

Programa de Especialización en Econometría Aplicada

SEUPROS -UNI

Modelos de Supervivencia

Clase 4

Edinson Tolentino

MSc Economics

email: edinson.tolentino@gmail.com

Twitter: @edutoleraymondi

Universidad Nacional de Ingeniería

17 de junio de 2023



1 Introducción

2 Preguntas

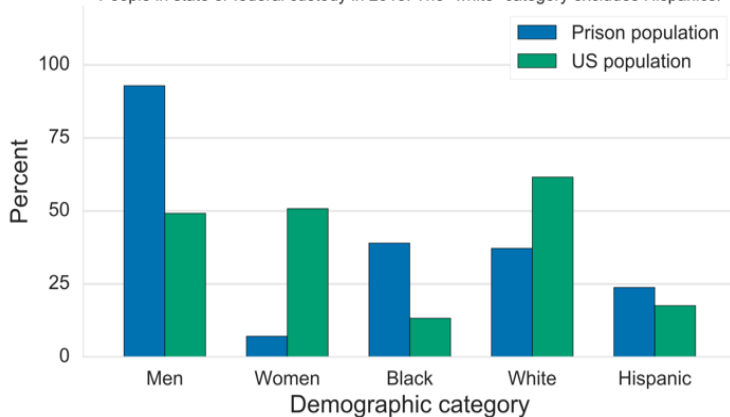
- Pregunta 1
- Pregunta 2
- Pregunta 3
- Pregunta 4
- Pregunta 5

Introducción



Demographics of the US Prison Population

People in state or federal custody in 2015. The "white" category excludes Hispanics.



Data source: <https://www.bjs.gov/content/pub/pdf/p15.pdf> (US Department of Justice)

ChartYourWorld.org

Introducción



- Los hombres afroamericanos poseen altas tasas de encarcelación respecto a los demás grupos en Estados Unidos.
- La tasa de **reincidencia (recidivism)** es de 50% en los Estados Unidos y son más altas entre los afroamericanos
- El presente taller explora los determinantes de **recidivism (reincidencia)** (desde el punto de vista de un ex-convicto liberado de su sistema de prisión)

Introducción



- Dada la información de la data **BD4.dta** contiene 1,145 presidiarios del sistema de prision de Estados Unidos.
- La propuesta del ejercicio es investigar los factores que determinan el tiempo de reincidir en un delito (tiempo de falla, failure time) luego de ser liberado. Esto es conocido en la literatura criminología como **reincidencia**.

Variables	Descripción
<i>durat_i</i>	número de meses dada la libertad del exconvicto del sistema de prisión
<i>d_i</i>	== 1 , si el exconvicto fue arrestrado, = 0 otro caso
<i>age_i</i>	edad del exconvicto en meses
<i>married_i</i>	== 1 si el exconvicto se encuentra casado, = 0, otro caso
<i>black_i</i>	== 1, si el exconvicto es negro, = 0 otro caso
<i>drugs_i</i>	== 1, si el exconvicto tiene historial de drogas, = 0 otro caso
<i>alcohol_i</i>	== 1, si el exconvicto tiene historial de alcohol, = 0 otro caso
<i>priors_i</i>	numero de condenas previas del exconvicto
<i>rules_i</i>	numero de reglas violadas durante su periodo en prision
<i>tserve_i</i>	tiempo de servicio en prision (en meses) antes de su liberacion

- Un investigador busca analizar el tamaño del tiempo (o duración) meses hasta que la policía vuelva a arrestar a un ex preso por un delito (o delitos).

Introducción



- La propuesta del ejercicio es investigar los **factores** que determina el tiempo de re-encarcelación (failure time)
- Hubo 552 liberados que fueron arrestados nuevamente por reincidir (38 % de liberados).

Variables	Descripción
<i>durat_i</i>	número de meses dada la libertad del exconvicto del sistema de prisión
<i>d_i</i>	== 1 , si el exconvicto fue arrestado, = 0 otro caso
<i>age_i</i>	edad del exconvicto en meses
<i>married_i</i>	== 1 si el exconvicto se encuentra casado, = 0, otro caso
<i>black_i</i>	== 1, si el exconvicto es negro, = 0 otro caso
<i>drugs_i</i>	== 1, si el exconvicto tiene historial de drogas, = 0 otro caso
<i>alcohol_i</i>	== 1, si el exconvicto tiene historial de alcohol, = 0 otro caso
<i>priors_i</i>	numero de condenas previas del exconvicto
<i>rules_i</i>	numero de reglas violadas durante su periodo en prision
<i>tserved_i</i>	tiempo de servicio en prision (en meses) antes de su liberacion

Introducción



- *d*, es la variable censurada

Variables	Descripción
<i>durat_i</i>	número de meses dada la libertad del exconvicto del sistema de prisión
<i>d_i</i>	== 1, si el exconvicto fue arrestrado, = 0 otro caso
<i>age_i</i>	edad del exconvicto en meses
<i>married_i</i>	== 1 si el exconvicto se encuentra casado, = 0, otro caso
<i>black_i</i>	== 1, si el exconvicto es negro, = 0 otro caso
<i>drugs</i>	== 1, si el exconvicto tiene historial de drogas, = 0 otro caso
<i>alcohol_i</i>	== 1, si el exconvicto tiene historial de alcohol, = 0 otro caso
<i>priors_i</i>	numero de condenas previas del exconvicto
<i>rules_i</i>	numero de reglas violadas durante su periodo en prison
<i>tsserved_i</i>	tiempo de servicio en prison (en meses) antes de su liberacion

Introducción



Introducción



Cuadro: Estadísticas descriptivas

	Personas	Promedio	Mediana	Min.	Max.	Std
max(time until return, follow)	1445	55.37	71.00	1.00	81.00	27
1 if failure; 0 if censored	1445	0.38	0.00	0.00	1.00	0
in months	1445	345.44	307.00	198.00	933.00	121
=1 if married when incarcerated	1445	0.26	0.00	0.00	1.00	0
=1 if black	1445	0.49	0.00	0.00	1.00	0
=1 if drug history	1445	0.24	0.00	0.00	1.00	0
=1 if alcohol problems	1445	0.21	0.00	0.00	1.00	0
prior convictions	1445	1.43	0.00	0.00	28.00	3
rules violations in prison	1445	1.19	0.00	0.00	27.00	2
time served, rounded to months	1445	19.18	12.00	0.00	219.00	21

Fuente: Exercise 6.

Elaboración: Autor

- El promedio de duración no condicional del periodo de libertad después de estar encarcelado es de 55.4 meses
- Aproximadamente, 38 % de la muestra ha sido arrestrado nuevamente (presumiblemente por cometer nuevamente un crimen)
- La variable age (edad) es expresado en meses, entonces en promedio la persona liberada tiene 29 años

Introducción



- La mediana del tiempo de duración en la muestra es de 71 meses
- La mediana del tiempo de duración para el grupo de datos no censurados, aquellos que salen del estado (los que son nuevamente puestos en prision) es de 19 meses
- Por tanto, se entiende que la mitad de aquellos casos que fallan, lo hicieron dentro de un año y medio luego de su liberación.
- Nota: El tiempo promedio de falla para el grupo no censurado es de 24 meses.

```
centile durat
```

Variable	Obs	Percentile	Centile	— Binom. Interp. — [95% Conf. Interval]	
durat	1,445	50	71	71	72

```
centile durat if _d==1
```

Variable	Obs	Percentile	Centile	— Binom. Interp. — [95% Conf. Interval]	
durat	552	50	19	17	21

Introducción



- Todos estas serán covariables que se utilizarán en el modelo.

Variables	Descripción
<i>durat_i</i>	número de meses dada la libertad del exconvicto del sistema de prisión
<i>d_i</i>	== 1 , si el exconvicto fue arrestado, = 0 otro caso
<i>age_i</i>	edad del exconvicto en meses
<i>married_i</i>	== 1 si el exconvicto se encuentra casado, = 0, otro caso
<i>black_i</i>	== 1, si el exconvicto es negro, = 0 otro caso
<i>drugs</i>	== 1, si el exconvicto tiene historial de drogas, = 0 otro caso
<i>alcohol_i</i>	== 1, si el exconvicto tiene historial de alcohol, = 0 otro caso
<i>priors_i</i>	numero de condenas previas del exconvicto
<i>rules_i</i>	numero de reglas violadas durante su periodo en prison
<i>tserve_d</i>	tiempo de servicio en prision (en meses) antes de su liberacion

Introducción



Antes de estimar cualquier modelo de duración, primero tenemos que decirle a Stata que vamos a realizar un análisis de supervivencia. Para ello usamos el comando **stset** (Survival Time Set).

Syntax

```
stset timevar [,failure(failvar)]
```

donde **timevar** es la variable duración y **failvar** describe los eventos. Luego se sugiere utilizar el comando **stsum** para resúmenes estadísticos de la data.

Seguido, es importante observar la función de supervivencia. Para ello, nos apoyamos con el comando **sts** que gráfica el estimador de Kaplan-Meier:

Syntax

```
sts graph [,by(varname)]
```

donde **varname** es una variable categórica que indica una característica de las unidades de análisis (por ejemplo, tamaño empresarial para el análisis de supervivencia de empresas).

Pregunta 1



- Explique ¿Cuál es el tiempo de falla (failure time) para esta aplicación en particular?

Pregunta 1: solución



- El tiempo de falla (*failure time*) en el presente caso esta dado por el tiempo de ser re-arrestado (*ex-convicto*) despues de haber sido liberado del sistema de prision de USA.
- El tiempo es medido en meses
- En esta aplicación, existe un **riesgo unico**
- El **riesgo** en cuestion se entiende como el **riesgo** de reincidir en el delito, por tanto, evidencia el estado de reencarcelación.
- El fenomeno es conocido en la literatura criminología como **recidivism (reincidencia)**

Pregunta 1: solución



```
.      stset durat, failure(_d)

      failure event:  _d != 0 & _d < .
obs. time interval:  (0, durat]
exit on or before:  failure

-----+-----
1,445  total observations
      0  exclusions

-----+-----
1,445  observations remaining, representing
      552  failures in single-record/single-failure data
80,013  total analysis time at risk and under observation
               at risk from t =          0
               earliest observed entry t =      0
               last observed exit t =      81
```

Pregunta 1: solución



```
.      stset durat, failure(_d)

      failure event:  _d != 0 & _d < .
obs. time interval:  (0, durat]
exit on or before:  failure
```

```
1,445 total observations
      0 exclusions
```

```
1,445 observations remaining, representing
      552 failures in single-record/single-failure data
80,013 total analysis time at risk and under observation
              at risk from t =      0
              earliest observed entry t =      0
              last observed exit t =     81
```

- Existen 552 observaciones que representa el **fallares (fallar)**
- Por otro lado, 893 sobrevivientes
- Entonces, existen 552 observaciones no censuradas y 893 observaciones censuradas en este caso.
- Existen 80,013 **periodo de riesgo**

Pregunta 1: solución



```
. ltable durat, hazard interval(0(1)80)
```

		Beg.	Cum.	Std.		Std.		
Interval	Total	Failure	Error	Hazard	Error	[95% Conf. Int.]		
1	2	1445	0.0055	0.0020	0.0056	0.0020	0.0017	0.0094
2	3	1437	0.0159	0.0033	0.0105	0.0027	0.0052	0.0158
3	4	1422	0.0256	0.0042	0.0099	0.0026	0.0047	0.0151
4	5	1408	0.0346	0.0048	0.0093	0.0026	0.0042	0.0143
5	6	1395	0.0457	0.0055	0.0115	0.0029	0.0059	0.0172
6	7	1379	0.0581	0.0062	0.0131	0.0031	0.0071	0.0192
7	8	1361	0.0706	0.0067	0.0133	0.0031	0.0072	0.0195
8	9	1343	0.0817	0.0072	0.0120	0.0030	0.0061	0.0179
9	10	1327	0.0941	0.0077	0.0137	0.0032	0.0073	0.0200
10	11	1309	0.1093	0.0082	0.0169	0.0036	0.0099	0.0240
11	12	1287	0.1170	0.0085	0.0086	0.0026	0.0035	0.0137
12	13	1276	0.1266	0.0087	0.0110	0.0029	0.0053	0.0168
13	14	1262	0.1370	0.0090	0.0120	0.0031	0.0059	0.0180
14	15	1247	0.1481	0.0093	0.0129	0.0032	0.0066	0.0192
15	16	1231	0.1640	0.0097	0.0189	0.0039	0.0112	0.0266
16	17	1208	0.1716	0.0099	0.0091	0.0028	0.0037	0.0146
17	18	1197	0.1779	0.0101	0.0075	0.0025	0.0026	0.0125
18	19	1188	0.1889	0.0103	0.0136	0.0034	0.0069	0.0202
19	20	1172	0.1952	0.0104	0.0077	0.0026	0.0027	0.0127
20	21	1163	0.2007	0.0105	0.0069	0.0024	0.0021	0.0117
21	22	1155	0.2097	0.0107	0.0113	0.0031	0.0052	0.0175
22	23	1142	0.2145	0.0108	0.0061	0.0023	0.0016	0.0107
23	24	1135	0.2256	0.0110	0.0142	0.0035	0.0072	0.0212

- Existen 552 observaciones que representa el **failures (fallar)**
- Por otro lado, 893 sobrevivientes
- Entonces, existen 552 observaciones no censuradas y 893 observaciones censuradas en este caso.
- Existen 80,013 **periodo de riesgo**

Pregunta 2



- ¿Cuál es la tasa de riesgo bajo una metodología no paramétrica y sin covariables (Kaplan-Meier) para los cinco meses? interprete los valores estimados.

Pregunta 2: solución



- La tasa de Kaplan-Meier es definida como:

$$\hat{\theta}(T_k) = \frac{h_k}{n_k}$$

- Donde h_k es el numero de **falla** del estado en el tiempo k
- Luego, n_k es el conjunto de riesgo en el tiempo k (conjunto de personas que se mantiene en el analisis)
- El conjunto de riesgo comprende todas aquellas unidades en riesgo de **salir del estado (o fallar)** en el momento relevante.

Pregunta 2: solución



tab durat			
max(time until return, follow)	Freq.	Percent	Cum.
1	8	0.55	0.55
2	15	1.04	1.59
3	14	0.97	2.56
4	13	0.90	3.46
5	16	1.11	4.57
6	18	1.25	5.81
7	18	1.25	7.06
8	16	1.11	8.17
9	18	1.25	9.41
10	22	1.52	10.93

- Mes uno:

$$\hat{\theta}(T_1) = 8/1445 = 0.0055$$

- Mes dos:

$$\hat{\theta}(T_2) = 15/1437 = 0.0104$$

- Recuerde h_k es el número de salida y n_k el el conjunto de riesgo (conjunto de riesgo)

$$\hat{\theta}(T_1) = h_k/n_k$$

Pregunta 2: solución



- La tasa de riesgo (hazard) de **Kaplan-Meier** para los primeros 5 meses son:

- Mes uno:

$$\hat{\theta}(T_1) = 8/1445 = 0.0055$$

- Mes dos:

$$\hat{\theta}(T_2) = 15/1437 = 0.0104$$

- Mes tres:

$$\hat{\theta}(T_3) = 14/1422 = 0.0098$$

- Mes cuatro:

$$\hat{\theta}(T_4) = 13/1408 = 0.0092$$

- Mes cinco:

$$\hat{\theta}(T_5) = 16/1395 = 0.0115$$

- Existe una probabilidad de 0.55 por ciento para un preso, que luego de ser liberado sea nuevamente arrestrado por un delito dentro de un mes posterior a su liberación.
- Existe una probabilidad de 1.04 por ciento para un preso liberado que sobrevivio al primer mes (que se mantienen sin realizar delitos al primer mes), vuelva a ser arrestado nuevamente antes de terminar el segundo mes.
- Existe una probabilidad de 0.98 por ciento para un preso liberado que sobrevivio al segundo mes (que se mantienen sin realizar delitos al segundo mes), vuelva a ser arrestado nuevamente antes de terminar el tercer mes.

Pregunta 2

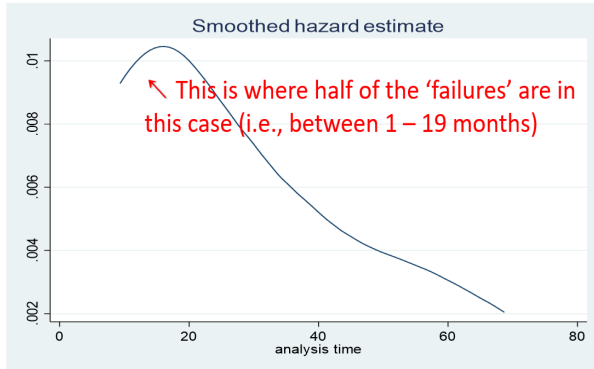


- Grafique la tasa de riesgo (hazard rate) suavizada de Kaplan-Meier (ver STATA)
¿Qué es lo que usted concluye del grafico?

Pregunta 2: solución



- Tasa de riesgo (hazard rate) de **Kaplan - Meier** recurrencia



Pregunta 3



- Realice la estimación de la función exponencial sobre la tasa de riesgo a través de la siguiente forma $\theta_{it} = \exp(X_i' \beta)$, dado que X es el conjunto de variables explicativas (definidas en la tabla líneas arriba). Interprete los coeficientes estimados para la edad(age) , raza negros (black), sentencias de condena previas (priors). Gráfique la línea base de hazard estimada para este modelo.

Pregunta 3: solución



- El modelo **exponencial de hazard** puede ser expresada como:

$$\theta_{it} = \exp(\beta_0 + \beta_1 age_i + \beta_2 married_i + \beta_3 black_i + \dots$$

$$\dots + \beta_4 drugs_i + \beta_5 alcohol_i + \beta_6 priors_i + \beta_7 rules_i + \beta_8 tserved_i)$$

- Podríamos observar que ninguna de las variables explicativas **cambia respecto al tiempo**
- Todas las covariables exhiben realizaciones para individuos previo a la fecha de su liberación para que no cambie mientras se acerca al tiempo de falla.

Pregunta 3: solución



- El logaritmo del modelo **exponencial de hazard** puede ser expresada como:

$$\log_e(\theta_{it}) = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{married}_i + \beta_3 \text{black}_i + \dots$$

$$\dots + \beta_4 \text{drugs}_i + \beta_5 \text{alcohol}_i + \beta_6 \text{priors}_i + \beta_7 \text{rules}_i + \beta_8 \text{tserved}_i$$

- Para la k^{th} variable binaria :

$$(\exp(\hat{\beta}_k) - 1) \times 100$$

- Para la k^{th} variable continua :

$$\frac{\partial \log(\hat{\beta}_k)}{\partial X_k} = \hat{\beta}_k$$

Pregunta 3: solución



Pregunta 3: solución



Cuadro: Ecuación (1)

	Exponencial	
_t		
age	-0.0034***	(0.0005)
married	-0.1531	(0.1079)
black	0.4578***	(0.0879)
alcohol	0.4808***	(0.1050)
drugs	0.2877***	(0.0981)
priors	0.0892***	(0.0133)
rules	0.0368**	(0.0164)
tserve	0.0105***	(0.0018)
_cons	-4.5903***	(0.1701)
Observaciones	1445	
Log-L	-1653.2	

Fuente: Exercise 6.

Elaboración: Autor

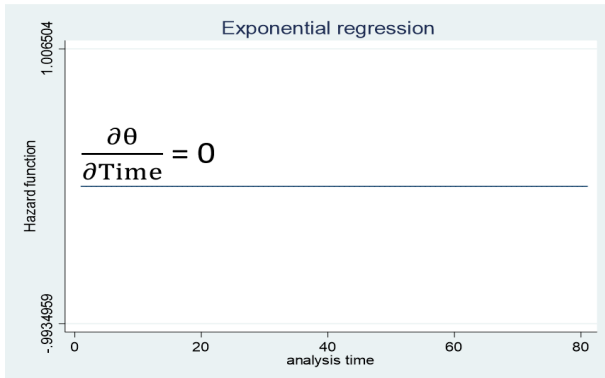
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

- En promedio y *ceteris paribus*, un incremento de un año más de edad del individuo reduce el riesgo (hazard) de cometer nuevamente un crimen (reincidencia) por 4.1 % (por ejemplo, $0.0034 \times 12 = 0.0408$ dado los doce meses al año)
- En promedio y *ceteris paribus*, el riesgo (hazard) de cometer nuevamente un crimen (reincidencia) para presos de raza negra (afroamericanos) es $[exp(0.4578) - 1] \times 100 = 58.1\%$ mayor respecto a los presos de raza no oscura.

Pregunta 3: solución



- Existe una **tasa de riesgo de duración constante (constante hazard duration)** sobre el riesgo del tiempo pasado en el estado (es decir, la libertad) es cero.



Pregunta 4



- Estime la función de Hazard de Weibull de la siguiente forma:

$$\theta_{it} = \exp(X_i' \beta) (\alpha t^{\alpha-1})$$

Nota: El estimado de α es denotado como p en la regresión de STATA para este presente modelo

- Comente los resultados obtenidos
- Use el nivel de significancia de 0.05 para testear la siguiente proposición:

$$H_o : \ln(\alpha) = 0$$

$$H_a : \ln(\alpha) < 0$$

¿Qué proposición se encuentra sobre este test? Demuestre como la varianza muestral para este test-estadístico se deriva. Interprete el valor del tes.

Pregunta 4.1: solución



- El modelo de Hazard Weibull ser expresada como:

$$\theta_{it} = \exp(\beta_0 + \beta_1 age_i + \beta_2 married_i + \beta_3 black_i + \dots$$

$$\dots + \beta_4 drugs_i + \beta_5 alcohol_i + \beta_6 priors_i + \beta_7 rules_i + \beta_8 tserve_i) \underbrace{\left[\alpha t^{\alpha-1} \right]}_{\text{Hazard—linea—base}}$$

- Si $\alpha = 1 \rightarrow$ dependencia constante de duración
- Si $\alpha < 1 \rightarrow$ dependencia negativa de duración
- Si $\alpha > 1 \rightarrow$ dependencia positiva de duración

Pregunta 4.1: solución



Cuadro: Ecuación (1)

Weibull		
_t		
age	-0.0032***	(0.0005)
married	-0.1517	(0.1079)
black	0.4368***	(0.0880)
alcohol	0.4610***	(0.1049)
drugs	0.2757***	(0.0981)
priors	0.0872***	(0.0134)
rules	0.0356**	(0.0165)
tserved	0.0098***	(0.0018)
_cons	-3.7975***	(0.2111)
/		
ln_p	-0.2178***	(0.0389)
Observaciones	1445	
Log-L	-1636.2	
p	0.804266	

Fuente: Exercise 6.

Elaboracion: Autor

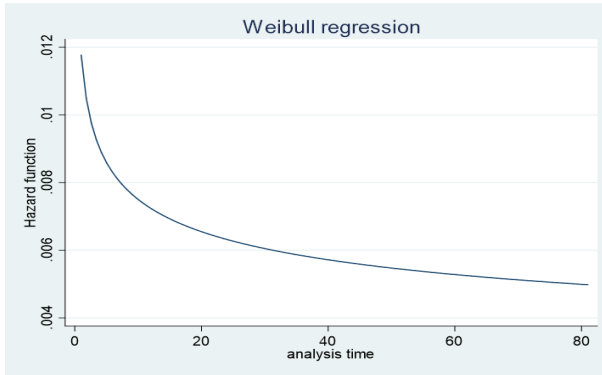
***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

- 1 Los estimadores de covariables son similares sobre aquellos estimados en el modelo Exponencial de Hazard
- 2 Como luce el gráfico de riesgo (hazard) dado este modelo?
- 3 Recuerde que **p** aquí es el α (valor de parametro de Weibull)

Pregunta 4.1: solución



- Dado $\hat{\alpha} < 1$
- La línea base de riesgo (baseline hazard) decrete monotonamente con respecto al tiempo
- El parámetro estimado de $\hat{\alpha}$ (conocido como de shape parameter) será estadísticamente diferente de la unidad?



Pregunta 4.2: solución



- Usando el nivel de significancia de 0.05 para testear la siguiente proposición:

$$H_o : \ln(\alpha) = 0$$

$$H_a : \ln(\alpha) < 0$$

- ¿Cuál es la proposición bajo el test?
- Demuestre como la varianza muestral para este test estadístico es redivado en el presente caso

Pregunta 4.2: solución



- La hipótesis nula y alternativa son expresada:

$$H_o : \ln(\alpha) \text{ versus } H_a : \ln(\alpha) < 0$$

- Note que $H_o : \ln(\alpha) = 0$ es equivalente a $\alpha = 1$
- La pregunta ahora es como derivar la varianza muestral en este caso?
- Respuesta : podemos usar el metodo delta

Pregunta 4.2: solución



- Si $\Delta = \ln(\alpha)$ entonces:

$$\frac{\partial \Delta}{\partial \alpha} = \frac{1}{\alpha}$$

- Esta expresión para estimar la varianza muestral estará dado por:

$$\text{Var}(\hat{\Delta}) = \left(\frac{\partial \Delta}{\partial \alpha} \right)^2 \text{Var}(\hat{\alpha})$$

$$\text{Var}(\hat{\Delta}) = \left(\frac{1}{\alpha} \right)^2 \text{Var}(\hat{\alpha})$$

- Insertamos los valores empiricos estimados y obtenemos:

$$\text{Var}(\hat{\Delta}) = \left(\frac{1}{0.8042664} \right)^2 (0.000979564804)$$

$$\text{Var}(\hat{\Delta}) = 0.001514374616$$

Pregunta 4.2: solución



- El resultado del test-t asintotico para la proposición propuesta dado por:

$$t = \frac{\ln(0.8042664)}{\sqrt{0.001514374616}} = -5.59745$$

- El valor critico de una cola al 0.05 es de **-1.64**
- La hipotesis nula plantea **la duración es constante** , dado los resultados dicha hipotesis es rechazada por la información
- Los resultados implican la existencia de una **dependencia duración negativa**
- Cuanto más tiempo esté en el estado de libertad, es menos probable que lo deje (es decir, reincidiendo).

Pregunta 5



- Estime el modelo de propociones y compare los resultados con los previos modelos usando el modelo de Weibull en la pregunta anterior. ¿Cuál es la diferencia en los supuestos comprendidos en el modelo Cox y el modelo de regresión de Weibull?

Pregunta 5: solución



- ¿Cuál es la diferencia en los supuestos comprendidos en el modelo Cox y el modelo de regresión de Weibull?
- Dado la forma funcional del modelo de Cox

$$\theta_i(X; t) = \exp(\beta X_i) \lambda_0(t)$$

- Esto permite una flexibilización de la línea base de hazard (riesgo) para el modelo de Proporciones Cox (Cox PH)

Pregunta 5: solución



Cuadro: Ecuación (1)

	Cox	
age	-0.0031***	(0.0005)
married	-0.1552	(0.1079)
black	0.4172***	(0.0881)
alcohol	0.4457***	(0.1049)
drugs	0.2689***	(0.0981)
priors	0.0865***	(0.0134)
rules	0.0359**	(0.0166)
tserved	0.0093***	(0.0018)
Observaciones	1445	
Log-L	-3818.8	

Fuente: Exercise 6.

Elaboracion: Autor

***, **, * denote statistical significance at the 1 %, 5 % and 10 % levels respectively for zero.

- Incorporación de la linea base de Hazard de manera más flexible (no-paramétrica) respecto al Weibull.
- Nota: termino no constante es timado por el modelo de Cox
- Los estimadores son ampliamente similares sobre aquellos obtenidos en el modelo Weibull

Pregunta 5: solución



Cuadro: Ecuación (1)

	Exp		Weibull		Cox	
main						
age	-0.0034***	(0.0005)	-0.0032***	(0.0005)	-0.0031***	(0.0005)
married	-0.1531	(0.1079)	-0.1517	(0.1079)	-0.1552	(0.1079)
black	0.4578***	(0.0879)	0.4368***	(0.0880)	0.4172***	(0.0881)
alcohol	0.4808***	(0.1050)	0.4610***	(0.1049)	0.4457***	(0.1049)
drugs	0.2877***	(0.0981)	0.2757***	(0.0981)	0.2689***	(0.0981)
priors	0.0892***	(0.0133)	0.0872***	(0.0134)	0.0865***	(0.0134)
rules	0.0368**	(0.0164)	0.0356**	(0.0165)	0.0359**	(0.0166)
tserved	0.0105***	(0.0018)	0.0098***	(0.0018)	0.0093***	(0.0018)
_cons	-4.5903***	(0.1701)	-3.7975***	(0.2111)		
/						
ln_p			-0.2178***	(0.0389)		
Observaciones	1445		1445		1445	
Log-L	-1653.2		-1636.2		-3818.8	

Fuente: Exercise 6.

Elaboracion: Autor

***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively for zero.

Pregunta 5: solución

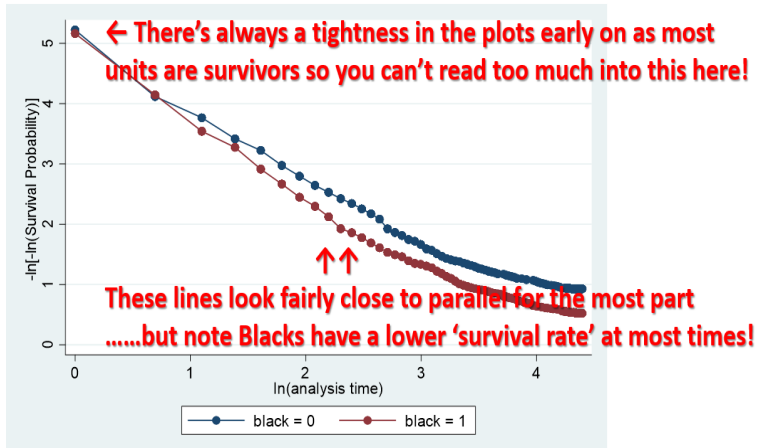


- **Investigar el supuesto de proporciones de hazard**
- Nosotros ahora graficamos e investigamos explícitamente el test para el supuesto de proporciones de hazard en el el modelo de Cox.
 - ① Verificamos el supuesto de riesgos proporcionales usando la gráfica de doble logarítmica de las probabilidades de supervivencia contra el logaritmo del tiempo de falla. La cual es conocido como **log-log plot**
 - ② Específicamente, la cantidad $\log_e [\log_e (\text{survivor} - \text{probability})]$ es graficada en relación al tiempo de falla.
- Lo que buscamos en estos gráficos es evidencia de líneas paralelas, por lo que no hay cruces ni convergencia o divergencia de estas líneas para cada covariables.

Pregunta 5: solución



- Gráfico de $\log_e [\log_e (\text{survivor} - \text{probability})]$ en relación al $\log_e (\text{Tiempo} - \text{de} - \text{falla})$ para Negros y No-negros (Black y No-Black)



Pregunta 5: solución

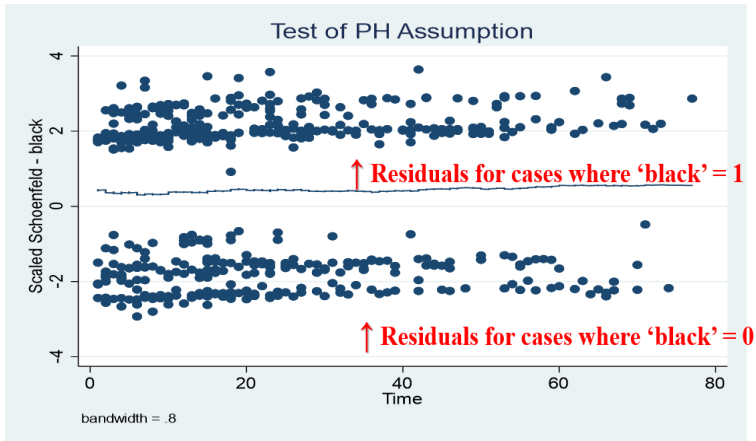


- Podemos calcular un conjunto de residuos del modelo Cox PH.
- Los más populares son los residuos de Schoenfeld.
 - ① Esto se calcula solo para aquellas observaciones que fallaron o salieron del estado (es decir, los 552 casos de falla aquí).
 - ② En lugar de un solo residuo para cada individuo, tenemos residuos separados para cada individuo para cada covariable.
- Estos residuos de Schoenfeld se pueden graficar en relación al tiempo de falla y sus valores deben ser independientes del tiempo para que se satisfaga el supuesto de riesgos proporcionales.

Pregunta 5: solución



- Gráfico de the Schoenfeld Residuals en relación al tiempo por Black



Pregunta 5: solución



- Una prueba estadística formal se basa en los residuos de Schoenfeld y es análoga a probar si el coeficiente de pendiente de una regresión de los residuos de Schoenfeld escalados en el tiempo es estadísticamente diferente de cero o no.
- La prueba se conoce como prueba de Grambsch y Therneau (1994) para riesgos proporcionales.
- La hipótesis nula para la prueba es: **el modelo de PH de Cox estimado satisface el supuesto de riesgos proporcionales.**

```
. estat phtest
```

Test of proportional-hazards assumption

Time: Time

	chi2	df	Prob>chi2
global test	12.23	8	0.1414

- La prueba global (o general) para las ocho covariables no rechaza la hipótesis nula de la independencia de los residuos de Schoenfeld y el tiempo de falla con un valor prob de **0,1414**.
- Esto confirma que la suposición de una formulación de riesgos proporcionales está justificada por los datos de esta solicitud.