



**UNIVERSIDAD AUTONOMA DEL ESTADO DE MÉXICO**



**FACULTAD DE ECONOMÍA**

**LICENCIATURA EN ACTUARÍA.**

**“APLICACIÓN DEL MODELO DE REGRESIÓN DE COX  
PARA DETERMINAR LA PROBABILIDAD DE  
SUPERVIVENCIA DE PEQUEÑAS Y MEDIANAS  
EMPRESAS DE LA REGIÓN DEL BAJÍO”**

**TESIS**

**QUE PARA OBTENER EL TITULO DE  
LICENCIADO EN ACTUARIA**

**PRESENTA:**

**IRVING OLIVER GONZÁLEZ GUTIÉRREZ**

**DIRECTOR DE TESIS:**

**Dr. en A. OSWALDO GARCÍA SALGADO**

**REVISORES:**

**Dra. En C.A. GABRIELA MUNGUÌA VÀZQUEZ**

**Dr. En I. RAÙL DE JESÙS GUTIÉRREZ**

**Toluca, México.2016**



## CONTENIDO

INTRODUCCIÓN.....	1
CAPÍTULO 1. MARCO TEÓRICO .....	5
1.1INTRODUCCIÓN .....	7
1.2 BIBLIOMETRÍA .....	7
1.3 MARCO HISTORICO: ANÁLISIS DE REGRESIÓN DE COX. ....	8
1.4 MARCO REFERENCIAL. ....	9
CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRAFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO. ....	21
2.1 INTRODUCCIÓN.....	23
2.2 LOCALIZACIÓN.....	23
2.3 SITUACIÓN DEMOGRÁFICA.....	25
2.4 REMESAS .....	26
2.5 DESEMPLEO. ....	28
2.6 PODER ADQUISITIVO. ....	31
2.7 ZONAS ECONÓMICAS Y CRECIMIENTO ECONÓMICO.....	32
2.8 LA REGIÓN DEL BAJÍO EN LOS MEDIOS DE COMUNICACIÓN .....	34
2.9 PEQUEÑAS Y MEDIANAS EMPRESAS DEL BAJÍO .....	37
CAPÍTULO 3.INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX. ....	39
3.1 INTRODUCCIÓN .....	41
3.2 ANÁLISIS DE SUPERVIVENCIA .....	41
3.3 MODELO DE REGRESIÓN COX.....	43
3.4 FUNCION DE RIESGO.....	44
3.5 MODELO DE RIESGO PROPORCIONAL .....	46
3.6 DATOS CENSURADOS .....	48
3.7 ESTIMACION DE PARAMETROS .....	50
3.7.1 FUNCION DE MAXIMA VEROSIMILITUD PARCIAL.....	52
3.8 ESTIMACIÓN DEL COEFICIENTE .....	54
3.9 ESTIMACIÓN DE LA FUNCION DE SUPERVIVENCIA .....	56
3.10 VALIDACIÓN DEL MODELO.....	58
3.10.1 PRUEBA DE WALD .....	58
3.10.2 PRUEBA DE RAZÓN DE VEROSIMILITUDES .....	58

3.10.3 PRUEBA DE PUNTAJES .....	59
3.11 INTERPRETACIÓN DEL MODELO.....	59
CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJIO. .....	61
4.1 INTRODUCCIÓN .....	63
4.2 DESCRIPCIÓN DE LA BASE DE DATOS.....	63
4.2.1 CLASIFICACIÓN POR CAPITAL MAYORITARIO.....	64
4.2.2 CLASIFICACIÓN DE EMPRESAS POR CERTIFICACIÓN.....	67
4.2.3 EDAD DEL GERENTE.....	70
4.2.4 DESCRIPCIÓN POR GÉNERO.....	71
4.2.5 GERENTE FUNDADOR .....	73
4.2.6 NIVEL DE FORMACIÓN ACADÉMICA DEL GERENTE .....	77
4.2.7 NÚMERO DE EMPLEADOS.....	80
4.2.8 CLASIFICACIÓN DE LAS EMPRESAS POR SECTOR.....	81
4.3 ANÁLISIS PORCRUCE DE VARIABLES.....	84
4.4 DESARROLLO DELMODELO DE REGRESIÓN DE COX.....	86
4.6 ESTIMACIÓN DE PARÁMETROS.....	87
4.5 INTERPRETACIÓN DE VARIABLES.....	90
4.7 COMPARACIÓN DE GRAFICAS DE SUPERVIVENCIA.....	92
CONCLUSIONES.....	95
BIBLIOGRAFÍA.....	99
ANEXOS.....	104
ANEXO A. REMESAS POR AÑO.....	104
ANEXO B. ENCUESTA NACIONAL DE OCUPACIÓN Y EMPLEO.....	105
ANEXO C. CÓDIGO EN R.....	106
ANEXO D. DATOS DE LA ENCUESTA 2011-2012.....	107

# INTRODUCCIÓN

---

Un país con crecimiento económico requiere de inversión, competencia y con ello el surgimiento de nuevas empresas que sirvan para estimular el mercado. Para que una nación se mantenga sana económicamente es forzoso el constante surgimiento de nuevas empresas que puedan competir en un ambiente democrático y de igualdad para evitar los monopolios anárquicos que desestabilizan el bienestar social. Todo esto con la finalidad de garantizar el desarrollo económico de los habitantes.

En la zona del bajo, la actividad económica y el creciente interés de los inversionistas extranjeros han creado mejores condiciones económicas para la población de esa localidad, así como incentivado las oportunidades de desarrollo, pero también, el crecimiento de las grandes empresas consolidadas dificulta el avance de las nuevas, debido a que se enfrentan en condiciones desiguales y en muchos de los casos son competencia directa.

Aunado a los retos externos que enfrentan las pequeñas y medianas compañías, existen variables propias que afectan directamente en su desarrollo o que de cierta manera son más fáciles de controlar. La subsistencia de una empresa no radica en la localización o tipo de competencia a su alrededor sino en su forma de organización, es por eso que en este estudio, el objetivo es identificar aquellas características internas de las compañías que favorecen a su crecimiento y determinar las principales variables que afectan en su rentabilidad a lo largo de los años.

Para la elaboración de este tratado se utilizaron datos reales de la encuesta de tipo transversal realizada el año 2011-2012 a pequeñas y medianas empresas de la región del Bajo donde se les pregunto acerca de sus actividades comerciales y características internas de la compañía.

# INTRODUCCIÓN

---

Así pues las hipótesis a comprobar son:

- Es posible construir un modelo general que determine la probabilidad de supervivencia de pequeñas y medianas empresas en México.
- Es posible identificar aquellas variables internas de la empresa que favorecen o afectan a su desarrollo.
- Demostrar que la técnica de regresión de Cox se puede llevar al campo financiero y económico.
- El modelo de regresión de Cox se puede aplicar a eventos probabilísticos como una herramienta útil y confiable.

La tesis se compone de cuatro capítulos. El primero trata sobre aquellos estudios teóricos que han utilizado la técnica de regresión de Cox, se expone la importancia que tiene en diferentes áreas y su consolidación con otras.

El capítulo dos se centra en el objeto de estudio, la zona del bajío, en esta sección se realiza una recopilación de información económica y demográfica, se consideran datos macroeconómicos y se comparan algunas estadísticas con la situación general del país, al final se expondrán algunas perspectivas o pronósticos que diferentes instituciones, editoriales o periódicos consideran para el territorio.

El título del tercer capítulo es “Introducción a la técnica de regresión de Cox”, que tiene por objetivo conocer específicamente la técnica de estimación no paramétrica de supervivencia, así como familiarizarse con los conceptos que implica su estudio y su análisis. Para cumplir con este objetivo, se explica de manera sistemática en qué consiste este método, iniciado con conceptos introductorios y la formulación del problema. Posteriormente se explica la

# INTRODUCCIÓN

---

metodología de construcción y selección de variables para el modelo, la estimación de los parámetros y las pruebas estadísticas del modelo.

Una vez entendida la referencia teórica del tema de investigación, en el capítulo 4 se describen las variables consideradas que se introducirán en el modelo, se explicaran las razones por las que fueron elegidas y para efectos de un mejor conocimiento de la información se hace estadística descriptiva y de supervivencia de cada variable y algunas combinaciones entre ellas, así como su diferencia estadística entre cada estrato de cada variable.

En el mismo capítulo, se desarrollara el modelo de supervivencia con apoyo del software estadístico R-Project versión 3.0.1 para obtener el modelo de regresión de Cox, se analizaran e interpretaran los resultados y se verifica la utilidad de esta técnica.

De esta manera se elabora la conclusión final del problema planteado inicialmente, demostrando la autenticidad o falsedad de las hipótesis propuestas.

# INTRODUCCIÓN

---

# CAPÍTULO 1. MARCO TEÓRICO

---



### 1.1 INTRODUCCIÓN

Para el desarrollo de este capítulo, se muestran las investigaciones publicadas en revistas científicas en línea en las que se pudo acceder. En dicha exploración, se consideraron aquellos artículos de acceso libre en la web.

Se realiza un breve recorrido histórico del desarrollo de la técnica que aquí se utilizará y finalmente se expone de manera general y particular algunos de los estudios que han utilizado en diferentes áreas del conocimiento.

### 1.2 BIBLIOMETRÍA

La búsqueda se hizo teniendo como prioridad examinar información y documentos sobre el Modelo Cox, junto con esta técnica, se inspecciono por algunas disciplinas afines a la actuaría como finanzas, economía y estadística.

Buscando la palabra “Cox Model” en la revista Springer Link, se encuentran variados documentos, un total de 127,688, repartidos entre artículos y capítulos de libros, 81,274 y 45,217 respectivamente, la mayoría de estos han sido publicados en inglés (124,674).

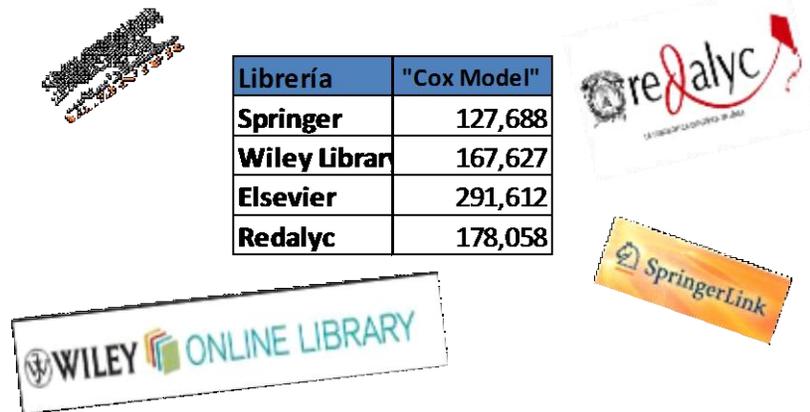
La aplicación de este método va desde medicina (48,047), ciencias biomédicas (27,012), matemáticas, economía y estadística, estas últimas en la posición 9, 11 y 13 de documentos escritos en la publicación, un total de 7,207, 6,059 y 5,767 respectivamente.

No se encontró ningún resultado para el área de finanzas ni seguros. Filtrándolo por el área económica se encuentran siete resultados, todos con acceso restringido.

## CAPÍTULO 1. MARCO TEÓRICO

A continuación se muestra en la tabla 1.1 los resultados que arrojó el buscador de cada una de las revistas en las que se averiguo.

**Tabla 1. 1 Número de artículos encontrados por librería.**



Librería	"Cox Model"
Springer	127,688
Wiley Librar	167,627
Elsevier	291,612
Redalyc	178,058

**Fuente:** Elaboración propia con base en el número de publicaciones por revista.

Es necesario destacar que la mayoría de los artículos son restringidos, sobre todo aquellos en los que la investigación es realizada para áreas que no son asociadas a las ciencias médicas.

Aunque el buscador de la mayoría de las revistas científicas no se puede filtrar por ramas como en Springer Link, la mayoría de estas investigaciones se han realizado para las ciencias médicas y solo una minoría ha sido para el campo actuarial o económico.

### 1.3 MARCO HISTORICO: ANÁLISIS DE REGRESIÓN DE COX.

El modelo semi-paramétrico proporcional Hazard o de riesgo y el modelo análogo multiplicativo que involucra covariantes del tiempo fue introducido por Cox (1972) para posteriormente ser desarrollado con bases teóricas por el mismo Cox (1975). Rápidamente fue reconocido como base para el análisis de supervivencia, particularmente en el área médica, epidemiológica y la demografía.

## CAPÍTULO 1. MARCO TEÓRICO

---

Los documentos de análisis más rigurosos sobre la procedencia de máxima verosimilitud parcial y la conexión entre la verosimilitud de Cox y el perfil de funciones de verosimilitud surgieron por Bailey (1983,1984), Jacobsen (1984) y Johansen (1983) y las estimaciones de la función base acumulativa  $H_0(t)$  y la función de supervivencia  $S_0(t)$  fueron dirigidas por Breslow (1974) (Lawless, 2002).

Información y documentos teóricos sobre esta técnica se pueden encontrar desde un nivel introductorio como los de Collett (2003), Cox y Oakes (1984), Hosmer y Lemeshow (1999) y el texto de Miller(1981), los de nivel intermedio como Kalbfleisch y Prentice (2002), Klein y Moeschberger (2003), Lawless (2002), Lee y Wang (2003), Martinussen y Scheike (2006) y Therneau y Grambsch (2000), hasta los de nivel más avanzado como los de Andersen (1993) y Fleming y Harrington (1991) y más recientemente surgen técnicas de computación evolutiva conocida como algoritmos genéticos (Douglas et al., 2006).

De acuerdo con Domenéch (2006), “el análisis de supervivencia y en especial el modelo Cox han tenido una creciente aplicación en psicología (Griffin, Gardner, 1989; Singer y Willet, 1991)”.

### 1.4 MARCO REFERENCIAL.

Como se ha mencionado anteriormente, la aplicación de la regresión Cox ha sido utilizada en diversas áreas, no sólo se puede estudiar la muerte de un individuo, esta técnica es posible utilizarla (si se está interesado) para conocer el periodo de subsistencia de algún elemento. Se observó que el área en la que más se aplica esta técnica es en la medicina, ya que muchos investigadores de las ciencias médicas requieren conocer el tiempo de supervivencia de un paciente con determinada enfermedad o si se trata con un tratamiento específico, además desean saber los factores (covariables) que más influyen en el tiempo de supervivencia. Este tipo de estudios son un antecedente o pre ambulo para las

## CAPÍTULO 1. MARCO TEÓRICO

---

investigaciones actuariales que pueden servir para el cálculo de riesgo en el contrato de un seguro.

Otros estudios van desde la fusión de empresas hasta el riesgo de daño de un aparato los cuales a continuación se mencionan, de tal manera que servirían para el cálculo de la prima de un seguro de daños o la probabilidad de ruina de una empresa.

Douglas et al. (2006) publicaron el trabajo de investigación con título: "Aplicación de los algoritmos genéticos para estimar los parámetros en un modelo de regresión de Cox" basándose en un estudio previo hecho en Caracas, entre 1980 y el año 2000 por Borges (2002, 2005), el cual estima los parámetros de Cox para pacientes del servicio de diálisis peritoneal del Hospital Clínico Universitario de Caracas, tomando como muestra 246 individuos; de esta se determinan tres variables importantes, diabetes, edad e índice quetelet o índice de masa corporal.

Este estudio propone un método alternativo para estimar los parámetros del modelo de regresión de Cox utilizando las técnicas de algoritmos genéticos. El resultado fue que el algoritmo no da la seguridad de obtener un óptimo global, sin embargo da una mejor solución que el método clásico. También se aprecia que varios valores de los parámetros obtenidos se encuentran dentro del intervalo de confianza de la técnica clásica (Douglas et.al., 2006).

Ayala, Borges y Colmenares (2007) verificaron la validez del modelo de Cox en su aplicación para identificar una alerta temprana de quiebra de una institución financiera. En aquellos años, la legislación de Venezuela no permitía la quiebra de una empresa, sino que fomentaba la fusión con otra, por tal motivo esta actividad fue el evento súbito que se evaluó considerando 34 entidades bancarias y 15 razones financieras, obteniendo cuatro variables estadísticamente significativas.

## CAPÍTULO 1. MARCO TEÓRICO

---

El trabajo resultó satisfactorio y consistente debido al resultado de las pruebas de residuos de Schoenfeld, residuos de Martingala, residuos *deviance* y residuos de puntuación, quienes verificaron en su totalidad el modelo. Además, en este se introduce el análisis de supervivencia y la regresión de Cox como metodología para la detección de variables indicadoras de riesgo de fusión en las instituciones financieras.

Para Cruz et al. (2010), la utilidad que le da a la técnica de regresión en el sector industrial es para determinar las variables que influyen en la calidad de servicio de empresas outsourcing. Consideró 19 empresas de las cuales se obtuvieron las covariables que más influyen en la calidad de servicio, estas fueron; tipo de mantenimiento, existencia de repuestos en el país, entrenamiento a usuarios, complejidad tecnológica, distancia entre la empresa y hospital, y cantidad de visitas de la empresa al hospital para medir la calidad del servicio mediante el tiempo de parada de los equipos.

En esta investigación se deseaba un aumento en el riesgo de supervivencia, pues el evento de estudio es el tiempo en que el equipo estará detenido (TAT; variable de estudio definido con esas siglas por los autores). Esta investigación difiere de los otros estudios debido a que generalmente toda covariable que aumente el riesgo para la supervivencia es un efecto indeseado porque acorta el período de vida de un paciente o funcionamiento de un equipo.

Los autores concluyeron en que al obtener el modelo de Cox para el tiempo de espera de los equipos en función de las covariables estadísticamente significativas permite predecir a las partes que participan en la formulación de contratos de servicios de mantenimiento el comportamiento en función de las características de las empresas (Cruz et al., 2010).

## CAPÍTULO 1. MARCO TEÓRICO

---

En ciencias de la salud, Doménech (1992) realizó un estudio de recaída de heroínómanos<sup>1</sup> recién deshabituados. El autor utilizó el método Kaplan-Meier y actuarial para estimar las funciones de supervivencia y de riesgo mientras que para uso predictivo manejó la regresión de Cox. Tomó una muestra de 32 heroínómanos observados de enero de 1990 hasta junio de 1992 y consideró tres variables; trabajo estable, edad al consumo de la droga y duración del consumo de la heroína dividido en tres categorías por periodos de tiempo.

El autor describe que por el método Kaplan-Meier la mediana del tiempo de recaída es de 218 días, por el método actuarial se determina de 7.32 meses, poco más de 222 días, con la regresión de Cox determinó que para la edad, el efecto (sobre la tasa de riesgo) de empezar 5 años mas tarde a consumir drogas, es el mismo entre los 14 y 19 años que entre los 24 y 29 años (Domenech, 1992).

Otra aplicación en el área médica fue la realizada por Nogales y Porras (2009) donde emplearon el modelo de regresión de Cox a un caso de estudio del sida en España. El estudio se basa en una muestra de 1,366 mujeres diagnosticadas con SIDA en Andalucía entre 1982 y 2001 y se utilizaron cuatro variables para construir el modelo, estas fueron; edad, fecha de diagnóstico, provincia de residencia y categoría de transmisión.

La investigación manejó la técnica de Kaplan-Meier para comparar las curvas de supervivencia donde se observó que el grupo de 30-35 años presenta una mayor supervivencia, mientras que el grupo de más de 35 años es el que presenta una peor resistencia de vida.

Los autores compararon las curvas de supervivencia de Kaplan Meier divididas en tres grupos de acuerdo al año de diagnóstico 1982-1992, 1993-1995 y 1996-2001, se observa que existe una mayor supervivencia para el último grupo

---

<sup>1</sup> Persona adicta a la heroína.

<sup>2</sup>La prostatectomía (extracción de la próstata) es la cirugía para extirpar toda la glándula prostática y algunos

## CAPÍTULO 1. MARCO TEÓRICO

---

después de cinco años. Respecto a la categoría de transmisión el grupo de los homosexuales tiene una supervivencia menor que el resto de los grupos hasta aproximadamente los 50 meses desde el diagnóstico. Con relación a la provincia de residencia, Granada y Sevilla son las provincias con menor riesgo, con una supervivencia hasta los 5 años y probabilidad de 67% y 58% respectivamente mientras que Málaga es la de menor probabilidad al mismo periodo con un 44%.

El análisis multivariado permitió observar que la introducción del denominado TARGA (Terapia antirretroviral activa) en 1996, mejora la supervivencia de las mujeres diagnosticadas de SIDA ya que las mujeres diagnosticadas antes de ese año de 1993-1995 tienen un riesgo de morir 2.3 veces superior que las diagnosticadas después. En cuanto a la edad, las mayores de 35 años tienen un riesgo de morir de 1.8 veces superior que las menores de 30 y la provincia con mayor supervivencia es Granada, mientras que para la categoría de transmisión no se encontraron diferencias (Nogales y Porras, 2009).

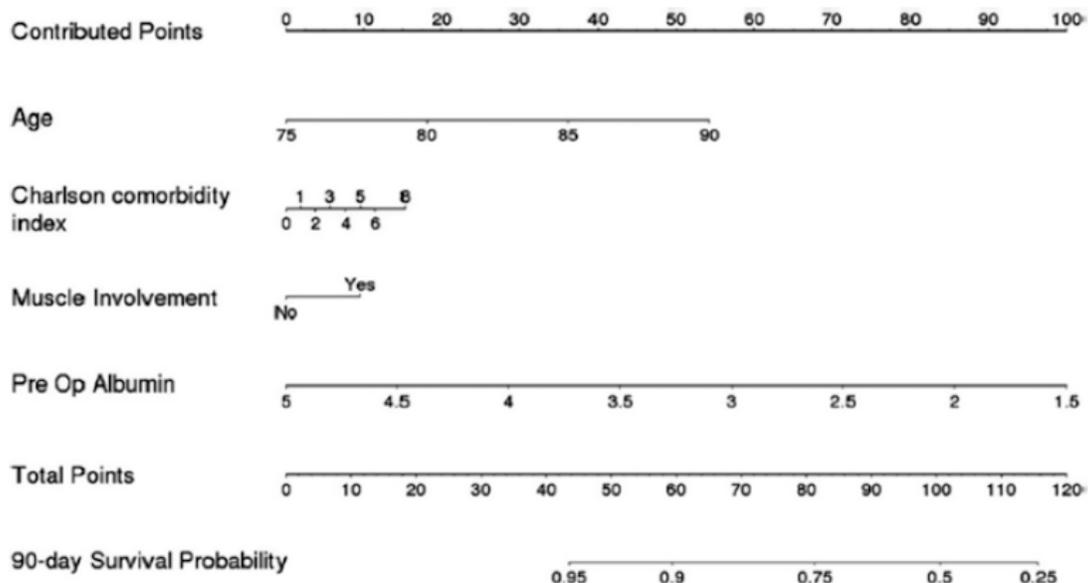
Siguiendo en línea con la revisión de los estudios médicos es importante destacar que la principal aplicación de la técnica de Cox ha sido realizada para estudios relacionados con el tratamiento del cáncer como los trabajos que a continuación se mencionan.

Morgan et al. (2011) desarrollaron un modelo estadístico multivariado con la finalidad de ayudar a la predicción de resultados después de la decisión clínica. El estudio tomo como muestra a 220 pacientes mayores de 75 años que se sometieron a un tratamiento en el Centro Médico de la Universidad de Vanderbilt (*Vanderbilt University Medical Center*) entre los años 2000 a 2008. Las variables obtenidas que se incluyeron son; edad, sexo, color de piel, albúmina preoperatoria, índice de comorbilidad Charlson CCI, etapa clínica, etapa patológica y el tipo de derivación urinaria.

## CAPÍTULO 1. MARCO TEÓRICO

Luego de realizar las pruebas necesarias, se consideraron cuatro de ocho variables estadísticamente significativas y la herramienta gráfica fue un nomograma mostrado en la Figura 1.1. La Figura se lee posicionándose sobre las características del paciente que tienen cierto puntaje definido por la línea superior y al realizar la suma de puntos se le asigna una probabilidad.

**Figura 1.1.** Nomograma predictivo de supervivencia a 90 días para pacientes en edad avanzada después de una cistectomía radical.



**Fuente:** Predicción de la probabilidad de supervivencia a 90 días de pacientes ancianos con cáncer de vejiga tratados con cistectomía radical (Morgan et al., 2011).

El objetivo de este estudio fue para determinar los candidatos a este tipo de cirugías observando la probabilidad de supervivencia en los siguientes 90 días críticos, después de la cirugía. Para aquellos que exceden esos días se consideran censurados (Morgan et al., 2011).

Dos años antes, Walzet al. (2009) se enfocaron en el estudio del cáncer de próstata con el propósito de identificar pacientes con riesgo de recurrencia de esta

## CAPÍTULO 1. MARCO TEÓRICO

---

enfermedad después de una prostectomía<sup>2</sup> radical, esto con la finalidad de reducir el riesgo de progresión.

El tratado fue hecho de 1992 al 2005 con un cohorte de 2,911 pacientes que se sometieron a la prostectomía. Las características que se consideraron al inicio del estudio para el modelo de regresión Cox fueron las siguientes; edad, antígeno prostático específico, gleason patológico, márgenes quirúrgicos, invasión de las vesículas seminales, extensión extra capsular y la invasión de los ganglios linfáticos.

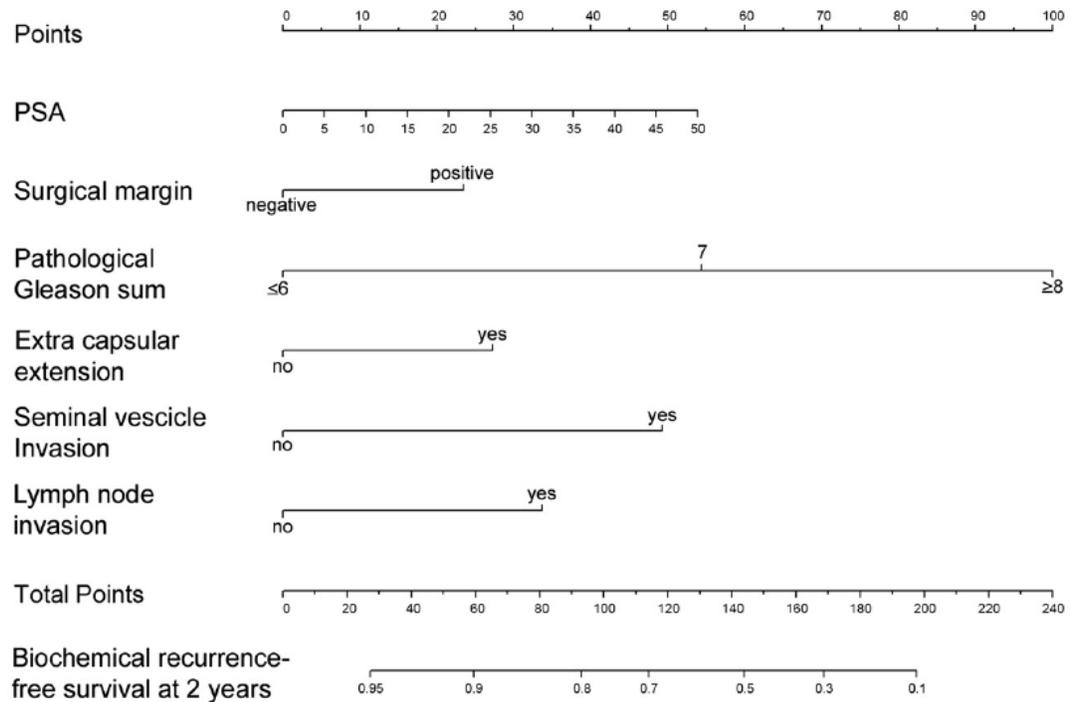
El resultado del estudio se esbozó en el nomograma de la Figura 1.2 el cual se lee de la misma manera que el anterior, nótese que de las variables antes mencionadas la única que no se encuentra es edad, esto porque fue una variable que no es estadísticamente significativa.

---

<sup>2</sup>La prostatectomía (extracción de la próstata) es la cirugía para extirpar toda la glándula prostática y algunos tejidos alrededor de esta. Se realiza con el fin de tratar el cáncer de próstata.

## CAPÍTULO 1. MARCO TEÓRICO

**Figura 1.2** Nomograma de predicción de la supervivencia libre de recurrencia biomédica a los 2 años después de prostatectomía radical.



**Fuente:** Nomograma de predicción de la probabilidad de recurrencia precoz tras la prostatectomía radical para el cáncer de próstata (Walzet al. 2009).

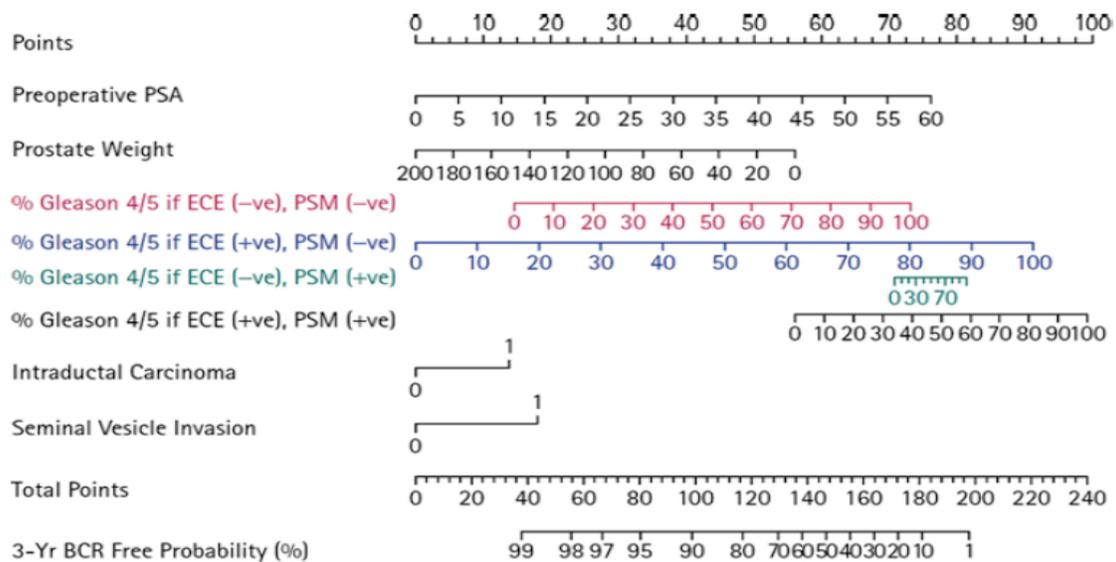
Un año más tarde, O'Brien et al. (2010) también estudiaron a pacientes que estuvieron expuestos a una prostatectomía radical, para este caso se buscó validar el nomograma de Kattan con una nueva muestra de 1,939 casos válidos para el análisis provenientes del Hospital de la Universidad de Australia del año 1998 al 2007.

Así mismo, el nomograma de Kattan establece la importancia de las variables clínico patológicas estas son; nivel de PSA sérico preoperatorio, gleason, extensión extra capsular (CEPE), la invasión de la vesícula seminal (SVI), márgenes quirúrgicos positivos (PSM) y la afectación de los ganglios linfáticos pélvicos (LNI) para predecir el riesgo de recurrencia bioquímica (BCR) tras la prostatectomía radical (DeMaris, 2013).

## CAPÍTULO 1. MARCO TEÓRICO

Además de las variables consideradas por el nomograma de Kattan, O'Brien et al. (2010) agregan el porcentaje de cáncer con los patrones de Gleason 4 y/ó 5, peso de la próstata, presencia o ausencia de carcinoma intraductal (PICD), y el volumen del tumor (DeMaris, 2013) para la inclusión al gráfico validados con la regresión de Cox.

**Figura 1. 3** Nomograma predictivo a tres años para determinar la recurrencia biomédica después de una prostatectomía radical.



**Fuente:** Recurrencia biomédica después de una prostatectomía radical (O'Brien, 2010).

Finalmente, los autores comentan que el estudio es útil para la predicción de fallos prematuros potencialmente temibles, identificando así los pacientes que pueden beneficiarse de la terapia adyuvante (O'Brien, 2010).

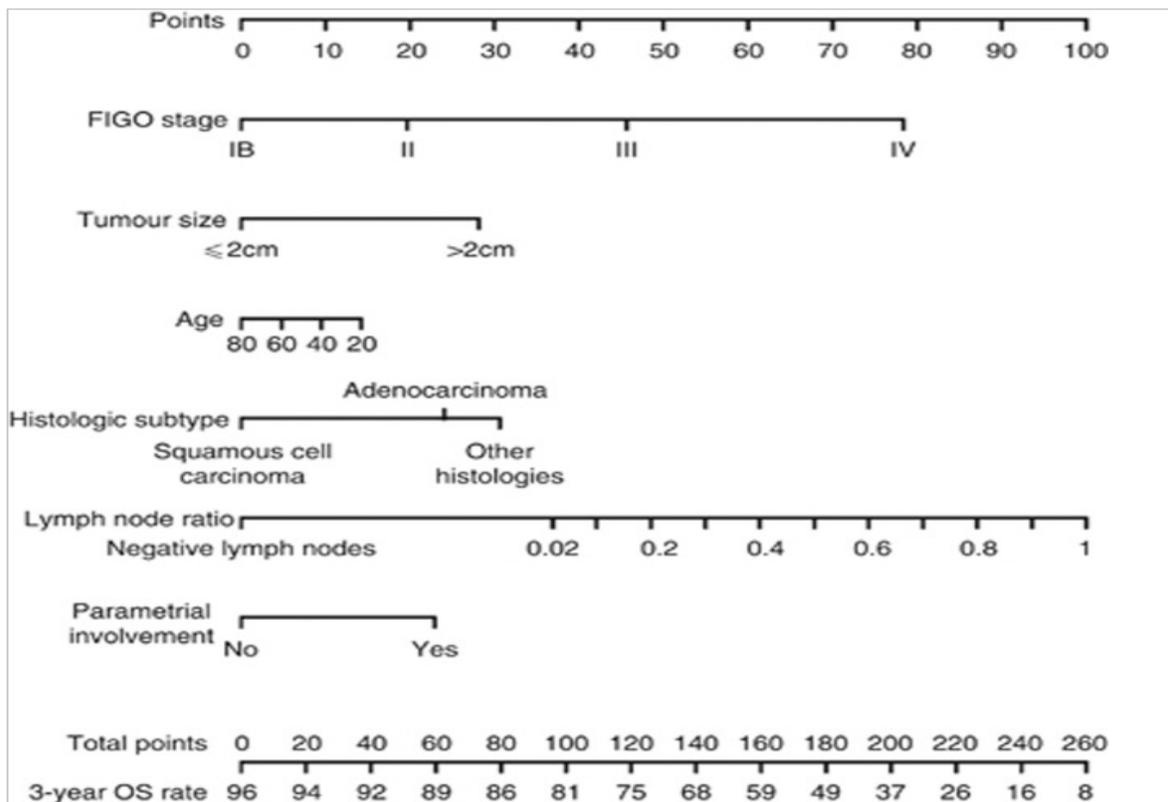
Otros estudios de supervivencia de Cox es la realizada por Polterauer et al. (2012) "Nomograma predictivo para la supervivencia de pacientes diagnosticados con cáncer cervical". El estudio tiene una muestra de 528 individuos diagnosticados con cáncer cervical en el Hospital de la Universidad de Viena e Innsbruck, Austria, durante el periodo de 1996 al 2009.

## CAPÍTULO 1. MARCO TEÓRICO

El modelo de regresión de Cox resultante seleccionó las siguientes variables estadísticamente significativas; edad, el estadio clínico con base a la Federación Internacional de Ginecólogos y Obstetras (FIGO; refleja el parámetro de pronóstico más importante en los pacientes con cáncer de cuello uterino), presencia de ganglios linfáticos positivos, la presencia de ganglios positivos relación de los ganglios linfáticos, participación parametrial y tamaño del tumor.

Polterauer et al. (2012) mencionan “El nomograma (Figura 1.4) se puede utilizar para predecir el pronóstico de los pacientes de forma individual y más precisa que la etapa FIGO y se basa en los siguientes seis parámetros fácilmente disponibles: estadio FIGO, tamaño tumoral, tipo histológico, relación de los ganglios linfáticos (LNR), la edad, y la participación del parametrio”.

**Figura 1. 4** Nomograma predictivo para la supervivencia de pacientes diagnosticados con cáncer cervical



**Fuente:** Nomograma predictivo para la supervivencia de pacientes diagnosticados con cáncer cervical (Polterauer et al., 2012).

## CAPÍTULO 1. MARCO TEÓRICO

---

Como se ha notado en las páginas anteriores, diversos autores utilizaron tanto la técnica de regresión de Cox como la del nomograma. Ambas han sido objeto de estudio para la creación de paqueterías especializadas. Campanu y Gonen (2011) escribieron el artículo “Construcción de un nomograma para estudios orientados a modelos de Cox utilizando R”.

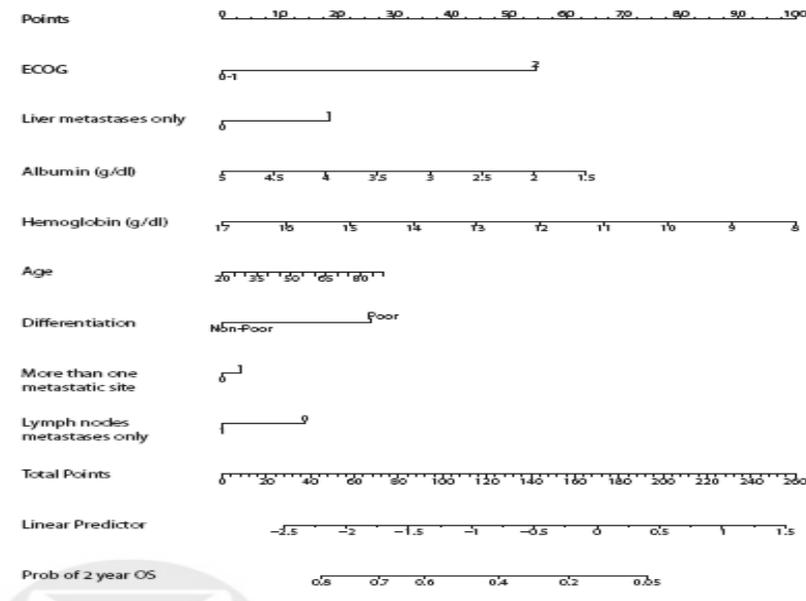
Esta investigación tuvo fundamentos en la de Power, Capanu, Kelsen y Shah “Desarrollo de un nomograma para predecir la supervivencia a 2 años a pacientes con cáncer gástrico metastásico” para construir el nomograma del modelo de regresión Cox, partiendo de la necesidad de utilizar un software especializado, en este caso se utilizó R-Project para facilitar la elaboración del nomograma.

Los datos son de aquellas personas que recibieron quimioterapia en el Centro de Cáncer en Memoria de Sloan-Kettering de enero de 1999 a julio de 2007, obteniendo un total de 985 pacientes.

Finalmente los autores mencionan que el modelo de Cox es usualmente usado para la construcción del nomograma, el cual modela la relación entre el conjunto de covariables y la función de riesgo de un particular tiempo de fallo. Para este caso el modelo se presenta en la Figura 1.5.

## CAPÍTULO 1. MARCO TEÓRICO

**Figura 1. 5** Nomograma clínico para pacientes con cáncer gástrico metastásico tratados con quimioterapia sistémica estimando la probabilidad de sobrevivir durante 2 años.



**Fuente:** Construcción de un nomograma para estudios orientados a modelos de Cox utilizando R (Campanu y Gonen, 2011).

Nótese que en este capítulo muchas de las investigaciones revisadas utilizan la técnica de regresión Cox como técnica de evaluación de las variables que impactan en un evento súbito y se utiliza como herramienta visual al nomograma, donde las áreas más importantes son la medicina y en mucho menor medida la financiera, económica y actuarial, que es donde se debe centrar mayor atención y donde nos enfocaremos en los siguientes capítulos.

**CAPÍTULO 2. ANÁLISIS  
ECONÓMICO, DEMOGRAFICO Y  
PERSPECTIVAS DE CRECIMIENTO  
DE LA REGIÓN DEL BAJÍO.**

---

## **CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.**

---

## **CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.**

---

### **2.1 INTRODUCCIÓN.**

El área del bajío es considerada ya una de las zonas económicas y estratégicas más importantes del país, debido a la gran cantidad de recursos que ha canalizado de inversionistas extranjeros y donde se presentan oportunidades macroeconómicas.

En este capítulo se presenta un esbozo de la magnitud e importancia tanto económica, geográfica y demográfica de la región del bajío, las cuales son factores para que empresas mexicanas, pequeñas y medianas, mejoren o fortalezcan su crecimiento.

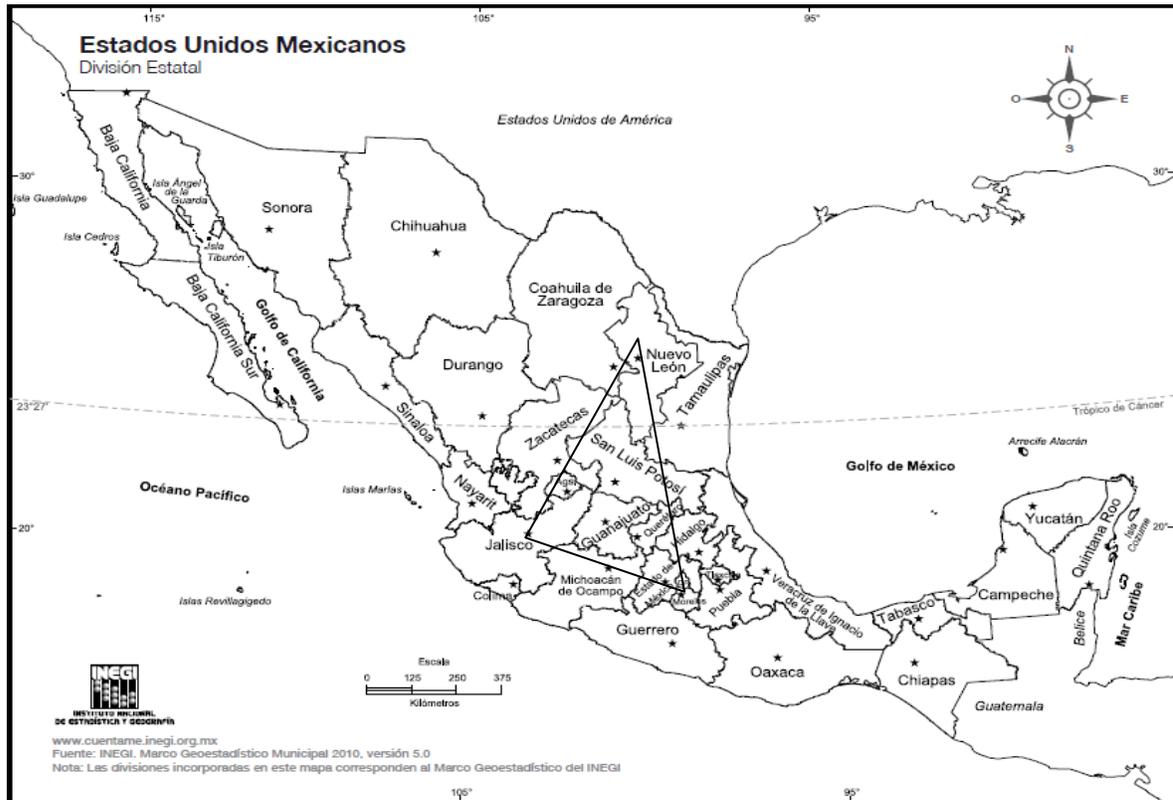
Adicional, en esta sección se muestran las perspectivas económicas que tienen diversos medios y expertos sobre la región y el impacto que tendrá sobre las Pymes en la zona.

### **2.2 LOCALIZACIÓN.**

El bajío mexicano está constituido por los estados de Guanajuato, Querétaro y Aguascalientes y por regiones de Jalisco, Michoacán y San Luis Potosí como se muestra en la figura 2.1. La zona se encuentra dentro de un triángulo formado por las principales ciudades del país en términos económicos de México; Ciudad de México, Guadalajara y Monterrey.

## CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.

**Figura 2. 1.** Mapa de la división política de los Estados Unidos Mexicanos; se enmarca la región del bajo.



**Fuente:** Instituto Nacional de Estadística y Geografía (INEGI) en línea [www.inegi.org.mx](http://www.inegi.org.mx) consultado el día 01 de febrero del 2015.

Estratégicamente, la zona está ubicada en una de las rutas más importantes del comercio mundial conectándose con Asia, Estados Unidos y el resto de América. Con Asia y Sudamérica porque tiene cercanía con el puerto michoacano de Lázaro Cárdenas, el cual tiene capacidad para manejar hasta 2 millones de TEUs<sup>3</sup> al año, por lo que hay un margen amplio de transporte de carga. El mismo puerto conecta a la red ferroviaria que incluye contratos con la Kansas City Southern que va a Estados Unidos y Canadá (Vela, 2015).

<sup>3</sup>Tráfico Marítimo de Contenedores o Unidad equivalente a veinte pies

## CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.

### 2.3 SITUACIÓN DEMOGRÁFICA.

El bajío contiene la totalidad de los municipios de Aguascalientes, Guanajuato y Querétaro. De acuerdo con la información obtenida del censo del INEGI del 2010 (Tabla 2.1), Aguascalientes cuenta con una población total de 1,184,996 habitantes, que representa un 1.05 por ciento del total de la población, a su vez, Guanajuato y Querétaro equivalen al 4.88 por ciento y 1.63 por ciento con 5, 486,372 y 1, 827,937 de habitantes respectivamente (INEGI, censo 2010).

Como se puede observar en la tabla 2.1, Jalisco, Michoacán y San Luis Potosí tienen un total de 7,350,682, 4,351,037 y 2,585,518 de pobladores respectivamente, de este total solo 26 municipios de San Luis Potosí con 1, 871,310 de habitantes, 26 de Michoacán con 629,240 habitantes y 43 de Jalisco con 5, 651,633 de habitantes son parte de la región de estudio.

Considerando el tamaño de la población de los municipios que conforman esta zona y sin tomar en cuenta algunos de Zacatecas, el total de personas que se encuentran es de 16, 651,488 que es un 14.82 por ciento del total en México.

**Tabla 2. 1.** Resumen de habitantes por estado en el bajío y aportación proporcional a la zona.

Estado	Censo 2010	Estado/Total México	Bajío	Región Bajío/Total México	Región Bajío/Total Bajío
México	112,336,538	100.00%	112,336,538	100.00%	
Aguascalientes	1,184,996	1.05%	1,184,996	1.05%	7.12%
Guanajuato	5,486,372	4.88%	5,486,372	4.88%	32.95%
Jalisco	7,350,682	6.54%	5,651,633	5.03%	33.94%
Michoacán d	4,351,037	3.87%	629,240	0.56%	3.78%
Querétaro	1,827,937	1.63%	1,827,937.00	1.63%	10.98%
San Luis Potd	2,585,518	2.30%	1,871,310	1.67%	11.24%
			<b>16,651,488</b>		

**Fuente:** Elaboración propia con información del INEGI censo 2010.

En promedio, el 73 por ciento de la población de los estados conforman la del bajío y la relación de hombres y mujeres en de 51% y 49% respectivamente.

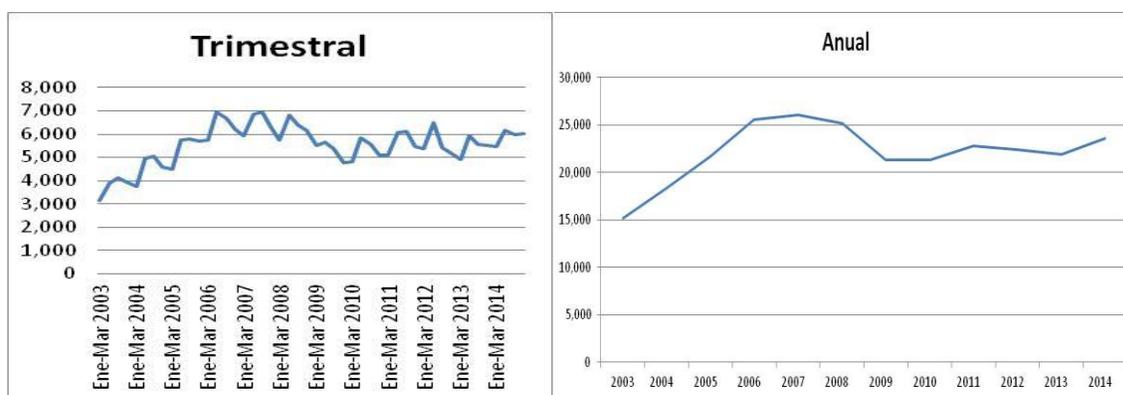
## CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.

De acuerdo con la Revista Por qué Bajío (2015), la zona posee cerca de 30 ciudades medianas con más de 100 mil habitantes cada una, León es la sexta ciudad del país con 1 millón 500 mil habitantes, le siguen San Luis Potosí y Querétaro con 1 millón 100 mil, Morelia y Aguascalientes con 750 mil, y finalmente Celaya e Irapuato con 500 mil habitantes aproximadamente cada una.

### 2.4 REMESAS

De acuerdo con el Banco de México, durante el 2014 las remesas tuvieron un incremento del seis por ciento con respecto del año anterior, el punto más alto se alcanzó en el 2007 antes de la crisis global del año siguiente, con este incremento se observa una recuperación paulatina para los siguientes periodos (Figura 2.2).

**Figura 2.** 2.Gráfica histórica de remesas en México, a la izquierda datos trimestrales, a la derecha datos anuales desde el 2003.



**Fuente:** Elaboración propia con base en datos del Banco de México en línea

[www.banxico.org.mx](http://www.banxico.org.mx).

El monto alcanzado para el año 2007 fue de 26,058.82 millones de dólares, posteriormente el mínimo se estableció en el 2010 con 21,303.88 millones. En el 2014 se tiene un registro de 23,647.28 millones.

Continuando con la figura 2.2 se aprecia que el incremento registrado de las remesas que los mexicanos residentes en el extranjero enviaron a México en

## **CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.**

---

estos meses muestra una importante mejoría considerando la caída de 2.4 por ciento registrada durante el 2013; y su tendencia sigue siendo a la alza.

Por otro lado se puede decir que 16 estados son los que captan el 80 por ciento del total del país, entre los que destacan Michoacán, Guanajuato y Jalisco. San Luis Potosí se encuentra en la posición número 11 de captación a total país, Querétaro y Aguascalientes se encuentran en el lugar 21 y 25 en captación de dólares (Ver Anexo A)

El primer lugar en captación es para Michoacán que se ha mantenido en esa posición desde el 2003, durante el 2014 recaudo un total de dos mil 244 millones de dólares aproximadamente, cifra que muestra un incremento del 10 por ciento con respecto al año anterior. Con dicho monto captó el 9.19 por ciento del total nacional.

Guanajuato sigue consolidando su segundo lugar nacional como captador de remesas con un incremento de 4.43 por ciento, esta razón de cambio anual es superado por 16 estados quienes se han desempeñado de mejor manera de un año a otro, pero aun así, el estado tiene una participación del 8.87 por ciento del total con un monto de dos mil 96 millones de dólares (Ver Anexo A).

La entidad jalisciense se colocó en el tercer lugar de captación a nivel nacional, con remesas que ascendieron de 1,755 millones en 2013 a 1,959 millones de dólares en 2014, cifra que muestra un incremento de 11.67 por ciento. A nivel nacional Jalisco captó el 8.29 por ciento de las remesas totales.

San Luis Potosí que aporta el 3.26 por ciento del total, tuvo un crecimiento del 8.89 por ciento, colocándose en el tercer lugar de incremento en la región del bajío.

## **CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.**

---

Aguascalientes, quien participa con el 1.37 por ciento de captación de remesas con respecto del total, incremento el monto en 5.98% al pasar de 305 millones en 2013 a 323 millones de dólares aproximadamente en 2014.

A su vez, Querétaro con un monto de 398.2 millones de dólares indica un decremento de 3.24 por ciento, destacando que fue la entidad con menor tasa de crecimiento.

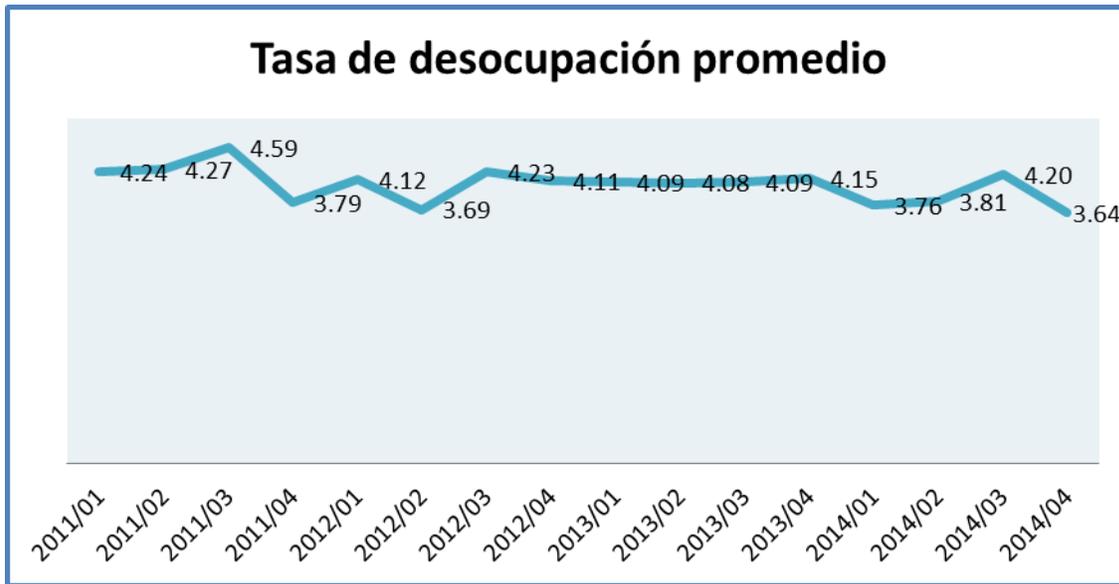
Nótese que los estados que integran la región del bajío captan cerca del 35 por ciento del total de remesas a nivel nacional, lo que indica una fuerte dependencia a este recurso y un alto índice de migración, sobre todo en Michoacán, Guanajuato y Jalisco.

### **2.5 DESEMPLEO.**

Con información del INEGI de la Encuesta Nacional de Ocupación y Empleo (ENOE, 2014), se muestra en la figura 2.3 la tasa promedio de desocupación de los estados que conforman la región del bajío. Como se aprecia en la gráfica el promedio ronda entre el tres y cuatro por ciento con tendencia a la baja para el último trimestre del 2014.

## CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.

**Figura 2.3.** Tasa de desocupación promedio.



**Fuente:** Elaboración propia realizada con información del INEGI [www.inegi.org.mx](http://www.inegi.org.mx).

En Aguascalientes la tasa cerró en 5.48 por ciento, mientras que en San Luis potosí se presentó la más baja con 2.83 por ciento, para Jalisco fue de 4.89 por ciento, Guanajuato 4.2 por ciento y Querétaro se estableció con una de las más altas con 5.45 por ciento(Ver Anexo B).

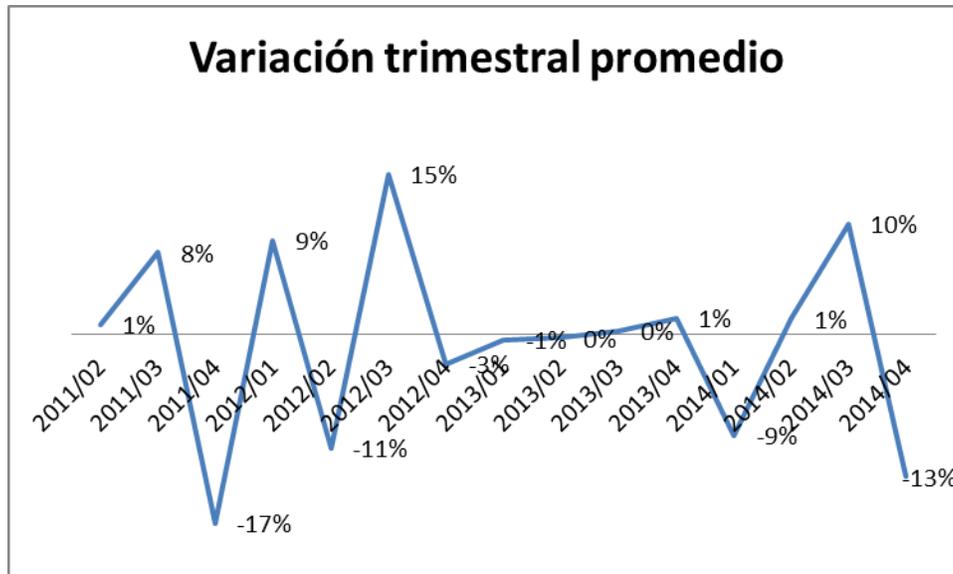
Respecto a la variación del último trimestre del 2013 al 2014, en todos los estados se registró un decremento en la tasa, Guanajuato fue quien presento mayor variación al respecto ya que paso de 5.15 por ciento a 4.20 por ciento, lo que es una reducción de 19 por ciento.

Después de Jalisco, Querétaro tuvo un decremento significativo en la variación de 14 por ciento, ya que anteriormente se mantenía en 6.34 y bajo a 5.45. Para Aguascalientes, San Luis Potosí y Michoacán, el porcentaje de variación para el mismo periodo mencionado anteriormente fue de -7, -9 y -9 por ciento respectivamente (Ver Anexo B).

## CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.

En General, para el último trimestre del 2014 la tasa de desocupación bajo para los estados de la región, mientras que a nivel nacional se mantuvo cerca del 5 por ciento, en la siguiente figura se observan las variaciones trimestrales desde el año 2011 al final del 2014.

**Figura 2.4.** Variación trimestral promedio



**Fuente:** Elaboración propia realizada con información del INEGI [www.inegi.org.mx](http://www.inegi.org.mx).

La figura 2.4 muestra variaciones drásticas trimestre con trimestre, que si bien en el último periodo mostrado hubo un decremento importante, no se percibe una tendencia, usualmente en el segundo y cuarto trimestre existen caídas importantes mientras que luego se perciben recuperaciones para los periodos consecutivos. Para validar la información se dejan los datos de tasa de desocupación en el Anexo B.

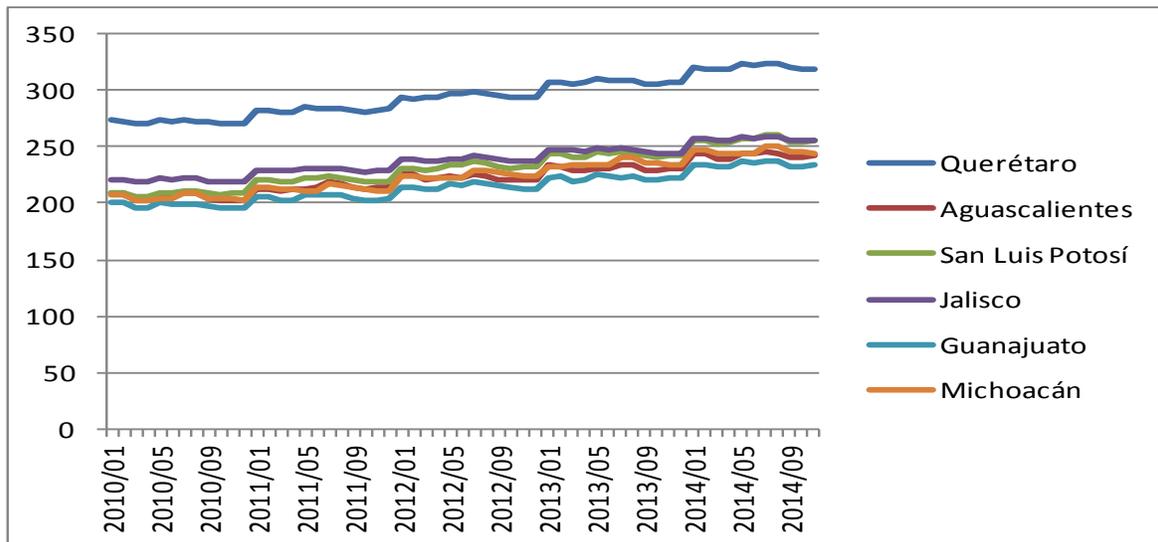
## CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.

### 2.6 PODER ADQUISITIVO.

Para realizar el análisis de este apartado se tomara como referencia el salario base de cotización del Instituto Mexicano de Seguridad Social (IMSS) por entidad federativa, de tal manera que tengamos una noción o parámetro del poder adquisitivo.

El salario de cotización de estas regiones se ha mantenido en crecimiento constante, con variaciones cercanas a la inflación con Querétaro siempre por encima de la media de estas entidades (Figura 2.5). Para el final del 2014 el promedio anual diario se registró de 320 para Querétaro, seguido de San Luis Potosí con 256 pesos y Jalisco con 255 pesos (Figura 2.6).

**Figura 2.5.** Evolución promedio del salario base de cotización de entidades pertenecientes al bajío. Año 2010 a 2014.



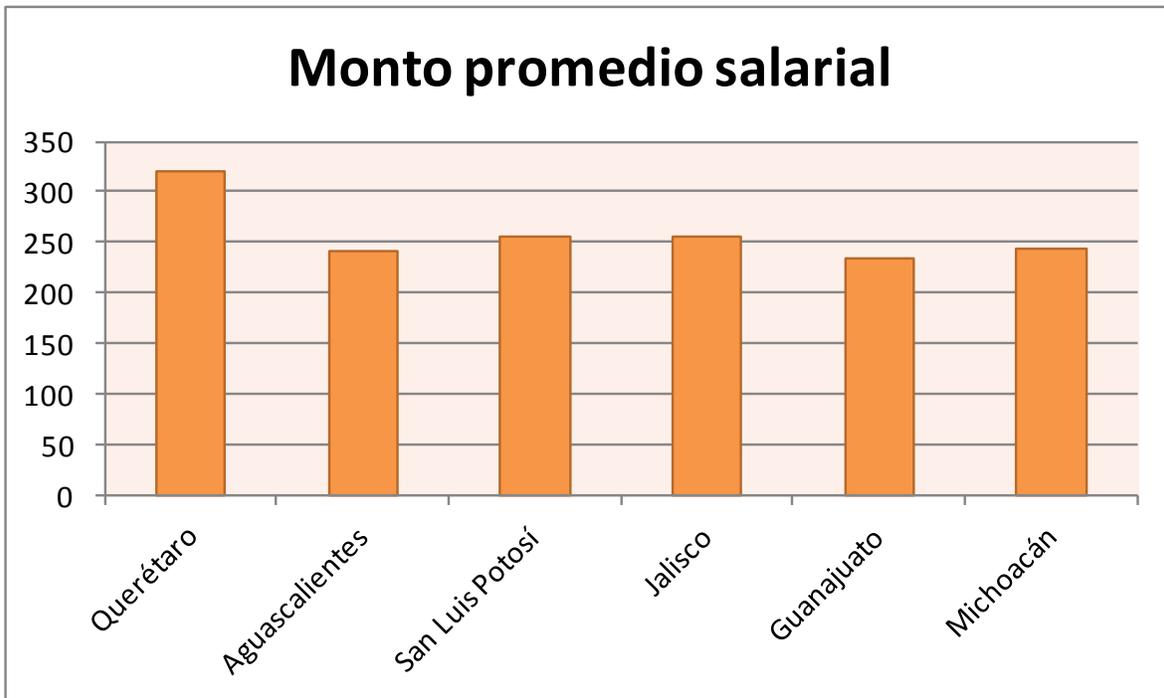
**Fuente:** Elaboración propia con datos del Instituto Mexicano de Seguridad Social en línea [www.imss.gob.mx](http://www.imss.gob.mx).

Aguascalientes y Michoacán mantienen un margen de diferencia cercano con 242 y 246, respectivamente. Quien se mantuvo en el último lugar de la zona fue Guanajuato, registró 234 pesos, muy por debajo del promedio de la zona (Ver Figura 2.6).

## CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.

La media para estas seis entidades fue de 260.41 pesos diarios, respecto al total del país es un 8% menos ya que este fue de 282.09 pesos. La entidad que registro monto mayor fue el Distrito Federal, al establecerse en 373.73 pesos diarios.

**Figura 2. 6.** Monto promedio salarial en estados del bajío.



**Fuente:** Elaboración propia realizada con información del IMSS en línea [www.imss.gob.mx](http://www.imss.gob.mx).

### 2.7 ZONAS ECONÓMICAS Y CRECIMIENTO ECONÓMICO.

El INEGI indicó que en el cuarto trimestre del 2014 el Indicador Trimestral de la Actividad Económica Estatal (ITAEE) sin considerar el petróleo, presentó un crecimiento de 3.35 por ciento respecto al mismo periodo del año pasado, nivel que fue superado por la mayoría de las entidades del bajío, con excepción de Michoacán que tuvo un crecimiento del 0.69 por ciento.

## **CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.**

---

De esta manera se observa que la economía de la región con mayor dinamismo en ese cuarto trimestre fue la queretana debido al fuerte crecimiento productivo del sector primario, el cual registró una tasa de 31.17 por ciento. Le siguieron el sector secundario con un aumento de 25.22 por ciento, mientras que el sector de servicios avanzó 6.19 por ciento.

En Aguascalientes también se observó un crecimiento importante de 12.47 por ciento, esto fue impulsado por el sector secundario con un crecimiento de 22.43 por ciento seguido del agropecuario con 5.77 por ciento y de servicios con 4.94 por ciento.

Al igual que Aguascalientes, Guanajuato fue impulsado por las actividades secundarias, resaltó el aumento de 15.25 por ciento en tanto que el sector servicios avanzó 3.95 por ciento y el primario tuvo una caída importante de 4.83 por ciento anual en el cuarto trimestre del 2014.

Para el caso de Jalisco y San Luis Potosí, la variación total fue de 4.39 y 3.79 por ciento respectivamente, esto es unos puntos por encima del promedio nacional. Para el caso de Jalisco el incremento más importante se presentó en el sector industrial, para San Luis Potosí se consideró el indicador sin el impacto que tiene el petróleo, de esto se tiene un incremento porcentual de 17.82 unidades en el sector primario y una caída de 26 puntos porcentuales.

Como se mencionó previamente, el estado de Michoacán tuvo el menor crecimiento de la región, esto derivado de las caídas en el sector industrial y de servicios, pero aún con esto la recuperación se observó en el sector agropecuario que creció un 13.64 por ciento y se mantuvo en el tercer lugar de este rubro de los seis estados analizados.

## CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.

A continuación se presenta un cuadro donde se muestran los crecimientos por estado y por sector para el último trimestre del 2014 con relación al último trimestre del año anterior.

**Tabla 2.2.** Crecimiento anual trimestral por sector.

Crecimiento 4to trimestre por sector 2013-2014				
	Total	Primarias	Secundarias	Terciarias
Aguascalientes	12.47%	5.77%	22.43%	4.94%
Guanajuato	7.32%	-4.83%	15.25%	3.95%
Jalisco	4.39%	1.00%	5.99%	2.89%
Michoacán	0.69%	13.64%	-3.40%	-0.11%
Querétaro	14.54%	31.17%	25.22%	6.19%
San Luis Potosí Sin petróleo	3.79%	17.82%	8.01%	-0.26%
Total País	3.35%			

**Fuente:** Elaboración propia con base de datos del INEGI en línea [www.inegi.org.mx](http://www.inegi.org.mx).

### 2.8 LA REGIÓN DEL BAJÍO EN LOS MEDIOS DE COMUNICACIÓN

Debido a su excelente ubicación geográfica, se ha desarrollado a lo largo de este territorio un importante corredor industrial que hoy alberga corporativos nacionales, poderosas multinacionales y su industria auxiliar.

Además, Guanajuato quien ha canalizado la mayor parte de las inversiones, se convirtió en el bastión económico más importante de la región. De acuerdo con información de la Secretaría de Desarrollo Económico Sustentable del estado de Guanajuato en su estudio titulado como "GTO el clúster más dinámico automotriz en México", su economía es la sexta con relación al PIB a nivel nacional y cuenta con una población de 5.7 millones de habitantes ubicándose en el 6to lugar del país, así como con una fuerza laboral de 2.5 millones de personas con un promedio de edad de 24 años.

Por la localización estratégica de Guanajuato, permite el acceso a 80 por ciento del mercado, 70 por ciento de comercio internacional, 70 por ciento de las exportaciones y 60 por ciento de la población del país, se localiza a un día

## **CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.**

---

manejando hacia el norte del país y a un día para el Golfo de México o el Océano Pacífico.

De la mano de la industria automotriz y aeroespacial, Guanajuato, Querétaro, Aguascalientes y San Luis Potosí desplazaron a las grandes ciudades del norte de la República mexicana en construcción y ampliación de parques industriales.

De acuerdo con la firma inmobiliaria Jones LangLaSalle citado de Méndez y Torres (2013), 80 por ciento de la superficie construida el año pasado se encuentra en los estados que conforman la región del Bajío, impulsada por la llegada de Bombardier, Eurocopter, Pirelli, Volkswagen, Honda, Mazda, Nissan y General Motors.

Ruíz y Méndez (2015) citando al Sistema Mexicano de Promoción de Parques Industriales (Simppi) mencionan que en el 2010 Guanajuato lideraba la región Bajío en número de parques industriales privados y públicos con un total de 23, seguido de Querétaro con 13 y de San Luis Potosí con diez.

Por su parte Baja California se coloca en primer lugar con 92, Nuevo León en segundo con 64, siguen Chihuahua con 39, Coahuila con 38, mientras que el Estado de México se coloca en quinto sitio con 34, Sonora en sexto con 33 complejos, en el séptimo lugar quedarían empatados Guanajuato y Tamaulipas con 23 parques cada uno; en el octavo se encuentra Puebla con 19, en el noveno Querétaro con 13 y en el décimo Sinaloa con 12 complejos (Ruíz y Méndez, 2015).

Mientras tanto la empresa Lintel prevé desarrollar seis polígonos industriales más en el estado de Guanajuato, con una inversión 50 millones de dólares y se prevé que genere de 15 mil a 25 mil empleos directos (Oropeza, 2015).

## **CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.**

---

Las inversiones más importantes que se han realizado recientemente en Celaya son las de la empresa automotriz Honda, que de acuerdo con Marvella (2015) “desde octubre de 2012 alrededor de dos mil empresas de todos los tamaños, 22 de ellas proveedoras del sector automotor de nivel uno, atraídas directamente por la armadora japonesa.”

La página web, monitor económico menciona que la zona del bajío ha desplazado a la zona norte del país como motor económico de la nación, “Aguascalientes, Guanajuato, Querétaro, San Luis Potosí y Jalisco suman más de cinco mil 400 millones de dólares en inversiones, el 60 por ciento de la cifra en los últimos cinco años dejando con mil 654 millones de dólares a las entidades de la frontera norte y donde Baja California, que fue el líder nacional ahora solamente suma poco más de 408 millones de dólares” (Vela, 2015).

Para Almanza (2015) en referencia a la industria automotriz francesa la Cámara Franco-Mexicana de Comercio e Industria capítulo Bajío estimó en el año 2015 la llegada de siete firmas francesas, dos a Querétaro y cinco más a otros estados de la región, indicó el presidente del organismo empresarial, Frederick Bron. Mencionó que los proyectos en Querétaro corresponden a empresas del sector automotriz, principalmente proveedoras de primero y segundo nivel.

El periódico uno más uno en su página web resume la relación económica entre México y Canadá y las condiciones favorables que se prevén; Las exportaciones mexicanas a Canadá en este 2015 serían del orden de 26 mil 78 millones de dólares, mientras que las ventas canadienses a México alcanzarían los nueve mil 849 millones de dólares. Lo que más exporta México a Canadá está en el sector automotriz, maquinaria eléctrica y productos electrónicos, minería, mientras que las exportaciones del sector aeroespacial se están integrando a las cadenas de producción.

## **CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.**

---

### **2.9 PEQUEÑAS Y MEDIANAS EMPRESAS DEL BAJÍO**

En el año del 2012, Manuel Molano Ruiz, director general del Instituto Mexicano para la Competitividad, indicó que las pymes tenían una supervivencia de entre 12 y 24 meses, por lo que era difícil que la banca pudiera otorgar préstamos a estas por considerarlas de alto riesgo (Becerril, 2012).

El INEGI en el boletín de prensa número 087/15, se publicó la investigación titulada como “Esperanza de vida de los negocios en México”<sup>4</sup>. El estudio se centró en dar seguimiento a las empresas desde el año del 2010 al 2015 basado en el Directorio Estadístico Nacional de Unidades Encuestadas (DENU), así como de los censos económicos, desde 1989 hasta 2009, este seguimiento se enfocó en verificar su continuidad en el mercado. La publicación identifica dos criterios para determinar la probabilidad de supervivencia de acuerdo a Edad que tiene al año del estudio la empresa y número de empleados y el otro que tiene que ver con los sectores económicos o actividad económica dividida en; Manufacturero, Comercial y Servicios privados no financieros.

Los resultados obtenidos se muestran en la tabla 2.3, en promedio, la esperanza de vida es de 7.7 años para cualquier empresa de cualquier sector, quien tiene una mayor probabilidad de supervivencia es el de manufactura seguido por los servicios financieros y al último comercio, por ende se tiene mayor esperanza de vida; para manufactura la esperanza es de 9.5 mientras que servicios privados tienen ocho años y comercio 6.6.

---

<sup>4</sup>Los resultados completos del estudio podrán ser consultados en la página de internet del Instituto: [www.inegi.org.mx](http://www.inegi.org.mx) en la sección de Investigación, y dentro de ésta en el apartado de Estadísticas Experimentales.

## CAPÍTULO 2. ANÁLISIS ECONÓMICO, DEMOGRÁFICO Y PERSPECTIVAS DE CRECIMIENTO DE LA REGIÓN DEL BAJÍO.

**Tabla 2. 3.** Probabilidad de supervivencia y esperanza de vida por sector de actividad en el primer año de vida de los negocios.

Sector	Probabilidad de supervivencia durante el primer año de vida	Esperanza de vida al nacer (años de vida futura)
Manufacturas	0.68	9.5
Comercio	0.62	6.6
Servicios privados no financieros	0.64	8.0
Los tres sectores	0.64	7.7

**Fuente:** INEGI, Esperanza de vida de los negocios en México 2014.

La probabilidad de muerte y esperanza de vida de acuerdo al número de empleados se muestra en la siguiente tabla 2.4:

**Tabla 2. 4.** Probabilidad de muerte y esperanza de vida en el primer año para negocios de 0-100 personas ocupadas.

Tamaño de los negocios (personas ocupadas)	Probabilidad de muerte al primer año de vida	Esperanza de vida al nacer (años futuros de vida)
0-2	0.38	6.9
3-5	0.32	8.4
6-10	0.17	15.0
11-15	0.18	20.7
16-20	0.15	23.3
21-30	0.12	22.0
31-50	0.09	22.0
51-100	0.07	22.0

**Fuente:** INEGI, Esperanza de vida de los negocios en México 2014.

Nótese que la probabilidad de desaparecer es mayor conforme los negocios son más pequeños, por ejemplo los de cero a dos personas ocupadas; cuatro de cada diez mueren durante el primer año de vida y su esperanza de vida al nacer es de casi 7 años. Esta probabilidad decrece conforme los negocios son más grandes y la esperanza de vida tiene un comportamiento creciente conforme aumenta el tamaño (Esperanza de vida de los negocios en México 2014, 2014)

Por último se comenta que lo que se pretendió en los párrafos de este capítulo fue dar un panorama general económico y demográfico de esta región para conocer el medio en el que se desarrollan las empresas que se analizan en esta investigación.

# CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---



### **3.1 INTRODUCCIÓN**

El enfoque principal de este capítulo es conocer los fundamentos teóricos de la técnica de supervivencia de regresión de Cox, empezando con el estudio de la función de riesgo y sus propiedades, además se describirán brevemente los diferentes tipos de modelos de supervivencia no paramétricos pero sin profundizar detalladamente en estos.

### **3.2 ANÁLISIS DE SUPERVIVENCIA**

Fox (2002) indica que el análisis de supervivencia, examina y modela un evento ocurrido en el tiempo. El evento común examinado es la muerte, por el cual el nombre de análisis de supervivencia. Sin embargo, el tiempo de supervivencia puede ser el tiempo transcurrido desde el inicio de un tratamiento hasta la desaparición de un trastorno o el fallo de un componente en una máquina. Como pudimos notar en el capítulo 1, existen gran diversidad de campos y aplicaciones donde se han realizado estas técnicas, en áreas como: medicina, ingeniería, sociología, psicología, entre otras.

Si bien es cierto que se ha utilizado en diversos campos, también se restringe para datos con ciertas características, como aquellas variables aleatorias con valores positivos. Estas variables generalmente representan el intervalo de tiempo transcurrido desde el inicio del seguimiento hasta el momento en el que se produce un determinado desenlace; este desenlace es el suceso objeto de estudio que se caracteriza por un cambio cualitativo brusco. Otra característica fundamental del análisis de supervivencia es que permite trabajar con los denominados datos incompletos, es decir, datos que contienen información parcial sobre el suceso estudiado (Palmer, 1988).

Los métodos utilizados para el análisis de supervivencia son esencialmente tres, estos son: Kaplan-Meier, el método actuarial y el modelo de regresión de Cox. El método Kaplan Meyer y el método actuarial son aproximaciones no

### **CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.**

---

paramétricas que se aplican cuando el tamaño de la muestra es grande porque agrupa los tiempos en intervalos (Domenech, 1992)

El método de estimación Kaplan Meyer fue desarrollado en 1958, denominado estimación de producto límite; calcula la probabilidad de supervivencia en el tiempo  $t_i$  aplicando la ley multiplicativa de las probabilidades, es decir, mediante el producto de la probabilidad de supervivencia en el tiempo  $t_{i-1}$  por la probabilidad condicionada de sobrevivir en el tiempo  $t_i$ .

$$S(t_i) = S(t_{i-1}) * S(t_i | t_{i-1}).$$

3. 1

Según Nogales (2009), este método no permite plantear un modelo de regresión para la supervivencia en función de variables explicativas, ya que con el método de Kaplan-Meier se asume que los grupos son homogéneos con respecto a todas las variables, pero esto no es siempre así.

Para el caso del método actuarial, que también se conoce como el método de la tabla de vida (*life-tablemethod*) tiene sus principios en los trabajos de Bohmer (1912), ha sido desarrollado principalmente con las aportaciones de Berkson y Gage (1950), Cutler y Ederer (1958) y Gehan (1969). Consiste en agrupar el tiempo de participación de cada sujeto en intervalos de tiempo predeterminados y, como el método de Kaplan-Meier, calcula las probabilidades de supervivencia mediante el producto de las estimaciones de las probabilidades condicionadas de supervivencia correspondientes a cada uno de los diferentes intervalos de tiempo.

“La agrupación en intervalos tiene el inconveniente de dar estimaciones poco precisas cuando el tamaño de la muestra no es suficientemente grande. Sin embargo para muestras grandes es más recomendable porque influye poco sobre las estimaciones y tiene la

## **CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.**

---

ventaja de presentar una curva de supervivencia muy clara y una tabla de fácil interpretación que sintetiza muy bien la evolución de la supervivencia de los sujetos estudiados” (Domenech, 1992).

Ahora bien, como se ha mencionado en el capítulo 1, el método de regresión de Cox fue introducido en 1972 por Cox, este modelo de supervivencia que examina la relación entre la supervivencia y una o más variables predictivas, usualmente denotadas como covariantes. El modelo puede comparar dos o más distribuciones de supervivencia de manera que se podrá evaluar estadísticamente su igualdad o diferencias y se considera semi-paramétrico por tener una parte no paramétrica y otra paramétrica.

### **3.3 MODELO DE REGRESIÓN COX.**

Cuando se requiere describir la relación entre los valores de dos o más variables se utilizan las técnicas estadísticas de regresión, es decir, el análisis de la relación entre la variable dependiente y la variable explicativa (independiente). Cuando más de una variable independiente necesita ser considerada, se llamara regresión múltiple.

El modelo de Cox es un método similar al de regresión múltiple, en este caso la variable dependiente es la función de riesgo dada en el tiempo en la cual Cox (1972) realizo un modelo de regresión que permite la evaluación no paramétrica de la relación entre este tipo de funciones.

### 3.4 FUNCION DE RIESGO.

La función de riesgo o también conocida como la tasa Hazard  $\lambda(t)$  representa la probabilidad condicionada por unidad de tiempo que tiene un sujeto de fallecer en un instante "t" si sobrevivió hasta el instante anterior (Domenech, 1992).

Sea X una variable aleatoria que denota la edad de fallo o muerte de un dispositivo o una persona respectivamente,  $f(x)$  denota su función de densidad y  $F(x)$  su función de distribución, de tal manera que  $F(x)=P(X \leq x)$  significa la probabilidad de que el dispositivo falle antes de la edad x.

Ahora con una nueva variable aleatoria T, representara el tiempo de vida que le resta al dispositivo, por tanto;

$$T=X-x$$

donde X es nuestra variable aleatoria anterior y x será la edad actual del elemento.

Ahora bien, la variable aleatoria T es una función que depende de la variable aleatoria X,  $T(X)$ . Para  $P [T(x) \leq t]=F_{T(x)}(t)$  se entiende como la probabilidad de que el tiempo que le resta de vida a una persona de edad X sea menor o igual a "t", por consiguiente la probabilidad resultante sigue como;

$$P[x < X \leq x + t | X > x] = \frac{F(x+t) - F(x)}{1 - F(x)} = 1 - \frac{S(x+t)}{S(x)} \quad 3.2$$

En ciencias actuariales, la función de probabilidad acumulada descrita anteriormente, se denota como  ${}_tq_x$  y la función de supervivencia que describe la

## CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---

probabilidad de que una persona de edad  $X$  muera después de “ $t$ ” años o sobreviva “ $t$ ” años es  ${}_t p_x$ , es decir, el complemento de  ${}_t q_x$ .

Para determinar la probabilidad que una persona muera en el instante siguiente, dado que ha sobrevivido hasta ese tiempo se plantea de la siguiente manera

$$\Pr[x < X \leq x + \Delta x | X > x] = \frac{F(x + \Delta x) - F(x)}{1 - F(x)} \cong \frac{f(x) * \Delta x}{S(x)} = \Delta x * \lambda_x \quad 3.3$$

La diferencia entre las ecuaciones 3.2 y 3.3, reside en la magnitud del tiempo, la función  $\lim_{\Delta x \rightarrow 0} \lambda_x = \frac{f(x)}{S(x)}$  es mejor conocida como fuerza de mortalidad ( $\lambda_x$ ); determina la probabilidad de muerte o fallo instantánea de un elemento de edad  $x$ . También llamada tasa de riesgo. De 3.3 se sigue que:

$$\lambda(t) = \frac{f(t)}{S(t)}, \quad t \geq 0 \quad 3.4$$

Si la variable  $T$  mide el tiempo de fallo,  $\lambda(t)$  representa la proporción esperada de individuos que sobreviviendo al tiempo  $t$  y mueren en el intervalo  $(t, t+\Delta t)$ .

### Algunas Propiedades (Bowers, 1997).

#### 1. Relación entre las distintas funciones y la función de riesgo.

	$f(t)=$	$F(t)=$	$S(t)=$	
$\lambda(t)$	$\lambda(t) * e^{-\int_0^t \lambda(t) dt}$	$1 - e^{-\int_0^t \lambda(t) dt}$	$e^{-\int_0^t \lambda(t) dt}$	3.5

#### 2. Función de riesgo acumulado

$$H(t) = \int_0^t \lambda(t) dt$$

3.  $\int_0^{\infty} \lambda(t) dt = \infty$

4. **Función de riesgo constante.** Una función constante significa que el riesgo de fallo no varía con el tiempo.
5. **Función de riesgo decreciente.** Una función decreciente muestra una disminución del riesgo de muerte a lo largo del tiempo.
6. **Función de riesgo creciente.** Muestra un aumento en el riesgo de muerte conforme avanza el tiempo.

### 3.5 MODELO DE RIESGO PROPORCIONAL

La regresión Cox consiste en obtener una función lineal de las variables independientes que permitan estimar, en función del tiempo, la probabilidad de que ocurra un suceso, además estima el valor de la supervivencia de un elemento “i” en un tiempo específico y modela la función de riesgo de ese elemento como producto de dos factores.

Sea un vector de p variables explicativas,  $\mathbf{X}=(X_1, X_2, \dots, X_p)$  que afectan a la respuesta (Ecuación 3.6).

$$\lambda_i(t) = \lambda_0(t)Z_i(\mathbf{X}, \beta) \tag{3.6}$$

Donde  $\beta$  denota los parámetros de la distribución, los cuales son necesarios estimar.

Suponga que se tienen distribuciones continuas para los tiempos de vida y son registrados con exactitud. Las tres parametrizaciones de  $Z_i$  pueden ser consideradas de la siguiente forma (Cox y Oakes, 1984).

### CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---

1. Log lineal.

$$Z(X; \beta) = e^{\beta X} = e^{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p} \quad 3.7$$

2. Lineal.

$$Z(X; \beta) = 1 + \beta X = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad 3.8$$

3. Logística.

$$Z(X; \beta) = \log(1 + \exp(\beta X)) = \log(1 + \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)) \quad 3.9$$

La función más común es el modelo log-lineal donde:

$$\log(\lambda_i(t)) = \log(\lambda_0(t)) + X\beta \quad 3.10$$

$$\log(\lambda_i(t)) = \log(\lambda_0(t)) + (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad 3.11$$

Lo cual provoca que sea una regresión de tipo lineal.

$$\log(\lambda_i(t)) = \alpha(t) + (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad 3.12$$

Esta regresión, deja a la función base de riesgo  $\alpha$  sin especificar o de nuevo a su equivalente  $\lambda_0(t)$ .

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k) \quad 3.13$$

De las ecuaciones 3.12 y 3.13 se considera al modelo como semi-paramétrico, ya que mientras la función base puede tomar cualquier forma, las covariantes entran al modelo lineal. La constante de la regresión representa la función base de riesgo cuando todas las  $x$ 's son 0.

## CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---

Considérese ahora dos observaciones “i” y “j” que difieren en los x-valores, con su correspondiente línea predictiva.

$$\delta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \quad 3.14$$

$$\delta_j = \beta_1 x_{j1} + \beta_2 x_{j2} + \dots + \beta_k x_{jk}$$

Para este tipo de regresión, se considera como la razón de riesgo de dos observaciones independiente del tiempo, llamado supuesto de proporción de las funciones de riesgo invariante al tiempo, este se ejemplifica en 3.15.

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)e^{\delta_i}}{\lambda_0(t)e^{\delta_j}} = \frac{e^{\delta_i}}{e^{\delta_j}} \quad 3.15$$

En la comparación del tiempo de falla de dos grupos, es de interés el caso en el que el riesgo de falla, en cualquier tiempo dado para un individuo en un grupo, es proporcional al riesgo en ese tiempo para un individuo en el otro grupo, a esto se le llama riesgos proporcionales. Esta hipótesis es el fundamento de un gran número de métodos de análisis en el área de supervivencia, por este motivo, en varios casos éste no es un modelo apropiado.

### 3.6 DATOS CENSURADOS

Dada la importancia que tiene identificar las características de los datos con que se trabaja, se explica brevemente dos tipos de censura debido a que la información de supervivencia se puede presentarse en diferentes formas:

- a) La censura ocurre cuando el tiempo de falla de un elemento de la muestra puede no ser observado en el horizonte de estudio.
- b) Cuando el elemento sale del estudio por causas ajenas al evento de interés.

### CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

En este trabajo se utilizará un tipo particular de datos, aquellos que presentan censura por la derecha. Una dato esta censurado por la derecha cuando el evento de interés sucede después de un tiempo preestablecido, cuando el sujeto abandona el estudio o muere por causas ajenas.

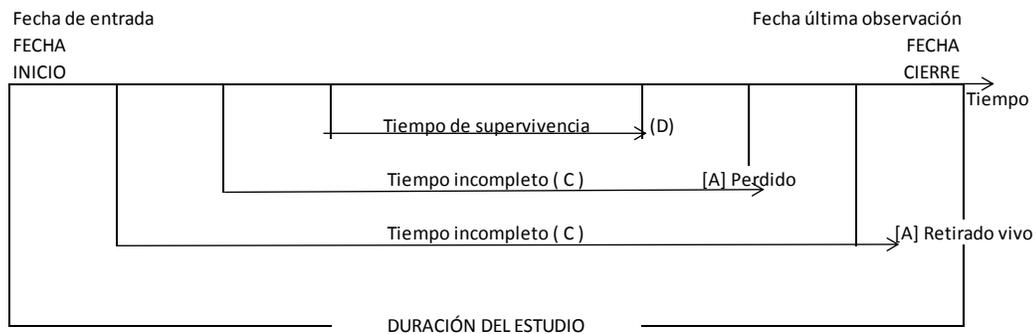
Se denota  $C_i$  el tiempo desde el comienzo hasta el final del estudio o hasta el abandono del estudio (tiempo de censura). Por tanto, el tiempo de seguimiento del individuo "i" es:

$$T_i = \min (T_i, D_i)$$

Un dato esta censurado por la izquierda cuando se sabe que el evento de interés sucedió antes del tiempo  $D_i$  pero el momento exacto es desconocido ya que sucedió antes del primer periodo de medida.

$$\text{Entonces } T_i = \max (T_i, D_i)$$

**Figura 3.1** Representación de los tiempos de participación de los sujetos



**Fuente:** Obtenido de Domenech, representación de los tiempos de participación de los sujetos (1992, pag 111).

### CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---

En la figura 3.1 se muestran gráficamente los tipos de datos donde el acontecimiento de interés está representado por la línea “D” para aquellos en los que se observó el evento de interés. Por lo general el punto del inicio del estudio o fecha de entrada no coincide con los tiempos de los sujetos cuando entran al estudio ya que muchas veces se van integrando progresivamente.

Los datos censurados están representados en las líneas A, los cuales son aquellos de los que se perdió el seguimiento o resultaron vivos al finalizar el estudio, estos también son conocidos como datos incompletos.

#### 3.7 ESTIMACION DE PARAMETROS

Sea una muestra de “n” observaciones,  $f(t_i | X_i, \beta)$  representa la probabilidad de supervivencia del individuo “i” al tiempo t dado los valores de las covariantes x y los parámetros de la regresión  $\beta$ , con X y  $\beta$  como vectores. Además se considera la función de supervivencia como  $S(t_i | X_i, \beta)$ .

Introduciremos el indicador para datos censurados  $C_i$ , el cual toma el valor de 0 si el dato o evento es censurado y 1 si la falla es observada.

De acuerdo con Lawless (2002), la función de máxima verosimilitud para datos censurado está dada por

$$L(\beta) = \prod_{i=1}^n [f(t_i | x_i, \beta)]^{c_i} \times [S(t_i | x_i, \beta)]^{1-c_i} \quad 3. 16$$

Atendiendo a una de las relaciones de la función de riesgo y función de densidad de (3.4) se tiene que;

### CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---

$$L(\beta) = \prod_{i=1}^n [\lambda(t_i | x_i, \beta) S(t_i | x_i, \beta)]^{c_i} [S(t_i | x_i, \beta)]^{1-c_i} = \prod_{i=1}^n [\lambda(t_i | x_i, \beta)]^{c_i} S(t_i | x_i, \beta) \quad 3.17$$

Podemos mostrar que

$$S(t_i | x_i, \beta) = S_0(t_i) e^{(x_i, \beta)} \quad 3.18$$

donde  $S_0$  es la función base de supervivencia. Sustituyendo 3.13 y 3.18 en 3.17 tenemos que

$$L(\beta) = \prod_{i=1}^n [\lambda_0(t_i) \exp(x_i, \beta)]^{c_i} S_0(t_i)^{\exp(x_i, \beta)} \quad 3.19$$

Aplicamos logaritmo a la función.

$$l(\beta) = \log(L(\beta)) = \sum_{i=1}^n c_i \log[\lambda_0(t) * \exp(x_i, \beta)] + \sum_{i=1}^n \log[S_0(t_i)^{\exp(x_i, \beta)}] \quad 3.20$$

Donde  $(x_i, \beta)$  es una combinación lineal de  $B_k$  y  $x_{ik}$ . Se sigue que

$$l(\beta) = \sum_{i=1}^n c_i (\log(\lambda_0(t)) + (x_i, \beta)) + \sum_{i=1}^n \log[S_0(t_i)] e^{(x_i, \beta)} \quad 3.21$$

De acuerdo al método de máxima verosimilitud, esta función tiene que maximizarse con respecto a los parámetros. Sin embargo, los parámetros del modelo de Cox no pueden ser estimados por el método de máxima verosimilitud al ser desconocida la forma específica de la función arbitraria de azar.

## **CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.**

La propuesta de Cox fue un método de estimación denominado verosimilitud parcial siendo las verosimilitudes condicionales y marginales casos particulares del anterior.

### **3.7.1 FUNCION DE MAXIMA VEROSIMILITUD PARCIAL**

El método de verosimilitud parcial se diferencia del método de verosimilitud ordinario en el sentido de que mientras el método ordinario se basa en el producto de las verosimilitudes para todos los individuos de la muestra, el método parcial se basa en el producto de las verosimilitudes de todos los sucesos ocurridos (Palmer, 1988).

Se propone utilizar la probabilidad condicional  $P [i=k | R (t_i)]$ , donde  $k$  representa el número de cambios hasta el tiempo  $t$  y  $R(t_i)$  representa el número de elementos en riesgo en el tiempo en que se dio el cambio  $k$  (Palmer, 1988), es decir, la probabilidad de que ocurran  $k$ -cambios dado que existen  $R(t)$  sujetos en riesgo en ese periodo.

Mostrando como ejemplo el siguiente ejercicio, suponiendo que el evento o el  $k_1$ -cambio sucedieron al elemento "a" para 3 elementos en riesgo "a", "b" y "c".

$$P [ i = k_1 | R( t_{k_1} ) ] = \frac{f_a(t_{k_1}) * S_b(t_{k_1}) * S_c(t_{k_1})}{f_a(t_{k_1}) * S_b(t_{k_1}) * S_c(t_{k_1}) + S_a(t_{k_1}) * f_b(t_{k_1}) * S_c(t_{k_1}) + S_a(t_{k_1}) * S_b(t_{k_1}) * f_c(t_{k_1})}$$

Aplicando (3.4) donde  $f_i(t_{k_1}) = \lambda_i(t_{k_1}) * S_i(t_{k_1})$

$$= \frac{\lambda_a(t_{k_1}) * S_a(t_{k_1}) * S_b(t_{k_1}) * S_c(t_{k_1})}{\lambda_a(t_{k_1}) * S_a(t_{k_1}) * S_b(t_{k_1}) * S_c(t_{k_1}) + S_a(t_{k_1}) * \lambda_b(t_{k_1}) * S_b(t_{k_1}) * S_c(t_{k_1}) + S_a(t_{k_1}) * S_b(t_{k_1}) * \lambda_c(t_{k_1}) * S_c(t_{k_1})}$$

Factorizando  $S_a(t_k)$ ,  $S_b(t_k)$  y  $S_c(t_k)$  de la ecuación da como resultado 3.22.

### CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---

$$= \frac{\lambda_a(t_{k_1})}{\lambda_a(t_{k_1}) + \lambda_b(t_{k_1}) + \lambda_c(t_{k_1})} \quad 3.22$$

De acuerdo con 3.6 y 3.7, entonces 3.22 es igual a

$$= \frac{e^{\beta x_a^t}}{e^{\beta x_a^t} + e^{\beta x_b^t} + e^{\beta x_c^t}} \quad 3.23$$

donde  $\beta$  es el vector de parámetros y “x” es el vector de observaciones en el tiempo t del elemento correspondiente.

De esta manera se deduce la formula general para todas las demás.

$$P [ i = k | R(t_i) ] = \frac{e^{\beta x_i^t}}{\sum_{j \in R(t_i)} e^{\beta x_j^t}} \quad 3.24$$

Esta función de probabilidad condicional es una alternativa a la función de densidad y para todas los casos de cambio de k se tiene que.

$$L(\beta) = \prod_k P [ i = k | R(t_i) ] = \prod_k \frac{e^{\beta x_i^t}}{\sum_{j \in R(t_i)} e^{\beta x_j^t}} \quad 3.25$$

Para efectos de esta investigación consideraremos dos casos, para datos completos sin empates y para casos incompletos. La primera se basa para datos en donde no existe censura en el evento de estudio de cada individuo y ha sucedido en tiempos diferentes. Se dice que no hay empates cuando el tiempo en que ocurre cada cambio en cada individuo es diferente para cada uno, es decir, si se tiene un estudio de observación continua, en donde su valor se registra de forma exacta.

## CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

Sean  $t_1 < t_2 < \dots < t_n$  los “n” tiempos observados correspondientes a los “n” sujetos de estudio, de tal manera que la función de máxima verosimilitud parcial va desde la primera observación hasta la última observación (n).

Para casos incompletos o censurados, se tiene una muestra de tamaño n donde se observan d fallas, por lo que se tienen n – d datos censurados.

### 3.8 ESTIMACIÓN DEL COEFICIENTE

La función de verosimilitud parcial (PL) viene dada por el producto de las probabilidades condicionales de 3.25.

Aplicando logaritmo se tiene.

$$l(\beta) = \log (L(\beta)) = \log \left( \prod_d \frac{e^{\beta x_i^t}}{\sum_{j \in R(t_i)} e^{\beta x_j^t}} \right) \quad 3.26$$

Por propiedades de logaritmos:

$$l(\beta) = \sum_d \log \left( \frac{e^{\beta x_i^t}}{\sum_{j \in R(t_i)} e^{\beta x_j^t}} \right) \quad 3.27$$

Obtenemos una diferencia:

$$l(\beta) = \sum_d \left[ \log (e^{\beta x_i^t}) - \log \left( \sum_{j \in R(t_i)} e^{\beta x_j^t} \right) \right] \quad 3.28$$

$$l(\beta) = \sum_d \left[ \beta x_i^t - \log \left( \sum_{j \in R(t_i)} e^{\beta x_j^t} \right) \right] \quad 3.29$$

Realizando la derivada respecto los coeficientes a estimar.

### CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---

$$U(\beta) = \frac{\partial}{\partial \beta} l(\beta) = \sum_{j=1}^d [x_i - \left( \frac{\sum_{j \in R(t_i)} e^{\beta x_j^t} x_j}{\sum_{j \in R(t_i)} e^{\beta x_j^t}} \right)] \quad 3.30$$

Hosmer y Lemeshow (1999) proponen

$$U(\beta) = \sum_{i=1}^m (x_i - X_{w_i}) \quad 3.31$$

Donde

$$X_{w_i} = \sum_{j \in R(t_i)} w_{ij}(\beta) x_j \quad 3.32$$

Y

$$w_{ij}(\beta) = \frac{e^{\beta x_j^t}}{\sum_{j \in R(t_i)} e^{\beta x_j^t}} \quad 3.33$$

Se sigue que la estimación de los valores de beta se obtendrá como solución del vector de la ecuación 3.31 igualada con 0.

$$U(\beta) = 0$$

$$\sum_{j=1}^d x_i = \sum_{j=1}^d \frac{e^{\beta x_j^t} x_j}{\sum_{j \in R(t_i)} e^{\beta x_j^t}} \quad 3.34$$

El valor de B que maximiza la ecuación, puede ser obtenido por algún método numérico, de acuerdo con Ross (2012) se puede realizar la iteración iniciando con  $x_0=0$  con la ecuación 3.34 y 3.35 que a continuación describiremos.

### CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---

$$I(\beta) = \frac{\partial^2}{\partial \beta \partial \beta} l(\beta) = \sum_{j=1}^d [\beta x_j^t - \left( \frac{e^{\beta x_j^t} x_j}{\sum_{j \in R(t_i)} e^{\beta x_j^t}} \right)] \times 2 \left( \frac{e^{\beta x_j^t}}{\sum_{j \in R(t_i)} e^{\beta x_j^t}} \right) \quad 3.35$$

Esta expresión (3.35) puede simplificarse como en 3.36 y su negativo es llamado información observada.

$$I(\beta) = \frac{\partial^2 l(\beta)}{\partial \beta_k^2} = - \sum_{i=1}^m \sum_{j \in R(t_i)} w_{ij}(\beta) (x_{jk} - X_{w_{jk}})^2 \quad 3.36$$

De acuerdo con Hosmer y Lemeshow (1990) los elementos fuera de las diagonales están definidos como:

$$\frac{\partial^2 l(\beta)}{\partial \beta_k \partial \beta_i} = - \sum_{i=1}^m \sum_{j \in R(t_i)} w_{ij}(\beta) (x_{jk} - X_{w_{jk}})(x_{jl} - X_{w_{il}}) \quad 3.37$$

El estimador de la varianza de los coeficientes estimados es la inversa de 3.36 evaluada en  $\hat{\beta}$

$$\widehat{Var}(\hat{\beta}) = I(\hat{\beta})^{-1} \quad 3.38$$

#### 3.9 ESTIMACIÓN DE LA FUNCION DE SUPERVIVENCIA

Para estimar la función de supervivencia, Hosmer y Lemeshow (1999) sugieren que la idea es imitar el argumento que utiliza Kaplan-Meyer utilizando la cantidad  $\alpha_i = 1 - (d_i / n_{i-1})$  como estimador de la función condicional al tiempo  $t_i$ . Lo que se puede ver como el número de vivos en el periodo "i" entre el número de muertos en el periodo "i-1". Esto es:

### CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.

---

$$\frac{S(t_i, x, \beta)}{S(t_{i-1}, x, \beta)} = \left\{ \frac{[S_0(t_i)]^{\exp(x'\beta)}}{[S_0(t_{i-1})]^{\exp(x'\beta)}} \right\} = \left\{ \frac{S_0(t_i)}{S_0(t_{i-1})} \right\}^{\exp(x'\beta)} = \alpha_i^{\exp(x'\beta)} \quad 3.39$$

\*Hosmer and Lemeshow (1999), pag 109. Applied survival analysis Regression Modeling of time to event data.

La estimación de la función condicional base de supervivencia es obtenida resolviendo la ecuación siguiente bajo algún método numérico.

$$\sum_{l \in D_i} \frac{\hat{\theta}_l}{1 - \alpha_i^{\hat{\theta}_l}} = \sum_{l \in R_i} \hat{\theta}_l \quad 3.40$$

$$\sum_{l \in D_i} \frac{e^{x'\beta}}{1 - \alpha_i^{e^{x'\beta}}} = \sum_{l \in R_i} e^{x'\beta}$$

donde

$$\hat{\theta}_l = e^{x'\beta}$$

$R_i$  denota el conjunto de sujetos en riesgo con tiempo de supervivencia  $t_i$ .

En caso de que no exista tiempos iguales de supervivencia,  $D_i$  contiene un sujeto y la solución es

$$\hat{\alpha}_i = \left[ 1 - \frac{\hat{\theta}_i}{\sum_{l \in R_i} \hat{\theta}_l} \right]^{\hat{\theta}_i^{-1}} = \left[ 1 - \frac{e^{x'\beta}}{\sum_{l \in R_i} e^{x'\beta}} \right]^{1/e^{x'\beta}} \quad 3.41$$

De esta manera, se obtiene la función base de supervivencia de la siguiente manera.

$$\hat{S}_0(t) = \prod_{t_1 \leq t} \hat{\alpha}_i \quad 3.42$$

### 3.10 VALIDACIÓN DEL MODELO

#### 3.10.1 PRUEBA DE WALD

Esta prueba está basada en la distribución asintóticamente normal de los estimadores de máxima verosimilitud. Esto es que para muestras grandes,  $B$  es un vector aleatorio de distribución normal con media  $\bar{b}$  y matriz de varianzas y covarianzas estimado por  $I^{-1}(b)$

Esta prueba tiene la hipótesis nula  $H_0: B = B_0$  para la cual la estadística de prueba es:

$$X_W^2 = (b - \beta_0)' I(b) (b - \beta_0) \quad 3.43$$

La cual tiene distribución Ji-cuadrada con  $p$  grados de libertad si  $H_0$  es cierta para muestras grandes.

#### 3.10.2 PRUEBA DE RAZÓN DE VEROSIMILITUDES

Esta prueba tiene la hipótesis nula  $H_0: B = B_0$  para la cual la estadística de prueba es:

$$X_{LR}^2 = 2[\ln(L(b)) - \ln(L(\beta_0))] \quad 3.44$$

## **CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.**

---

La cual tiene distribución Ji-cuadrada con  $p$  grados de libertad si  $H_0$  es cierta para muestras grandes.

### **3.10.3 PRUEBA DE PUNTAJES**

Esta prueba está basada en derivada parcial de la función de log verosimilitud parcial con respecto a  $\beta$ . Para muestras grandes, se tiene distribución asintótica normal con el vector cero por media y matriz de varianzas y covarianzas dada por  $I(\beta)$  cuando  $H_0$  es cierta. Aquí la hipótesis nula es  $H_0: \beta = \beta_0$ , con  $H_0$  como distribución Ji-cuadrada con  $p$  grados de libertad.

### **3.11 INTERPRETACIÓN DEL MODELO.**

La razón de riesgo es una medida de la supervivencia que puede verse a lo largo del horizonte del tiempo  $t$ , esta es obtenida mediante la división de dos funciones con características diferentes  $\exp(\beta \cdot x_1) / \exp(\beta \cdot x_2)$ . Donde  $\beta$  es el valor del coeficiente estimado y  $x_1, x_2$  son los valores de la variable  $X$  para dos sujetos.

En caso hipotético suponga que se estudia un cohorte de personas que es seguida durante ciertos años, en donde se estudia la muerte de los individuos. Bajo este supuesto se tiene el interés de observar el evento de muerte al final de la investigación.

Suponga que  $X$  es una variable binaria que denota el sexo de la persona (1=masculino; 2=femenino). Sea entonces la razón de  $x$  igual a 2, ésta razón de riesgo significa que a cualquier tiempo del estudio la tasa de muerte por unidad de tiempo de los hombres es el doble que las mujeres, es decir, la probabilidad de ocurrencia del evento de muerte se presenta dos veces más en los hombres que en las mujeres.

### **CAPÍTULO 3. INTRODUCCIÓN A LA TÉCNICA DE REGRESIÓN DE COX.**

---

Los coeficientes de regresión Cox indican la relación existente entre la variable explicativa correspondiente y la función de riesgo. Comúnmente un valor positivo del coeficiente supone un aumento en la función de riesgo para el sujeto, lo cual conlleva una relación negativa con el tiempo de cambio. Es decir, un coeficiente positivo indica un mayor riesgo de que exista un cambio.

Con esto se concluye la descripción teórica de la técnica que se utiliza en el capítulo cuarto de esta investigación. Esta técnica se aplica a un caso empírico que corresponde a las empresas del bajío que es la esencia de este trabajo.

**CAPÍTULO 4. CONSTRUCCIÓN  
DEL MODELO DE COX PARA LAS  
PYMES DEL BAJIO.**

---

## **CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO**

---

## **CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO**

---

### **4.1 INTRODUCCIÓN**

El objetivo de este capítulo es construir el modelo que pueda dar solución o respuesta a nuestras hipótesis planteadas principalmente para cumplir con los objetivos planteados. El modelo servirá para encontrar aquellas variables que influyen en la trascendencia y desarrollo de una empresa.

El apartado se divide en dos secciones, primeramente se introduce con una descripción de las variables consideradas para la construcción del modelo y se muestran las gráficas de supervivencia de Kaplan-Meier por estrato. La siguiente parte corresponde a la construcción del modelo mediante R-Project versión 3.0.1. En esta sección se estiman los parámetros, se determinan las variables estadísticamente significativas y se interpretan los resultados. Por último se muestra el modelo final junto con algunos resultados de interés y se realiza una comparación entre el modelo Kaplan-Meier y el modelo propuesto en esta tesis.

### **4.2 DESCRIPCIÓN DE LA BASE DE DATOS.**

Los datos considerados para este estudio son resultado de una recolección realizada para empresas medianas y pequeñas. La encuesta se efectuó a empresas del bajío obteniendo una muestra de 1034 de tipo PYME dentro del periodo de 2011 al 2012.

Las variables seleccionadas para la construcción del modelo son catalogadas como sigue:

1. Variable de tiempo (t), medida en años que la empresa tiene en operaciones.
2. Capital mayoritario.
3. Cuenta con certificación de calidad ISO 9000 o equivalente.
4. Edad del director ó gerente general.

## **CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO**

---

5. Género del director ó gerente general.
6. Gerente fundador.
7. Nivel de formación académica del gerente.
8. Número de empleados.
9. Sector en la que se desempeña la empresa.
10. Estatus de censura; 1 si el entrevistado contesto no es rentable en el momento en que se realizó la encuesta.

El modelo de análisis de supervivencia estima el tiempo que tardan las empresas para tener una caída en su rentabilidad o bien la probabilidad de tener una baja en su rentabilidad.

A continuación se describe cada una de las variables, el tipo de dato y la relación que tienen. Para tener una idea de la distribución de la base de datos, se presenta mediante estadística descriptiva y graficas de supervivencia de tipo Kaplan-Meier que nos proporcione un diagnostico objetivo de su comportamiento.

### **4.2.1 CLASIFICACIÓN POR CAPITAL MAYORITARIO**

Esta variable considera dos tipos de empresas, aquellas empresas en donde la familia tiene más del 50 por ciento del capital o de las acciones de la empresa y aquellas otras en las que una familia no es dueña mayoritaria, teniendo como resultado una variable de tipo dummy.

De acuerdo con la Fundación de Familias de Negocios (*Business Families Foundation*; BFF) en el año 2013, México ocupó el quinto puesto a nivel mundial con más empresas familiares. Éste estudio reveló que el 70% de este tipo de negocios no llegan al tercer año de vida y sólo el 12% de las empresas de tipo familiar permanece hasta la tercera generación.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

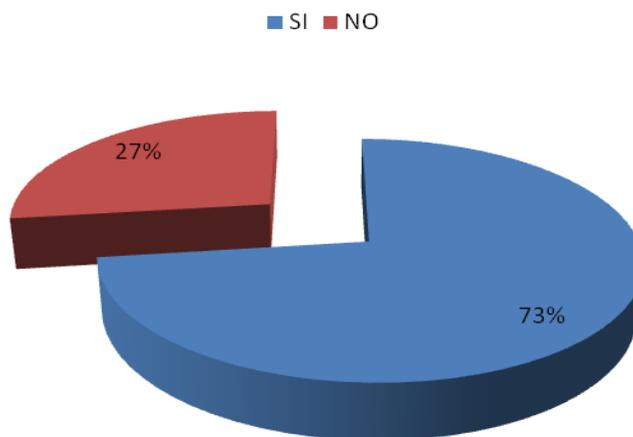
---

Por otra parte, la encuesta de empresas familiares en México realizada por Banamex y Profit Consulting, revelo que el 78 por ciento de este tipo de empresas no tiene una política clara sobre la contratación, despido y remuneración de socios de la familia. El 31 por ciento de las empresas contratan a los miembros de la familia por su parentesco y no por su capacidad laboral. El 33 por ciento de las empresas familiares llega a la segunda generación y sólo el 13 por ciento sobrevive a la tercera (Leyva, 2014).

Las conclusiones del estudio destacan la necesidad de encontrar soluciones para la continuidad de este tipo de empresas, por este motivo se dedujo la importancia del análisis de esta variable ya que podría repercutir en la rentabilidad.

Para esta investigación, la mayoría de las empresas tienen un capital mayoritario familiar, es decir, la familia es propietaria de al menos el 50% del capital de la empresa como se aprecia gráficamente en figura 4.1. En total se registraron 754 en las que predomina el capital familiar contra 280, lo que representa el 73% y 27% respectivamente.

**Figura 4. 1.** Distribución de la frecuencia relativa por tipo de clasificación de capital.

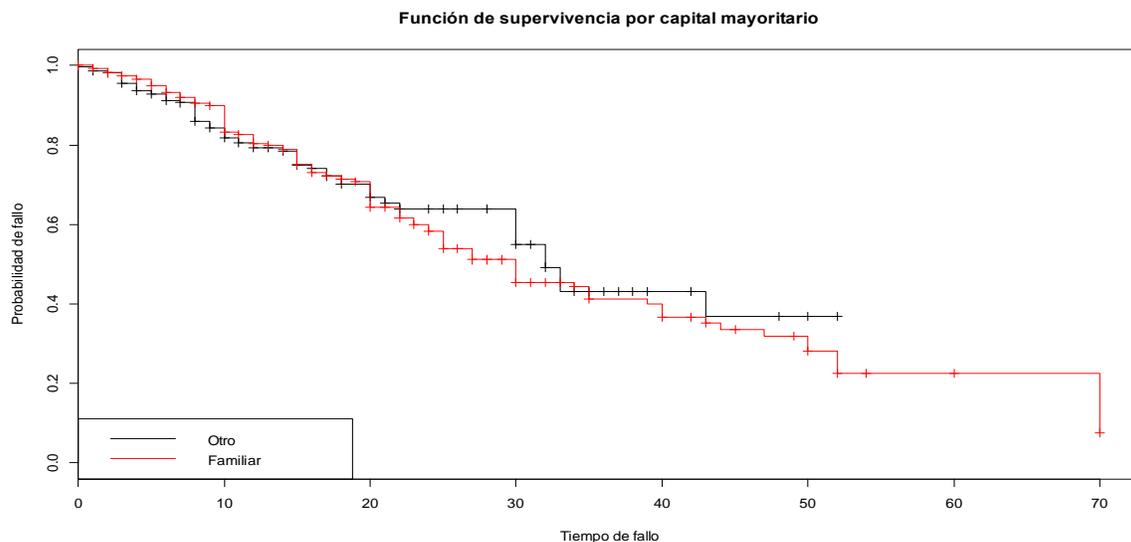


**Fuente:** Elaboración propia con datos obtenidos de la encuesta 2011-2012.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

En la figura 4.2 se muestra la gráfica comparativa de la curva de supervivencia de Kaplan Meier por estrato de la variable capital mayoritario. Se aprecia que después de los 23 años aproximadamente las curvas empiezan a separarse y se reencuentran hasta los 30 años aproximadamente.

**Figura 4. 2.** Grafica de supervivencia de Kaplan-Meier de clasificación por capital.



**Fuente:** Elaboración propia con R-Project.

En la figura 4.3 se muestra el resultado de corrida en R Project para los datos por capital mayoritario denotado como Fam para la lectura en el software, la parte enmarcada en rojo es el código para llamar a la prueba de puntajes o de Log-Rank y por debajo una tabla donde muestra la distribución por resultado y cálculos para el estadístico, Fam igual con cero es que la familia no es socio mayoritario de la empresa y uno en caso contrario.

El dato utilizado para verificar la igualdad o diferencia de curvas es el enmarcado en el óvalo rojo, la prueba Log-Rank que establece la igualdad de las curvas de supervivencia con un p-value igual a 0.732. Con este valor se demuestra que para este estudio no existe ninguna limitante para ningún tipo de empresa con esta variable, el hecho de que la familia intervenga o no en las decisiones de la empresa no repercute en la rentabilidad de la empresa.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

**Figura 4. 3.** Prueba Log-Rank por la variable clasificación por capital mayoritario de la consola R.

```
> plot(survfit(Surv(Años, NoRenta) ~ Fam), lty=c(1, 1, 1, 1), col=1:4)
> survdiff(Surv(Años, NoRenta) ~ Fam)
Call:
survdiff(formula = Surv(Años, NoRenta) ~ Fam)

      N Observed Expected (O-E)^2/E (O-E)^2/V
Fam=0 280      71     73.5    0.0855    0.117
Fam=1 754     237    234.5    0.0268    0.117

Chisq= 0.1  on 1 degrees of freedom, p= 0.732
```

**Fuente:** Elaboración propia en R-Project

### 4.2.2 CLASIFICACIÓN DE EMPRESAS POR CERTIFICACIÓN.

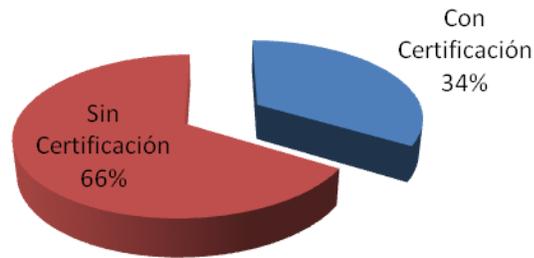
La rentabilidad de las empresas depende de la calidad de productos y servicios y la manera actual de identificar estos factores es mediante certificaciones que se otorgan para medirlo, una de ellas es la certificación ISO 9000. Esta certificación requiere estandarizar las actividades del personal, asegurar la calidad del servicio, medir la eficacia y eficiencia de la organización, entre otras actividades.

Es por esto que mediremos o evaluaremos el impacto que tiene el hecho de contar con alguna certificación que acredite la calidad de la empresa. La variable utilizada es categorizada como una variable dummy, para aquellas empresas que cuentan con dicha certificación o equivalente y para aquellas que no cuentan con tal, ya sea que está en trámite o simplemente no la tienen.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

**Figura 4. 4.** Distribución por empresas certificadas.



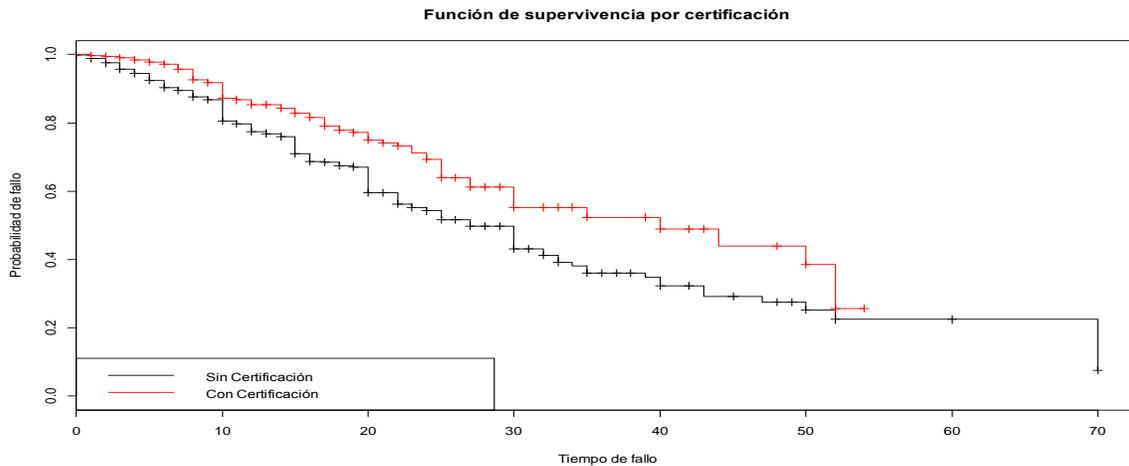
**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

Como se observa en la Figura 4.4, la mayoría de las empresas encuestadas no cuentan con una certificación ISO 9000 o equivalente, de hecho estas representan un 66% de la muestra lo que corresponde a 686 encuestas del total. Aquellas que cuentan con este reconocimiento son tan solo el 34% o lo equivalente a 384 compañías.

En la siguiente grafica de curvas de supervivencia, se nota cierta diferencia entre las que cuentan con certificación y las que no lo poseen. Aquellas que no tienen la certificación muestran una probabilidad de supervivencia menor a las que tienen tal (Figura 4.5).

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

**Figura 4. 5.** Curvas de supervivencia de Kaplan-Meier por certificación.



**Fuente:** Elaboración propia en R Project versión 3.0.1.

La diferencia estadística la podemos notar mediante la prueba de puntajes en la figura 4.6, se demuestra que para un nivel de significancia al cinco por ciento se comprueba que las curvas de supervivencia son estadísticamente diferentes una de la otra lo cual indica que al estar certificado se está asegurando la calidad del producto.

**Figura 4. 6.** Prueba de Log-Rank de la consola de R.

```
> plot(survfit(Surv(Años,NoRenta)~Cer),lty=c(1,1,1,1),col=1:4)
> survdiff(Surv(Años,NoRenta)~Cer)
Call:
survdiff(formula = Surv(Años, NoRenta) ~ Cer)

      N Observed Expected (O-E)^2/E (O-E)^2/V
Cer=0 686      229      198      4.80      14
Cer=1 348       79      110      8.66      14

Chisq= 14 on 1 degrees of freedom, p= 0.00018
```

**Fuente:** Elaboración propia, R Project versión 3.0.1.

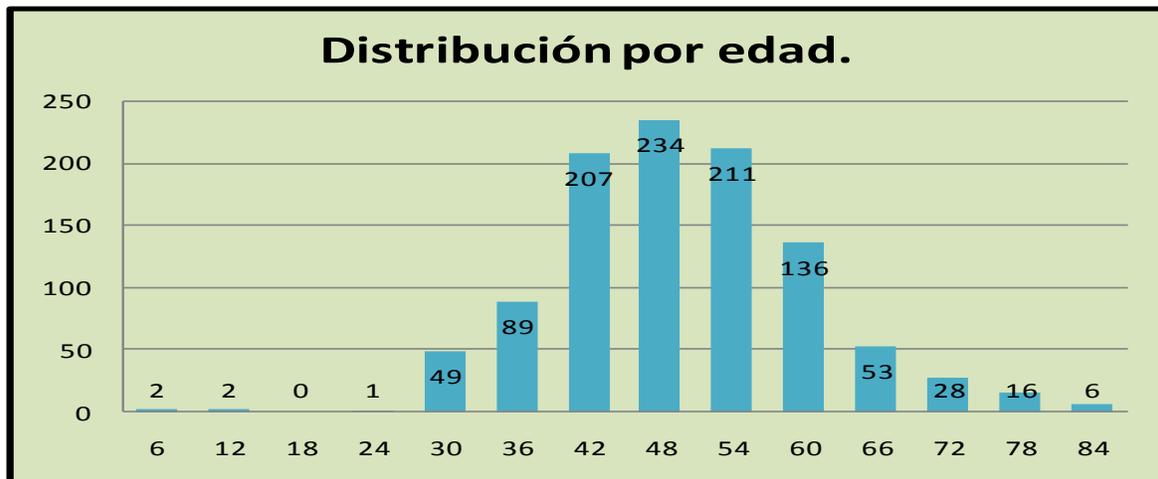
## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

### 4.2.3 EDAD DEL GERENTE.

La variable edad del gerente es de tipo continua que tiene que ser mayor que cero. Está es considerada para tener una unidad de medida que refleje la experiencia con la que cuenta la persona que está al frente de la empresa. En ciertas ocasiones no es determinante la edad para el rendimiento o elaboración de alguna actividad pero esta será usada para tener una unidad de medida de la experiencia.

En la Figura 4.7 se muestra la distribución por edad de los directores o gerentes de la empresa, nótese que la mayoría se encuentra entre la edad de 42 y 48 años, existen muy pocos casos de personas jóvenes con menos de 35 años. La edad promedio es de 47 años y los datos atípicos posiblemente corresponden a datos de encuestas mal levantadas.

**Figura 4. 7** Distribución por edad de los directores al frente de la empresa.



**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

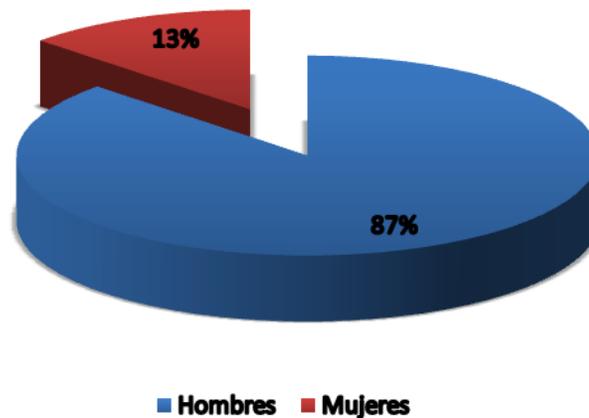
## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

### 4.2.4 DESCRIPCIÓN POR GÉNERO.

La Figura 4.8 nos muestra un predominio de empresas que son dirigidas por hombres, siendo estos un total de 901, lo que representa el 87 por ciento de la muestra. El resto es de 133 empresas dirigidas por mujeres.

**Figura 4. 8**Distribución por género de directores al frente de la empresa.

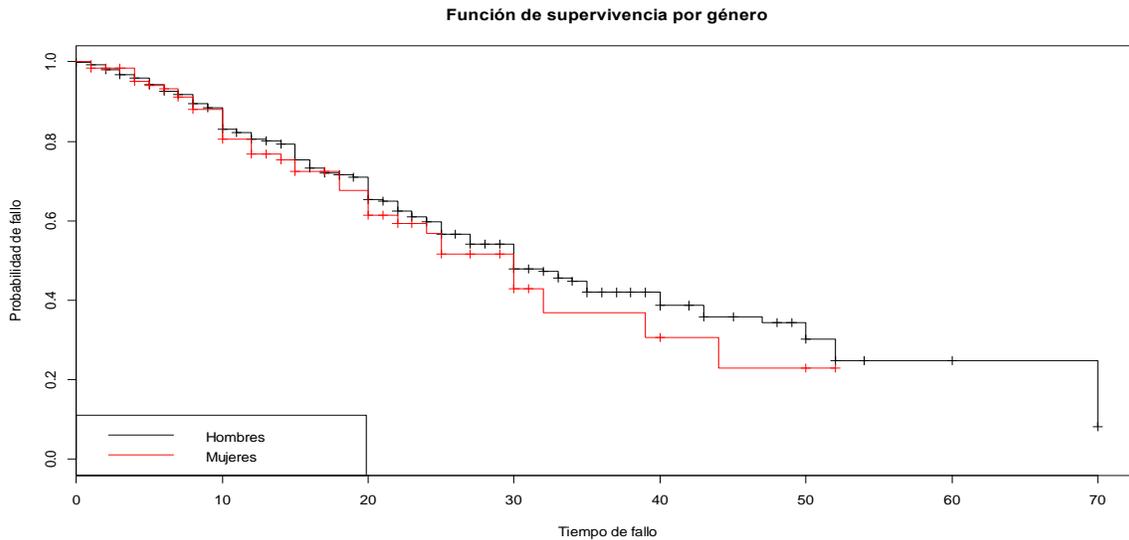


**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

En la Figura 4.9 se observa que aquellas empresas que son dirigidas por mujeres tienen ligeramente una menor supervivencia o mayor riesgo en la caída de la rentabilidad de la empresa a partir de los 30 años de vida de la institución.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

**Figura 4. 9.** Graficas de supervivencia de Kaplan-Meier por género.



**Fuente:** Elaboración propia mediante R-Project.

Al realizar la prueba estadística de puntajes que a continuación se muestra en la Figura 4.10 de la consola de R, se tiene un p-value igual a 0.427, lo que significa que las gráficas de supervivencia no son estadísticamente diferentes. Este resultado da muestra de que la diferencia de género no es válida para diferenciar el éxito de una empresa, por tanto, en este apartado se resalta la importancia de que en términos económicos para el rubro de dirección empresarial tanto hombres como mujeres son igual de capaces al momento de la toma de decisiones.

**Figura 4. 10.** Prueba de Log Rank de la consola R-Project.

```
> plot(survfit(Surv(Años,NoRenta)~Gen),lty=c(1,1),col=1:2)
> survdiff(Surv(Años,NoRenta)~Gen)
Call:
survdiff(formula = Surv(Años, NoRenta) ~ Gen)

      N Observed Expected (O-E)^2/E (O-E)^2/V
Gen=1  901     266     270.5   0.0738     0.63
Gen=2  133      42      37.5   0.5321     0.63

Chisq= 0.6 on 1 degrees of freedom, p= 0.427
```

**Fuente:** Elaboración propia mediante R-Project.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

### 4.2.5 GERENTE FUNDADOR

El profesor Noam Wasserman de la Harvard Business School ha pasado la última década estudiando el impacto que tiene el fundador sobre una compañía (citado de Fortune magazine, 25/02/2012), al respecto mencionan que:

“En un estudio de 460 nuevas empresas o startups estadounidenses, encontró que, en promedio, aquellas en las que los CEO<sup>5</sup> fundadores seguían siendo los principales tomadores de decisiones eran menos valiosas que las empresas gestionadas por CEO externos. Dicho en palabras llanas, las habilidades necesarias para inventar un nuevo producto o servicio son distintas a las que se necesitan para administrar un negocio, y pocas personas poseen ambas.

Quienes apoyan a los fundadores argumentarán que, aunque a veces se puede enseñar a un soñador a ser un enérgico operador empresarial, rara vez puedes enseñar a un hombre de negocios a soñar.

Tal vez lo más importante, los CEO fundadores miran a largo plazo. Mientras que los CEO conservan su trabajo por unos ocho años en promedio, los CEO fundadores procuran a menudo manejar sus compañías de por vida y construir un legado para sí mismos mientras cambian el mundo con un nuevo producto o servicio. Como escribe Ben Horowitz”. (Este texto pertenece a la edición de Fortune del 25 de febrero de 2012).

Por este motivo, se decidió estudiar ésta variable por lo que implica o podría impactar para la rentabilidad de la empresa si es el caso. La variable

---

<sup>5</sup>Chief executive officer por sus siglas en Inglés o en español director ejecutivo en jefe.

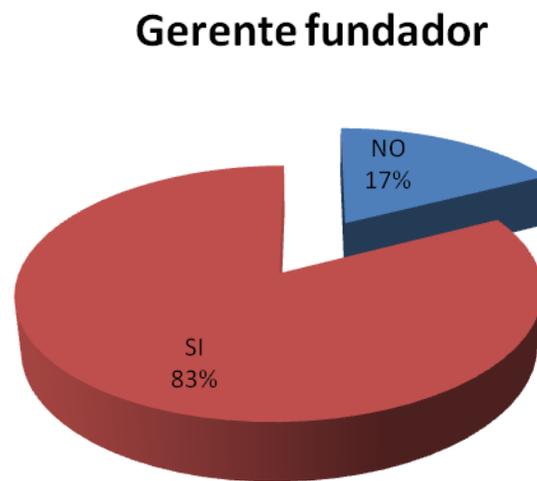
## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

utilizada es por tanto de tipo dummy, que representa el tipo de gerente donde toma el valor de uno si es fundador y cero en el otro caso.

Obsérvese en la Figura 4.11 que la mayoría de las empresas tienen en su estructura a uno de los socios fundadores al frente de la compañía, un 83 por ciento predominante.

**Figura 4. 11.** Distribución por tipo de gerente.



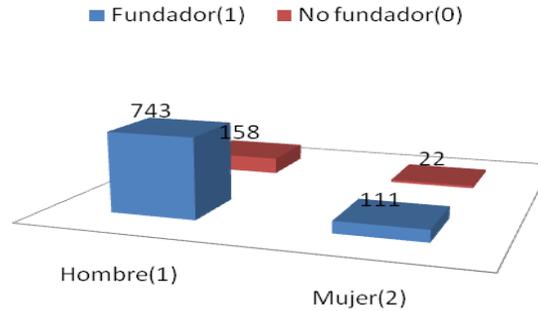
**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

Ahora bien, en la figura siguiente (Figura 4.12) se puede observar que las empresas que prevalecen son emprendidas por hombres, esta situación es evidente cuando hablamos de la cantidad de empresas dirigidas por género. Existe una razón de 72 por ciento de empresas emprendidas por hombres, mientras que por las mujeres son apenas el 11 por ciento del total.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

**Figura 4. 12.** Relación por género y tipo de gerente.



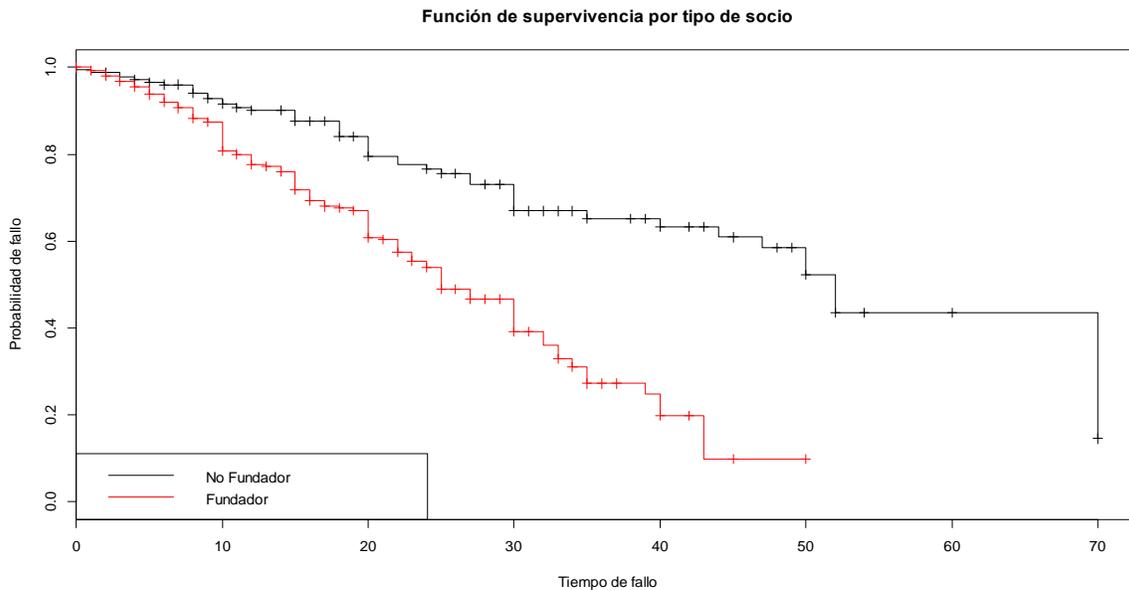
**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

Además, el 83 por ciento de los hombres que son gerentes han sido parte de la fundación, y el 82 por ciento de las mujeres gerentes, han sido parte también de su constitución. Es decir, en términos porcentuales prácticamente es la misma razón entre hombres y mujeres quienes deciden crear una empresa.

A continuación en la figura 4.13 se muestran las líneas de supervivencia destacando la diferencia entre las empresas que no conservan a sus fundadores y las que sí, las primeras tienen un menor riesgo en la baja de su rentabilidad que las otras. Después de los 43 años de edad de las empresas, aquellas con su fundador al frente se estabilizan, mientras que las otras perduran por más tiempo.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

**Figura 4. 13.** Distribución de supervivencia de Kaplan – Meyer por tipo de gerente.



**Fuente:** Elaboración propia mediante R-Project.

La prueba de puntajes en la figura 4.14 confirma lo anteriormente planteado con un p-value de  $2.58 \cdot e^{-09}$ , lo que significa que se rechaza la hipótesis nula de igualdad de las funciones de supervivencia (para el nivel de significancia del 5%).

**Figura 4. 14.** Prueba Long-Rank de la consola de R-Project por tipo de gerente.

```
> plot(survfit(Surv(Años, NoRenta) ~ Fun), lty=c(1,1), col=1:2)
> survdiff(Surv(Años, NoRenta) ~ Fun)
Call:
survdiff(formula = Surv(Años, NoRenta) ~ Fun)

      N Observed Expected (O-E)^2/E (O-E)^2/V
Fun=0 180      49    92.4    20.42    35.5
Fun=1 854     259   215.6     8.76    35.5

Chisq= 35.5 on 1 degrees of freedom, p= 2.58e-09
```

**Fuente:** Elaboración propia mediante R-Project.

## **CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO**

---

### **4.2.6 NIVEL DE FORMACIÓN ACADÉMICA DEL GERENTE**

Otro aspecto importante para determinar el éxito de una empresa es el nivel educativo de los recursos humanos. En este caso nos enfocaremos en el nivel de formación del gerente, quien es el encargado de las principales decisiones para guiar el rumbo que seguirá la empresa.

La clasificación por esta categoría está dada en tres niveles, el primero representa a aquellos con estudios de nivel básico, bachillerato o técnico, es decir, aquellas que a lo más alcanzaron el nivel medio superior.

La segunda categoría es para aquellas personas que estudiaron al menos tres años la licenciatura, los que lo hicieron con menos de este periodo se consideran en el básico. Y la última categoría es para aquellos que tiene estudios de posgrado.

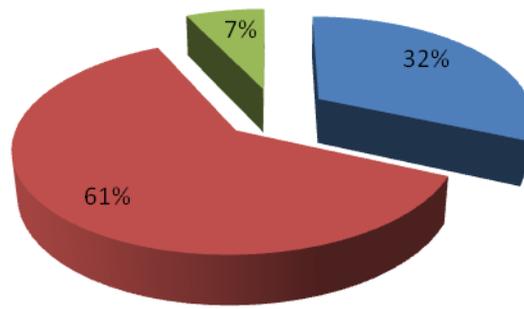
Se observa en la figura 4.15 que la mayoría son dirigidas por aquellos con estudios universitarios siendo estos 633, posteriormente le siguen las personas con estudios básicos con un total de 328 empresas y por último están los que tienen posgrado, estos últimos corresponden al siete por ciento del total, es decir, 73 empresas.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

**Figura 4.15.** Distribución por nivel de formación del gerente.

### Nivel de formación del gerente

■ Estudios básicos, bachillerato, técnico= 328 ■ Estudios universitarios= 633 ■ Posgrado= 73

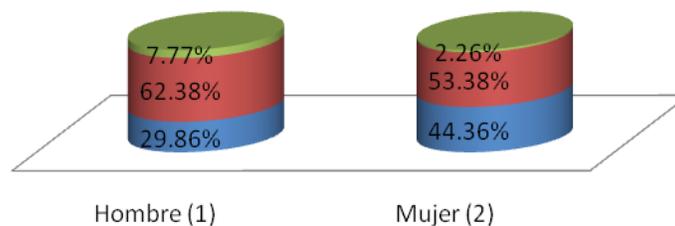


**Fuente:** Elaboración propia con datos obtenidos de la encuesta 2011-2012.

El porcentaje de nivel de estudios por género se muestra a continuación en la Figura 4.16, para el caso de las mujeres existen tendencias cercanas de dirección de una empresa con estudios básicos que con universitarios. Adicional, la relación entre gerentes fundadores con estudios universitarios predomina frente a los demás.

**Figura 4. 16.**Distribución de la formación académica del gerente por género.

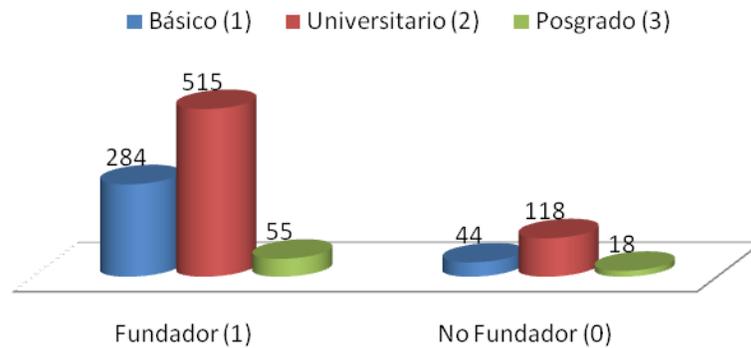
■ Básico (1) ■ Universitario (2) ■ Posgrado (3)



**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

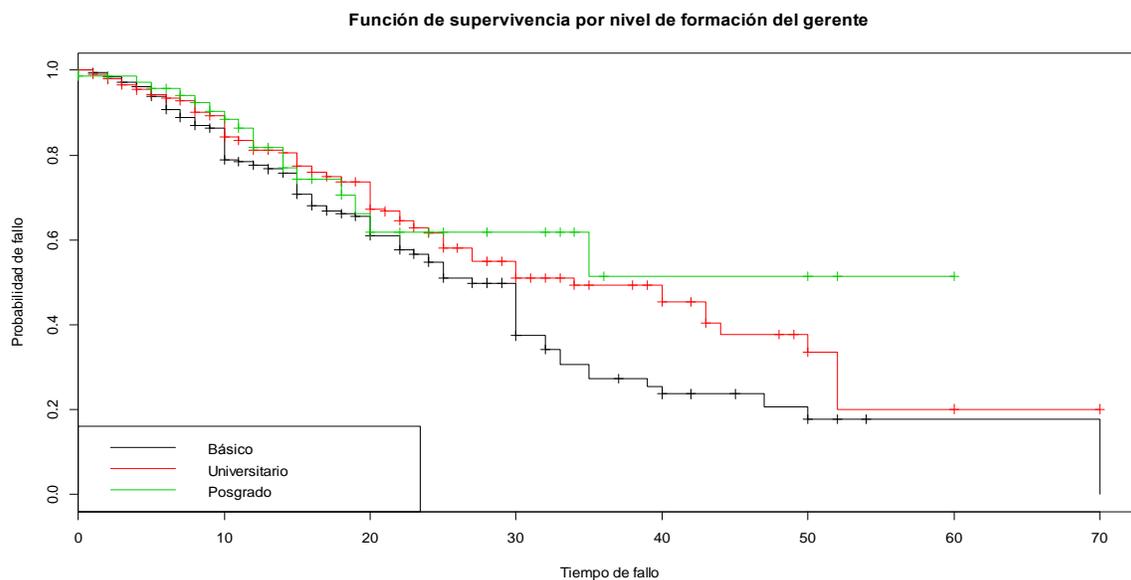
**Figura 4. 17.** Distribución de nivel académico por tipo de gerente.



**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

De manera visual, la gráfica de supervivencia (Figura 4.18) modela mayor probabilidad de persistencia si el nivel de estudios es más alto, éstas curvas empiezan a separarse o a diferenciarse en mayor medida después del año 20.

**Figura 4. 18.** Grafica de supervivencia de Kaplan- Meyer por nivel académico del gerente.



**Fuente:** Elaboración propia en R Project.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

La variable utilizada en R es llamada como “For” donde uno indica el nivel básico, dos el universitario y tres posgrado. Como se mencionó anteriormente, las curvas de supervivencia se muestran diferentes y es confirmado por la prueba de puntajes, esta prueba tiene un p valor de 0.0169(Figura 4.19) lo que indica que mientras más elevado es el nivel de preparación del gerente es más probable la supervivencia de la empresa a lo largo de los años.

**Figura 4. 19.** Prueba de log Rank de la consola de R-Project por nivel académico.

```
> plot(survfit(Surv(Años,NoRenta)~For),lty=c(1,1,1,1),col=1:4)
> survdiff(Surv(Años,NoRenta)~For)
Call:
survdiff(formula = Surv(Años, NoRenta) ~ For)

      N Observed Expected (O-E)^2/E (O-E)^2/V
For=1 328      124    101.6      4.96      7.73
For=2 633      167    184.2      1.60      4.16
For=3  73       17     22.3      1.25      1.41

Chisq= 8.2 on 2 degrees of freedom, p= 0.0169
```

**Fuente:** Elaboración propia en R Project.

### 4.2.7 NÚMERO DE EMPLEADOS

Otra característica importante para el desarrollo de una empresa, es el número de empleados, esta variable de estudio servirá para verificar la consolidación de cualquier empresa. Generalmente las empresas que tienen más empleados han encontrado el equilibrio de desarrollo a largo plazo.

Esta variable es de tipo entero que debe tener valores mayores que 0 en condiciones reales. En la gráfica siguiente (figura 4.20) se muestra que la mayoría de las empresas encuestadas tienen menos de 32 empleados.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

**Figura 4. 20.** Distribución de las Pymes por número de empleados.



**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

### 4.2.8 CLASIFICACIÓN DE LAS EMPRESAS POR SECTOR.

Esta variable está dividida por cuatro variables categóricas;

1. Manufactura
2. Construcción
3. Comercio
4. Servicios.

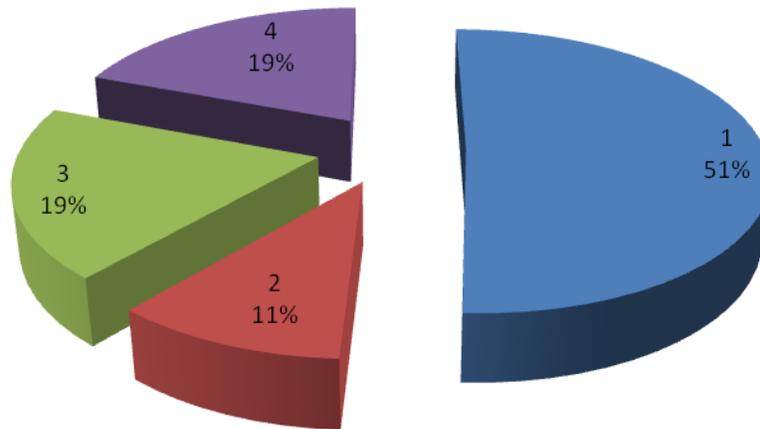
La clasificación de las empresas por este tipo es importante para entender el riesgo de cada uno de los sectores, por motivos de ciclos económicos algunos sectores son más afectados que otros.

A continuación se muestra la distribución de tipos de empresa encuestadas para esta investigación, la mayor parte de estas fueron manufactureras, esto es un 51 por ciento, después le siguen comercio y servicios y las que menos se registraron fueron las de construcción con un 11 por ciento del total (Figura 4.21).

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

**Figura 4. 21.** Distribución de las Pymes por sector.

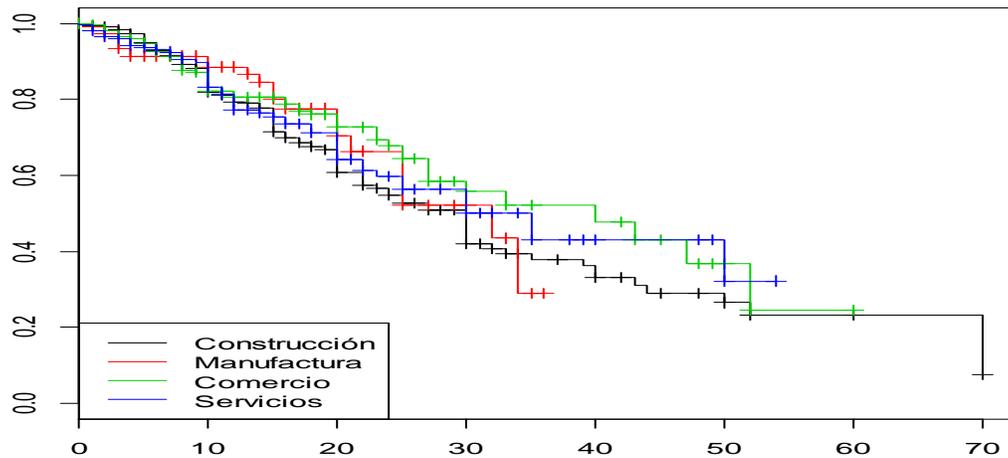


**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

En la gráfica de función de supervivencia por sector (Figura 4.22), se aprecia que las empresas constructoras tienen menor probabilidad de supervivencia, puede que este factor sea así por las últimas crisis que han golpeado fuertemente a este sector en especial la crisis hipotecaria del 2008. Por otra parte, las empresas comerciales son menos propensas a recaer en la rentabilidad ya que la venta de sus productos es constante, no así la variación en sus precios.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

Figura 4. 22. Grafica de supervivencia de Kaplan Meyer por sector.



Fuente: Elaboración propia en R Project.

En los primeros años, las gráficas de supervivencia se muestran iguales, posteriormente la volatilidad va diferenciando una curva de otra aunque la diferencia no es notable, la equivalencia de estas curvas se demuestra en la Figura 4.23.

Figura 4. 23. Consola de R. Prueba de Log Rank.

```
> plot(survfit(Surv(Años,NoRenta)~Sec),lty=c(1,1,1,1),col=1:4)
> survdiff(Surv(Años,NoRenta)~Sec)
Call:
survdiff(formula = Surv(Años, NoRenta) ~ Sec)

      N Observed Expected (O-E)^2/E (O-E)^2/V
Sec=1 525      172    157.0    1.4368    3.065
Sec=2 109       24     25.8    0.1308    0.149
Sec=3 201       55     65.9    1.8144    2.404
Sec=4 199       57     59.2    0.0849    0.109

Chisq= 3.6 on 3 degrees of freedom, p= 0.305
```

Fuente: Elaboración propia R Project

La prueba de puntajes muestra un p valor de 0.308, por tanto, no existe evidencia empírica que demuestre que las curvas de supervivencia sean diferentes. Esto es que el tipo de sector en donde se desarrolle una empresa no implica que tenga mayor éxito que en otro.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

### 4.3ANÁLISIS PORCRUCE DE VARIABLES

Ahora bien, realizando un cruce de variables por capital mayoritario, tipo de socio, nivel de formación y género, estas últimas características del gerente, se obtiene la siguiente Figura.

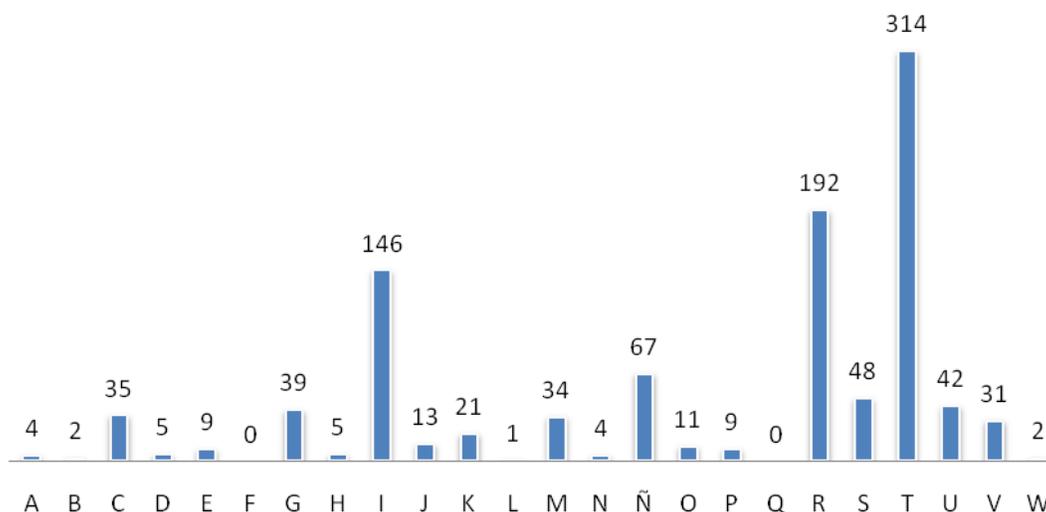
**Figura 4.24. Cruce de variables.**

Grupo	Capital	Socio	Estudios	Género	No.	%	
A	Familiar(1) 280	No Fundador(0) 55	Básicos(1) 6	H(1)	4	0.39%	
B				M(2)	2	0.19%	
C			Universitarios(2) 40	H(1)	35	3.38%	
D				M(2)	5	0.48%	
E			Posgrado(3) 9	H(1)	9	0.87%	
F				M(2)	0	0.00%	
G		Fundador(1) 225	Básicos(1) 44	H(1)	39	3.77%	
H				M(2)	5	0.48%	
I			Universitarios(2) 159	H(1)	146	14.12%	
J				M(2)	13	1.26%	
K			Posgrado(3) 22	H(1)	21	2.03%	
L				M(2)	1	0.10%	
M		No Familiar(0) 754	No Fundador(0) 125	Básicos(1) 38	H(1)	34	3.29%
N					M(2)	4	0.39%
Ñ	Universitario(2) 78			H(1)	67	6.48%	
O				M(2)	11	1.06%	
P	Posgrado(3) 9			H(1)	9	0.87%	
Q				M(2)	0	0.00%	
R	Fundador(1) 629		Básicos(1) 240	H(1)	192	18.57%	
S				M(2)	48	4.64%	
T			Universitario(2) 356	H(1)	314	30.37%	
U				M(2)	42	4.06%	
V			Posgrado(3) 33	H(1)	31	3.00%	
W				M(2)	2	0.19%	

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

Para entender mejor la tabla anterior se realizó la siguiente grafica que nos mostrara más detalladamente el comportamiento de estas variables agrupadas.

**Figura 4.25.** Distribución absoluta por grupo.



**Fuente:** Elaboración propia con datos de la encuesta 2011-2012.

De acuerdo a la Figura 4.25, se muestran algunos sucesos importantes, algunas de ellas se notaron a lo largo de este capítulo, el resto se analizara a continuación.

Se describen seis sucesos destacados; en el grupo T, muestra el predominio de ciertas variables, corresponde a las empresas no familiares las cuales el gerente es parte de la fundación, con estudios universitarios y además es hombre. En segundo lugar se encuentra el grupo R, esta se diferencia del anterior porque el nivel de estudios del gerente es básico. En tercer lugar se localiza el grupo I, en este caso la diferencia con el T es que es una empresa de tipo familiar.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

En el siguiente sitio se encuentra el grupo Ñ, lo que representa el 6.48 por ciento de la muestra, esta se diferencia de la primera por que el gerente no es fundador.

En el quinto sitio están aquellas empresas que son dirigidas por mujeres con estudios básicos (Grupo S) y las que tienen estudios universitarios en el sexto (Grupo U), ambos grupos muestran que son parte de la fundación de la empresas de una empresa que no es dueño una familia.

### 4.4DESARROLLO DELMODELO DE REGRESIÓN DE COX.

Una vez que se han conocido y analizado las variables que serán introducidas en el modelo de regresión de Cox, el modelo se ejecutara mediante el software R Project. En los siguientes apartados se muestra el proceso de selección de variables y ejecución del modelo óptimo que dé solución a los objetivos planteados.

Las variables referidas de aquí en adelante para el análisis de resultados arrojadas por el método se enlistan en la tabla siguiente.

**Tabla 4. 1. Traducción de las variables introducidas en el modelo.**

Nombre o ID en el software	Descripción.
Cer	La empresa cuenta con certificación de calidad. Considera los siguientes valores; 0 si no cuenta con Certificación 1 si cuenta con Certificación
Eda	Edad del gerente o director. Toma valores de 0 a infinito.
Fam	La familia cuenta con el capital mayoritario. Considera los

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

	<p>siguientes valores;</p> <p>0 si familia no es dueña de la mayor parte del capital de la empresa</p> <p>1 si la familia es dueña de la mayor parte del capital de la empresa</p>
For	<p>Nivel académico del gerente. Esta variable toma los siguientes valores;</p> <p>1 si tiene estudios básicos, bachillerato, técnico.</p> <p>2 si tiene estudios universitarios (3 o más años).</p> <p>3 si cuenta con Posgrado</p>
Fun	<p>El gerente es socio fundador de la empresa. Considera los siguientes valores;</p> <p>0 si el gerente no es socio fundador</p> <p>1 si el gerente es socio fundador.</p>
Gen	<p>Genero del gerente o director. Considera los siguientes valores;</p> <p>0 si el género es femenino</p> <p>1 si el género es masculino</p>
Num	<p>Número de empleados. Considera valores enteros positivos.</p>
Sec	<p>Sector en el que se desarrolla la empresa. Toma los siguientes valores.</p> <p>1 Empresa manufacturera o del sector primario.</p> <p>2 Empresa constructora.</p> <p>3 Empresa comercial.</p> <p>4 Empresa de servicios.</p>

### 4.6 ESTIMACIÓN DE PARÁMETROS.

A continuación en la figura 4.26 se muestra la salida de R la cual corresponde al resumen del modelo de supervivencia de Cox, con los años de la empresa como tiempo de fallo y la variable NoRenta como indicador de censura de una muestra de 1034 empresas con 308 eventos de fallo, mostrada en la sexta línea de la figura.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

En la tabla se muestra en la primera columna el nombre de la variable codificada en R, en la segunda el valor de los coeficientes estimados, posteriormente la tasa de riesgo relativo, así como el error estándar, el valor z de una distribución normal y su p-valor correspondiente encerrado en el ovalo rojo. El intervalo de confianza para las respectivas tasas de riesgo se muestran en las últimas dos columnas de la tabla de abajo al 95 por ciento de confianza.

**Figura 4.26.**Resultados de la primer corrida en R.

```
> summary(cox1)
Call:
coxph(formula = Surv(Años, NoRenta) ~ Eda + Sec + For + Cer +
      +Fam + Fun + Gen + Num)

n= 1034, number of events= 308

      coef exp(coef) se(coef)      z Pr(>|z|)
Eda -0.031933  0.968572  0.005975 -5.344 9.08e-08 ***
Sec -0.024381  0.975913  0.049695 -0.491 0.623695
For -0.156620  0.855029  0.104665 -1.496 0.134552
Cer -0.280590  0.755338  0.134177 -2.091 0.036511 *
Fam -0.053662  0.947752  0.140680 -0.381 0.702871
Fun  1.226079  3.407841  0.189131  6.483 9.01e-11 ***
Gen -0.127213  0.880546  0.170200 -0.747 0.454804
Num -0.006912  0.993112  0.002017 -3.427 0.000611 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
Eda    0.9686    1.0324    0.9573    0.9800
Sec    0.9759    1.0247    0.8853    1.0758
For    0.8550    1.1696    0.6964    1.0497
Cer    0.7553    1.3239    0.5807    0.9825
Fam    0.9478    1.0551    0.7194    1.2487
Fun    3.4078    0.2934    2.3523    4.9371
Gen    0.8805    1.1357    0.6308    1.2292
Num    0.9931    1.0069    0.9892    0.9970

Concordance= 0.662 (se = 0.019 )
Rsquare= 0.097 (max possible= 0.972 )
Likelihood ratio test= 105.2 on 8 df, p=0
Wald test               = 84.23 on 8 df, p=6.883e-15
Score (logrank) test = 87.25 on 8 df, p=1.665e-15
```

**Fuente:** Elaboración propia en R Project.

Los valores-p enmarcados en el ovalo rojo indican que las covariables; Sec, For, Fam y Gen no contribuyen significativamente al modelo a una significancia del 10 por ciento. De esto se ve que para esta muestra no influye la diferencia de género, la clasificación por empresa familiar no es trascendente, el sector tampoco

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÓ

es útil para este análisis y que es equivalente la experiencia con los estudios académicos.

Nótese que las variables antes mencionadas son aquellas en las que la diferencia de curvas de supervivencia (prueba de puntajes) no son estadísticamente diferentes a excepción de la variable de nivel académico del gerente. Una vez quitando estas variables, se realiza el modelo con aquellas que hacen diferencia.

**Figura 4. 27.** Resultados de la segunda corrida en R.

```
> summary(coxph(Surv(Años, NoRenta)~Eda+Cer+Fun+Num)
Call:
coxph(formula = Surv(Años, NoRenta) ~ Eda + Cer + Fun + Num)

n = 1034, number of events = 308

      coef exp(coef)    se(coef)      z Pr(>|z|)
Eda -0.031554  0.968938  0.005920 -5.330 9.80e-08 ***
Cer -0.306793  0.735803  0.132506 -2.315 0.020595 *
Fun  1.252040  3.497469  0.187911  6.663 2.68e-11 ***
Num -0.006751  0.993272  0.001969 -3.428 0.000607 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
Eda    0.96889    1.0321    0.9578    0.9802
Cer    0.7358    1.3591    0.5675    0.9540
Fun    3.4975    0.2859    2.4199    5.0548
Num    0.9933    1.0068    0.9894    0.9971

Concordance= 0.659 (se = 0.019 )
Rsquare= 0.094 (max possible= 0.972 )
Likelihood ratio test= 102.2 on 4 df, p=0
Wald test = 81.42 on 4 df, p=1.11e-16
Score (logrank) test = 84.44 on 4 df, p=0
```

**Fuente:** Elaboración propia en R Project

De la figura 4.27 se observa que las cuatro variables consideradas son estadísticamente significativas, el valor p de la razón de verosimilitud indica que es un modelo adecuado considerando todas las covariables al ser menor que 0.01. El estadístico de Wald  $p=0$  prueba que cada coeficiente beta es diferente de cero, es decir, se rechaza la hipótesis nula de:

$$H_0 = \beta_i = 0; i = 1, 2, \dots, p$$

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

Finalmente, el p valor de la prueba de puntos hace ver que existen diferencias en las curvas de supervivencia de dos o más grupos, de tal manera que si se agrupan las curvas por estratos, estas no son iguales al ser p-valor=0<0.01.

Se concluye que el modelo es significativo en cualquiera de los tres criterios, prueba de razón de verosimilitud, prueba de Wald y prueba de puntajes - score o log-Rank-.

Por consiguiente el modelo final se presenta de la siguiente manera:

$$\lambda(t; X) = \lambda_0(t) \times e^{b_1 \text{Edad} + b_2 \text{Numero de empleados} + b_3 \text{Certificación} + b_4 \text{Fundador}} \quad 4.1$$

### 4.5 INTERPRETACIÓN DE VARIABLES.

De acuerdo con Palmer (1988), un valor positivo del coeficiente aumenta el valor de la función de riesgo, lo cual conlleva a una relación negativa con el tiempo de cambio. El coeficiente positivo indica un mayor riesgo de que se produzca un cambio. Los coeficientes negativos indican un menor riesgo, lo que hacen ser factores de protección.

El coeficiente B1 se interpreta de la siguiente manera;

$$e^{b_1 * \text{Edad}} = e^{-0.031554 * \text{Edad}} = 0.968938^{\text{Edad}} \quad 4.2$$

Indica que para el incremento de un año, el logaritmo de la función de riesgo se reduce en 0.311 unidades aproximadamente, es decir, el riesgo de una recaída en la rentabilidad se reduce a una tasa de  $(1-0.968938)=0.03106$  o 3.106 por ciento por cada incremento de un año en la edad del gerente. Lo que nos muestra el valor de la experiencia por año.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

---

El coeficiente B2 se interpreta de la misma manera;

$$e^{b_1 * Num} = e^{-0.006751 * Num} = 0.993272^{Num} \quad 4.3$$

Al incremento de un empleado, el logaritmo de la función de riesgo se reduce en 0.993272 unidades aproximadamente, es decir, el riesgo de una recaída en la rentabilidad se reduce a una tasa de  $(1-0.993272)=$  o 0.6728 por ciento por cada incremento de un año en la edad del gerente.

Coeficiente B3;

$$\frac{\lambda_{CC}}{\lambda_{SC}} = \frac{e^{b_3 * 1}}{e^{b_3 * 0}} = e^{b_3 * Certificación} = e^{-0.306793 * 1} = 0.7358 \quad 4.4$$

El coeficiente negativo indica que las empresas con certificación tienen mejor pronóstico de supervivencia que las que no lo tienen.

También indica que el riesgo de recaída en la rentabilidad para las empresas con certificación es 0.7358 veces que el de las empresas sin certificación.

El inverso de este valor indica el riesgo estimado para las empresas con certificación, es decir, el riesgo de recaída en la rentabilidad es 1.3590 veces más alto para empresas sin certificación.

Coeficiente B4;

$$\frac{\lambda_F}{\lambda_{NF}} = \frac{e^{b_4 * 1}}{e^{b_4 * 0}} = e^{b_4 * Fundador} = e^{1.252} = 3.4373 \quad 4.5$$

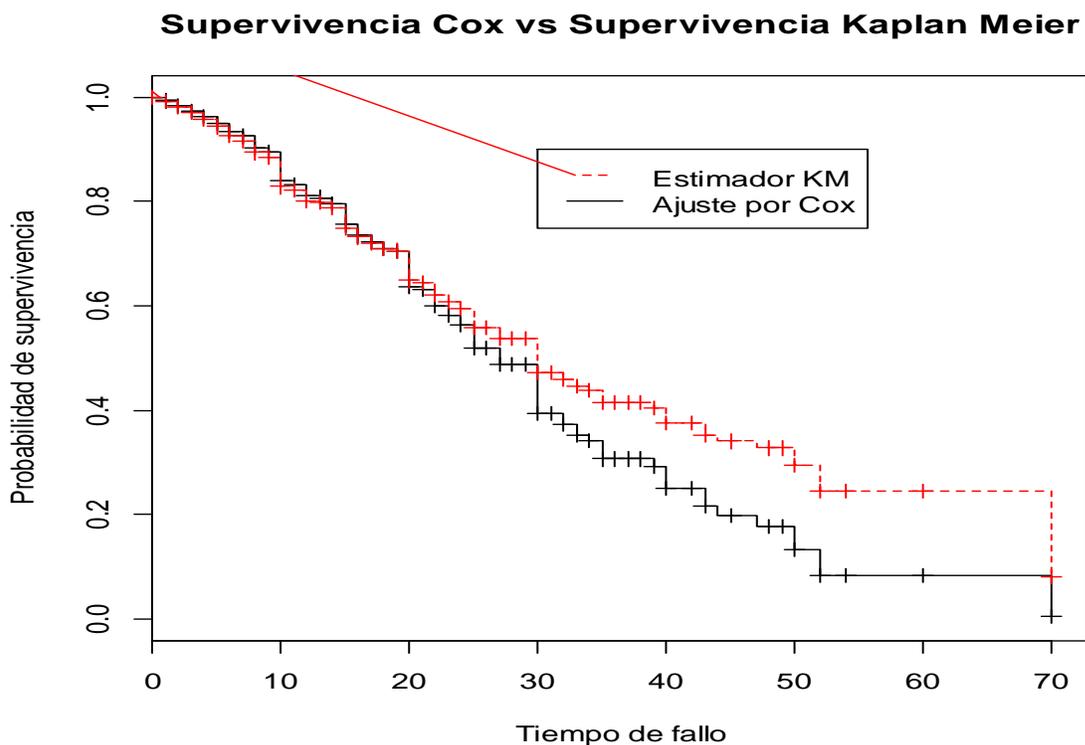
## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

En este caso, el riesgo de recaída es 3.4373 veces mayor para aquellos socios fundadores que para aquellos que no lo son.

### 4.7 COMPARACIÓN DE GRÁFICAS DE SUPERVIVENCIA.

A continuación se presentan las gráficas de supervivencia por el método de Kaplan-Meier y la de regresión de Cox, además se muestran algunos valores para estos dos esquemas.

**Figura 4. 28.** Comparación de graficas de supervivencia.



**Fuente:** Elaboración propia en R Project.

Para la curva Kaplan-Meier existe una mayor probabilidad de supervivencia que con la de Cox, esta diferencia podría ser debido a que el primero considera a todos los elementos de la población homogéneos, es decir no estudia las posibles causas de fallo.

## CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO

En la tabla 4.2 se muestran los valores por cada uno de los métodos. En esta se puede observar que en los primeros veinte años no existen valores que pudieran diferenciar una curva de otra, esta diferencia inicia a partir del año 20 y posteriormente se va haciendo evidente con los años posteriores.

**Tabla 4. 2.** Valores de supervivencia por los métodos de Cox y de Kaplan-Meier.

Tiempo	KM	Cox	Tiempo	KM	Cox
0	0.999	0.99917	20	0.648	0.63662
1	0.991	0.99251	21	0.645	0.6326
2	0.981	0.98397	22	0.62	0.59983
3	0.969	0.97328	23	0.607	0.58238
4	0.958	0.96303	24	0.594	0.56437
5	0.943	0.94935	25	0.559	0.51837
6	0.927	0.93472	27	0.538	0.48709
7	0.917	0.92473	30	0.473	0.39343
8	0.893	0.90261	32	0.46	0.37346
9	0.884	0.89398	33	0.446	0.35253
10	0.828	0.83905	34	0.438	0.34121
11	0.821	0.83137	35	0.414	0.30689
12	0.801	0.81063	39	0.405	0.29299
13	0.798	0.80721	40	0.377	0.25093
14	0.788	0.79648	43	0.353	0.21604
15	0.75	0.75525	44	0.341	0.19794
16	0.732	0.73448	47	0.328	0.17873
17	0.721	0.72225	50	0.295	0.13359
18	0.71	0.70929	52	0.246	0.08394
19	0.705	0.70383	70	0.082	0.00643

**Fuente:** Elaboración Propia.

## **CAPÍTULO 4. CONSTRUCCIÓN DEL MODELO DE COX PARA LAS PYMES DEL BAJÍO**

---

# CONCLUSIONES

---

Para concluir, se hará una recapitulación de lo que se analizó a lo largo de esta tesis, y por su puesto lo que se logró con esta investigación con ayuda de la técnica de regresión de Cox.

Esta investigación es considerada del campo de la estadística de métodos de supervivencia, donde se estudiaron las pequeñas y medianas empresas del bajío derivado de una encuesta realizada en el año 2011-2012, que fue de tipo transversal.

El problema consistió en estimar el tiempo de supervivencia de las empresas, debido a que un estudio longitudinal implican más recursos, se trató con esta información para estimar la probabilidad de pasar de un estado a otro, es decir, de tener un riesgo de baja en la rentabilidad, esto implica el efecto de diversas variables: años de la empresa, capital mayoritario, certificación de calidad ISO 9000 o equivalente, edad del director o gerente general, género del director o gerente general, gerente fundador, nivel de formación académica del gerente, número de empleados y/o sector en el que se desempeña la empresa.

La hipótesis a comprobar era; “Es posible construir un modelo general que determine la probabilidad de supervivencia de pequeñas y medianas empresas en México, y en específico en la región del Bajío” con el propósito de medir la capacidad de sobrevivencia de una empresa Pyme en la región del Bajío.

Los resultados mostraron que el modelo de regresión de Cox cumplió con las expectativas matemáticas para estudiar la supervivencia de las empresas, que si bien no es nuevo, si es innovador al ser tratada para estos temas, ya que en estudios actuariales se enfoca a supervivencia de individuos y pocas veces en

## CONCLUSIONES

---

estudios relacionados con entidades como organizaciones o empresas, que para esta investigación son Pymes.

Durante el tratamiento de graficas de supervivencia por grupo y su diferencia estadística, resulto ser eficiente pues este análisis nos da un preámbulo de las variables que influncian en la supervivencia de una Pyme.

Lo anterior no implica que se reste eficacia e importancia a el método seleccionado para determinar las variables, que fue escogido por ser uno de tipo poco convencional en el tratamiento de estos temas y resulta más interesante por conjugar dos funciones, una paramétrica y otra de tipo no paramétrica, que si bien tratadas por separado aportan información, cuando se trabaja con ambas al mismo tiempo muestran detalles más útiles.

Los resultados arrojados por la ejecución del modelo de regresión de Cox con el software estadístico R Project versión 3.0.1, muestran los efectos de las variables antes citadas y se pudieron determinar los principales factores que miden el riesgo de supervivencia para una Pyme en la región de Bajío. Las variables que se consideraron importantes fueron cuatro, dos propias del gerente; edad, si es fundador, y dos de la empresa; cuenta con alguna certificación y el número de empleados de esta.

Finalmente, se demuestra que esta técnica puede ser una herramienta financiera, actuarial y empresarial adecuada para datos en que la suposición de riesgos proporcionales es válida; esto implica que la estructura de este modelo impone proporcionalidad entre funciones de riesgo para diferentes niveles de covariables; o sea, para dos individuos diferentes, la razón de dos funciones de riesgo es constante respecto al tiempo.

## CONCLUSIONES

---

Quedando demostradas nuestras hipótesis que afirmaban que:

- Es posible construir un modelo general que determine la probabilidad de supervivencia de pequeñas y medianas empresas en México, demostrándose está a través de la técnica de Cox.
- Es posible identificar aquellas variables que determinan el desarrollo de una empresa.
- Demostrar que la técnica de regresión de Cox se puede llevar al campo financiero y empresarial, basándose de un modelo actuarial.
- El modelo de regresión de Cox se puede aplicar a eventos probabilísticos como una herramienta útil y confiable para diversas áreas.

Con esto se concluye esta investigación, demostrando que el análisis de supervivencia es un tema de importancia en diferentes campos del saber, pero sobre todo es importante la innovación de estas técnicas en el campo actuarial, económico y empresarial.

## BIBLIOGRAFÍA

---

# BIBLIOGRAFÍA

---

## ARTICULOS Y REVISTAS.

- Ayala M.A.; Borges R.E. y Colmenares L.G. (2007). Verificación de los supuestos del modelo de Cox. Caso de estudio: banca comercial venezolana 1996-2004; Red de Revistas Científicas de América Latina, el Caribe, España y Portugal, 23, 27-43.
- Campanu M. y Gonen M. (2011). Building a Nomogram for Survey-Weighted Cox Models Using R; Journal of Statistical Software, 23.
- Cox D. R. (1975). Partial Likelihood; Biometrika, Department of mathematics, Imperial College., 62, 269-276.
- Cox. D.R. (1972). Regression Models and Life-Tables; Journal of the Royal Statistical Society, 34, 187-220.
- Cruz A.M.; Sandra P. Usaquén Perilla, Nidia N. y Vanegas Pabón. (2010). Aplicación Del Modelo De Cox Para Estimar El Desempeño De Proveedores De Servicio De Mantenimiento.
- Domenéch J.M. (1992). Una aplicación del análisis de la supervivencia en ciencias de la salud; Anuario de Psicología, Universidad Autónoma de Barcelona, 55, 109-141.
- Douglas Rivas; Maldonado J. L.; Borges R.; Colmenares G. (2006). Aplicación de los algoritmos genéticos para estimar los parámetros en un modelo de regresión de Cox; Red de Revistas Científicas de América Latina, el Caribe, España y Portugal, Economía, 57-74.

## BIBLIOGRAFÍA

---

- Morgan T.; Keegan K. A.; Barocas D.; Ruhotina N.; Phillips S.; Chang S.; Penson D.; Clark P.; Smith J. y Cookson M. (2011). Predicting the Probability of 90-Day Survival of Elderly Patients With Bladder Cancer Treated With Radical Cystectomy.
- Nogales C. E. y Porras A. M. (2009). Aplicación del Modelo de riesgos proporcionales de Cox a pacientes con sida en España; Revista Investigación Operacional, 30, 214-222.
- O'Brien B. A.; Cohen R.J.; Wheeler T. M. y Moorin R.E. (2010). A post-radical-prostatectomy nomogram incorporating new pathological variables and interaction terms for improved prognosis. BJU International. 107. 389-395.
- Allison P. (1982). Discrete-Time Methods for the Analysis of Event Histories; American Sociological Association, 13, 61-98.
- Polterauer, Grimm, Hofstetter, Concin, Natter, Stuardza, Pötter, Marth, Reinthaller y Heinze. (2012). Nomogram prediction for overall survival of patients diagnosed with cervical cancer; British Journal of Cancer, 107, 918-924.
- Stephen J. Walters. (2009). what is a Cox Model?; Medical Statistics, Ed. 2da, School of Health and Related Research, University of Sheffield.
- Walz J.; Felix K.H; Chun; Klein E. A.; Reuther A.; Saad F.; Graefen M.; Huland H. y Karakiewicz P.I. (2008). Nomogram Predicting the Probability of Early Recurrence After Radical Prostatectomy for Prostate Cancer; The journal of urology, 181, 601-608.

## BIBLIOGRAFÍA

---

### LIBROS

- Bowers N. L.; Gerber H. U.; Hickman J. C.; Jones D. A.; Nesbitt C.J.; (1997), Actuarial Mathematics. Ed. 2da. The Society Of Actuaries.
- Cox D.R. y Oakes D. (1984). Analysis of survival data. Chapman & Hall / CRC
- DeMaris A. y Selman S. H. (2013). Converting Data Into Evidence: A Statistics Primer For The Medical Practitioner. Springer Link.
- Fox J. (2002). Cox Proportional-Hazards Regression for Survival Data.
- Hosmer D. W. y Lemeshow S. (1999). Applied Survival Analysis, Regression Modeling of Time to Event Data. John Wiley&Sons, Inc.
- Lawless. J.F. (2002). Statistical Models And Methods For Lifetime Data, Ed. 2. John Wiley & Sons, Inc.
- Palmer A. L. y Losilla J. M. (1988), El modelo de azar proporcional: la regresión de Cox. Barcelona: Universidad Autónoma de Barcelona.
- Prentice Ross L. y Kalbfleisch J.D. (2012). "The Statistical Analysis of Failure Time Data", Ed. WILEY SERIES IN PROBABILITY AND STATISTICS, Cap 4. Relative Risk (Cox) Regression Models, Cap 6 Likelihood Construction and Further Models.

## BIBLIOGRAFÍA

---

### PUBLICACIONES

- Borges P. (2005). Análisis de Supervivencia utilizando el lenguaje R; Simposio de Estadística, Colombia.
- Therneau T. & Grambsch P. (2000). Modeling Survival Data: Extending the Cox Model. U.S.A.

### PUBLICACIONES ELECTRONICAS

- Almanza L. (2015). Automotriz atraerá a siete francesas. Periódico EL ECONOMISTA [versión electrónica], publicado 01.02.2015, <http://eleconomista.com.mx/estados/queretaro/2015/02/01/automotriz-atraera-siete-francesas>
- Almanza L. (2015). Pymes energéticas brotan, Periódico EL ECONOMISTA [versión electrónica], publicado 02.02.2015. <http://eleconomista.com.mx/estados/queretaro/2015/02/02/pymes-energeticas-brotan>
- Becerril I. (2012). La mayoría de las Pymes no sobrevive más de 2 años, periódico EL FINANCIERO [versión electrónica] publicado 16.01.2012. [http://biblioteca.iiec.unam.mx/index.php?option=com\\_content&task=view&id=15629&Itemid=146](http://biblioteca.iiec.unam.mx/index.php?option=com_content&task=view&id=15629&Itemid=146)
- Instituto Nacional de Estadística y Geografía (INEGI). (2014). Esperanza de vida de los negocios. Boletín de prensa núm. 087/15, publicado 18.02.2015, Aguascalientes. [http://www.inegi.org.mx/saladeprensa/boletines/2015/especiales/especiales2015\\_02\\_38.pdf](http://www.inegi.org.mx/saladeprensa/boletines/2015/especiales/especiales2015_02_38.pdf)

## BIBLIOGRAFÍA

---

- [http://www.inegi.org.mx/prod\\_serv/contenidos/espanol/bvinegi/productos/nueva\\_estruc/702825068431.pdf](http://www.inegi.org.mx/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825068431.pdf)
- Leyva J. (2014). Empresas familiares, vulnerables por falta de profesionalismo; estudio. Periódico EL FINANCIERO [versión electrónica] publicado 17.05.2014. <http://www.elfinanciero.com.mx/economia/empresas-familiares-vulnerables-por-falta-de-profesionalismo-estudio.html>.
  - Marvella C. (2015). Industria del Bajío, con el acelerador a fondo. Periódico EL FINANCIERO [versión electrónica], publicado 11.02.2015. <http://www.elfinanciero.com.mx/bajio/industria-del-bajio-con-el-acelerador-a-fondo.html>.
  - Méndez E. & Torres A. (2013). Boom en El Bajío, nuevo polo industrial de México, Excélsior [versión electrónica], publicado 02.06.2013, <http://www.excelsior.com.mx/nacional/2013/06/02/902058#imagen-1>.
  - Oropeza P. (2015). Lintel desarrollará seis polígonos industriales en Guanajuato, periódico EL FINANCIERO [versión electrónica], publicado 05.02.2015, <http://www.elfinanciero.com.mx/bajio/lintel-desarrollara-seis-poligonos-industriales-en-guanajuato.html>.
  - Ruiz M. & Méndez V. (2015). Atrae el estado más parques industriales, am, publicado 26.01.2015, <http://www.am.com.mx/LEON/local/atrae-el-estado-mas-parques-industriales-176110.html>.
  - Vela R., Inversión de Japón en El Bajío supera los 5 mil millones de dólares. Monitor económico de Baja California, publicado 20.01.2015, <http://monitoreconomico.org/noticias/2015/jan/20/inversion-de-japon-en-el-bajio-supera-los-5-mil-millones-de-dolares/>

# ANEXOS

## ANEXO A. REMESAS POR AÑO.

Remesas en millones de dólares por entidad federativa 2003-2014. Fuente [www.banxico.org.mx](http://www.banxico.org.mx)

Título	2003	2008	2009	2010	2011	2012	2013	2014
TOTAL	15,138.69	25,144.99	21,306.33	21,303.88	22,802.97	22,438.32	22,302.75	23,647.28
Michoacán	1,787.47	2,448.86	2,132.28	2,144.50	2,245.06	2,209.36	2,048.72	2,244.03
Guanajuato	1,407.46	2,317.67	1,944.87	1,981.33	2,155.79	2,138.30	2,007.58	2,096.46
Jalisco	1,335.07	1,914.79	1,695.09	1,755.57	1,895.79	1,883.51	1,755.02	1,959.85
Distrito Federal	814.76	1,083.86	965.85	999.28	1,151.92	1,013.56	1,394.59	1,513.95
Estado de México	1,106.46	2,066.70	1,700.77	1,637.55	1,658.38	1,563.78	1,433.00	1,462.13
Puebla	853.88	1,615.67	1,374.88	1,371.22	1,469.64	1,403.25	1,334.56	1,338.65
Oaxaca	787.12	1,522.25	1,298.47	1,296.54	1,427.38	1,366.22	1,150.87	1,214.79
Guerrero	877.45	1,435.46	1,200.26	1,201.48	1,262.36	1,231.01	1,205.26	1,203.48
Veracruz	999.18	1,618.31	1,296.30	1,237.44	1,273.09	1,176.01	1,027.66	1,047.41
Tamaulipas	234.46	500.51	414.96	402.30	445.30	485.49	709.28	833.16
San Luis Potosí	403.50	760.75	626.76	629.47	700.80	738.70	707.04	769.88
Hidalgo	608.45	960.97	752.07	715.51	762.66	721.49	630.14	720.50
Zacatecas	402.37	681.55	573.30	581.71	625.45	654.45	633.80	700.17
Baja California	142.00	334.32	322.06	347.96	396.75	464.86	619.58	619.85
Nuevo León	189.17	323.77	292.99	283.98	308.92	340.03	597.15	614.47
Chihuahua	236.67	474.79	407.82	397.84	419.30	466.82	519.22	554.20
Morelos	373.16	622.59	548.12	554.86	586.82	561.27	514.51	527.66
Sinaloa	320.51	487.69	456.75	470.22	511.82	501.23	502.98	517.02
Chiapas	435.13	811.12	609.73	574.46	594.83	572.73	501.85	502.10
Durango	262.49	442.00	374.79	379.10	416.62	431.09	458.91	490.97
Querétaro	283.34	436.40	360.15	354.53	383.30	378.58	411.54	398.20
Coahuila	139.83	278.36	234.17	234.01	246.97	283.51	327.19	392.26
Nayarit	227.53	376.45	341.63	337.40	356.36	339.52	321.08	361.77
Sonora	128.33	310.96	278.70	292.02	326.92	326.76	341.16	337.11
Aguascalientes	260.27	332.33	282.21	293.92	306.32	332.67	305.60	323.88
Tlaxcala	149.19	305.21	258.86	258.52	274.55	253.24	217.07	218.86
Colima	103.74	184.66	164.80	171.52	183.82	180.17	183.30	216.77
Tabasco	85.89	156.02	114.41	111.34	111.73	111.26	117.17	130.96
Yucatán	60.41	136.12	109.94	112.69	117.81	119.19	125.43	129.32
Quintana Roo	52.85	97.35	85.57	86.80	92.08	93.32	100.84	105.04
Campeche	51.61	72.78	55.82	55.06	57.82	55.62	54.90	55.77
Baja California Sur	18.95	34.70	31.92	33.75	36.65	41.36	45.76	46.58

## ANEXOS

### ANEXO B. ENCUESTA NACIONAL DE OCUPACIÓN Y EMPLEO.

Fuente: INEGI.

Periodo	Población económicamente activa (PEA) Ocupada a/	Población económicamente activa (PEA)	Tasas de desocupación Michoacán de Ocampo a/	Tasas de desocupación Guanajuato a/ (Porcentaje)	Tasas de desocupación Querétaro	Aguascalientes a/ (Porcentaje)	San Luis Potosí a/ (Porcentaje)
2010/01	45,524,339.00	2,544,935.00	3.14	5.54	6.51	7.43	4.26
2010/02	46,597,624.00	2,535,508.00	4.04	5.83	6.95	6.41	4.21
2010/03	46,452,588.00	2,737,444.00	3.69	6.46	6.28	6.44	3.74
2010/04	45,911,934.00	2,566,784.00	3.63	5.75	7.12	6.93	4.37
2011/01	46,005,815.00	2,499,353.00	3.06	6.39	5.81	6.05	4.16
2011/02	46,891,586.00	2,590,526.00	2.93	6.01	6.53	6.20	4.01
2011/03	47,350,681.00	2,776,351.00	2.87	6.02	5.66	8.16	4.84
2011/04	48,307,467.00	2,465,029.00	2.88	5.12	5.25	6.32	3.35
2012/01	47,726,265.00	2,466,577.00	3.27	6.77	5.18	6.58	2.95
2012/02	49,003,380.00	2,473,798.00	3.13	5.89	4.12	5.95	3.14
2012/03	49,275,020.00	2,652,030.00	4.31	6.03	4.98	6.26	3.82
2012/04	48,822,271.00	2,495,728.00	4.11	5.59	5.78	6.01	3.10
2013/01	48,358,255.00	2,488,987.00	3.69	6.23	5.22	6.05	3.34
2013/02	49,296,229.00	2,599,636.00	4.27	5.87	5.69	4.73	3.78
2013/03	49,309,167.00	2,725,186.00	4.79	5.61	4.89	5.70	3.53
2013/04	49,945,599.00	2,425,287.00	4.37	5.16	6.35	5.87	3.13
2014/01	49,080,947.00	2,478,071.00	3.30	5.26	4.87	6.39	2.77
2014/02	49,301,557.00	2,535,195.00	3.21	4.73	5.93	5.83	3.20
2014/03	49,455,344.00	2,736,699.00	3.42	5.11	6.88	6.26	3.49
2014/04	49,823,798.00	2,284,602.00	3.99	4.20	5.45	5.48	2.84

## ANEXOC. CÓDIGO EN R.

### Preparación del Insumo

```
>datos<-read.table("datos.txt",header=T)
>attach(datos)
>View(datos)
>library(rms)
```

### Código usado en las Gráficas de supervivencia KM en R

```
>plot(survfit(Surv(Años,NoRenta)~Fam),col=c(1,2), main="Grafica de
supervivencia de KM por Familia")
```

### Prueba de puntajes y razón de verosimilitud

```
>survdif(Surv(Años,NoRenta)~Fam)
```

### Regresión de Cox

```
>coxph(Surv(Años,NoRenta)~Fam+Gen+Cap+Cer+Num+Eda+Fun)
>summary(coxph(Surv(Años,NoRenta)~Fam+Gen+Cap+Cer+Num+Eda+Fun))
>plot(survfit(modelo))
```

### Comparación de las gráficas de supervivencia.

```
>summary(survfit(Surv(Años,NoRenta)~1)) #Valores KM
>summary(survfit(modelo)) #Ver valores de Cox
>plot(survfit(coxph(Surv(Años,NoRenta)~Eda+For+Cer+Fun)),col=1,main="Superv
ivencia Cox vs Supervivencia Kaplan Meier",xlab="Tiempo de
fallo",ylab="Probabilidad de supervivencia",conf.int=FALSE,lty=1)
>lines(survfit(Surv(Años,NoRenta)~1),col=2,lty=2,conf.int=FALSE)
>legend(30,0.9,legend=c("Estimador KM", "Ajuste por Cox"),lty=c(2,1),col=c(2,1))
```

## ANEXOS

### ANEXO D. DATOS DE LA ENCUESTA 2011-2012.

Primeros 20 registros de la encuesta realizada a pequeñas y medianas empresas de la región del bajo.

Registro	Gen	For	Fun	Fam	Sec	Eda	Cer	Cap	NoRenta	Años	Num
1	1	2	0	0	3	50	0	0	0	16	90
2	1	2	1	0	4	50	0	0	0	1	10
3	1	2	1	1	1	38	0	0	0	10	52
4	1	1	1	1	1	39	0	0	0	9	50
5	1	2	1	1	4	46	0	0	1	11	17
6	1	2	0	1	1	38	0	0	0	5	18
7	1	1	0	1	4	48	0	0	0	17	52
8	1	2	1	1	4	52	1	0	0	2	100
9	1	2	0	0	4	31	1	0	0	6	28
10	1	1	1	1	4	80	0	0	0	14	7
11	1	2	1	1	1	46	0	0	0	9	10
12	1	1	1	1	4	38	0	0	0	3	10
13	1	2	0	1	1	35	0	0	1	8	12
14	1	1	1	1	1	43	0	0	1	15	5
15	1	1	0	1	1	48	0	0	0	42	5
16	1	2	1	1	1	55	1	1	1	30	6
17	2	1	1	1	1	38	0	1	0	10	12
18	1	1	0	1	3	48	0	0	0	42	5
19	2	1	1	1	3	38	0	1	0	10	12
20	1	2	1	1	4	50	1	0	0	30	7