

# LA DURACIÓN DEL TIEMPO DE BÚSQUEDA DEL PRIMER EMPLEO

**Josefina García Lozano**

Universidad Católica San Antonio

e-mail: [JGlozano@pdi.ucam.edu](mailto:JGlozano@pdi.ucam.edu)

**Mercedes Carmona Martínez**

Universidad Católica San Antonio

[Mcarmona@pdi.ucam.edu](mailto:Mcarmona@pdi.ucam.edu)

**Juan C. Gómez Gallego**

Universidad Católica San Antonio

[Jcandido@pdi.ucam.edu](mailto:Jcandido@pdi.ucam.edu)

## **Resumen**

El objetivo de esta ponencia se centra en el análisis de la duración de la búsqueda del primer empleo de los egresados universitarios, que es un aspecto clave a la hora de valorar el proceso de inserción laboral de los jóvenes titulados. Para llevarlo a cabo aplicamos la metodología del análisis de supervivencia, se estiman distintos modelos de duración en base a la información suministrada por una Encuesta de Inserción Laboral realizada por la Universidad Católica San Antonio (referida a los años 2001-2002) y tomando los datos de los egresados de las titulaciones de Administración y Dirección de Empresa, Diplomado en Enfermería y Licenciado en Publicidad y Relaciones Públicas.

Se contrasta la influencia que los diferentes niveles de competencia adquiridos en el proceso formativo pueden ejercer sobre el tiempo de búsqueda del primer empleo significativo. Para ello se estiman modelos de función de riesgo proporcional (Cox, 1972) y modelos log-lineales, analizando el efecto de las covariables consideradas sobre la duración. Los resultados confirman que para cada tipo de titulación existen determinadas competencias que adquiridas en un nivel suficiente facilitan la inserción laboral del universitario.

*Palabras clave:* análisis de supervivencia, modelos log-lineales, competencias, inserción laboral.

*Área temática:* Métodos cuantitativos

## 1.- Introducción

El desempleo constituye uno de los fenómenos socioeconómicos más preocupantes a los para la mayor parte de los países desarrollados y, en particular, la población juvenil se enfrenta cada día a mayores dificultades para acceder a un empleo. Este fenómeno es particularmente importante para aquellos que se incorporan por primera vez al mercado laboral a pesar del descenso de la población juvenil y de la menor participación laboral de este grupo poblacional (véanse Dolado *et al.*, 2000; Aguilar y Navarro, 2004), de manera que la búsqueda del primer empleo se convierte en un factor clave en el proceso de inserción de los jóvenes y su análisis adquiere una especial relevancia.

Los estudios sobre los determinantes de la duración del desempleo son abundantes en la literatura socioeconómica desde la década de los ochenta. Para un compendio de esta literatura puede verse, por ejemplo, Atkinson y Micklewright (1991). Sin embargo, existen pocos trabajos en nuestro país que utilicen el análisis de supervivencia para analizar la duración del desempleo juvenil. Cabe destacar el estudio de Ahn y Ugidos (1995), realizado a partir de la Encuesta de Condiciones de Vida y Trabajo (1985), y los de Ahn *et al.* (1997) o Alba (1998) que emplean la Encuesta de Población Activa (EPA) y más recientemente, Davia y Smith (2001) que usan el Panel de Hogares de la Unión Europea (1994) para comparar las transiciones al empleo de los jóvenes en España y en el Reino Unido. Las investigaciones sobre la duración del período de búsqueda del primer empleo resultan aún más escasas, quizás por la falta de información estadística adecuada. Así, Cañada (1999) estima modelos de duración a partir de los ficheros de la EPA enlazada, Salas (1999) analizan la situación de los titulados de las Universidades de Granada, Albert *et al.* (2003) emplean el módulo específico de transición de la educación al mercado laboral incluido en la EPA en el segundo trimestre del año 2000. y Aguilar y Navarro (2004) analiza la duración de la búsqueda del primer empleo, a partir de una encuesta representativa del colectivo juvenil a nivel nacional, la Encuesta Sociodemográfica (INE, 1991).

Por otra parte, en el marco del proceso de Convergencia Europea de Educación Superior, se ha profundizado y dado forma al concepto de competencia aplicado al campo de la educación superior. Desde la Declaración de Bolonia, y en gran medida, a impulsos del proyecto Tuning algunas universidades se han abocado a las tareas de, primero, de-

finir el concepto de competencia, y luego, de definir lo que consideran competencias generales y competencias específicas, estas últimas ligadas a las diferentes titulaciones.

El proyecto Tuning define la competencia como:

“una combinación dinámica de atributos, en relación a conocimientos, habilidades, actitudes y responsabilidades, que describen los resultados del aprendizaje de un programa educativo o lo que los alumnos son capaces de demostrar al final de un proceso educativo”.

Las competencias generales definidas en el marco del proyecto Tuning europeo, apuntan al logro de determinados resultados de formación general, de tipo transversal, que sintonizan fuertemente con los objetivos de cada Titulación en el seno de los objetivos y razón de ser de cada Universidad. En esta primera etapa, aunque con diferentes denominaciones, un grupo representativo de las competencias genéricas para todos los estudiantes de grado de Universidad podría ser:

Comunicación (comunicar y presentar argumentos orales y escritos)

Uso de las tecnologías de la información y la comunicación

Competencia matemática (interpretar y presentar información numérica)

Competencia social (trabajo con otros como resultado del desarrollo de habilidades interpersonales)

Competencia de aprendizaje (mejorar el propio aprendizaje, desarrollo de habilidades de estudio y de investigación, búsqueda de información, capacidad de planificar, gestionar y reflexionar sobre el propio aprendizaje)

Resolución de problemas

Este trabajo se realiza con datos sobre los egresados de la Universidad Católica San Antonio de Murcia y donde se han definido y valorado las cinco competencias transversales aproximando por las siguientes conceptos: conocimientos teóricos, conocimientos prácticos, motivación para el trabajo, capacidad para el trabajo en equipo y satisfacción en el trabajo. El objetivo que se pretende es valorar como factores determinantes del tiempo que tardan los graduados en encontrar el primer empleo significativo, el nivel alcanzado en cada una de las competencias consideradas.

La estructura del trabajo es la siguiente. En la segunda sección, se describen la metodología así como los datos y las variables utilizadas en el análisis empírico. En la tercera sección se especifican los modelos de duración estimados, mientras que en la cuarta se

interpretan los resultados obtenidos. Por último, se resumen las principales conclusiones.

## **2.- Metodología.**

Se conoce como análisis de duración o de supervivencia al análisis de datos en el que se estudia la distribución del 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 como la exponencial, Weibull o Gamma, entre otras. Una segunda característica asociada a este tipo de datos es la censura<sup>1</sup>. Como tercera característica señalamos que para obtener duraciones de individuos hay que realizar un seguimiento de estos a lo largo de un período determinado, y puede ocurrir que haya 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.

Las primeras referencias se pueden situar en los años donde se analizaba el tiempo de duración de un suceso utilizando la metodología de las tablas de vida mediante la estimación Producto Límite derivada por Kaplan y Meier (1958). Es en el año 1972, cuando aparece un trabajo relevante para estos datos, el modelo de regresión de función de riesgo proporcional propuesto por Cox. 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.

---

<sup>1</sup> La duración de un evento está censurada cuando su tiempo de fallo no ha sido observado dentro del período de estudio.

Dentro de estos modelos se situarían los propuestos por Stute (1993), cuyos parámetros de interés pueden estimarse sin especificar la distribución de la duración.

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, 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  $t$ :

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T \leq t + \Delta t / T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} .$$

La función de densidad, la función de riesgo y la función de supervivencia están claramente relacionadas mediante:

$$f(t) = S(t)\lambda(t) \quad \text{y} \quad \lambda(t) = \frac{-d \ln S(t)}{dt} .$$

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  $H(t) = -\ln S(t)$ .

### ***2.1.- Modelos de duración para poblaciones homogéneas.***

Por población homogénea entendemos una población donde todos los individuos presentan las mismas características o lo que significa lo mismo, 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 definidas 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 del 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.

El método de Kaplan-Meier o del producto límite es un procedimiento donde se estima la curva de supervivencia calculando la supervivencia como producto de probabilidades condicionadas, 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  $0 = t_0 < t_1 < \dots < t_n$ , 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:

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

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.

Si notamos por:

$n_t$ : 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  $t$ .

$$p_t = \frac{n_t - d_t}{n_t} = 1 - \frac{d_t}{n_t}$$

$p_i$  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:

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

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 más muestras tienen funciones de supervivencia similares. Dos procedimientos no paramétricos frecuentemente 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 del 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  $S(t) = e^{-\lambda t}$ ,  $t > 0$ ,  $\lambda > 0$ . Esta es la función de supervi-

vencia 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  $\lambda$  es grande indica alto riesgo y pequeña supervivencia; mientras que si  $\lambda$  es pequeña indica bajo riesgo y alta supervivencia.

Se obtiene que el estimador máximo-verosímil de  $\lambda$  es:

$$\hat{\lambda} = \frac{d}{\sum_{i=1}^n t_i}$$

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

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

donde  $\chi_g^2(x)$  es el cuantil x de una  $\chi^2$  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  $\lambda(t) = \alpha\beta t^{\beta-1}$   $0 \leq t < \infty$ , donde los parámetros  $\alpha$  (parámetro de escala) y  $\beta$  (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  $\beta$ , la función de riesgo crece o decrece de forma monótona.

Para  $\lambda(t) = \alpha\beta t^{\beta-1}$   $0 \leq t < \infty$  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:

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

donde  $\Gamma$  es la función gamma de Euler.

El estimador máximo-verosímil del parámetro  $\alpha$  es:

$$\hat{\alpha} = \frac{d}{\sum_{i=1}^n t_i^\beta}$$

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  $\lambda(t) = \frac{e^\theta kt^{k-1}}{1 + e^\theta t^k}$  siendo la función de

supervivencia  $S(t) = (1 + e^\theta t^k)^{-1}$  y siendo la función de densidad  $f(t) = \frac{e^\theta kt^{k-1}}{(1 + e^\theta t^k)^2}$  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  $\beta$ . 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  $X_1, X_2, \dots, X_p$ , cuyos 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 riesgo se puede expresar como el producto de una función de  $t$  y otra función que únicamente depende de  $X_1, X_2, \dots, X_p$ . En particular si:  $\lambda(x, \beta) = e^Z$  siendo  $Z$  la combinación li-

near:  $Z = \sum_{j=1}^p \beta_j X_j = \beta_1 X_1 + \dots + \beta_p X_p$  tenemos el modelo de regresión de Cox.

El análisis consistirá entonces en estimar los parámetros desconocidos  $\beta_1, \dots, \beta_p$ . Observemos que, si las estimaciones de todos los parámetros fueran nulas, significaría que las variables  $X_1, X_2, \dots, X_p$  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:  $S(t) = \exp\left\{-\int_0^t \lambda(s, x) ds\right\}$ . 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  $B_1, \dots, B_p$ , estimaciones de los parámetros desconocidos  $\beta_1, \dots, \beta_p$  es el de máxima verosimilitud. A partir de  $B_1, \dots, B_p$ , la

$$\hat{Z} = \sum_{j=1}^p B_j X_j = B_1 X_1 + \dots + B_p X_p$$
 estimación de Z será: y en consecuencia, la estimación de  $\lambda(t, \beta)$  será:  $\hat{\lambda}(t, \beta) = e^{\hat{Z}} = (e^{B_1})^{X_1} \dots (e^{B_p})^{X_p}$ . 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:  $-2LL = -2$  ( *logaritmo de la verosimilitud* ).

### 3.- Resultados.

Se ha aplicado el modelo descrito a los alumnos de Enfermería que finalizaron sus estudios en 2000 en la Universidad Católica San Antonio de Murcia. Son un total de 80 egresados a los que se ha encuestado, obteniendo así información acerca de las siguientes variables:

Meses transcurridos desde que finalizaron sus estudios hasta que encontraron el primer empleo (t).

Importancia de las competencias teóricas adquiridas en su periodo de formación, valoradas de 0 a 5 por el encuestado (CT).

Importancia de las competencias prácticas adquiridas en su periodo de formación, valoradas de 0 a 5 por el encuestado (CP).

Importancia de los hábitos de trabajo en equipo, valorados de 0 a 5 por el encuestado (EQ). Esta variable está directamente relacionada con lo que antes denominábamos “habilidades sociales”.

Satisfacción del encuestado en el trabajo, valorada de 0 a 5 por el encuestado (SAL).

Motivación del encuestado para el trabajo que realiza, valorada de 0 a 5 por el encuestado (MO).

Las valoraciones medias de cada una de los aspectos citados varían considerablemente según el tiempo que el encuestado ha tardado en encontrar su primer empleo. Así, se puede comprobar (tabla 1) que la valoración media que los egresados hacen de las com-

petencias adquiridas en su periodo formativo es mucho más alta si han tardado menos de seis meses en encontrar empleo que si han tardado más de este tiempo; y además, la motivación (MO) es la única variable que desciende de forma continua a medida que se incrementa el tiempo necesario para encontrar el empleo.

*Tabla 1. Valores medios de las variables.*

<b>Tiempo tardado para encontrar empleo</b>	<b>CT</b>	<b>CP</b>	<b>EQ</b>	<b>SAL</b>	<b>MO</b>	<b>n</b>
<b>Menos de 3 meses</b>	3.197	3.131	3.656	2.574	3.230	61
<b>De 3 a 6 meses</b>	4.000	3.818	4.455	3.000	3.000	11
<b>De 6 a 9 meses</b>	2.400	2.600	2.800	2.000	1.400	5
<b>De 10 a 12 meses</b>	1.333	1.000	1.667	1.333	1.333	3

Por otra parte, es importante señalar la elevada correlación existente entre las cinco variables consideradas (tabla 2), especialmente entre competencias teóricas (CT) y prácticas (CP), trabajo en equipo (EQ) y remuneración (SAL). La variable motivación (MO) tiene una menor correlación con todas ellas.

*Tabla 2. Matriz de correlaciones.*

	<b>CT</b>	<b>CP</b>	<b>EQ</b>	<b>SAL</b>	<b>MO</b>
<b>CT</b>	1.0000	0.9912	0.9848	0.9507	0.6682
<b>CP</b>		1.0000	0.9947	0.9693	0.6973
<b>EQ</b>			1.0000	0.9795	0.6862
<b>SAL</b>				1.0000	0.7194
<b>MO</b>					1.0000

Las tablas, 3, 4 5, 6 y 7 contienen los valores estimados por aplicación del Método de Kaplan-Meier. 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.

TABLA3. Población Global

	Me- dia	Medi- ana	Per- c 25	Perc . 75	I.C. Me- dia	Des . Típ.
Glo- bal	3.63	2.00	6.00	1.00	2.72 - 4.54	0.4 6

TABLA 5. Competencia 2: Conocimientos Teóricos.

	Me- dia	Medi- ana	Per- c 25	Perc. 75	I.C. Me- dia	Des. Típ.
Es- tra- to 0	5.44	2.00	9.00	1.00	2.41- 8.48	1.55
Es- tra- to 3	3.70	3.00	5.00	2.00	2.24- 5.16	0.75
Es- tra- to 4	2.89	1.00	6.00	1.00	1.60- 4.19	0.66
Es- tra- to 5	3.23	2.00	5.00	1.00	1.77- 4.69	0.74

TABLA 6. Competencia 3: Conocimientos prácticos

TABLA 4. Competencia 1: Motivación para el Trabajo

	Me- dia	Medi- ana	Per- c 25	Perc . 75	I.C. Me- dia	Des . Típ.
Es- tra- to 0	5.88	2.00	9.0 0	1.00	2.58 - 9.17	1.6 8
Es- tra- to 1	5.33	2.45	9.0 0	2.00	1.36 - 9.31	2.0 3
Es- tra- to 2	3.67	4.00	6.0 0	1.00	0.82 - 6.51	1.4 5
Es- tra- to 3	4.73	5.00	6.0 0	1.00	2.99 - 6.46	0.8 9
Es- tra- to 4	2.40	1.00	3.0 0	1.00	1.19 - 3.61	0.6 2
Es- tra- to 5	1.91	1.00	2.0 0	1.00	1.01 - 2.80	0.4 6

TABLA . Competencia 4: Trabajo en equipo

	Me- dia	Medi ana	Per c 25	Perc . 75	I.C. Me- dia	Des . Típ.		Me- dia	Medi ana	Per c 25	Perc . 75	I.C. Me- dia	Des . Típ.
Es- trato 0	5.44	2.00	9.0 0	1.00	2.41 - 8.48	1.6 8	Es tra to 0	5.44	2.00	9.0 0	1.00	2.41 - 8.48	1.5 5
Es- trato 2	3.00	1.00	5.0 0	1.00	0.00 - 6.92	1.4 5	Es tra to 4	3.85	4.00	6.0 0	1.00	2.45 - 5.25	0.7 1
Es- trato 3	4.70	4.00	6.0 0	2.00	2.86 - 6.54	0.8 9	Es tra to 5	2.96	1.00	4.0 0	1.00	1.96 - 3.97	0.5 1
Es- trato 4	1.94	1.00	2.0 0	1.00	1.22 - 2.66	0.6 2							
Es- trato 5	3.69	3.00	6.0 0	1.00	2.03 - 5.36	0.4 6							

La tabla 8 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 8.- Log Rank Statistic

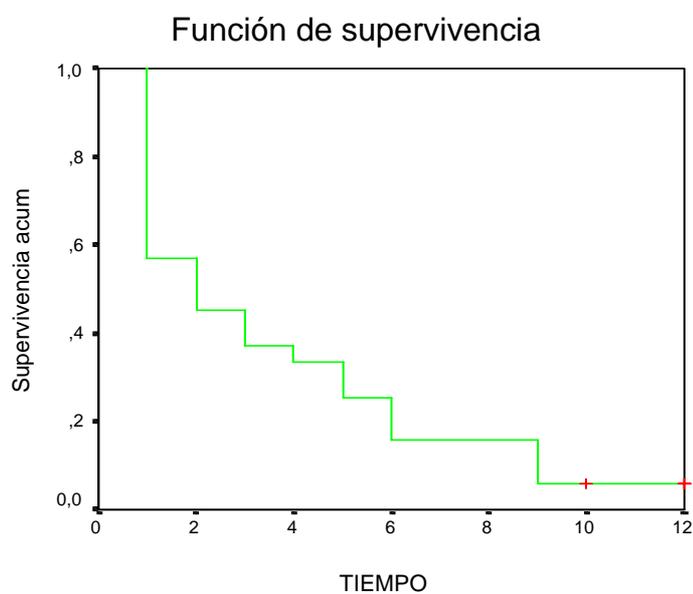
Tarone-Ware Statistic

Fac- tor	0	1	2	3	4	Fac- tor	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	0,06	0,51			3	0,64	0,05	0,42		

	(0,24)	(0,81)	(0,47)				(0,42)	(0,82)	(0,51)		
<b>4</b>	2,68 (0,10)	1,21 (0,27)	0,62 (0,42)	3,18 (0,07)		<b>4</b>	2,37 (0,12)	2,40 (0,12)	1,07 (0,30)	4,49 (0,03)	
<b>5</b>	4,68 (0,03)	3,20 (0,07)	1,53 (0,21)	6,06 (0,01)	0,24 (0,62)	<b>5</b>	3,15 (0,07)	3,57 (0,058)	1,55 (0,21)	5,60 (0,01)	0,09 (0,76)

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 las competencias consideradas:

*Gráfica 1 Población Global*



*Gráfica 2. Competencia 1: Motivación para el Trabajo*

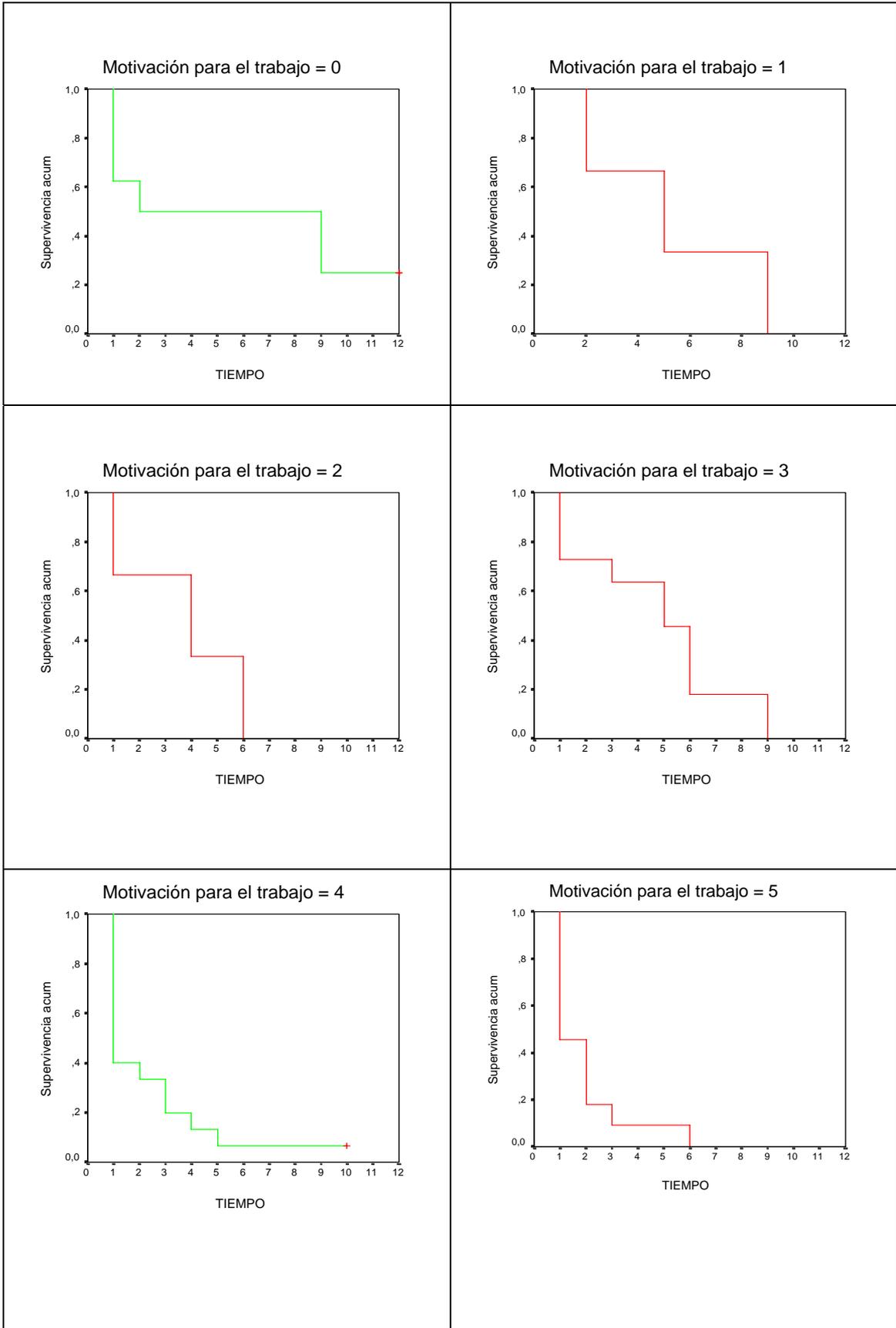
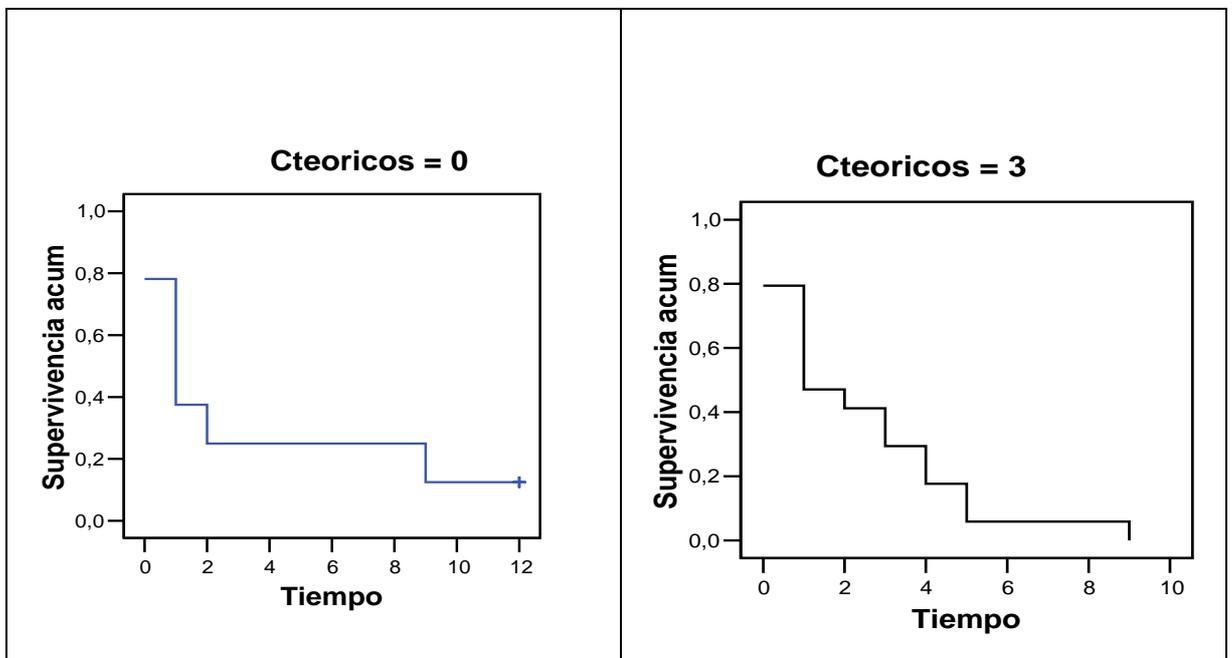


Gráfico 3. Competencia: Conocimientos Teóricos



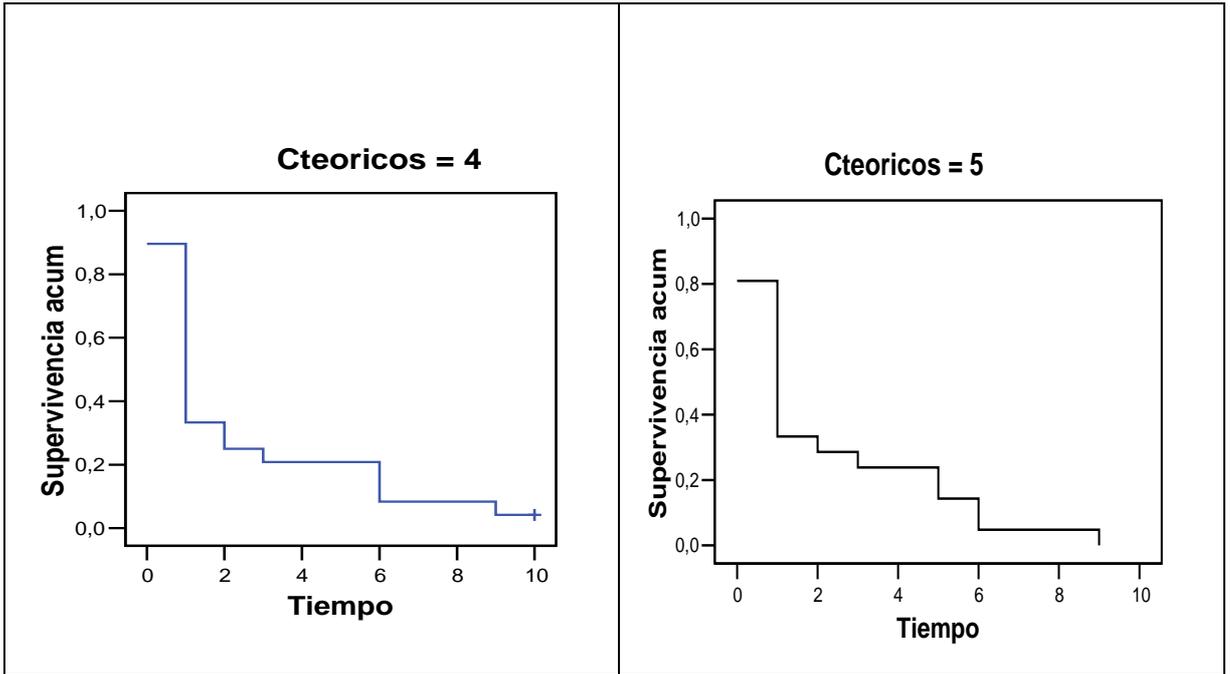
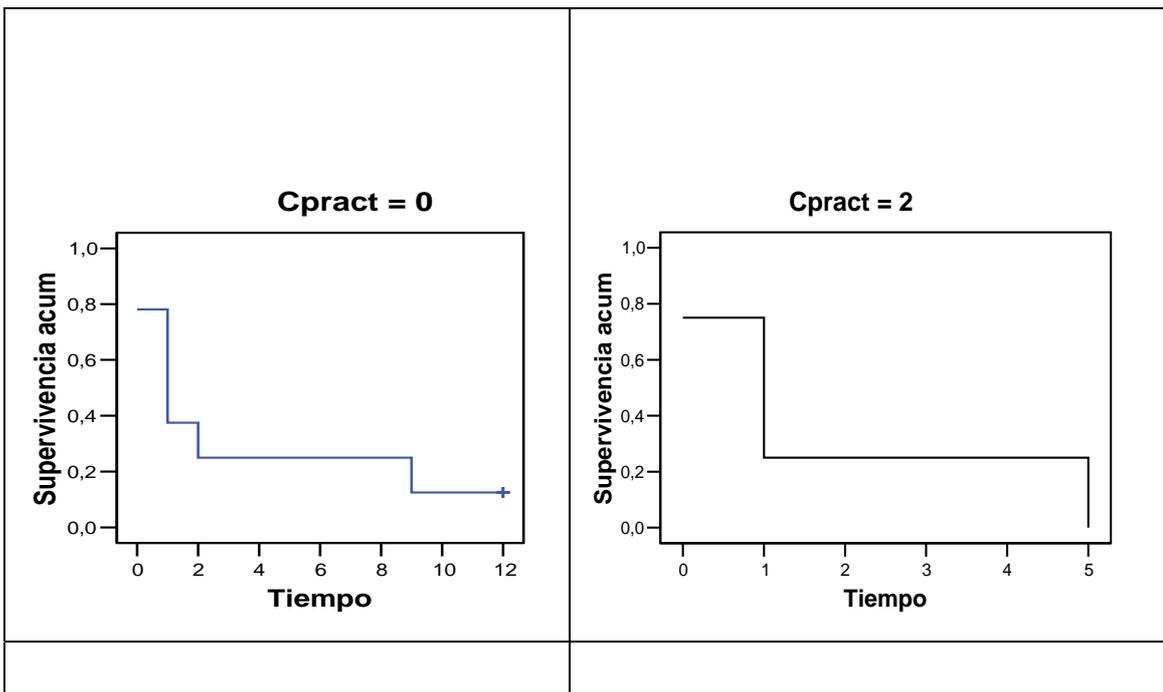


Gráfico 4. Competencia : Conocimientos Prácticos



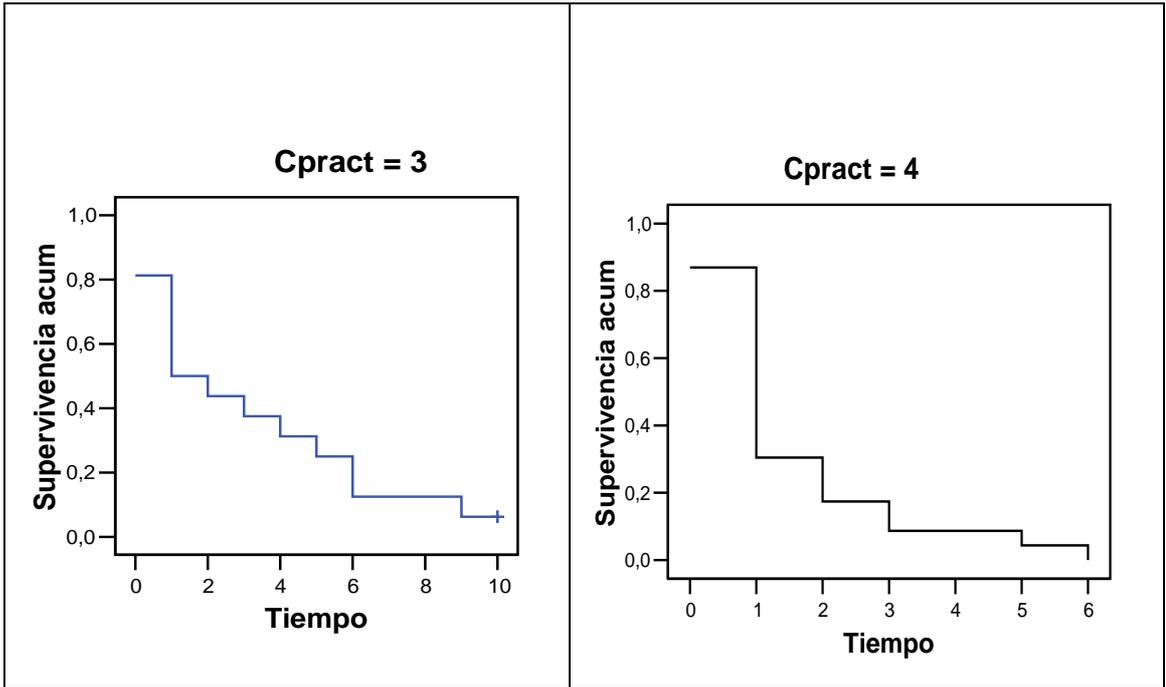
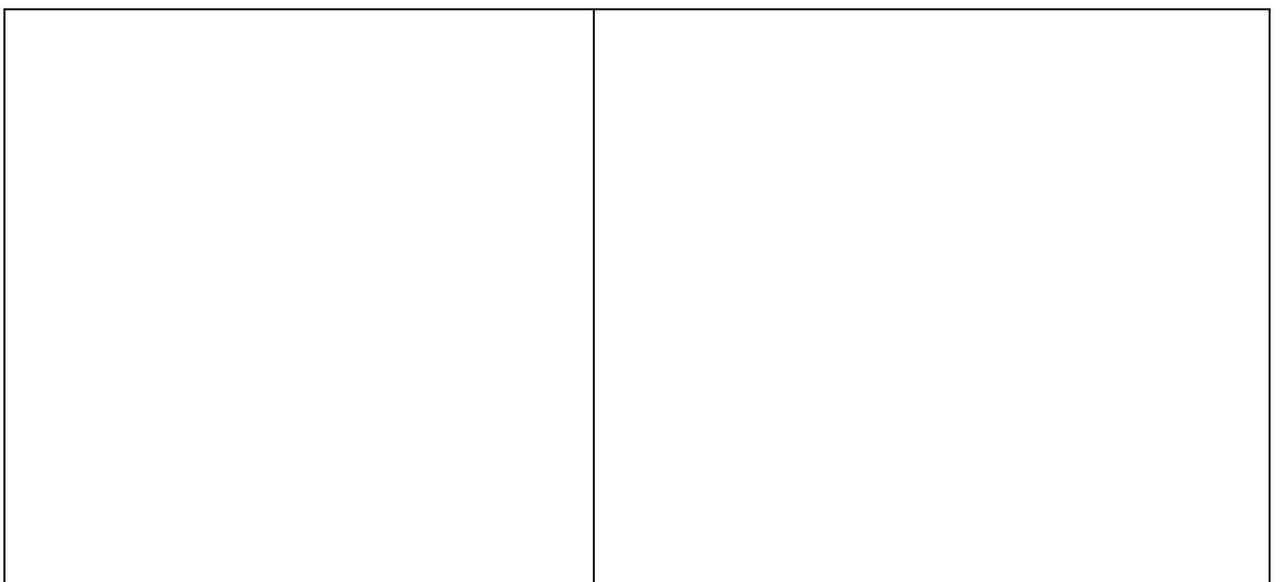
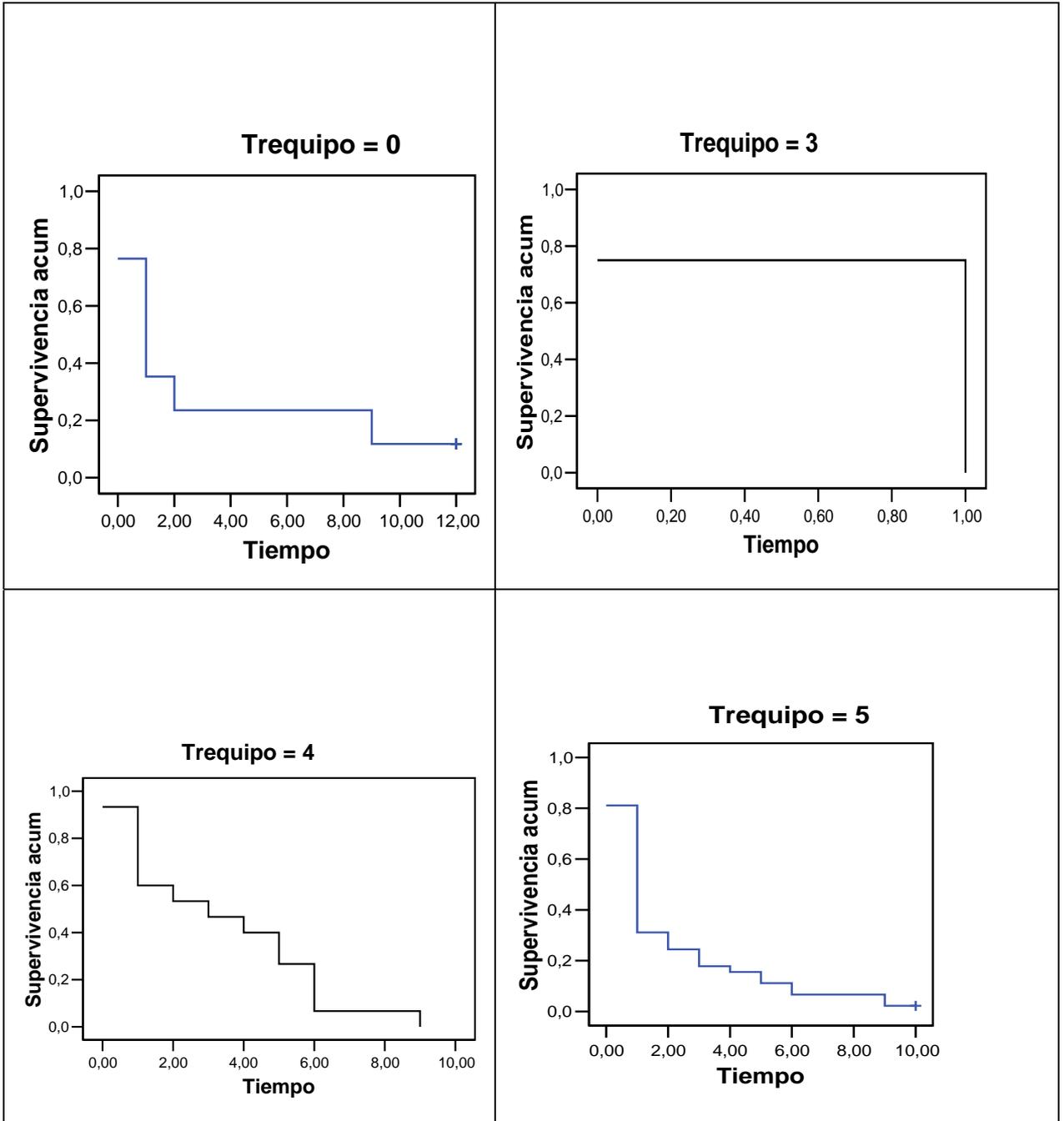


Gráfico 5. Competencia : Trabajo en equipo





*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 9: Variables en la ecuación*

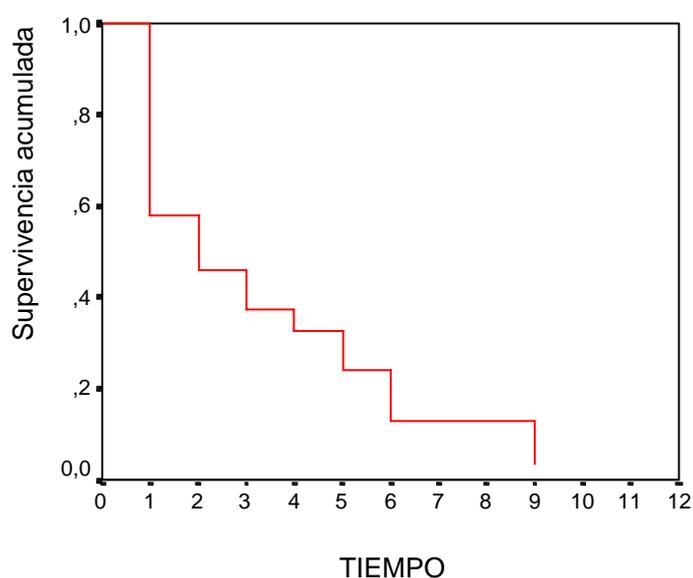
		<b>B</b>	<b>ET</b>	<b>Wald</b>	<b>Gl</b>	<b>Sig.</b>	<b>Exp(B)</b>	<b>IC para exp(B)</b>	
<b>Paso 1</b>	<b>C. Teórico</b>	-.145	.229	.402	1	.526	.865	0.55	1.35
	<b>C. Práctico</b>	-.115	.165	.487	1	.485	1.122	0.81	1.55
	<b>T. Equipo</b>	-.056	.191	.085	1	.771	1.057	0.72	1.53
	<b>Satisf. Salario</b>	-.136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

*Tabla 10: Medias de Covariables*

<b>Conocimientos Teóricos</b>	<b>Conocimientos-Práctico</b>	<b>Trabajo Equipo</b>	<b>Satisfacción Salario</b>	<b>Motivación Trabajo</b>
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:

**Función de supervivencia  
en media de covariables**



#### **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 transcurre 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, 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.

Se presentan las características estimadas para cada subpoblación definida para los niveles de las diferentes competencias. 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 señalar la validez de la metodología expuesta para modelizar este tipo de variables de tan significativa importancia socioeconómica.

## **5.- Referencias bibliográficas**

- Aguilar, M.I.; Navarro, M.L. y de la O, C. (2002): “Diferencias regionales en el comportamiento laboral de las jóvenes españolas”. *Revista de Estudios Regionales*, nº 62, págs. 123-146.
- Ahn, N. y Ugidos, A. (1995): “Desempleo juvenil en España: qué determina su duración”. *Ekonomiaz*, nº 31-32, págs. 245-265.
- Albert C.; Juárez, J.P.; Sánchez, R. y Toharia, L. (2000): “La transición de la escuela al mercado de trabajo en España: años noventa”. *Papeles de Economía Española*, nº 86, págs. 42-58.

- Corrales, H. y Rodríguez, B. (2003): “La transición del sistema educativo al mercado laboral. Análisis de los factores determinantes del primer desempleo”. *Comunicación presentada a las XII Jornadas de Economía de la Educación*, Madrid, Septiembre de 2003.
- Cox, D. (1972): “Regression models and life-tables”. *Journal of the Royal Statistical Society, Series B*, 34, págs. 187-220.
- Cox, D. y Oakes, D. (1985): *Analysis of survival data*. New York: Chapman and Hall.
- Davia, M.A. (1999): “La inserción laboral de los jóvenes españoles: una primera exploración con el Panel de Hogares de la Unión Europea”. *Comunicación presentada en las III Jornadas de Economía Laboral*. Oviedo, Septiembre de 1999.
- Grrotings, Peter. (1994), From qualification to competence: what are we talking ab *European Journal Vocational Training*. Thessaloniki.
- Kalbfleisch, J. D. & Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data*, John Wiley and Sons: New York.
- Kaplan, E. L: & Meier, P. (1958), “Nonparametric estimation from incomplete observations”, *Journal of the American Statistical Association*, 53, 457-481.
- Le Boterf, Guy, (1991), *L’ingenierie des competences*.
- Lee, E. T. (1980), *Statistical Methods for Survival Data Analysis*, Lifetime Learning Publications: Belmont, California.
- Mertens,Leonard.(1996) *Competencia Laboral: sistemas, surgimiento y modelos*.
- Orbe, J. & Ferreira, E. & Núñez-Antón V. (2001), “Modelling the duration of firms in Chapter 11 bankruptcy using a flexible model”, *Economics Letters* 71, 35-42.
- Orbe, J. & Ferreira, E. & Núñez-Antón V. (2002), “Censored partial regression”, *Biostatistics* 4, 109-121.
- 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.
- Orbe, J. & Ferreira, E. & Núñez-Antón V. (2003), “Modelling the duration of firms in Chapter 11 bankruptcy using a flexible model”, *Economics Letters* 71, 35-42
- Orbe, J. (1999), “Un modelo de regresión parcial censurado para análisis de supervivencia”, Tesis doctoral. Universidad del País Vasco.

- Petrongolo, B. y San Segundo, M.J. (1999): “¿Incentiva el paro juvenil la escolarización secundaria?”. *Ekonomiaz* nº 43, págs. 11-37.
- Salas, M. (1999): Inversión en capital humano: demanda, búsqueda de empleo y rendimientos privados de la educación superior. Una aproximación metodológica al estudio del mercado de trabajo de los titulados. *Tesis Doctoral*, Universidad de Granada.
- Stute, W. (1993), “Consistent estimation under random censorship when covariables are present”, *Journal of Multivariate Analysis*, 45,89-103.