ANÁLISIS DE SUPERVIVENCIA. MODELIZACIÓN DEL TIEMPO DE BÚSQUEDA DEL PRIMER EMPLEO.

GARCÍA LOZANO, Josefina jglozano@pdi.ucam.edu

GARCÍA FERNÁNDEZ , Pedro pgfernandez@pdi.ucam.edu

GÓMEZ GALLEGO, Juan C.

FUENTES LORCA, Rosana

Universidad Católica San Antonio de Murcia

RESUMEN

El objetivo de este trabajo es aplicar la metodología tradicional del análisis de supervivencia para modelizar el tiempo invertido por los egresados universitarios en encontrar el primer empleo significativo. Comenzaremos considerando al conjunto de egresados como un conjunto de individuos homogéneos, sin introducir el posible efecto de otras variables. En este punto, utilizaremos el método de estimación no-parametrico propuesto por Kaplan y Meier .

Posteriormente y usando los métodos tradicionales para poblaciones heterogéneas, modelos de función de riesgo proporcional propuestos por Cox (1972), introduciremos en el modelo variables exógenas que recojan ciertas características de los egresados, por ejemplo, el nivel adquirido por el egresado en los diferentes tipos de competencias transversales, estimando su posible efecto sobre el tiempo de supervivencia.

Palabras clave: análisis de supervivencia, modelos de duración, competencias de egresados.

1.- Introducción

Se conoce como análisis de duración o de supervivencia al análisis de datos en el que se recoge el periodo de tiempo que transcurre desde un punto de partida fijado hasta el momento en que acontece un suceso particular.

Los datos de duración, debido a sus particulares características, no pueden analizarse con las técnicas habituales utilizadas en otros tipos de datos. Normalmente los datos presentan una asimetría positiva, lo cual hace razonable no suponer una distribución normal. Resulta más adecuado considerar otro tipo de distribuciones que se adaptan mejor a estos datos, como la exponencial, Weibull o Gamma, entre otras. Una segunda característica asociada a este tipo de datos es la censura. La duración de un evento está censurada cuando su tiempo de fallo no ha sido observado dentro del período de estudio.

Como tercera característica tenemos que para obtener duraciones de individuos hay que realizar un seguimiento de estos a lo largo de un período determinado, y es lógico suponer que vamos a tener una serie de individuos con información incompleta en este período de seguimiento. Por último es posible que las características del individuo y de su entorno evolucionen en lugar de permanecer constantes, con lo cual tendríamos variables dependientes del tiempo.

Dentro de la abundante literatura existente para esta metodología y haciendo un repaso cronológico, podríamos comenzar alrededor de los años 50 donde se analizaba el tiempo de duración de un suceso utilizando la metodología de las tablas de vida. En esta época tenemos la estimación Producto Límite derivada por Kaplan y Meier (1958). Es en el año 1972, cuando aparece otro trabajo relevante para estos datos, el modelo de regresión de función de riesgo proporcional propuesto por Cox. Es uno de los artículos más citados de toda la literatura científica en este contexto, y del cual han derivado multitud de trabajos. Más adelante aparecen libros que tratan el análisis de duración, entre los que destacamos, Kalbfleisch y Prentice (1980), Lee (1980), Lawless (1982), Cox y Oakes (1984) y Collet (1994). Partiendo del modelo de Cox, se han derivado y analizado diferentes modelos, como los modelos multiestado, que recogen la posibilidad de tener varios tiempos de fallo, los modelos "frailty", que recogen la posibilidad de no tener independencia entre los individuos o la posibilidad de que los individuos analizados no formen un grupo homogéneo dadas las variables explicativas recogidas en el modelo. La existencia de variables explicativas dependientes del tiempo, ya recogida por Cox en su artículo, ha generado otro tipo de modelos que recogen este efecto. La otra gran clase de modelos de duración para poblaciones heterogéneas estaría formada por los modelos de duración acelerada también conocidos como modelos log-lineales. Dentro de estos modelos se situarían los propuestos por Stute (1993), cuyos parámetros de interés pueden estimarse sin especificar la distribución de la duración.

En el periodo de tiempo comprendido entre los años 1960- 2000, este modelo se ha aplicado fundamentalmente en el campo de la salud, (modelos de supervivencia), y en el industrial, (modelos de fiabilidad). En este trabajo pretendemos aplicarlo al campo socioeconómico para modelizar el tiempo que invierte un egresado hasta encontrar el primer empleo significativo.

Ahora bien, la teoría sobre competencias profesionales aunque reciente, nos presenta la posibilidad de que a un mayor nivel de competencias incorporadas puede influir positivamente en el tiempo que tarda un egresado en encontrar empleo. La competencia comporta siempre un conjunto de conocimientos, procedimientos, actitudes y capacidades que son personales y se complementan entre si, de manera que el individuo pueda actuar con eficiencia frente a las situaciones profesionales. La adquisición, transición y realización de las competencias está tanto en los procesos formales como en procesos informales de la vida cotidiana o profesional.

Distintos autores Mertens (1989), Alex (1991), Le Boterf (1991), hacen referencia como mínimo a dos grandes grupos de competencias: unas de carácter especifico de un determinado puesto de trabajo o función laboral, y otras de carácter transversal que son demandadas y aplicables en contextos mas amplios.

Las competencias transversales son aquellas competencias genéricas, comunes a la mayoría de profesiones y que se relacionan con la puesta en practica integrada de aptitudes, rasgos de personalidad, conocimientos adquiridos y también valores. Son las competencias que responden a las nuevas alfabetizaciones, que incluyen manejo de equipos informáticos, de sistemas de información, idiomas y otras habilidades y actitudes que permiten a la persona ser multifuncional.

La universidad aparece como una plataforma que debe mejorar su capacidad de respuesta en la transición del universitario al mercado laboral. Esta etapa no siempre implica el logro de una situación estable y relacionada con los procesos de cualificación seguidos, de ahí que el estudio de las demandas sociales hayan sido objeto de reflexión los últimos años. La complejidad y amplitud de este campo nos lleva a delimitar el propósito de este trabajo:

- ✓ Valorar cuanto tiempo tardan los graduados en encontrar empleo significativo, según su nivel de competencias adquirido.

El presente trabajo se ha estructurado en tres apartados. En el primero se hace una revisión de la metodología utilizada “análisis de supervivencia”. En el segundo se presentan los resultados obtenidos al aplicar la metodología a una cohorte de egresados de la Diplomatura en Enfermería de la Universidad Católica San Antonio UCAM, y por último en el tercero se presentan las conclusiones obtenidas.

2.- Metodología.

Supongamos que la variable aleatoria no negativa T que representa a las duraciones de un grupo de individuos homogéneo (no existen variables explicativas relacionadas con la variable duración) tiene función de densidad f(t), siendo t una realización de T. Sean $F(t)$ y $S(t)$ las funciones de distribución y de supervivencia de la variable aleatoria T.

La *función de riesgo o tasa de azar o razón de fallo* $\lambda(t)$, se define como la tasa instantánea de fallo o muerte en T = t, condicionada a que el individuo ha sobrevivido hasta el momento t: $\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \le T \le t + \Delta t / T \ge t)}{\Delta t} = \frac{f(t)}{S(t)}$. La función de densidad, la función de riesgo y la función de supervivencia están claramente relacionadas mediante: $f(t) = S(t)\lambda(t)$ y $\lambda(t) = \frac{-d \ln S(t)}{dt}$. Otra función que resulta útil es la *función integrada o acumulada de riesgo* $H(t) = \int_0^t \lambda(s)\, ds$ para la cual $S(t) = e^{-H(t)}$ por lo que $H(t) = -\ln S(t)$.

El objetivo del análisis de supervivencia es estimar las funciones de supervivencia y de riesgo a partir de los tiempos de supervivencia observados.

2.1.- Modelos de duración para poblaciones homogéneas.

Por población homogénea entendemos una población donde todos los individuos presentan las mismas características o las diferentes características de estos no son relevantes para el análisis de las duraciones. En estos casos el estudio de las duraciones se realiza mediante la estimación de las funciones presentadas anteriormente. Básicamente la estimación de estas funciones se puede realizar de dos formas: métodos paramétricos y métodos no paramétricos.

a) Estimaciones no paramétricas de la función de supervivencia.

En el caso en que la distribución de la variable tiempo de supervivencia sea desconocida, los métodos de estimación no paramétricos más utilizados son el estimador de Kaplan y Meier (1958), o la estimación de las tablas de vida conocida también como estimación actuarial. Mediante estos dos procedimientos se estiman las funciones de distribución, supervivencia y riesgo del tiempo de duración sin realizar ningún supuesto sobre la forma de la distribución. Haremos referencia al estimador de Kaplan y Meier.

Procedimiento de Kaplan-Meier.

Kaplan y Meier proponen el método conocido como del producto límite al objeto de resolver los problemas planteados por la ausencia de información.

El estimador de la curva de supervivencia propuesto por Kaplan y Meier se basa en el mismo principio que el actuarial: *calcular la supervivencia como producto de probabilidades condicio-*

nadas, pero llevando la partición del tiempo de estudio en intervalos al caso extremo de considerar que cada intervalo contenga sólo la observación correspondiente a un individuo, sea ésta fallo o censura. Si los datos observados corresponden a los tiempos $0 = t_0 < t_1 <....< t_n$, se considera la partición determinada por los intervalos $(t_{i-1}, t_i]$, al que pertenece el instante de tiempo t_i pero no el t_{i-1} y además el interior de estos intervalos está siempre libre de censuras, que sólo ocurrirán, en su caso, en un extremo.

Si llegan n_i individuos con vida al intervalo, $(t_{i-1},t_i]$, el estimador de la probabilidad de fallo en ese intervalo, condicionada a haber sobrevivido hasta entonces, será:

$$q_i = \begin{cases} \dfrac{1}{n_i} & si\ \ en\ t_i\ \ se\ produce\ un\ \ fallo \\ 0 & si\ \ en\ t_i\ \ se\ produce\ una\ censura \end{cases}$$

Los intervalos que no contienen fallos no contribuyen a la construcción de S(t), ya que para ellos la estimación de la probabilidad condicionada de supervivencia en el intervalo es igual a 1. La existencia de censuras sí influye en el número de individuos expuestos al riesgo de fallar al comienzo del intervalo siguiente, que se ve disminuido en una unidad.

Utilizaremos la siguiente notación:

n_t: Número de individuos que llegan al comienzo del intervalo.

d_t : Número de fallecidos en el intervalo (t, t +1] .

p_t : Proporción de individuos que han sobrevivido al instante t. $p_t = \dfrac{n_t - d_t}{n_t} = 1 - \dfrac{d_t}{n_t}$

p_i es, por tanto, la probabilidad condicional de sobrevivir el i-ésimo tiempo, habiendo sobrevivido hasta el (i-1)-ésimo (antes denotado por $S_{i/i-1}$).

La probabilidad de supervivencia después del instante «t_i» será:

$$S(t_i) = p_1 \cdot p_2 \cdots p_i = \prod_{j=1}^{i} \left(1 - \frac{d_j}{n_j}\right) = S(t_{i-1}) \cdot \left(1 - \frac{d_i}{n_i}\right) = S(t_{i-1}) \cdot p_i$$

Cuando t = 0, S(0) = 1; es decir, todos los individuos comienzan vivos el estudio.

Las estimaciones conseguidas con el método de las tablas de vida coincidirán con las estimaciones de Kaplan y Meier al aumentar el número de intervalos hacia infinito y reducir la longitud de los intervalos. Esta sería otra posibilidad para obtener el estimador de Kaplan-Meier (Collet (1994) Cap.2), por eso diremos que el estimador de Kaplan-Meier es el caso límite del estimador por tablas de vida. La varianza del estimador (Orbe (1999) Cap 3) está dada por la expresión:

$$\operatorname{var}\left[\hat{S}(t)\right] \approx \left[\hat{S}(t)\right]^2 \sum_{j=1}^{r} \frac{d_j}{n_j\left(n_j - d_j\right)} \text{ para } t \in [t_r, t_{r+1}).$$

Las propiedades asintóticas de este estimador han sido analizadas en primer lugar por Kaplan y Meier (1958). Posteriormente, Breslow y Crowley (1974) analizan la distribución asintótica para el caso de censura aleatoria. Stute y Wang (1993) analizan la consistencia del estimador, derivando las condiciones necesarias para la consistencia y, posteriormente Stute (1995) analiza la distribución asintótica de este estimador para el caso de censura aleatoria. Estos trabajos, además de analizar estas propiedades para la función de distribución, también van a proponer estimadores, derivando sus propiedades, para la media y otros momentos de la variable duración, entre otras funciones. En la tesis doctoral "Un modelo de regresión parcial censurado para análisis de supervivencia" de Jesús Orbe Lizundia y tomando como referencia el trabajo de Stute y Wang (1993) se desarrolla el análisis de la consistencia para la estimación de la función de distribución.

Una vez estimada la función de supervivencia de la variable duración, ésta puede ser utilizada, mediante procedimientos gráficos, para tratar de ver la posible distribución a la que se ajustan los datos. Muchas veces resulta interesante determinar si dos o mas muestras tienen funciones de supervivencia similares. Dos procedimientos no paramétricos comúnmente usados son el test de Wilcoxon y el test de Savage. Ambos son una generalización de los tests de rango para el caso censurado. La idea de estos tests consiste en medir la diferencia entre el número de fallos observados y el número de fallos esperados bajo la hipótesis nula de igualdad, de forma que, si la diferencia es significativa, rechazaremos la hipótesis nula.

b) **Estimaciones paramétricas de la función de supervivencia**.

Cuando la distribución de la variable tiempo de supervivencia sea conocida, las inferencias basadas en la parametrización de dicha distribución serán más precisas o eficientes. Si la distribución de probabilidad asumida es correcta, los errores estándar de los estimadores en las aproximaciones paramétricas son menores. Además estas aproximaciones permiten realizar inferencias poblacionales no limitándose a la muestra analizada como en el caso de las alternativas puramente no paramétricas. Teóricamente cualquier distribución para la cual S(0)=1 puede utilizarse como distribución de supervivencia. Sin embargo, existen ciertas familias de distribuciones específicamente útiles, para ajustarse a los datos de un problema de análisis de supervivencia. Supongamos que los datos siguen un modelo de probabilidad determinado. El modelo más sencillo es el que supone que la tasa de riesgo no varía en el tiempo, es decir $\lambda(t) = \lambda$ para $0 \leq t < \infty$. En este caso, la probabilidad condicionada de que un individuo muera en un intervalo de tiempo determinado (lo suficientemente pequeño), estando vivo en t, será la misma con independencia del momento en el que se observe el individuo. Esta característica se conoce como *pérdida de memoria*. A partir de esta suposición se obtiene $S(t) = Ke^{-\lambda t}$ siendo K la constante de integración. La condición S(0)=1 implica que necesariamente K=1, y la

solución que se obtiene es $S(t) = e^{-\lambda t}, \quad t > 0, \quad \lambda > 0$. Esta es la función de supervivencia de la distribución exponencial, porque la función de densidad es $f(t) = \lambda(t)S(t) = \lambda e^{-\lambda t}$.

El problema de la distribución exponencial es que, salvo en procesos industriales, es difícilmente sostenible que la supervivencia se defina por una tasa de riesgo constante. La distribución exponencial queda caracterizada por la función de riesgo Si λ es grande indica alto riesgo y pequeña supervivencia; mientras que si λ es pequeña indica bajo riesgo y alta supervivencia.

Se obtiene que el estimador máximo-verosimíl de λ es: $\hat{\lambda} = \dfrac{d}{\sum_{i=1}^{n} t_i}$

El intervalo de confianza para λ al nivel de confianza $1-\alpha$ es:

$$\left(\frac{\hat{\lambda}\chi^2_{2d}\left(1-\frac{\alpha}{2}\right)}{2d}, \frac{\hat{\lambda}\chi^2_{2d}\left(\frac{\alpha}{2}\right)}{2d} \right)$$

donde $\chi^2_g(x)$ es el cuantil x de una χ^2 con g grados de libertad.

Existen otras distribuciones alternativas, entre las cuales, la más utilizada es la distribución de *Weibull*. Los modelos obtenidos son una importante generalización de la distribución exponencial y permiten una dependencia temporal del riesgo. La función de riesgo toma la forma $\lambda(t) = \alpha\beta t^{\beta-1} \quad 0 \le t < \infty$, donde los parámetros α (*parámetro de escala*) y β (*parámetro de forma*) son dos constantes positivas.Si $\beta = 1$, la función de riesgo es constante, con lo que los tiempos de supervivencia siguen una distribución exponencial. Para otros valores de β, la función de riesgo crece o decrece de forma monótona .

Para $\lambda(t) = \alpha\beta t^{\beta-1} \quad 0 \le t < \infty$ tenemos que: $S(t) = e^{-\alpha t^{\beta}}$ por lo que $f(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}}$ y estamos ante la función de densidad de la variable *Weibull.*

La esperanza matemática y la varianza son respectivamente:

$$E[T] = \Gamma\left(\frac{1+\frac{1}{\beta}}{\alpha}\right) \qquad Var[T] = \frac{1}{\alpha^2}\left[\Gamma\left(1+\frac{2}{\beta}\right) - \Gamma^2\left(1+\frac{1}{\beta}\right)\right]$$

donde Γ es la función gamma de Euler.

Puede demostrarse que el estimador máximo-verosímil del parámetro α es: $\hat{\alpha} = \dfrac{d}{\sum_{i=1}^{n} t_i^{\beta}}$

Para hallar $\hat{\beta}$ hay que resolver una ecuación no lineal, y en consecuencia deberemos de resolverla mediante un método iterativo como el de Newton-Raphson.

Existen otros modelos típicos en el análisis de la supervivencia, como por ejemplo, el *modelo log-logístico*, cuya función de riesgo es $\lambda(t) = \frac{e^{\theta} k t^{k-1}}{1 + e^{\theta} t^{k}}$ siendo la función de supervivencia $S(t) = \left(1 + e^{\theta} t^{k}\right)^{-1}$ y siendo la función de densidad $f(t) = \frac{e^{\theta} k t^{k-1}}{\left(1 + e^{\theta} t^{k}\right)^{2}}$ que es la función de densidad de una variable log-logística.

2.2.- Modelos de duración para poblaciones heterogéneas.

El análisis, hasta ahora, no ha tenido en cuenta la posible heterogeneidad de la población. Una forma de recoger la heterogeneidad consiste en la introducción de variables regresoras en el modelo.

El efecto de las variables regresoras sobre la distribución de la duración dependerá de la especificación adoptada en el modelo. En los modelos de regresión ordinaria el efecto de las variables explicativas consistirá en aumentar ó disminuir la media de la distribución de la variable dependiente. En el análisis de duración, el modelo más utilizado para recoger el efecto de las variables explicativas relacionadas con la variable duración es el modelo de función de riesgo proporcional derivado por Cox (1972). La característica fundamental de estos modelos es que diferentes individuos tienen funciones de riesgo proporcionales, es decir, la razón de las funciones de riesgo entre dos individuos con distintos vectores de variables regresoras no depende de t. Por tanto la función de riesgo podrá expresarse como un producto entre una función que depende de la duración y otra que depende del vector de variables regresoras. Esto es: $\lambda(t, x) = \lambda_0(t) \cdot \lambda(x, \beta)$ donde $\lambda_0(t)$ es conocida como la función de riesgo básica. Por tanto, se puede apreciar que el efecto de las variables regresoras consiste en multiplicar a la función de riesgo por un factor de escala. La forma funcional $\lambda(x, \beta)$ habitualmente elegida es $\lambda(x, \beta) = \exp(x\beta)$. Con esta especificación garantizamos la no negatividad de la función de riesgo sin imponer restricciones sobre los parámetros β. Una de las razones fundamentales para el uso tan extensivo de este modelo se debe a la posibilidad de estimar el modelo sin suponer una distribución concreta para la duración, por lo general desconocida. Es decir, vamos a disponer de un modelo muy flexible ya que puede ser estimado sin especificar una forma funcional concreta para la función de riesgo básica. En la regresión de Cox, a diferencia de los métodos anteriores, se supone que existe un conjunto de variables independientes $X_1, X_2, \ldots, X_p$, cuyos valores valores influyen en el tiempo que transcurre hasta que ocurre el suceso final. El modelo que se postula es: $\lambda(t, x) = \lambda_0(t) \cdot \lambda(x, \beta)$. Es decir, se supone que la función de ries-

go se puede expresar como el producto de una función de t y otra función que únicamente depende de $X_1,X_2;.....,X_p$. En particular si: $\lambda(x,\beta) = e^Z$ siendo Z la combinación lineal: $Z = \sum_{j=1}^{p} \beta_j X_j = \beta_1 X_1 + ... + \beta_p X_p$ tenemos el modelo de regresión de Cox.

El análisis consistirá entonces en estimar los parámetros desconocidos $\beta_1,....,\beta_p$. Observemos que, si las estimaciones de todos los parámetros fueran nulas, significaría que las variables $X_1,X_2;...,X_p$ no influyen en el tiempo transcurrido hasta que ocurre el suceso final. En dicho caso, la función $h(x,\beta)$ sería igual a 1 y en consecuencia $\lambda(t,x) = \lambda_0(t) \cdot \lambda(x,\beta) = \lambda_0(t)$. La función de supervivencia, S(t), probabilidad de que el suceso final no ocurra hasta pasado un período de tiempo superior o igual a t, puede obtenerse, mediante una relación matemática, directamente a partir de la función de riesgo: $S(t) = \exp\left\{ -\int_0^t \lambda(t,x)\, ds \right\}$. Por ello, una vez estimados los parámetros del modelo, además de la estimación de la función de riesgo se obtendrá la estimación de la función de supervivencia para cada instante t.

El criterio para obtener los coeficientes $B_1,.....,B_p$, estimaciones de los parámetros desconocidos $\beta_1,....., \beta_p$ es el de máxima verosimilitud. A partir de $B_1,.....,B_p$, la estimación de Z será: $\hat{Z} = \sum_{j=1}^{p} B_j X_j = B_1 X_1 + ... + B_p X_p$ y en consecuencia, la estimación de $\lambda(t,\beta)$ será: $\hat{\lambda}(t,\beta) = e^{\hat{Z}} = \left(e^{B_1}\right)^{X_1} .....\left(e^{B_p}\right)^{Xp}$. Luego para valores fijos de los restantes términos, cuanto mayor sea el coeficiente B_i mayor será la estimación de $\lambda(t,\beta)$ o, lo que es lo mismo, la de $\lambda(t, x)$. En otras palabras, mayor será la probabilidad estimada de que el suceso final ocurra en un pequeño intervalo $(t, t + \Delta t)$, supuesto que no ha ocurrido antes del instante t. Comprobar la bondad del ajuste es analizar cuán probables son los resultados muestrales a partir del modelo ajustado. La probabilidad de los resultados obtenidos se denomina verosimilitud. Para comprobar si la verosimilitud difiere de 1 (que el modelo se ajusta perfectamente a los datos) se utiliza el estadístico: -2LL= -2 (l*ogaritmo de la verosimilitud* **).**

3.- Resultados.

3.1. Método de Kaplan-Meier.

Las tablas 1, 2, 3, 4 y 5 contienen los valores estimados de las características más importantes de las funciones de supervivencia para la población global y para las subpoblaciones definidas por los diferentes niveles en las competencias consideradas

TABLA 1. Población Global

	Media	*Mediana*	*Perc 25*	*Perc. 75*	*I.C. Media*	*Des. Típ.*
Global	3.63	2.00	6.00	1.00	2.72-4.54	0.46

TABLA 2. Competencia 1: Motivación para el Trabajo

	Media	*Mediana*	*Perc 25*	*Perc. 75*	*I.C. Media*	*Des. Típ.*
Estrato 0	5.88	2.00	9.00	1.00	2.58-9.17	1.68
Estrato 1	5.33	2.45	9.00	2.00	1.36-9.31	2.03
Estrato 2	3.67	4.00	6.00	1.00	0.82-6.51	1.45
Estrato 3	4.73	5.00	6.00	1.00	2.99-6.46	0.89
Estrato 4	2.40	1.00	3.00	1.00	1.19-3.61	0.62
Estrato 5	1.91	1.00	2.00	1.00	1.01-2.80	0.46

TABLA 3. Competencia 2: Conocimientos Teóricos.

	Media	*Mediana*	*Perc 25*	*Perc. 75*	*I.C. Media*	*Des. Típ.*
Estrato 0	5.44	2.00	9.00	1.00	2.41-8.48	1.55
Estrato 3	3.70	3.00	5.00	2.00	2.24-5.16	0.75
Estrato 4	2.89	1.00	6.00	1.00	1.60-4.19	0.66
Estrato 5	3.23	2.00	5.00	1.00	1.77-4.69	0.74

TABLA 4. Competencia 3: Conocimientos prácticos

	Media	*Mediana*	*Perc 25*	*Perc. 75*	*I.C. Media*	*Des. Típ.*
Estrato 0	5.44	2.00	9.00	1.00	2.41-8.48	1.68
Estrato 2	3.00	1.00	5.00	1.00	0.00-6.92	1.45
Estrato 3	4.70	4.00	6.00	2.00	2.86-6.54	0.89
Estrato 4	1.94	1.00	2.00	1.00	1.22-2.66	0.62
Estrato 5	3.69	3.00	6.00	1.00	2.03-5.36	0.46

TABLA 5. Competencia 4: Trabajo en equipo

	Media	*Mediana*	*Perc 25*	*Perc. 75*	*I.C. Media*	*Des. Típ.*
Estrato 0	5.44	2.00	9.00	1.00	2.41-8.48	1.55
Estrato 4	3.85	4.00	6.00	1.00	2.45-5.25	0.71
Estrato 5	2.96	1.00	4.00	1.00	1.96-3.97	0.51

La tabla 6 contiene los valores de los estadísticos de contraste Log- Rank y Tarone-Ware, y los niveles de significación para la comparación entre las diferentes funciones de supervivencia según los niveles adquiridos por el egresado en la competencia motivación para el trabajo.

Tabla 6

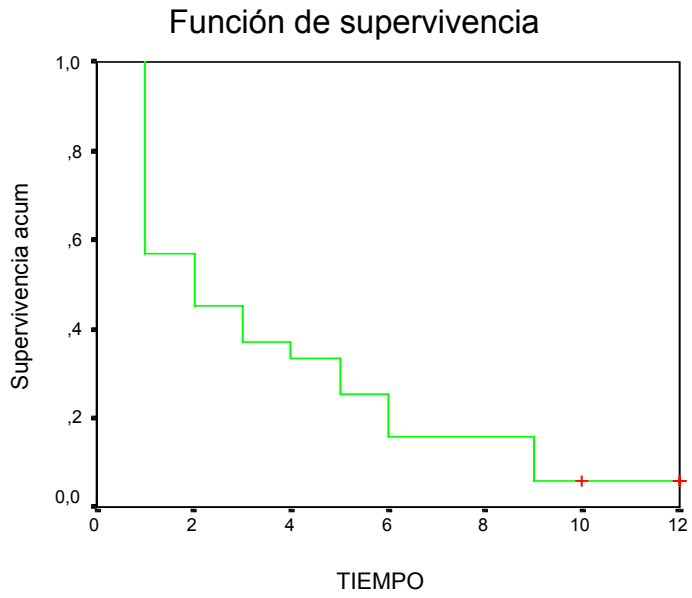
	Log Rank Statistic						*Tarone-Ware Statistic*					
Factor	**0**	**1**	**2**	**3**	**4**		**Factor**	**0**	**1**	**2**	**3**	**4**
1	0,19 (0,66)						**1**	0,02 (0,87)				
2	1,04 (,30)	0,48 (0,48)					**2**	0,58 (,44)	0,44 (0,50)			
3	1,37 (0,24)	0,06 (0,81)	0,51 (0,47)				**3**	0,64 (0,42)	0,05 (0,82)	0,42 (0,51)		
4	2,68 (0,10)	1,21 (0,27)	0,62 (0,42)	3,18 (0,07)			**4**	2,37 (0,12)	2,40 (0,12)	1,07 (0,30)	4,49 (0,03)	
5	4,68 (0,03)	3,20 (0,07)	1,53 (0,21)	6,06 (0,01)	0,24 (0,62)		**5**	3,15 (0,07)	3,57 (0,058	1,55 (0,21)	5,60 (0,01)	0,09 (0,76)

3.2.- Gráficas de las funciones de supervivencia.

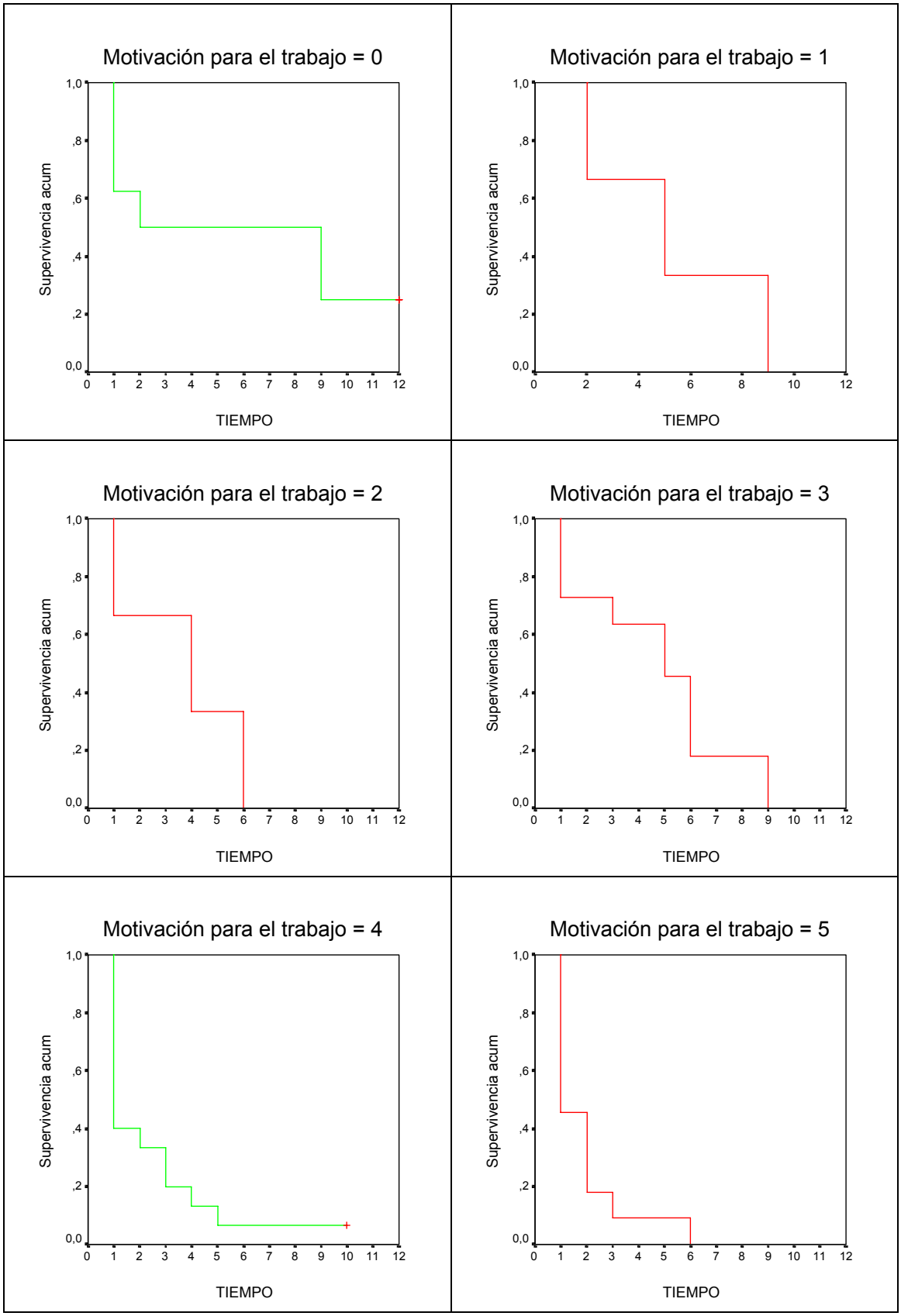
Las gráficas siguientes corresponden a las estimaciones de las funciones de supervivencia para la población global y para las subpoblaciones definidas por los diferentes niveles en la competencia Motivación para el trabajo

Gráfica 1

Población Global



Gráfica 2. Competencia 1: Motivación para el Trabajo



3.2.- Regresión de Cox.

Con el modelo de Cox pretendemos valorar el poder explicativo que las competencias desarrolladas en el egresado tienen en la función de supervivencia.

Los resultados sobre la estimación de parámetros, variables significativas, etc., aparecen en las siguientes tablas:

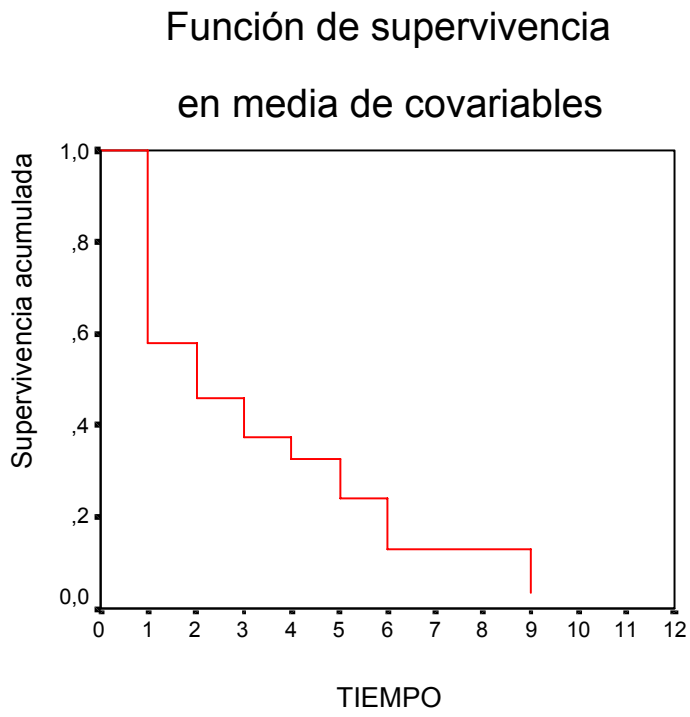
Tabla 6: Variables en la ecuación

		B	ET	Wald	Gl	Sig.	Exp(B)	IC para exp(B)	
Paso 1	**C. Teórico**	-.145	.229	.402	1	.526	.865	0.55	1.35
	C. Práctico	-.115	.165	.487	1	.485	1.122	0.81	1.55
	T. Equipo	-.056	.191	.085	1	.771	1.057	0.72	1.53
	Satisf. Salario	-.136	.178	.583	1	.445	.873	0.61	1.23
	Motivación Trabajo	-.316	.145	4.718	1	.030	1.371	1.03	1.82
Paso 2	**C. Teórico**	-.107	.187	.326	1	.568	.899	0.62	1.29
	C. Práctico	.125	.162	.599	1	.439	1.133	0.82	1.55
	Satisf. Salario	-.129	.177	.530	1	.467	.879	0.62	1.24
	Motivación Trabajo	.318	.145	4.776	1	.029	1.374	1.03	1.82
Paso 3	**C. Práctico**	.064	.123	.270	1	.603	1.066	0.83	1.35
	Satisf. Salario	-.151	.173	.763	1	.382	.859	0.61	1.20
	Motivación Trabajo	.295	.141	4.365	1	.037	1.343	1.01	1.77
Paso 4	**Satisf. Salario**	-.111	.157	.503	1	.478	.895	0.65	1.21
	Motivación Trabajo	.307	.141	4.737	1	.030	1.359	1.03	1.79
Paso 5	**Motivación Trabajo**	.234	.098	5.657	1	.017	1.264	1.04	1.53

Tabla 7: Medias de Covariables

C. Teórico	C. Práctico	Trabajo Equipo	Satisf. Salario	Motiva Trabajo
3.353	3.275	3.824	2.725	3.078

En este caso la curva de supervivencia para los valores medios de las variables explicativas tiene la forma siguiente:



4.- Conclusiones.

El objetivo de este trabajo, además de presentar las ideas básicas de la metodología clásica en el análisis de supervivencia, es justificar su validez para modelizar la duración del tiempo que trascurre hasta que un egresado obtiene el primer empleo significativo.

Los datos utilizados corresponden a la cohorte de egresados, 1997-2000, de la titulación " Diplomatura en Enfermería de la Universidad Católica San Antonio de Murcia". El grupo de egresados responde a las características "ser mujer" (90%), de edad media 25 años, y que en un 85% encuentra su primer empleo significativo antes de los 15 meses.

Utilizando parte de la información disponible hemos aplicado un análisis no paramétrico para realizar una primera descripción de la variable "tiempo de duración hasta el primer empleo significativo" para el supuesto de población homogénea. Se estiman, por Kaplan Meier, las características de la variable modelizada, ver tabla 1, y se obtiene que con una probabilidad del 95%,

el intervalo (2.72, 4.54) contiene al tiempo medio que un egresado tarda en encontrar el primer empleo.

La tabla 2 contiene las características estimadas para cada subpoblación definida por el nivel de competencia "motivación para el trabajo". Es de destacar que el tiempo medio estimado para el nivel de máxima motivación es significativamente inferior al de nivel mínimo. En el resto de covariables consideradas no se aprecia una influencia significativa en el tiempo de búsqueda de empleo. La afirmación anterior se confirma con los valores de la significación (0,0306) de la tabla "Long Rank Statistic".

En segundo lugar, se ha estimado el modelo de Regresión de Cox considerando como covariables las siguientes cinco competencias: conocimientos teóricos, conocimientos prácticos, motivación para el trabajo, satisfacción con el salario y trabajo en equipo. La única variable significativa en el modelo es "motivación para el trabajo" con significación 0,013. El coeficiente en la ecuación para la variable significativa es 0,234 lo que significa que un incremento de un punto en el nivel de motivación para el trabajo repercute en una disminución del 23,4% del tiempo que tarda el egresado en encontrar el primer empleo.

Por último destacar:

- ✓ La singularidad del tipo de egresados considerados, ya que, después de un periodo no demasiado elevado, las observaciones son completas y no censuradas.
- ✓ La validez de la metodología para modelizar este tipo de variables de tan significada importancia socioeconómica.

5.- Bibliografía

1. Breslow, N. & Crowley, J. (1974): "A large sample study of the life table and product limit estimates under random censorship". *The Annals of Statistics*, **2**, 437-453.
2. Collett, D. (1994), *Modelling Survival Data in Medical Research*, Chapman and Hall: London.
3. Cox, D. R. (1972), "Regression models and life-tables", *Journal of the Royal Statistical Society-Series* B, **34**, 187-220.
4. Cox,D. R. & Oakes, D. (1984), *Analysis of Survival Data*, Chapman and Hall: New York.
5. Grrotings, Peter. (1994), From qualification to competence: what are we talking ab *European Journal Vocational Training*. Thessaloniki.
6. Kalbfleisch, J. D. & Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, John Wiley and Sons: New York.
7. Kaplan, E. L: & Meier, P. (1958), "Nonparametric estimation from incomplete observations", *Journal of the American Statistical Association*, **53**, 457-481.
8. Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons: New York.
9. Le Boterf, Guy, (1991), L'ingenierie des competences.
10. Lee, E. T. (1980), *Statistical Methods for Survival Data Analysis*, Lifetime Learning Publications: Belmont, California.
11. Mertens,Leonard.(1996) Competencia Laboral: sistemas, surgimiento y modelos.

12. Orbe, J. (1999), "Un modelo de regresión parcial censurado para análisis de supervivencia" , Tesis doctoral. Universidad del País Vasco.
13. Orbe, J. & Ferreira, E. & Núñez-Antón V. (2001), "Modelling the duration of firms in Chapter 11 bankruptucy using a flexible model", *Economics Letters* **71**, 35-42.
14. Orbe, J. & Ferreira, E. & Núñez-Antón V. (2002), "Censored partial regression ", *Biostatistics* **4**, 109-121.
15. Orbe, J. & Ferreira, E. & Núñez-Antón V. (2002), "Comparing proportional hazards and accelerated failure time models for survival analysis", *Statistics in medicine* **21**, 3493-3510.
16. Orbe, J. & Ferreira, E. & Núñez-Antón V. (2003), "Modelling the duration of firms in Chapter 11 bankruptucy using a flexible model", *Economics Letters 71*, 35-42
17. Stute, W. & Wang, J. L. (1993), "The strong law under random censorship", *The Annals of Statistics*, **21**, 1591-1607.
18. Stute, W. (1993), "Consistent estimation under random censorship when covariables are present", *Journal of Multivariate Analysis*, **45**,89-103.
19. Stute, W.(1995a), "The central limit theorem under random censorship", *The Annals of Statistics*, **23**, 422-439.