 es el tamaño de muestra de cada uno de los grupos

- Sean $t_1 < t_2 < \cdots < t_k$ los tiempos (distintos) en que ocurre el evento, para cualquiera de los dos grupos.

Prueba de Log rank

- Siguiendo la lógica del estimador de Kaplan-Meier se construyen intervalos $[0, t_1), [t_1, t_2), \ldots, [t_{k-1}, t_k), \ldots$.
- En el intervalo $[t_j, t_{j+1})$ observamos el siguiente escenario

	Grupo 0	Grupo 1	Total
Evento	d_{0j}	d_{1j}	d_j
No evento	$n_{0j} - d_{0j}$	$n_{1j} - d_{1j}$	$n_j - d_j$
En riesgo	n_{0j}	n_{1j}	n_j

donde d_j y n_j son el número de eventos en el tiempo t_j y el número de personas en riesgo al inicio del intervalo, respectivamente.

Cáncer de ovario

```
> mod1 <- summary(survfit(Surv(futime,fustat)~rx,data=ovarian))
> mod1
Call: survfit(formula = Surv(futime, fustat) ~ rx, data = ovarian)

                rx=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   59     13       1    0.923  0.0739        0.789        1.000
  115     12       1    0.846  0.1001        0.671        1.000
  156     11       1    0.769  0.1169        0.571        1.000
  268     10       1    0.692  0.1280        0.482        0.995
  329      9       1    0.615  0.1349        0.400        0.946
  431      8       1    0.538  0.1383        0.326        0.891
  638      5       1    0.431  0.1467        0.221        0.840

                rx=2
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  353     13       1    0.923  0.0739        0.789        1.000
  365     12       1    0.846  0.1001        0.671        1.000
  464      9       1    0.752  0.1256        0.542        1.000
  475      8       1    0.658  0.1407        0.433        1.000
  563      7       1    0.564  0.1488        0.336        0.946
```

- Los intervalos generados: $[0, 59), [59, 115), \ldots, [563, 638), \ldots$

Cáncer de ovario

- Para el intervalo $[59, 115)$ tenemos

	Grupo 1	Grupo 2	Total
Evento	1	0	1
No evento	12	13	25
Total	13	13	26

- Para el intervalo $[115, 156)$ tenemos

	Grupo 1	Grupo 2	Total
Evento	1	0	1
No evento	11	13	24
Total	12	13	25

Prueba de Log Rank

- Supongamos que no existe diferencia entre ambos grupos (la hipótesis nula es cierta), entonces

	Grupo 1	Grupo 2	Total
Evento	?		d_j
No evento			
Total	n_{1j}		n

- Si asumimos que n, n_{1j} y d_j son fijos, se tiene

$$d_{1j} \mid n_{1j}, n_j, d_j \sim Hipergeometrica(n_{1j}, d_j, n_j)$$

Prueba de Log rank

- El número esperado de eventos en el grupo 1 será

$$E[d_{1j} \mid n_j, d_j, n_{1j}] = \frac{d_j}{n_j} n_{1j}$$

- La diferencia entre lo que observamos y esperamos (bajo el supuesto que las funciones de supervivencia son iguales) es

$$w_j = O_j - E_j = d_{1j} - \frac{d_j}{n_j} n_{1j}$$

- La diferencia global esta dada por

$$U = \sum_{j=1}^{k} w_j = O - E$$

donde k es el número de tiempos distintos donde ocurre al menos un evento (en las muestras combinadas).

Cáncer de ovario

- Para el intervalo $[59, 115)$ tenemos

	Grupo 1	Grupo 2	Total
Evento	1	0	1
No evento	12	13	25
Total	13	13	26

entonces $E_1 = (1/26)(13) = 0.5$ y $O_1 = 1$

- Para el intervalo $[115, 156)$ tenemos

	Grupo 1	Grupo 2	Total
Evento	1	0	1
No evento	11	13	24
Total	12	13	25

entonces $E_2 = (1/25)(12) = 0.48$ y $O_2 = 1$

Prueba de Log rank

- Si H_0 es cierto, entonces esperariamos que $U \approx 0$. Es decir que
$$O \approx E$$
- Si el riesgo instantáneo es menor en el grupo 1 (de manera consistente a lo largo del tiempo), entonces
$$U = O - E <\approx 0$$
- Si el riesgo instantáneo es mayor en el grupo 1 (de manera consistente a lo largo del tiempo), entonces
$$U = O - E >\approx 0$$

Prueba de Log rank

- La varianza de U esta dada por

$$Var[U] = \sum_{j=1}^{k} \left[\frac{d_j(n_j - d_j)d_{jl}(d_j - d_{jl})}{n_j^2(n_j - 1)} \right]$$

- El estadístico de Log rank esta dada por

$$X^2 = \frac{U^2}{Var[U]}$$

y tiene una distribución $\chi^2_{(1)}$ bajo la hipótesis nula.

Prueba de Log Rank

- Región de rechazo con un error tipo I de 0.05

$$\{x^2 \in [0, \infty)\ ,\ x^2 > \chi^2_{1,0.95}\} = \{x^2 \in [0, \infty)\ ,\ x^2 > 3.84\}$$

- El valor de p esta dado por

$$p = P(\chi^2_1 > x^2_{obs})$$

  donde x^2_{obs} es el estadístico de Log Rank calculado de los datos.

- Una formula aproximada para la prueba de Log Rank es

$$X^2 \approx \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Prueba de Log rank: Cáncer de ovario

```
> mod1 <- survdiff(Surv(futime,fustat)~rx,data=ovarian)
> mod1
Call:
survdiff(formula = Surv(futime, fustat) ~ rx, data = ovarian)

      N Observed Expected (O-E)^2/E (O-E)^2/V
rx=1 13        7     5.23     0.596      1.06
rx=2 13        5     6.77     0.461      1.06

 Chisq= 1.1  on 1 degrees of freedom, p= 0.303

> 1-pchisq(mod1$chisq,1)
[1] 0.3025911
```

- En el grupo con el tratamiento 1 ($rx = 1$), se han observado siete muertes. Sin embargo, si creemos que no hay diferencia entre los tratamientos, esperariamos observar cinco muertes.
- Concluimos que no hay suficiente diferencia entre los eventos observados y esperados para concluir que existe diferencia en la supervivencia entre ambos grupos.

Prueba de Log rank: Cáncer cervical

```
> survdiff(Surv(time=time,event=rec)~Group,data=cancer)
Call:
survdiff(formula = Surv(time = time, event = rec) ~ Group, data = cancer)

                 N Observed Expected (O-E)^2/E (O-E)^2/V
Group=Abandoned 19        7     4.59     1.270       1.9
Group=Completed 44        8    10.41     0.559       1.9

 Chisq= 1.9  on 1 degrees of freedom, p= 0.168
```

- **Interpretación**: No existe evidencia de que mujeres que completaron la histerectomía están a mayor o menor riesgo de tener un episodio recurrente de cáncer cervical en comparación con aquellas que no completaron el procedimiento ($p = 0.168$).

Prueba de Log rank: Cáncer cervical

```
> survdiff(Surv(time=time, event=dead)~Group, data=cancer)
Call:
survdiff(formula = Surv(time = time, event = dead) ~ Group, data = cancer)

                 N Observed Expected (O-E)^2/E (O-E)^2/V
Group=Abandoned 19       15     15.1  0.000486  0.000724
Group=Completed 44       35     34.9  0.000210  0.000724

 Chisq= 0  on 1 degrees of freedom, p= 0.979
```

- **Interpretación**: No existe evidencia de que mujeres que completaron la histerectomía tengan un mayor tiempo de vida en comparación con aquellos que no lo completaron ($p = 0.979$).

Prueba de Log Rank: Extensión

- Supongamos que estamos comparando $p+1$ poblaciones (en lugar de dos). Es decir

$$H_0 : S_0(t) = S_1(t) = \cdots = S_p(t) \ , \ \forall t \geq 0$$

o equivalentemente

$$H_0 : \lambda_0(t) = \lambda_1(t) = \cdots = \lambda_p(t) \ , \ \forall t \geq 0$$

- Entonces en el intervalo $[t_j, t_{j+1})$ tenemos

	Grupo 0	Grupo 1	...	Grupo p	Total
Evento	d_{0j}	d_{1j}	...	d_{pj}	d_j
No evento	$n_{0j} - d_{0j}$	$n_{1j} - d_{1j}$	...	$n_{pj} - d_{pj}$	$n_j - d_j$
En riesgo	n_{0j}	n_{1j}	...	n_{pj}	n_j

Prueba de Log Rank: Extensión

- El vector de número de eventos observados menos el número de eventos esperados (bajo la hipótesis nula) es

$$\begin{pmatrix} d_{1j} - (d_j/n_j)n_{1j} \\ d_{2j} - (d_j/n_j)n_{2j} \\ \vdots \\ d_{pj} - (d_j/n_j)n_{pj} \end{pmatrix}$$

- Si sumamos sobre todos los intervalos

$$U = \begin{pmatrix} \sum_{j=1}^{k} \left[d_{1j} - (d_j/n_j)n_{1j}\right] \\ \sum_{j=1}^{k} \left[d_{2j} - (d_j/n_j)n_{2j}\right] \\ \vdots \\ \sum_{j=1}^{k} \left[d_{pj} - (d_j/n_j)n_{pj}\right] \end{pmatrix}$$

Prueba de log rank: Extensión

- Note que tomamos un vector de dimensión p pues si incluimos los $p+1$ elementos sería redundante
- Tenemos que definir una matriz de varianza-covarianza de U, denotada por V de dimensión $p \times p$ con

$$
\begin{aligned}
V_{ll} &= \sum_{j=1}^{k}\left[\frac{d_j(n_j-d_j)d_{jl}(d_j-d_{jl})}{n_j^2(n_j-1)}\right] \\
V_{ls} &= \sum_{j=1}^{k}\left[\frac{d_j(n_j-d_j)d_{jl}d_{js}}{n_j^2(n_j-1)}\right]
\end{aligned}
$$

para $l \neq s$

Prueba de Log rank: Extensión

- La prueba de log rank es

$$T = U^t V^{-1} U$$

y tiene distribución $\chi^2_{(p)}$ bajo la hipótesis nula donde p es el número de grupos menos uno.

- **Nota**: Este estadístico es numericamente idéntico sin importar que p grupos escogemos.

Prueba de Log rank: HIV

- Recordemos que que el tratatamiento con aciclovir no tuvo un efecto en reducir el riesgo de adquirir VIH.
- Nos podemos plantear como análisis secundario qué factores incrementan el riesgo de acquirir el virus: ¿Edad?

```
> hiv$nage <- cut(hiv$age,breaks=c(18-1,24,30,38,max(hiv$age)))
> survdiff(Surv(time=time,event=cen)~factor(nage),data=hiv)
Call:
survdiff(formula = Surv(time = time, event = cen) ~ factor(nage),
    data = hiv)

                         N Observed Expected (O-E)^2/E (O-E)^2/V
factor(nage)=(17,24] 467       37     22.1   10.0745    13.629
factor(nage)=(24,30] 428       22     20.7    0.0793     0.105
factor(nage)=(30,38] 432       19     20.5    0.1041     0.137
factor(nage)=(38,79] 440        7     21.7    9.9926    13.440

 Chisq= 20.3  on 3 degrees of freedom, p= 0.000148
```

- Existe evidencia que el riesgo de adquirir VIH está asociado con la edad ($p < 0.001$)

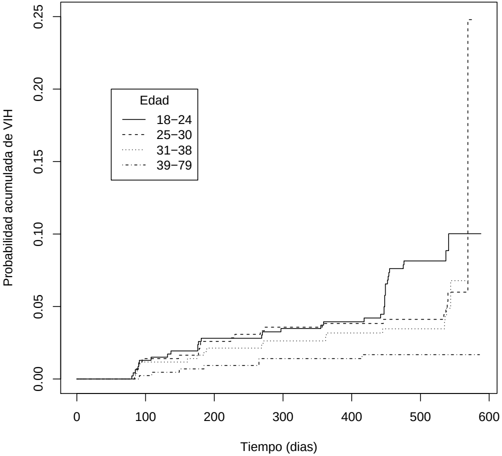

Figura 3: Estudio: VIH

Outline

Prueba de Log rank estratificado

- Supongamos que existe heterogeneidad entre las poblaciones que deseamos comparar.
- Estudio HPTN 039
  - ¿Es de esperarse que los hombres de EEUU y de Perú se comporten de manera diferente?
  - Como esta diferencia puede afectar nuestra observación de la relación entre edad y tiempo a la adquisición de VIH
- La prueba de hipótesis en este caso es

$$H_0 : S_{0j}(t) = \cdots = S_{pj}(t) \ , \ \forall t > 0 \ , \ j = 1, \ldots, s$$

donde $j = 1, \ldots, s$ denota los diferentes estratos y $p + 1$ es el número de poblaciones a comparar

Prueba de Log rank estratificado

- La prueba de Log rank en ese caso es

$$T^s = \left(\sum_{h=1}^{s} U^{(h)}\right)^t \left(\sum_{h=1}^{s} V^{(h)}\right)^{-1} \left(\sum_{h=1}^{s} U^{(h)}\right)$$

y tiene una distribución $\chi^2_{(p)}$ bajo la hipótesis nula.

- **Nota**: La prueba de log rank estratificada consiste en calcular la prueba de log rank dentro de cada estrato y despues combinarlo como un estimador global

Prueba de Log rank: HIV

```
> survdiff(Surv(time=time,event=cen)~factor(nage)+strata(country),data=hiv)
Call:
survdiff(formula = Surv(time = time, event = cen) ~ factor(nage) +
    strata(country), data = hiv)

                        N Observed Expected (O-E)^2/E (O-E)^2/V
factor(nage)=(17,24] 467       37     23.4    7.8417   11.5649
factor(nage)=(24,30] 428       22     21.5    0.0121    0.0166
factor(nage)=(30,38] 432       19     20.4    0.0900    0.1186
factor(nage)=(38,79] 440        7     19.7    8.2010   12.5225

 Chisq= 19.2  on 3 degrees of freedom, p= 0.00025
```

- **Interpretación**: Despues de estratificar por país (EEUU y Perú), se reafirman las evidencias de que la edad está asociada con el tiempo a adquirir VIH ($p < 0.001$).

Prueba de Log rank: Cáncer

```
> cancer$nage <- cut(cancer$Age,breaks=c(20,40,60,max(cancer$Age)))
> survdiff(Surv(time=time,event=rec)~Group+strata(nage),data=cancer)
Call:
survdiff(formula = Surv(time = time, event = rec) ~ Group + strata(nage),
    data = cancer)

                 N Observed Expected (O-E)^2/E (O-E)^2/V
Group=Abandoned 19        7     4.63     1.216      1.85
Group=Completed 44        8    10.37     0.542      1.85

 Chisq= 1.9  on 1 degrees of freedom, p= 0.174
```

- **Interpretación**: Despues de estratificar por edad (menores de 40, entre 41 y 60 y más de 60 años), no se observó diferencia en el riesgo de recurrencia de cáncer entre las mujeres que se sometieron a una histerectomía y aquellas que la abandonaron ($p = 0.174$).

Outline

1. Motivación
   - HPTN 039
   - Cáncer cervical

2. Prueba de Log Rank
   - Prueba de Log Rank estratificado
   - Prueba de log rank ponderado

Prueba de log rank ponderado

- La prueba de log rank tiene mayor poder para detectar diferencias donde los riesgos son proporcionales entre las poblaciones:

$$\frac{\lambda_1(t)}{\lambda_0(t)} = c \ , \ \forall t > 0$$

- Qué sucede si
  - El efecto de un tratamiento se nota al inicio y después de cierto tiempo, las personas tratadas y no tratadas muestran un riesgo similar.
  - El efecto de un tratamiento demora en observarse, es decir las diferencias serán mínimas al principio y se incrementan al final.

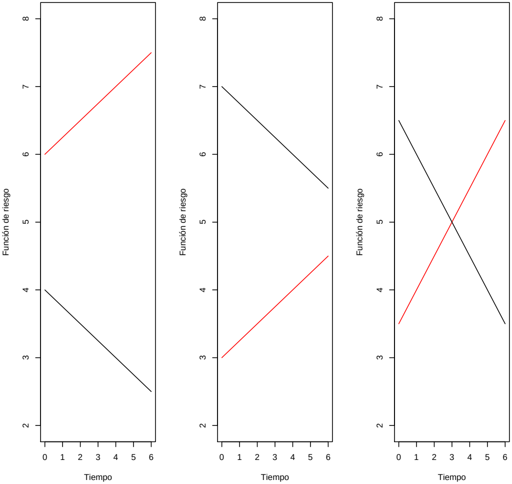

Figura 4: Funciones de riesgo para dos poblaciones

Prueba de log rank ponderado

- Definimos el estadístico de log rank ponderado

$$U(g) = \sum_{j=1}^{k} g_j w_j$$

donde $g_1, \ldots, g_k$ son pesos escogidos de acuerdo al problema para enfatizar (o no) las diferencias medidas por los w_i'.

- Ejemplos:
  - Gehan-Breslow: $g_j^G = n_j$
  - Harrington-Fleming (usado por R) [3]:

$$g_j^{(H)} = \left\{ \prod_{i \leq j} \left[1 - \frac{d_i}{n_i} \right] \right\}^{\rho} = \left\{ \hat{S}^{KM}(t_j) \right\}^{\rho} \ , \ \rho \in R$$

Prueba de log rank ponderado

- Si suponemos que los riesgos son iguales en todas las poblaciones (H_0), entonces:

$$E[U(g)] = 0 \quad , \quad Var[U(g)] = \sum_{j=1}^{k} g_j^2 W_j$$

- Si $p+1$ es el número de poblaciones que estoy comparando, entonces el estadístico de log rank ponderado está dado por:

$$U(g) V(g)^{-1} U(g)$$

que tiene una distribución $\chi^2_{(p)}$ bajo la hipótesis nula

Prueba de Log rank ponderado: Cáncer cervical

```
> survdiff(Surv(time=time, event=rec)~Group, rho=0, data=cancer)

                 N Observed Expected (O-E)^2/E (O-E)^2/V
Group=Abandoned 19        7     4.59     1.270       1.9
Group=Completed 44        8    10.41     0.559       1.9

 Chisq= 1.9  on 1 degrees of freedom, p= 0.168

> survdiff(Surv(time=time, event=rec)~Group, rho=1, data=cancer)

                 N Observed Expected (O-E)^2/E (O-E)^2/V
Group=Abandoned 19     5.85     3.78     1.135         2
Group=Completed 44     6.23     8.31     0.517         2

 Chisq= 2  on 1 degrees of freedom, p= 0.157
```

- **Interpretación**: Aún al incluir ponderaciones que enfatizan diferencias tempranas, no se observan diferencias en los riesgos a tener una recurrencia de cáncer entre mujeres que abandonaron la histerectomía y aquellas que se la realizaron ($p = 0.157$).

Referencias I

[1] C. Celum, A. Wald, J. Hughes, J. Sanchez, S. Reid, S. Delany-Moretlwe, F. Cowan, M. Casapia, A. Ortiz, J. Fuchs, S. Buchbinder, B. Koblin, S. Zwerski, S. Rose, J. Wang, and L. Corey. Effect of aciclovir on hiv-1 acquisition in herpes simplex virus 2 seropositive women and men who have sex with men: a randomised, double-blind, placebo-controlled trial. *Lancet*, 371(9630): 2109–2119, 2008.

[2] H. Gray, E. Seifert, V. Sal y Rosas, F. Nicandri, W. Koh, and B. Goff. The abandoned radical hysterectomy for cervical cancer: clinical predictors and outcomes. *Obstet Gynecol Int*, 2010.

[3] D. Harrington and T. Fleming. A class of rank test procedures for censored survival data. *Biometrika*, 69: 553–566, 1982.