

# Construcción de Tablas de Rotación Mediante Estadística No Paramétrica

Carlos Fernando LOZANO NATHAL, Sergio RAMOS LOYA,  
José Enrique PÉREZ SALVADOR

---

## RESUMEN

En la valuación de beneficios, es necesario contar con estimadores de rotación apropiados que minimicen el error y la volatilidad en los costos. En muchos casos, la información disponible es insuficiente por lo que se debe contar con herramientas que permitan la construcción de estimadores precisos con dicha información. La Estadística No Paramétrica ofrece modelos de sobrevivencia y técnicas para construir estimadores de rotación estadísticamente creíbles. El presente documento describe algunas de esas herramientas y muestra el proceso de construcción de una tabla de rotación que pueda ser utilizada como base para la producción de tablas personalizadas de acuerdo a las características de cada organización.

---

## 1. INTRODUCTION

El Comité de Investigación y Desarrollo (CID) de Nathal Actuarios y Consultores, S.A. de C.V., estableció la tarea de construir una tabla de rotación que fuera utilizada como base para producir tablas personalizadas en las valuaciones actuariales.

La información provino de empresas del sector privado de diferentes industrias y áreas geográficas. No fueron consideradas empresas muy grandes que pudieran dominar el comportamiento de la tabla de rotación.

La sección 2 habla de la rotación y de sus causas.

La sección 3 describe la información fuente utilizada para construir la tabla así como de consideraciones hechas en la etapa de procesamiento de datos.

La sección 4 menciona brevemente el modelado, los modelos de sobrevivencia, los tipos de modelos y describe

# Non-parametric Statistics in the Construction of Turnover Tables

Carlos Fernando LOZANO NATHAL, Sergio RAMOS LOYA,  
José Enrique PÉREZ SALVADOR

---

## ABSTRACT

In pension and employee benefits valuations, it is necessary to have suitable turnover estimators that minimize cost error and volatility. In many cases, diverse tools should be considered to construct precise estimators when historical information is insufficient. Non-parametric statistics provides survival models and techniques to build credible turnover estimators with incomplete data. This document describes some of these tools and shows the process of building a standard turnover table in which information particular to an organization can be integrated to produce customized tables.

---

## 1. INTRODUCTION

The Research and Development Committee of Nathal Actuarios took on the task of constructing a standard turnover table that could be used as a basis to produce custom tables in actuarial valuations.

The information used for this purpose came from Mexican private companies from different industries and geographical areas but excludes large companies that could influence the general tendency of the turnover table.

Section 2 refers to turnover and its causes.

Section 3 describes the source of the data used to construct the table and considerations made at the data processing stage.

Section 4 briefly mentions modeling, survival models, model types and describes three estimation models: empirical, moment estimation for incomplete data and

tres modelos: empírico, estimador de momento para datos incompletos y aproximación del estimador de producto-límite (Kaplan-Meier) para conjuntos grandes de datos.

La sección 5 expone los resultados obtenidos con los tres modelos.

En la sección 6 se encuentran las conclusiones y futuras consideraciones.

## 2. LA ROTACIÓN Y SUS CAUSAS

El término rotación, puede entenderse en dos sentidos diferentes:

- (Entradas y salidas) Intercambio de personas entre la organización y el ambiente, se define por el número de personas que ingresan y salen de la organización (Chiavenato 1994). Proceso en el que los empleados dejan la organización y tienen que ser reemplazados (Mathis y Jackson 1999, 93).
- (Salidas) Separación (voluntaria o involuntaria) de empleados de una organización (US Bureau of Labor Statistics 2010).

El último será ocupado para el presente documento.

En la literatura actuarial, las hipótesis demográficas comprenden comúnmente cuatro causas de rotación o decrementos: Muerte, invalidez, retiro y terminación. Las tablas de mortalidad e invalidez varían por edad y género. Las tablas de retiro están basadas en edad principalmente. Las tablas de terminación varían generalmente por edad y servicio y comprenden causas de separación por otros motivos que muerte, invalidez y retiro. Las tablas deben establecerse de acuerdo a las características de cada empresa. Dependiendo de la cantidad de información, las tablas pueden variar por otros factores como tipo de trabajo, industria o estatus. Marilyn Oliver (2009) describe aspectos importantes a considerar en el uso o construcción de hipótesis demográficas.

## 3. FUENTE Y PROCESAMIENTO DATOS

La información fue compilada año con año de las valuaciones actuariales practicadas a 209 empresas mexicanas privadas de diferentes tamaños e industrias, en el período comprendido entre Noviembre de 2002 y Agosto de 2010. Las empresas aportaron información de dos a nueve años de experiencia sin especificar las causas de salida.

product limit (Kaplan-Meier) approximation for large data sets.

Section 5 presents the results obtained from the three models.

Section 6 covers the conclusions y future considerations.

## 2. TURNOVER AND ITS CAUSES

Turnover have two different meanings:

- (Outputs and inputs) Exchange of employees between the organization and the environment, defined by the number of people entering and leaving the organization (Chiavenato 1994). Process in which employees leave the organization and have to be replaced (Mathis y Jackson 1999, 93).
- (Output) Separation of an employee from an establishment (voluntary, involuntary, or other) (US Bureau of Labor Statistics 2010).

The latter will be used for the rest of the document.

In actuarial literature, demographic assumptions commonly comprise four main causes or decrements: mortality, disability, retirement, and termination. Mortality and disability tables vary in age and gender. Retirement tables are based primarily on age. Termination tables generally vary according to age and service and include causes other than death, disability and retirement. Tables must be set according to the characteristics of each company. Depending on the amount of information, tables may vary according to additional factors like job type, industry, or status. Marilyn Oliver (2009) describes important aspects to consider in the use or construction of demographic assumptions.

## 3. DATA SOURCE AND PROCESSING

Year-to-year data was separately compiled from valuations performed on 209 Mexican private companies, which differ in size and industry, between November 2002 and August 2010. Companies provided information from two to nine years of experience without mentioning the reason of separation.

Following recommendations from the ASOP 23 "Data

Siguiendo las recomendaciones del ASOP 23 "Data Quality" (Actuarial Standards Board 2004) y con ayuda de una aplicación personalizada, la información fue verificada chequeando consistencia en nombres, genero, fechas y salarios. El cruce, o reconciliación de datos, fue logrado usando un programa con cuatro niveles de búsqueda eliminando duplicación de datos y alcanzando la calidad deseada para la generación de la tabla de rotación.

La Tabla 1 contiene un resumen de la información comparada con dos estudios que consideran, substancialmente, más información: Steve Koop (1997) y Edward W. Frees (2003).

**Tabla 1. Resumen de datos**

Estudio	Cantidad	Período	Años de exposición
N10-2010	209 cías.	2002-2010	122,000
Frees-2003	112 planes	1994-2000	1,700,000
Koop-1997	38 planes	1989-1994	3,065,000

#### 4. MODELOS DE SOBREVIVENCIA

Los modelos de sobrevivencia miden el tiempo hasta que un evento ocurra (muerte, enfermedad, recuperación, falla, etc.) y están relacionados con los Procesos de Conteo; la diferencia más importante radica en que los últimos pueden considerar eventos múltiples como sería la salida por segunda vez de un empleado dado que haya sido recontratado. El análisis de sobrevivencia puede ser encontrado en libros de bioestadística como Lee & Wang (2003) o Kleinbaum & Klein (2005) y en libros actuariales como Klugman, Panjer & Willmot (2008), y London (1997). Para los propósitos de este documento, el estudio se enfocará en la tasa de rotación,  $q_x$ , en vez de la función de sobrevivencia  $S(x)$ .

La siguiente lista menciona, de manera no exhaustiva, algunos de los modelos de sobrevivencia que ofrece la Estadística para los datos observados:

- No paramétricos: Empírico, Estimación de Momento, Estimación de Producto-Límite (Kaplan-Meier), Estimador de Nelson-Åalen.
- Semiparamétricos: Estimación de Densidad de Kernel, Estimador de Riesgos Proporcionales de Cox.
- Paramétricos: Weibull, Exponencial, Gamma, Log Logística, Lognormal, Gamma generalizada, Pareto,

Quality" (Actuarial Standards Board 2004) and with support from a customized application, data was verified by checking consistency in names, gender, dates, and salaries. Data reconciliation was achieved thanks to a program containing four search levels thus eliminating duplicates and achieving the desired quality to produce the turnover table.

Table 1 contains a summary of the information compared to two studies considering substantially more information: Steve Koop (1997) and Edward W. Frees (2003).

**Table 1. Data summary**

Study	Number	Period	Years of exposure
N10-2010	209 co.	2002-2010	122,000
Frees-2003	112 plans	1994-2000	1,700,000
Koop-1997	38 plans	1989-1994	3,065,000

#### 4. SURVIVAL MODELS

Survival models measure time up to a certain event (death, illness, recovery, fault, etc.) and are related to counting processes; the main difference being that these processes can consider multiple events such as a second separation when an employee has been rehired. Survival analysis can be found in biostatistics books such as Lee & Wang (2003), or Kleinbaum & Klein (2005), and actuarial books such as Klugman, Panjer & Willmot (2008), and London (1997). For the purpose of this document, the approach will focus on the turnover rate,  $q_x$  instead of the survival function  $S(x)$ .

Some of the survival models offered by Statistics for the observed data are mentioned in the following non-exhaustive list:

- Nonparametric: Empirical, Moment Estimation, Product-limit Estimation (Kaplan-Meier), Nelson-Åalen Estimator.
- Semiparametric: Kernel Density Estimation, Cox proportional hazards estimation.
- Parametric: Weibull, Exponential, Gamma, Log Logistics, Lognormal, Generalized Gamma, Pareto, Gompertz, Makeham, Generalized Linear Models (Logit Model).

## Gompertz, Makeham, Modelos Lineales Generalizados (Modelo Logit).

Los datos pueden ser completos en donde se conoce el tiempo de entrada y de salida, o incompletos en donde se desconoce la entrada o la salida. El estudio puede hacerse punto por punto en donde el número de elementos iguala al número de registros, o agrupado por intervalos (e.g. años de edad o de servicio).

En los estudios de rotación es común contar con datos incompletos; cuando se delimita el período de estudio, se trunca la edad ingreso si es menor a la edad al inicio del estudio y se censura la edad de baja si es mayor a la fecha de fin de estudio.

Cuando se cuenta con una gran cantidad de datos, la estimación puntual puede demandar gran cantidad de cálculos que aportan poco valor si la función de distribución sólo se requiere para pocos puntos. Tal es el caso en la construcción de tablas de rotación, por lo que el enfoque se dará será el de datos agrupados e incompletos. Los modelos empleados para el presente estudio son el Empírico, Estimación de Momento para datos incompletos y la aproximación del estimador de Kaplan-Meier para conjuntos grandes de datos incompletos.

### 4.1. Estimador Empírico

Se define como la tasa de número de eventos entre el número de vidas expuestas.

Sean:

- $x$  = edad redondeada a cero posiciones
- $d_x$  = número de salidas totales en el intervalo  $[x, x+1)$
- $n_x$  = número de casos expuestos en  $[x, x+1)$

El estimador empírico de salidas de edad  $x$  por todas las causas,  $\hat{q}_x^{(\tau)}$ , esta dado por el cociente:

$$\hat{q}_x^{(\tau)} = \frac{d_x}{n_x}$$

### 4.2. Estimador de Momento para datos incompletos

La idea básica del estimador de momento es la de generar el cociente de salidas entre vidas *programadas* en el intervalo  $[x, x+1)$ . A continuación se presenta resumen del modelo expuesto por Dick London (1997, 113).

Sean:

- $0 \leq r_j < 1$  la fracción de tiempo de entrada en  $[x, x+1)$

Data is called complete when the time of entry and leaving are known and incomplete when the time of entry or leaving is unknown. The study can be done point by point when the number of elements equals the number of records, or grouped at intervals (e.g. years of age or service).

In turnover studies is common to have incomplete data; for instance, when the study period is defined, the age of entry is truncated if it is lesser than the age at the start of the study, and the age of separation is censored if it is greater than the age at the end of study.

When there is a large amount of data, point estimation may require many calculations that add little value if the distribution function is only needed at a few points. Such is the case in the construction of turnover tables, so that the study will deal with grouped incomplete data sets. The models used for this study are Empirical, Moment Estimation for incomplete data, and Kaplan-Meier type approximation for large data sets.

### 4.1. Empirical estimation

Defined as the ratio of number of events to number of lives exposed to the risk.

Let:

- $x$  = rounded age to zero positions
- $d_x$  = number of separations in the interval  $[x, x+1)$
- $n_x$  = number of exposed employees in  $[x, x+1)$

The empirical estimator of separation,  $\hat{q}_x^{(\tau)}$ , is:

$$\hat{q}_x^{(\tau)} = \frac{d_x}{n_x}$$

### 4.2. Moment estimation for incomplete data sets

The basic idea of the moment estimation is to generate the ratio of separations to scheduled lives in the interval  $[x, x+1)$ . The following is a summary of the model presented by Dick London (1997, 113).

Let:

- $0 \leq r_j < 1$  the fraction of time of entry in  $[x, x+1)$

$0 < s_j \leq 1$  la fracción de tiempo programado de salida en  $[x, x+1)$

$0 < s_j \leq 1$  the fraction of time of scheduled time of separation in  $[x, x + 1)$

El estimador de momento por todas las causas a edad  $x$  está dado por:

$$\hat{q}_x^{(\infty\tau)} = \frac{d_x}{\sum_{j=1}^{n_x} (s_j - r_j)}$$

The empirical estimator for all causes at  $x$  is given by:

$$\hat{q}_x^{(\infty\tau)} = \frac{d_x}{\sum_{j=1}^{n_x} (s_j - r_j)}$$

Se debe notar que la fracción de tiempo programado de salida es  $s_j = 1$  cuando ocurre una salida.

It should be noted that the fraction of time of schedule separation is  $s_j = 1$  when there is a separation.

### 4.3. Aproximación del estimador de Kaplan-Meier para conjuntos grandes de datos incompletos

### 4.3. Kaplan-Meier estimator approximation for large and incomplete data sets

El Estimador de Producto Límite de Kaplan y Meier (1958) es comúnmente usado en la construcción de curvas de sobrevivencia. El modelo que se resume a continuación proviene de Klugman, Panjer y Willmot (2008, 366-369).

The product-limit estimator of Kaplan and Meier (1958) is commonly used in the construction of survival curves. The model summarized below comes from Klugman, Panjer and Willmot (2008, 366-369).

La idea general de este estimador es obtener la mayor cantidad de información de las observaciones disponibles a través de probabilidad condicional.

The general idea of this estimator is to get as much information as possible from the available observations through conditional probability.

**Ejemplo:** Considere el número de estudiantes por generación como se muestra en la Tabla 2.

**Example:** Consider the number of students per generation as shown in Table 2.

Tabla 2. Estudiantes por generación

Año	2008	2009
0	10	20
1	8	17
2	7	

Table 1. Students per generation

Year	2008	2009
0	10	20
1	8	17
2	7	

Queremos estimar la proporción de estudiantes que permanecen por dos años o más:

We want to estimate the proportion of students who remain for two or more years:

$$P[X \geq 2] = \hat{S}(2) = \frac{7}{10} = 0.70$$

$$P[X \geq 2] = \hat{S}(2) = \frac{7}{10} = 0.70$$

La técnica de Kaplan-Meier utiliza la segunda muestra, por un año, para contribuir a la estimación:

The Kaplan-Meier technique uses the second sample, for one year, to contribute to the estimator:

$$P[X \geq 2] = P[X \geq 2 | X \geq 1] \cdot P[X \geq 1]$$

$$P[X \geq 2] = P[X \geq 2 | X \geq 1] \cdot P[X \geq 1]$$

$$\hat{S}(2) = \hat{S}(2|1) \cdot \hat{S}(1)$$

$$\hat{S}(2) = \hat{S}(2|1) \cdot \hat{S}(1)$$

$$\hat{S}(2) = \frac{7}{8} \cdot \frac{8 + 17}{10 + 20} = 0.73$$

$$\hat{S}(2) = \frac{7}{8} \cdot \frac{8 + 17}{10 + 20} = 0.73$$

□

□

Al agrupar los datos, se necesitan hipótesis acerca de la ubicación de los datos dentro del intervalo. Asumiendo que

When grouping data, assumptions are required about the data location within the interval. Assuming that the points

los puntos de truncamiento y censuramiento ocurren uniformemente en el intervalo y sean:

$e_x$  = número de empleados observados por primera vez en  $[x, x+1]$   
 $d_x$  = número de salidas (no censuradas) en  $(x, x+1]$   
 $u_x$  = número de observaciones censuradas en  $(x, x+1]$   
 $r_x$  = grupo en riesgo en el intervalo  $(x, x+1]$

El estimador de Kaplan-Meier de salidas a edad  $x$  por todas las causas esta dado por:

$$\hat{q}_x^{(\tau)} = \frac{d_x}{\frac{e_x - u_x}{2} + \sum_{j=0}^{x-1} [e_j - (d_j + u_j)]} = \frac{d_x}{r_x}$$

Los extremos del intervalo difieren porque el truncamiento se puede dar en el extremo izquierdo del primer intervalo pero no en el extremo derecho del último intervalo. Lo opuesto sucede para el censuramiento.

En el numerador se encuentran las salidas no censuradas. El primer elemento del denominador corresponde a la hipótesis de que las entradas y salidas censuradas se distribuyen uniformemente durante el intervalo. El segundo elemento comprende a todas las entradas registradas menos todas las salidas registradas (censuradas y no censuradas).

Los algoritmos desarrollados para el estudio mostraron un menor tiempo de cálculo para los estimadores de Kaplan-Meier que para los estimadores de momento.

#### 4.4. Aproximación del estimador bivariado de Kaplan-Meier para conjuntos grandes de datos incompletos

De acuerdo al reporte Pension Plan Termination and Retirement Study (Frees 2003), las tablas basadas en edad y antigüedad son los mejores estimadores de tasas de terminación seguidas de tablas basadas en antigüedad solamente. El algoritmo para producir estimadores bivariados se basó principalmente en Ling & Ying (1993) y en Bae, Choi, Park & Kim (2005) en donde se encuentra un breve resumen de los estimadores bivariados de Kaplan-Meier.

La parte más sensible se encuentra en la deducción del grupo de riesgo por edad y antigüedad,  $r_{xy}$ . La aproximación empleada fue la de establecer cuatro clases por antigüedad con puntos de corte en 2, 5 y 10 años:

$$\begin{aligned} 0 &\leq y < 2 \\ 2 &\leq y < 5 \\ 5 &\leq y < 10 \end{aligned}$$

of truncation and censoring occur uniformly in the interval and letting:

$e_x$  = number of employees that were first observed in  $[x, x + 1]$   
 $d_x$  = number of (uncensored) separations in  $(x, x + 1]$   
 $u_x$  = number of censored observations in  $(x, x + 1]$   
 $r_x$  = risk group in the interval  $(x, x + 1]$

The Kaplan-Meier estimator of separations at age  $x$  from all causes is given by:

$$\hat{q}_x^{(\tau)} = \frac{d_x}{\frac{e_x - u_x}{2} + \sum_{j=0}^{x-1} [e_j - (d_j + u_j)]} = \frac{d_x}{r_x}$$

The interval endpoints differ because truncation is possible at the left end of the first interval but not at the right end of the last interval. The reverse is true for censoring.

The number of separations is found in the numerator. The first element in the denominator corresponds to the hypothesis that entries and censored separations are uniformly distributed in the interval. The second element comprises all the entries observed less all the separations observed (censored and uncensored).

The algorithms developed for the study showed less computation time for the Kaplan-Meier estimators over the moment estimators.

#### 4.4. Bivariate Kaplan-Meier estimator approximation for large and incomplete data sets

According to the Pension Plan Termination and Retirement Study (Frees 2003), tables based on age and service are the best predictors of termination rates, followed by service alone. The algorithm used to produce bivariate estimators was based mainly on Ying & Ling (1993) and Bae, Choi, Park & Kim (2005), where a brief summary of the bivariate Kaplan-Meier estimators can be found.

The most sensitive part lies in determining the group exposed to risk by age and service,  $r_{xy}$ . Four service classes with breakpoints in 2, 5 and 10 years were used in order to compute the approximation:

$$\begin{aligned} 0 &\leq y < 2 \\ 2 &\leq y < 5 \\ 5 &\leq y < 10 \end{aligned}$$

$$10 \leq y < \infty$$

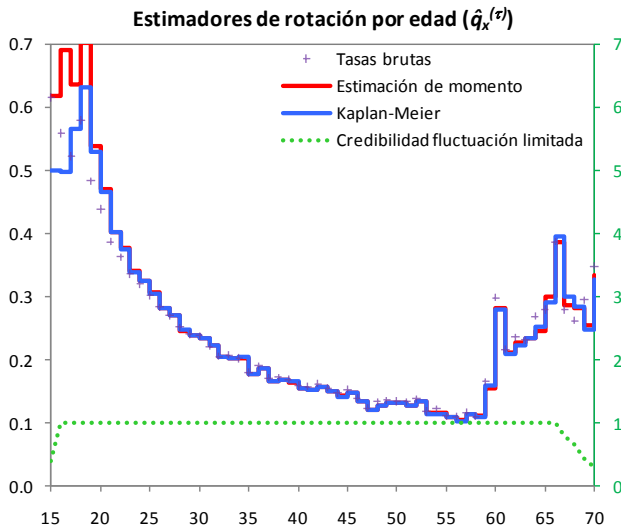
Al aplicar los puntos de corte a cada registro, se crean cuatro cadenas con nuevas edades de truncamiento y censuramiento. En caso de tener intersección vacía se genera una cadena vacía. Se reúnen las cuatro series de cadenas y se aplica la aproximación univariada de estimadores de Kaplan-Meier a cada una.

Es recomendable establecer procedimientos de verificación en cada etapa del proceso, desde la validación de datos, determinación de antigüedades y edades de truncamiento/censuramiento (dado que una empresa puede tener un período de observación diferente a otra), hasta la generación de grupos de riesgo y tasas de rotación.

## 5. CONSTRUCCIÓN DE TABLA DE ROTACIÓN N10

En una primera etapa, se generaron los estimadores descritos previamente con la información validada. Los resultados se muestran en la Figura 1.

Figura 1.



La línea inferior punteada representa el factor de credibilidad  $Z$  (multiplicado por  $0.1$  para tener mayor claridad). Un valor  $Z=1$  ( $0.1$  en la gráfica) significa que la información es creíble y que el estimador es estable basado en

$$P \left[ \frac{|\hat{q}_x - q_x|}{q_x} \leq r \right] \geq p$$

Es decir, el estimador es estable si la diferencia del estimador relativa al valor hipotético es más pequeña que un error de tolerancia  $r$  con gran probabilidad  $p$ . Comúnmente se utilizan valores de  $r=5\%$  y  $p=90\%$ . Para el

$$10 \leq y < \infty$$

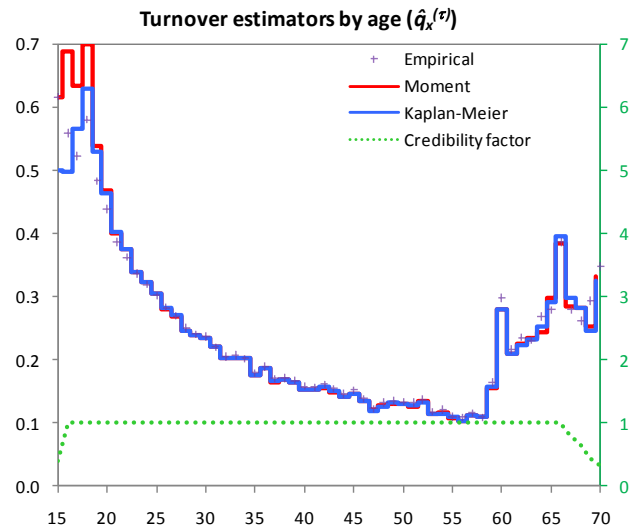
Four strings are created when the breakpoints are applied to each record, generating new truncated and censored ages. If the intersection is empty, then an empty string is created. Gather the four string sets and apply the univariate Kaplan-Meier approximation to each one of them.

It is recommended to establish verification procedures at every stage of the process, from validating data, establishing service and truncation/censoring ages (given that a particular company may have a different observation period to another company), to generating risk groups and turnover rates.

## 5. N10 TURNOVER TABLE CONSTRUCTION

In the first stage, the estimators described above were generated with the validated data set. The results are shown in Figure 1.

Figure 1.



The bottom dotted line represents the credibility factor  $Z$  (multiplied by  $0.1$  for clarity purposes). A value  $Z=1$  ( $0.1$  in the graphic) means that the information is credible and the estimator is stable based on

$$P \left[ \frac{|\hat{q}_x - q_x|}{q_x} \leq r \right] \geq p$$

That is, the estimator is stable if the difference of the estimator relative to the hypothetical value is smaller than an error tolerance of  $r$  with big probability  $p$ . Common values being  $r=5\%$  and  $p=90\%$ . In this case, the values were

estudio se utilizaron valores de  $r=4\%$  y  $p=99\%$ .

Los estimadores fueron sometidos a pruebas estadísticas en una segunda etapa. El ploteo p-p (relativo a la distribución empírica para comparar los modelos) muestra buen ajuste en ambos modelos y la prueba  $\chi^2$  muestra un resultado ligeramente mejor en los estimadores de Kaplan-Meier que en los de la estimación de momento. Los resultados son mostrados en la Figura 2. Los estimadores de Kaplan-Meier fueron elegidos por presentar mejores resultados y cálculos más simples.

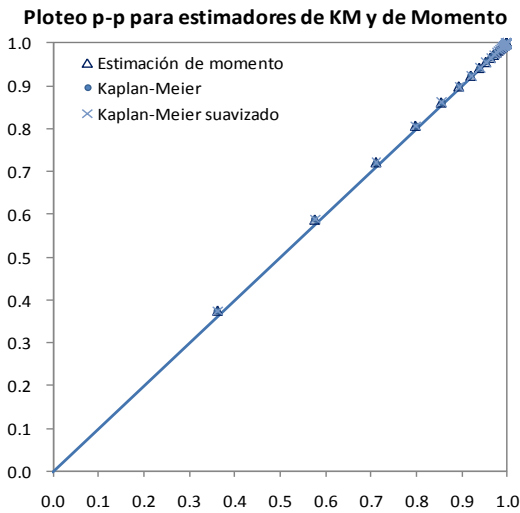
Al estudiar una secuencia de estimadores por edad, se decidió emplear la técnica de suavización de trazadores, encontrada en Klugman, Panjer & Wilmot (2003), la cual ofrece un balance adecuado entre ajuste y suavización. Otro método comúnmente usado es el de Whittaker basado en diferencias finitas y se encuentra en London (1985).

$r=4\%$  and  $p=99\%$ .

The estimators were tested In a second stage. A p-p plot (relative to the empirical distribution to compare the models) shows good fit in both models, and the  $\chi^2$  test shows that the Kaplan Meier estimators have slightly better results than the moment estimators. The results are shown in Figure 2. Kaplan-Meier estimators were chosen because they showed better results and simpler calculations.

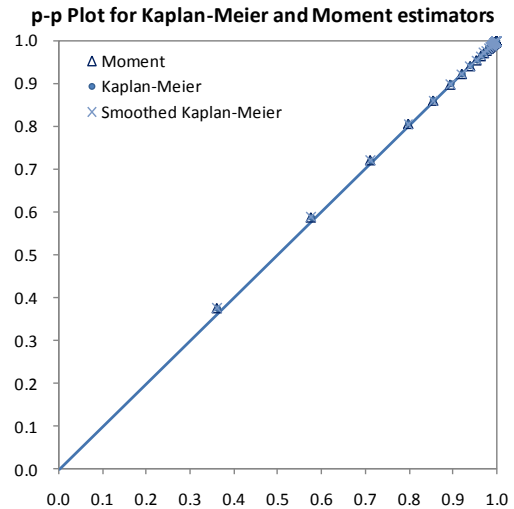
Since we are studying a sequence of successive estimators by age, we decided to use the smoothing spline technique, found in Klugman, Panjer & Wilmot (2003), which offers an appropriate balance between fit and smoothness. Another common method is Whittaker graduation based on finite differences; it can be found in London (1985).

**Figura 2. Resultados de pruebas estadísticas**



Prueba	Estimación de momento	Kaplan-Meier	Kaplan-Meier suavizado
$\chi^2$	93.9833	93.1984	117.0293

**Figura 2.**



Test	Moment estimation	Kaplan-Meier	Smooth Kaplan-Meier
$\chi^2$	93.9833	93.1984	117.0293

Los estimadores de Kaplan-Meier fueron suavizados en dos secciones diferentes: La primera contempla edades menores a 60 años y la segunda contempla edades mayores o iguales a 60 años. La razón es que el *retiro* es más abrupto que la *terminación*, lo cual requiere de una curva con mayor ajuste.

Los estimadores suavizados fueron probados encontrando buen ajuste en el ploteo p-p pero no en la prueba  $\chi^2$ , lo

Kaplan-Meier estimators were smoothed in two sections: the first covers ages below 60 years and the second covers ages greater than or equal to 60 years. The reason is that retirement has more abrupt termination, which requires a greater fit.

The smoothed estimators were tested finding good fit on the p-p plot but not on the  $\chi^2$  test, which is explained by



que se explica por la suavización concedida. La gráfica pasa entre los puntos en vez de sobre los puntos.

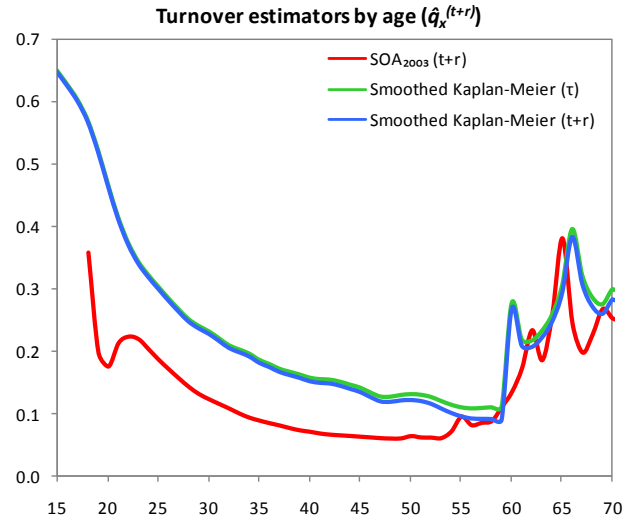
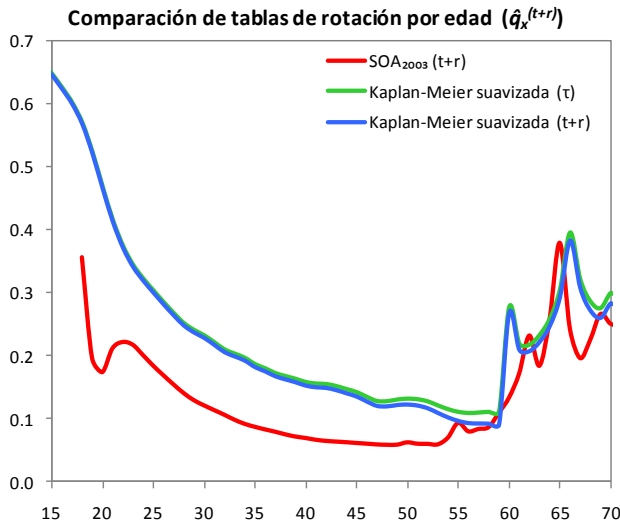
the smoothing granted. The function crosses among the points instead of over the points.

Dado que no se conocen las causas de salida, la tercera etapa comprende la separación de decrementos tomando las tablas base de mortalidad EMSSA09 y la de invalidez EISS97 que son las tablas de experiencia mexicana más recientes. La Figura 3 muestra una comparación de la con la del estudio de la SOA (Frees 2003) para decrementos de terminación y retiro.

Given that the separation causes remain unknown, the third stage dealt with breakdown of decrements taking the most recent standard tables of mortality EMSSA09 and disability EISS97 from Mexican experience. The result is shown in Figure 3 along with the SOA (Frees 2003) table for termination plus retirement.

Figura 3.

Figura 3.



La gráfica SOA 2003 presenta un pico a los 55 años que es la edad mínima de retiro de algunos planes privados, el siguiente pico es a los 62 años que es la edad mínima de retiro de la seguridad social en Estados Unidos, y el siguiente pico es encontrado a los 65 años. Para México, el primer pico se encuentra a los 60 años que es la edad mínima de retiro de seguridad social y el siguiente a los 66 años en vez de los 65 años que es la edad normal de retiro. Se requiere un estudio más profundo para determinar las causas de este desfase. Una explicación probable puede ser el desconocimiento de los trámites necesarios para obtener la pensión, que generalmente toman varios meses.

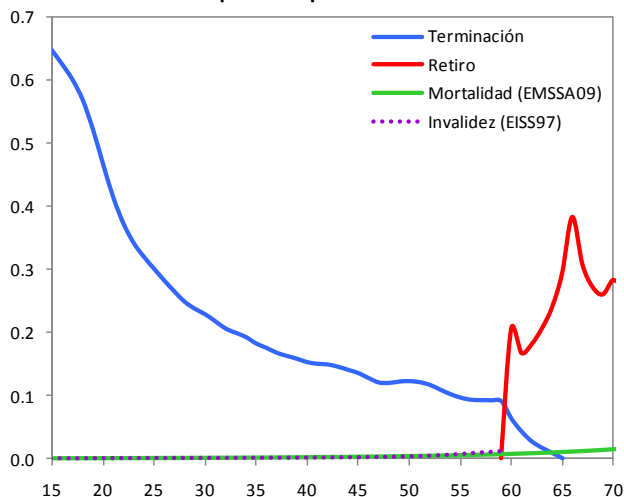
SOA 2003 graph shows a peak at 55 years, which is the minimum retirement age in some pension plans, the next peak at 62 years represents the minimum social security retirement age in the United States, and the next peak is found at 65 years. Mexican turnover shows the first peak at 60 years, which is the minimum social security retirement age, and a second peak at 66 years instead of the normal retirement age, at 65 years. Further study is necessary to find out the causes for delaying. A probable explanation may be the lack of knowledge regarding the requirements to obtain the social security pension, which usually takes several months.

Por otra parte, algunos empleados cumplen el requisito de edad mínima de retiro de la seguridad social, pero no el de número mínimo de contribuciones semanales, por lo que se aplicó una hipótesis de porcentaje de empleados que cumplen con los criterios de retiro para las edades 60 a 65 años. La Figura 4 muestra la separación por decrementos. N10 y NR10 corresponden a las tablas de terminación y retiro respectivamente.

Some employees meet the early retirement age for social security but not the minimum number of weekly contributions; as such, an assumption regarding the percentage of employees meeting the criteria was applied for ages between 60 and 65 years. Figure 4 shows the decrements separation. N10 and NR10 represent, respectively, the termination and retirement tables.

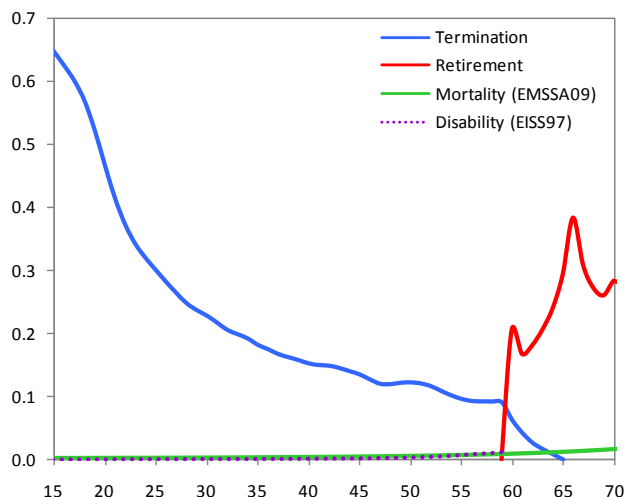
**Figura 4.**

**Separación por decrementos**



**Figura 4.**

**Decrements**



La Figura 5 muestra la gráfica de terminación por años de servicios para empleados menores a 60 años. Se debe notar que el factor de credibilidad se alcanza en el rango de edades de 1 a 25 años, después, decrece abruptamente. Se encuentra una reducción de tasas a los 10, 20 y 30 años de servicio. No hay suficiente información después de los 30 años de servicio el cual es un punto sensible al valorar beneficios. Las obligaciones por beneficios por terminación serán mayores si el cola sube y las obligaciones por beneficios al retiro serán mayores si la cola baja.

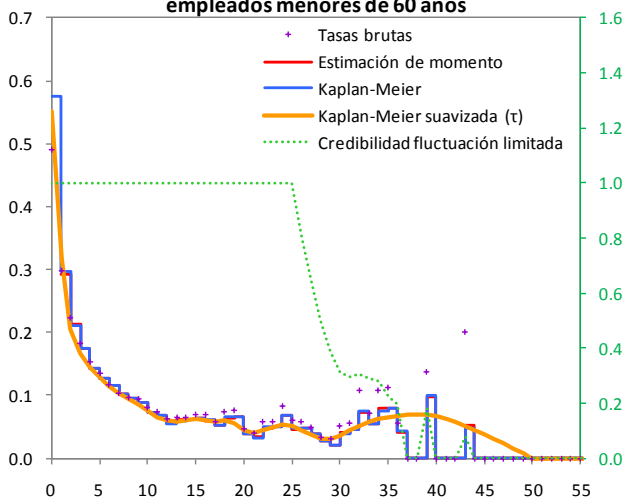
Figure 5 shows the termination table by service for employees under 60 years old. Note that full credibility is achieved for ages from 1 to 25 years and then decreases sharply. Reduction of termination rates are found in 10, 20 and 30 years of service. There is not enough information passing 30 years of service which is a sensitive point when valuating benefits. The liability for termination benefits will be higher if the tail goes up and the liability for retirement benefits will be higher if the tail goes down.

Cuando no se cuenta con la información de decrementos, el promedio de tasas mortalidad e invalidez de los empleados de cada grupo debe descontarse de la estimación para separar la tasa de terminación.

When decrement information is not provided, employee average of mortality and disability rates of every group should be discounted from the estimator in order to separate the termination rates.

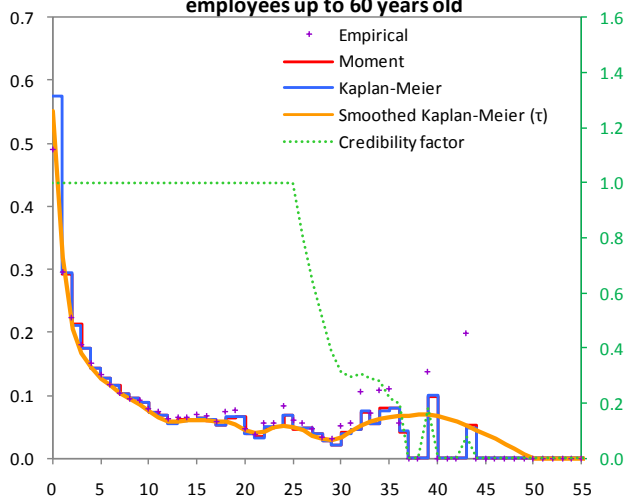
**Figura 5.**

**Estimadores de rotación por antigüedad ( $\hat{q}_y^{(r)}$ ) para empleados menores de 60 años**



**Figura 5.**

**Turnover estimators by service ( $\hat{q}_y^{(r)}$ ) for employees up to 60 years old**



Según el documento de Marilyn Oliver (2009, 36), en la

According to the document of Marilyn Oliver (2009), the

práctica se piensa que el mejor predictor de tasas de terminación es el de tablas selectas que varíen por edad y años de servicio por un período selecto de 3 a 5 años (a veces 10) y a partir de ahí, sólo por edad.

Para el presente estudio, se construyeron aproximaciones de los estimadores bivariados de Kaplan-Meier por edad y grupos de antigüedad de  $0 \leq y < 2$ ,  $2 \leq y < 5$ ,  $5 \leq y < 10$ ,  $0 \leq y < \infty$ , los cuales son mostrados en la Figura 6.a. La Figura 6.b muestra que se tiene información creíble hasta aproximadamente los 60 años por lo que solamente se determinaron las tasas de terminación y no de retiro. Se debe notar que los factores de credibilidad están multiplicados por una constante para efectos de claridad:

Grupo	Constante
$0 \leq x < 2$	4/10
$2 \leq x < 5$	3/10
$5 \leq x < 10$	2/10
$10 \leq x < \infty$	1/10

best practical predictor of termination rates is generally thought to be select and ultimate tables that vary by age and service during a select period, generally lasting 3 to 5 (and sometimes 10) years in practice, and by age only thereafter.

In this study, approximations to bivariate Kaplan-Meier estimators were constructed by age and service groups  $0 \leq y < 2$ ,  $2 \leq y < 5$ ,  $5 \leq y < 10$ ,  $0 \leq y < \infty$ , which are shown in Figure 6.a. Figure 6.b shows that there is credible information approximately up to 60 years which allow us to construct termination rates only. It should be noted that the credibility factors are multiplied by a constant for clarity purposes:

Group	Constant
$0 \leq x < 2$	4/10
$2 \leq x < 5$	3/10
$5 \leq x < 10$	2/10
$10 \leq x < \infty$	1/10

Figura 6.a

Terminación por edad y antigüedad ( $\hat{q}_{x,y}^{(t)}$ )  
(Kaplan-Meier Bivariado)

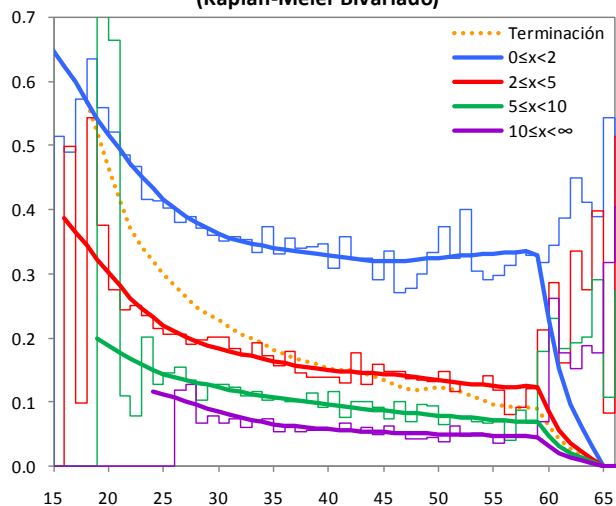


Figura 6.a

Termination by age and service ( $\hat{q}_{x,y}^{(t)}$ )  
(Bivariate Kaplan-Meier)

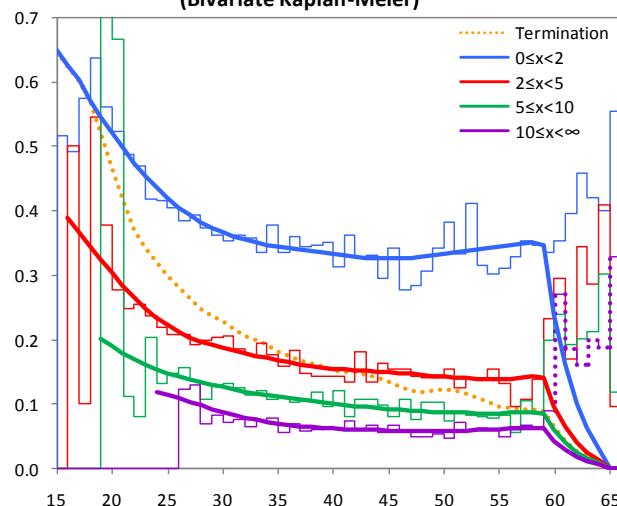


Figura 6.b

Terminación por edad y antigüedad ( $\hat{q}_{x,y}^{(t)}$ )  
Factor de credibilidad (Z)  $r=4\%$ ,  $p=99\%$

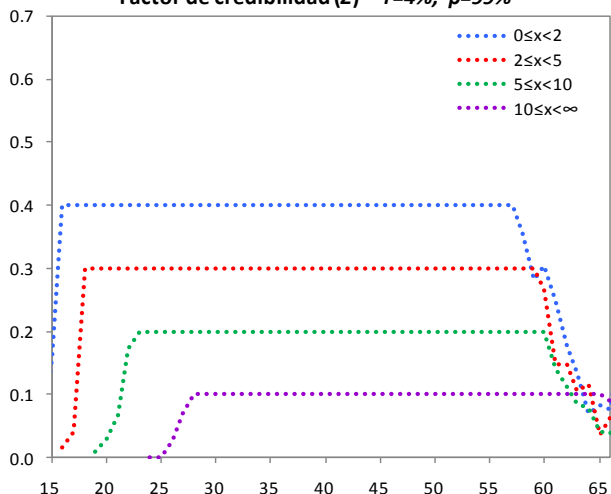
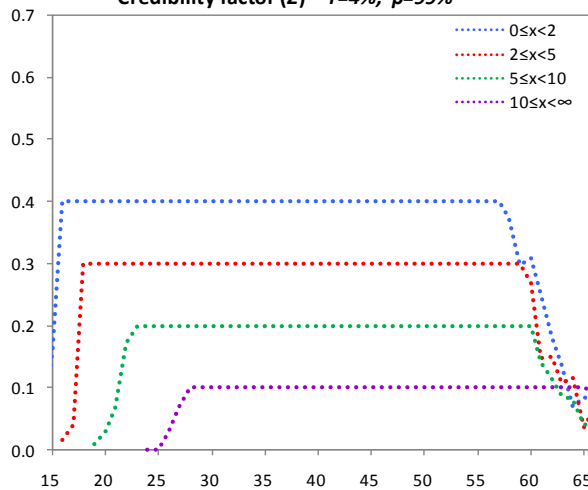


Figura 6.b

Termination by age and service ( $\hat{q}_{x,y}^{(t)}$ )  
Credibility factor (Z)  $r=4\%$ ,  $p=99\%$



Cada organización tiene una composición de empleados casi única. Se pueden construir tablas de rotación para empresas con suficiente información histórica y complementarlas con información de tablas base como N10 y NR10. Cuando se tiene información limitada, se pueden emplear técnicas de calibración como multiplicar por el cociente de datos Observados/Esperados, aunque una práctica más conservadora sería calibrar hasta el extremo del intervalo de confianza. Dichas técnicas salen del alcance de este documento.

## 6. CONCLUSIONES

De los modelos expuestos, los estimadores de Kaplan-Meier tienen mayor ajuste y se generan con cálculos más sencillos que los de estimador de momento. Ambos modelos arrojan buenos resultados.

Una vez que se generan los estimadores es recomendable aplicar graduarlos para producir una progresión ordenada de tasas. En el presente estudio se emplearon dos graduaciones por observar tendencias diferentes antes y después de los 60 años, el pre-retiro y el retiro.

Después de graduar los valores, se debe considerar la separación de rotación por diferentes decrementos. Para la experiencia mexicana se emplearon la EMSSA09 como mortalidad, EISS97 como invalidez y una proporción de cumplimiento de requisitos de retiro para separar la terminación del retiro.

Como ventajas de los procedimientos mostrados en el presente estudio se encuentran:

- Mayor explotación de información con los datos

Every organization has an employee structure almost unique. Turnover tables can be constructed for organizations with enough historical data available and complemented with standard tables such as N10 and NR10. When data is limited, techniques like calibrating with Actual/Expected ratios may be used, although a more conservative approach would be to calibrate up to the bound of the confidence interval. Such techniques are beyond the scope of this document.

## 6. CONCLUSIONS

Of those models applied in this document, Kaplan-Meier estimators fit better and can be compute with simpler algorithms than the moment estimators. Both models present good results.

Once the estimates are generated is advisable to apply a graduation in order to construct a sequence of successive estimators. In the present study two graduations with different tendencies were produce before and after the age 60, pre-retirement and retirement.

After graduating the values, the table should be separate into different decrements. EMSSA09 mortality table, EISS97 disability table, and a eligibility proportion of retirement were used for Mexican experience in order disaggregate termination and retirement..

Some of the advantages are:

- Greater use of incomplete data available.
- Better fit than parametric models
- Spreadsheets can be used to construct the models

incompletos disponibles.

- Mayor ajuste que con modelos paramétricos.
- Pueden construirse en hojas de cálculo.

Las limitaciones encontradas fueron:

- Debe enfatizarse que un buen resultado dependen de la cantidad y la calidad de los datos.
- Cuando la información no cuenta con causas de salida, se deben emplear cálculos extras para la separación de decrementos.

Las tablas generadas con estas técnicas pueden servir como base para ajuste particular de cada empresa. Se recomienda hacer un mayor uso de la teoría de la credibilidad y, cuando sea posible, modelos lineales generalizados.

Se recomienda hacer una revisión anual para contar con mayor información y obtener modelos más fieles, así como para hacer un análisis de dinámica de rotación a través de diferentes años.

## BIBLIOGRAFÍA

Actuarial Standards Board. *Actuarial Standard of Practice No. 23, "Data Quality"*. EUA: American Academy of Actuaries, 2004.

Bae, Whasoo, Hyemi Choi, Byenong U. Park, & Choongrak Kim. «Smoothing Techniques For The Bivariate Kaplan-Meier Estimator.» *Communications in Statistics — Theory and Methods*, 34, 2005: 1659-1674.

Chiavenato, Idalberto. *Recursos Humanos*. 2ed. Colombia: McGraw Hill, 1994.

Frees, Edward W. *Pension Plan Termination and Retirement Study*. Society of Actuaries, 2003.

Kaplan, E. L., & Paul Meier. «Nonparametric Estimation from Incomplete Observations.» *Journal of American Statistical Association*, 53, 1958: 457-481.

Kleinbaum, David G., & Mitchel Klein. *Survival Analysis, A Self-Learning Text*. 2ed. Atlanta, Georgia: Springer, 2005.

Klugman, Stuart, Harry Panjer, & Gordon Willmot. *Loss Models: From Data to Decisions*. 3ed. New York: Wiley, 2008.

Koop, Steve. *Pension Plan Turnover Table Construction*.

Some limitations are:

- It should be emphasized that good results depend on data quantity and quality.
- Additional computation is required when decrements are not provided

These techniques allow constructing standard tables that can be used to customized tables for each company. It is recommended to use credibility theory and, when applicable generalized linear models.

It is recommended that an annual review to have more information and more faithful models, as well as an analysis of rotational dynamics across different years.

## BIBLIOGRAPHY

Actuarial Standards Board. *Actuarial Standard of Practice No. 23, "Data Quality"*. EUA: American Academy of Actuaries, 2004.

Bae, Whasoo, Hyemi Choi, Byenong U. Park, & Choongrak Kim. «Smoothing Techniques For The Bivariate Kaplan-Meier Estimator.» *Communications in Statistics — Theory and Methods*, 34, 2005: 1659-1674.

Chiavenato, Idalberto. *Recursos Humanos*. 2ed. Colombia: McGraw Hill, 1994.

Frees, Edward W. *Pension Plan Termination and Retirement Study*. Society of Actuaries, 2003.

Kaplan, E. L., & Paul Meier. «Nonparametric Estimation from Incomplete Observations.» *Journal of American Statistical Association*, 53, 1958: 457-481.

Kleinbaum, David G., & Mitchel Klein. *Survival Analysis, A Self-Learning Text*. 2ed. Atlanta, Georgia: Springer, 2005.

Klugman, Stuart, Harry Panjer, & Gordon Willmot. *Loss Models: From Data to Decisions*. 3ed. New York: Wiley, 2008.

Koop, Steve. *Pension Plan Turnover Table Construction*. Society of Actuaries, 1997.

Lee, Elisa T., & John Wenyu Wang. *Statistical Methods for*

- Society of Actuaries, 1997.
- Lee, Elisa T., & John Wenyu Wang. *Statistical Methods for Survival Data Analysis*. 3ed. Oklahoma City: Wiley, 2003.
- Lin, D. Y., & Zhiliang Ying. «A Simple Nonparametric Estimator of the Bivariate Survival Function Under Univariate Censoring.» *Biometrika*, Vol. 80, No. 3, 1993: 573-581.
- London, Dick. *Graduation: The Revision of Estimates*. Connecticut: ACTEX Publications, 1985.
- London, Dick. *Survival Models And Their Estimation*. 3ed. Winstead, Connecticut: ACTEX Publications, 1997.
- Mathis, Robert L., & John H. Jackson. *Human Resource Management*. 9ed. Mason, Ohio: South-Western Educational Publishing, 1999.
- Oliver, Marilyn. *Assessment And Selection Of Actuarial Assumptions For Measuring Pension Obligations*. Society of Actuaries, 2009.
- US Bureau of Labor Statistics. *Glossary*. 2010. <http://www.bls.gov/bls/glossary.htm> (30 de Marzo, 2010).
- Survival Data Analysis*. 3ed. Oklahoma City: Wiley, 2003.
- Lin, D. Y., & Zhiliang Ying. «A Simple Nonparametric Estimator of the Bivariate Survival Function Under Univariate Censoring.» *Biometrika*, Vol. 80, No. 3, 1993: 573-581.
- London, Dick. *Graduation: The Revision of Estimates*. Winstead, Connecticut: ACTEX Publications, 1985.
- . *Survival Models And Their Estimation*. 3ed. Winstead, Connecticut: ACTEX Publications, 1997.
- Mathis, Robert L., y John H. Jackson. *Human Resource Management*. 9ed. Mason, Ohio: South-Western Educational Publishing, 1999.
- Oliver, Marilyn. *Assessment And Selection Of Actuarial Assumptions For Measuring Pension Obligations*. Society of Actuaries, 2009.
- US Bureau of Labor Statistics. *Glossary*. 2010. <http://www.bls.gov/bls/glossary.htm> (March 30th, 2010).