

MODELO PARA LA MEDICIÓN DEL RIESGO DE INSOLVENCIA EMPRESARIAL: UN CASO DE ESTUDIO EN LAS EMPRESAS PYME DE COLOMBIA

Investigadores principales:

JAIME ESPINOSA PEÑA
Magíster en Finanzas y Ciencias Computacionales

EDUARDO ARTURO CRUZ TREJOS
Magíster en Administración Económica y Financiera

SERGIO ARISTIZÁBAL HERNÁNDEZ
Ingeniero Financiero

Auxiliares de investigación:

VICTORIA ANDREA AGUIRRE MORALES
FAYSURI FLÓREZ DÍAZ
JOHANA LÓPEZ MONTOYA
DANITZA MARIANA MARTÍNEZ ARMIJO
VANESSA NARVÁEZ OSORIO
JESSICA ALEJANDRA OSORIO CASTRO
SENDY DAYANA PARRA BECERRA
MANUELA VILLEGAS BECERRA

UNIVERSIDAD LIBRE SECCIONAL PEREIRA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA FINANCIERA
2013

MODELO PARA LA MEDICIÓN DEL RIESGO DE INSOLVENCIA EMPRESARIAL: UN CASO DE ESTUDIO EN LAS EMPRESAS PYME DE COLOMBIA

Investigadores principales:

JAIME ESPINOSA PEÑA
Magíster en Finanzas y Ciencias Computacionales

EDUARDO ARTURO CRUZ TREJOS
Magíster en Administración Económica y Financiera

SERGIO ARISTIZÁBAL HERNÁNDEZ
Ingeniero Financiero

Auxiliares de investigación:

VICTORIA ANDREA AGUIRRE MORALES
FAYSURI FLÓREZ DÍAZ
JOHANA LÓPEZ MONTOYA
DANITZA MARIANA MARTÍNEZ ARMIJO
VANESSA NARVÁEZ OSORIO
JESSICA ALEJANDRA OSORIO CASTRO
SENDY DAYANA PARRA BECERRA
MANUELA VILLEGAS BECERRA

UNIVERSIDAD LIBRE SECCIONAL PEREIRA
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA FINANCIERA
2013

Nota de aceptación:

Firma del Presidente del Jurado

Firma del Jurado

Pereira, 15 de abril de 2013

RESUMEN

Este informe relata de manera concisa los resultados obtenidos en el desarrollo del proyecto: "Modelo para la medición del riesgo de insolvencia para empresas PYME en Colombia"; así como también se muestra la metodología empleada para el desarrollo y es comparada con el diseño metodológico planificado con antelación al inicio de actividades investigativas.

La primera parte del documento muestra la descripción problemática en la medición del riesgo de insolvencia y su importancia para la toma de decisiones en el panorama nacional por medio de la citación de diferentes experiencias o estado del arte en el país en cuanto a la medición del riesgo de insolvencia. Seguidamente se hace la fundamentación teórica de los conceptos y herramientas estadísticas y jurídicas para la estructuración de modelos para la medición del riesgo de insolvencia. La tercera parte del documento muestra en forma detallada la metodología empleada en cada una de las fases ejecutorias del proyecto de investigación, empezando por la caracterización de la base de datos de las empresas del sector real y terminando con la explicación de la construcción del modelo de medición del riesgo de crédito por medio de la regresión logística. Inmediatamente son mostrados, de manera sintetizada, los resultados de la caracterización de la base de datos de las empresas; la caracterización de las variables más importantes para el comportamiento de cada uno de los subsectores estudiados; el ajuste de variables a distribuciones teóricas continuas; y, finalmente, la estructuración y las tablas de análisis del modelo de regresión logística. Finalmente, se plantean algunas conclusiones por la ejecución del proyecto y las referencias bibliográficas que sustentan todo el desarrollo del proyecto de investigación.

Paralelamente se estuvo trabajando en dos actividades: la creación de un complemento para *Excel* que realiza análisis financiero de manera automatizada y el tema central del proyecto que era la medición del riesgo de insolvencia. La herramienta surge por la necesidad de contar con un sistema de información que permitiera ver la información de los estados financieros de las empresas registradas en las superintendencias y cámaras de comercio de Colombia, así como permitir el análisis financiero de estos mismos.

ABSTRACT

This report describes concisely the results of the project: "Model for measuring the risk of insolvency for SME companies in Colombia", and also shows the methodology used for development and is compared with study design planned prior to the commencement of research activities.

The first part of the paper shows the description problem in measuring the risk of insolvency and its importance for decision making in the national scene through the citation of different experiences or state of art in the country in terms of risk measurement insolvency. Subsequently the theoretical concepts and legal and statistical tools for structuring models for measuring the risk of insolvency. The third part of the paper shows in detail the methodology used in each of the executory phase of the research project, starting with the characterization of the database of actual companies and ending with the explanation of the construction of the model credit risk measurement by means of logistic regression. Immediately are shown, so synthesized, the results of the characterization of database companies, the characterization of the most important variables for the behavior of each of the subsectors studied; adjustment to theoretical distributions continuous variables, and Finally, structuring and analysis tables logistic regression model. Finally, it presents some conclusions for project implementation and references that support the entire development of the research project.

In parallel, it worked on two activities: the creation of an Excel add-in for performing automated financial analysis and the focus of the project was to measure the risk of insolvency. The tool arises from the need for an information system that would see the financial statements of companies registered with the superintendents and chambers of commerce of Colombia, as well as allowing the financial analysis of the same.

CONTENIDO

1. INTRODUCCIÓN	
1.1 DESCRIPCIÓN DEL PROBLEMA Y ESTADO DEL ARTE	11
1.1.1 Modelación del riesgo crediticio como elemento fundamental en el cálculo de la pérdida esperada en una institución financiera	13
1.1.2 Aplicación del modelo Altman a la inversión de reservas del Banco Central de Bolivia	14
1.1.3 Modelos para medir el riesgo de crédito de la banca	15
1.1.4 Aplicación del análisis discriminante y regresión logística en el estudio de la morosidad en las entidades financieras	17
1.1.5 Predicción del fracaso empresarial: una revisión	18
1.1.6 Un mapa de riesgo de crédito para el sistema financiero colombiano	20
1.1.7 Determinantes de fragilidad en las empresas colombianas	21
1.2 JUSTIFICACIÓN	23
1.3 OBJETIVOS DE LA INVESTIGACIÓN	23
1.3.1 Objetivo general	23
1.3.2 Objetivos específicos	24
2. MARCO REFERENCIAL	25
2.1 MARCO TEÓRICO	25
2.1.1 Regresión logística	25
2.1.2 Riesgo de crédito	28
2.1.3 Empresas PYME de Colombia	34
2.2 MARCO CONCEPTUAL	36
2.3 MARCO LEGAL Y NORMATIVO	39

3. METODOLOGÍA	40
3.1 TIPO DE INVESTIGACIÓN	40
3.2 FASES DE DESARROLLO	40
3.2.1 Fase I: Caracterización de la base de datos empresarial de la Superintendencia de Sociedades de Colombia 2003-2011	40
3.2.1.1 Construcción de la base de datos	41
3.2.1.2 Afinación de la base de datos	47
3.2.1.3 Relación de las tabla balance, resultados y flujo de efectivo	51
3.2.2 Fase II: Caracterización cuantitativa de las empresas a estudiar	54
3.2.3 Fase III: Ajuste de distribución de variables	59
3.2.4 Fase IV: Desarrollo del modelo de riesgo de crédito con la regresión logística	60
3.2.4.1 Definición preliminar de la muestra	60
3.2.4.2 Determinación del grupo de entrenamiento	61
3.2.4.3 Hipótesis	62
3.2.4.4 Selección de variables	62
4. RESULTADOS DE LA INVESTIGACIÓN	64
4.1 CARACTERIZACIÓN DE LA FUERZA EMPRESARIAL MÁS REPRESENTATIVAS DE COLOMBIA POR SECTOR, DEPARTAMENTO Y REGIÓN	64
4.2 CARACTERIZACIÓN CUANTITATIVAS DE LAS EMPRESAS ESTUDIADAS DENTRO DE SUS RESPECTIVOS SUBSECTORES	67
4.3 AJUSTE DE DISTRIBUCIÓN DE VARIABLES	69
4.4 MODELO DE RIESGO DE INSOLVENCIA A TRAVÉS DE LA REGRESIÓN LOGÍSTICA	70
4.4.1 Modelo de regresión logística 01	70

5. CONCLUSIONES

6. REFERENCIAS BIBLIOGRÁFICAS

LISTA DE FIGURAS

Figura 3.1 Cómo eliminar columnas en Excel	51
Figura 3.2 Definición de nombre a un rango en una hoja de cálculo Excel	53
Figura 3.3 Definición de nombre a un rango en una hoja de cálculo Excel	
Parámetros	54
Figura 3.4 Rótulos de la tabla definitiva Flujo de Efectivo	54
Figura 3.5 Cuentas del flujo de efectivo unidas a balance general y estado de resultados en la hoja de cálculo Excel	55
Figura 3.6 Mapa del proceso metodológico de la fase II	58

LISTA DE CUADROS

Cuadro 1.1 Movimiento de sociedades por departamento y nivel económico año 2009	14
Cuadro 1.2 Comparativo entre ventajas de modelos para la medición del riesgo de crédito	18
Cuadro 1.3 Comparativo entre desventajas de modelos para la medición del riesgo de crédito	19
Cuadro 1.4 Capacidades para enfrentar el fracaso	22
Cuadro 1.5 Esquema general del mapa de riesgo	24
Cuadro 2.1 Actividades económicas estudiadas	36
Cuadro 3.1 Lista grupos contenedores de la base de datos construida	45
Cuadro 3.2 Historial del salario mínimo en Colombia desde 1995 hasta 2012	49
Cuadro 3.3 Cuentas definitivas del Balance General	51
Cuadro 3.4 Cuentas definitivas del Estado de Resultados	52
Cuadro 3.5 Cuentas definitivas del Flujo de Efectivo	53
Cuadro 3.6 Tabla de regresión generada por el comando en la hoja de cálculo Excel	60
Cuadro 3.7 Proporción de empresas en reorganización según tamaño por activos	63
Cuadro 3.8 Grupo de entrenamiento compuesto por empresas en reorganización y por empresas aparentemente sanas	64
Cuadro 3.9 Codificación de variables para aplicar la Regresión Logística	66
Cuadro 4.1 Participación de empresas por sectores en los departamentos desde 2003 hasta 2011	67
Cuadro 4.2 Participación de empresas por sectores en las regiones desde 2003 hasta 2011	68

Cuadro 4.3 Participación de empresas por sectores, regiones y departamentos desde 2003 hasta 2011	69
Cuadro 4.4 Número de variables que quedaron en la ecuación final de regresión para cada subsector desde el año 1995 hasta el año 2010	69
Cuadro 4.5 Ajuste del comportamiento de las variables de los sectores estudiados	71
Cuadro 4.6 Resumen del procesamiento de los casos Regresión Logística modelo 01	67
Cuadro 4.7 Codificación de la variable dependiente modelo Regresión Logística modelo 01	73
Cuadro 4.8 Codificaciones de variables categóricas Regresión Logística modelo 01	73
Cuadro 4.9 Tabla de clasificación sin variables independientes Regresión Logística modelo 01	74
Cuadro 4.10 Constante en la ecuación Regresión Logística modelo 01	75
Cuadro 4.11 Variables que no están en la ecuación Regresión Logística modelo 01	75
Cuadro 4.12 Pruebas ómnibus sobre los coeficientes del modelo 01	76
Cuadro 4.13 Resumen del modelo 01	76
Cuadro 4.14 Tabla de clasificación Regresión Logística modelo 01	77
Cuadro 4.15 Variables en la ecuación Regresión Logística del modelo 01	77
Cuadro 4.16 Historial de iteraciones con variables más significativas en el modelo 02	73
Cuadro 4.17 Prueba ómnibus sobre los coeficientes del modelo 02	78
Cuadro 4.18 Bondad de ajuste de tres estadísticos al modelo 02	79
Cuadro 4.19 Tabla de contingencia para prueba de Hosmer y Lemeshow modelo 02	79
Cuadro 4.20 Estadísticos del modelo 02, factores de riesgo y/o protección	79

1. INTRODUCCIÓN

El riesgo de crédito es tratado en tres perspectivas: la cualitativa, cuantitativa y la mixta. Los expertos en riesgo trabajan de una manera mixta ya que tanto el aspecto cualitativo como el cuantitativo no son mutuamente excluyentes, por el contrario, el paradigma de la confrontación entre estas dos formas de adquisición de conocimiento requieren de una ayuda mutua para resultados más efectivos. Cuando un banco presta dinero a sus clientes, les exige una serie de requisitos para poder aspirar a recibir el dinero que solicitan; tales requisitos son tanto de características cualitativas y cuantitativas. Normalmente son más importantes los requisitos cualitativos como experiencia crediticia, referencias personales y comerciales, calificación en las centrales de riesgo, actividades económicas, estrato social, etc. Este modelo cualitativo es el que manejan las entidades crediticias de los países del tercer mundo.

1.1 DESCRIPCIÓN DEL PROBLEMA Y ESTADO DEL ARTE

La predicción de insolvencia en las empresas es un tema que no ha sido estudiado ampliamente en Colombia y que requiere un replanteamiento acerca de su importancia debido a que el fracaso empresarial es un problema que afecta negativamente a los emprendedores colombianos.

Los modelos utilizados actualmente en Colombia para la predicción de insolvencia y la determinación de eficiencia de las empresas son modelos que se aplican hace más de 30 años y que merecen ser modificados y complementados con las nuevas herramientas que brinda la informática, las técnicas de análisis estocástico y las teorías del aprendizaje artificial, como es el caso de las Redes Neuronales Artificiales. En el cuadro 1.1 se presentan estadísticas resumidas por el autor acerca del movimiento de sociedades en el año 2009.

La constitución y liquidación de empresas afecta directamente el desarrollo de la economía del país. Del cuadro 1.1 se puede deducir que, por ejemplo, si en un departamento hay una disminución en el número de empresas constituidas de un año para otro, no necesariamente habrá una disminución en el capital total del valor total que suman todas las compañías creadas, debido a la capitalización que se aporta según el tamaño de las empresas. Es menester aprovechar al máximo las estadísticas, ya que con la utilización conjunta del modelo que se pretende construir, las instituciones financieras y los datos históricos relacionados directamente con la insolvencia, se puede lograr un impacto considerable en la teoría económica colombiana.

Las instituciones financieras más importantes del país son las encargadas, principalmente, de generar investigaciones y crear modelos para el funcionamiento del sistema financiero. Estos modelos, normalmente, son traídos de otros países que poseen condiciones económicas diferentes y más avanzadas.

Cuadro 1.1 Movimiento de sociedades por departamento y nivel económico año 2009

NIVEL ECONÓMIC O	DEPARTAMENTOS	SOCIEDADES CONSTITUIDAS				SOCIEDADES LIQUIDADAS			
		CANTI DAD	VARIAC IÓN	PESOS	VARIAC IÓN	CANTI DAD	VARIA CIÓN	PESOS	VARIA CIÓN
Industrializ ados	Cundinamarca	18.014	-0,40%	2.960,5 4	51,73%	3.566	1,77%	756.000, 00	67,87 %
	Atlántico	1.439	- 34,60%	102.56 6,00	9,04%	392	17,40 %	35.416,0 0	61,94 %
	Antioquía	3.695	20,75%	290.98 9,00	43,60%	990	18,71 %	123.946, 00	16,90 %
	Valle de Cauca	2.757	0,99%	129.90 5,00	- 27,30%	856	8,90%	388.748, 00	303,7 0%
	Santa fé de Bogotá	17.654	-0,40%	2.664,4 9	51,73%	3.209	1,77%	680.400, 00	67,87 %
De crecimien to	Risaralda	467	-9,64%	38.324, 00	-9,02%	156	-59,90%	6.790 00	- 59,90%
	Caldas	364	13,40 %	25.664, 81	71,70 %	127	-11,81%	8.301 07	- 86,18%
	Bolivar	-	-	-	-	-	-	-	-
	Santander	578	- 31,03 %	22.942, 52	- 53,68 %	210	32,91%	17.36 1,00	184,98 %
Basadas en el sector primario	Tolima	375	- 24,55 %	18.461, 00	16,40 %	141	-22,95%	5.974 00	- 10,37%
	Huila	435	- 15,04 %	19.439, 00	0,80%	95	-28,57%	6.373 00	- 56,44%
	Cesar	271	27,23 %	7.640,0 0	65,73 %	48	152,63%	5.003 00	764,08 %
	Quindío	235	20,51 %	14.508, 00	31,52 %	91	8,33%	8.510 00	- 19,79%
	Córdoba	243		12.820, 00	- 59,60 %	103		8.410 00	- 77,40%
	Norte de Santander	939	-1,78%	59.204, 20	20,22 %	154	-8,88%	7.285 00	66,67%
	Boyacá	429	- 33,18 %	52.581, 00	140,98 %	99	-11,61%	20.96 3,50	114,94 %
	Nariño	202	- 14,04 %	21.543, 00	173,60 %	49	-12,50%	3.291 00	68,34%

Fuente: DANE adaptado por los autores. Cifras en millones de pesos colombianos.

Cuadro 1.1 Continuación

NIVEL ECONÓMICO	DEPARTAMENTOS	SOCIEDADES CONSTITUIDAS				SOCIEDADES LIQUIDADAS			
		CANTIDAD	VARIACIÓN	PESOS	VARIACIÓN	CANTIDAD	VARIACIÓN	PESOS	VARIACIÓN
Fundamentadas en el agro y la minería	La Guajira	-	-	-	-	-	-	-	-
	Magdalena	366	-30,70%	9.297,00	-49,60%	82	7,32%	2.705,00	-49,50%
	Sucre	162		3.604,30	-46,40%	43		13.079,20	-6,90%
	Meta	367	-36,83%	19.151,00	-43,18%	99	-17,50%	3.755,00	-12,25%
	Arauca	34	-59,04%	3.056,60	-35,58%	16	60,00%	1.026,70	13,26%
	Cauca	177	-25,63%	12.207,00	-64,82%	51	-5,56%	2.048,00	-91,30%
	Casanare	321	-26,21%	21.352,00	37,96%	58	31,82%	729	-1,62%
	Caquetá	52	-35,80%	1.687,00	-16,11%	20	25,00%	1.698,00	28,64%
L789Bajo nivel de desarrollo	Chocó	65	54,76%	1.664,00	36,39%	35	218,18%	923	610,00%
	Amazonas	32	14,29%	2.156,00	35,60%	10	-16,67%	652	-8,56%
Totales		49.673	-9,08%	896.387	14,14%	10.700	15,55%	2.109.387	69,96%

Por ello, se deben crear modelos propios del país para evitar falencias y decisiones inoportunas de cualquier agente del sistema financiero. A continuación se detallan algunos estudios recientes en el tema de la medición del riesgo de insolvencia empresarial o riesgo de crédito desde la perspectiva de las entidades financieras.

1.1.1 Modelación del riesgo crediticio como elemento fundamental en el cálculo de la pérdida esperada en una institución financiera

“Para el desarrollo de la propuesta se identifican las variables que permitan realizar el trabajo empírico, utilizando diferentes metodologías para la cuantificación formal de la probabilidad de incumplimiento y se realiza un caso de aplicación para el sistema financiero colombiano” (Tamara, Aristizábal & Velasquez, 2010).

Inicialmente es importante determinar el concepto y las características del riesgo, planteando además un paralelo entre los acuerdos de Basilea y los de la superintendencia financiera de Colombia.

El riesgo es la probabilidad de pérdida que asume un agente económico como consecuencia del incumplimiento de las obligaciones contractuales. Para la modelación del riesgo crediticio se hace énfasis en la metodología sobre la pérdida esperada propuesta por Basilea, en donde se resume a grosso modo (Tamara, Aristizábal y Velázquez, 2010):

- La probabilidad de incumplimiento: Probabilidad de que en un lapso de 12 meses estos incurran en incumplimiento.
- La exposición al momento de incumplimiento: El valor expuesto del activo entendida como el saldo de la obligación al momento del cálculo de la pérdida esperada y
- La pérdida dado el incumplimiento: Es el deterioro económico en el que incurre la entidad en caso de que se materialice alguna de las situaciones de incumplimiento. Las cuales difieren en ciertos aspectos del incumplimiento planteado por la superintendencia financiera.

Para un diagnóstico más acertado acerca la medición de la pérdida esperada, los autores emplearon modelos como Logit, Probit, Análisis Discriminante, Diagramas de Árbol y el Cálculo de Indicadores, con el fin de realizar comparaciones y concluir al respecto. Para ello se empleó una base de datos de 1500 registros de las cuales 300 fueron tomadas como muestra. A partir de los cálculos se pudo determinar a qué clase de riesgo y en que categoría puede estar una persona natural que realiza una actividad económica determinada.

Los resultados arrojados en la modelación de riesgo crediticio como elemento fundamental en el cálculo de la pérdida esperada en una institución financiera llevaron a la conclusión de que los modelos utilizados fueron significativos y las variables más influyentes en el incumplimiento fueron los ingresos, activos, pasivos, edad, patrimonio y actividad económica de cada uno de los agentes.

1.1.2 Aplicación del modelo Altman a la inversión de reservas del Banco Central de Bolivia

“La crisis financiera desatada durante los años 2007-2009, supuso la quiebra de una de las entidades más importantes de los mercados financieros como Lehman Brothers, poniendo de manifiesto la necesidad de mejorar el monitoreo del riesgo crediticio de las entidades financieras en general” (Ballón, 2010).

El modelo de Altman es un sencillo método de análisis de la fortaleza financiera de una empresa, puede llegar a predecir la probabilidad de quiebra establecida por una combinación de indicadores financieros. La aplicación del modelo surge debido al nivel de reservas internacionales del Banco Central de Bolivia (BCB) ya que sobrepasaba los USD 8500 millones por lo que se hizo necesario contar con el mayor número de herramientas posibles para monitorear y controlar el riesgo crediticio.

Inicialmente para el análisis se empleó una muestra de 47 bancos para las gestiones 2008 y 2009 con un total de 94 observaciones; posteriormente se realizó un análisis discriminante con una variante Logit en la ecuación de compartimiento; para este objetivo los grupos seleccionados fueron los correspondientes a bancos con una calificación mínima de A y bancos fuera de lineamientos de inversión con una calificación A-. Además de un análisis discriminante, se emplearon 4 regresiones logísticas a los datos donde se midió el nivel de significancia individual de las variables y conjunta, logrando clasificar a cada uno de los bancos, entre los que cumplían con los lineamientos de inversión y los que no. El análisis de ambos modelos, análisis discriminante y regresión múltiple, sugiere un buen desempeño para los bancos debido al porcentaje significativo de ajuste del modelo.

Es necesario realizar este tipo de estudios ya que se reduce la incertidumbre en torno al concepto de que empresa puede llegar o está más propensa a la quiebra y a partir de esto tomar medidas correctivas aprovechando las ventajas del análisis tradicional de los indicadores financieros, combinándolos con las técnicas estadísticas más sofisticadas.

1.1.3 Modelos para medir el riesgo de crédito de la banca

“Este trabajo describe los principales modelos de determinación de riesgos de crédito de la banca, a fin de comparar y dar a conocer su utilidad en la administración del riesgo de crédito bancario” (Saavedra & Saavedra, 2009).

El riesgo de crédito es la probabilidad de que a su vencimiento, una entidad no haga frente, en parte o en su totalidad, a su obligación de devolver una deuda o rendimiento, acordado sobre un instrumento financiero, debido a quiebra, iliquidez o alguna otra razón. Una medida usual para calcular el riesgo específico de una empresa es a través de las calificaciones (rating) esta medida es útil para clasificar compañías según su riesgo de crédito.

Los principales factores que deben tomarse en cuenta para decidir si se otorga o no un crédito son:

- Capacidad: Consiste en evaluar la habilidad y experiencia en los negocios que tenga la persona o empresa.
- Capital: Analiza la situación financiera.
- Colateral: Elementos de que dispone el acreditado para garantizar el cumplimiento del pago en el crédito.
- Carácter: Cualidades de honorabilidad y solvencia moral que tiene el deudor para responder al crédito.

- Condiciones: Factores exógenos que pueden afectar la marcha del negocio del acreditado.

Cuadro 1.2 Comparativo entre ventajas de modelos para la medición del riesgo de crédito

Ventajas de los modelos

Tradicional-les o estáti-cos	Sistemas expertos	<ul style="list-style-type: none"> • Facilidad de obtención de datos • Facilidad de aplicación
	Sistemas de califi-cación	<ul style="list-style-type: none"> • Indicadores de los es-tados financieros
Modernos o dinámicos	KMV	<ul style="list-style-type: none"> • En la metodología se encuentra implícita la EDF • Toma en cuenta la di-versificación requerida en los portafolios de deuda • Probabilidad de in-cumplimiento individu-al
	CyrCE	<ul style="list-style-type: none"> • Uso en mercados emergentes • Se tiene información parametrizada en el mercado • Se obtienen límites de créditos individuales • Puede evaluar seg-mentos particulares de cartera

Fuente: Saavedra & Saavedra (2010)

Los modelos modernos de administración del riesgo intentan registrar la alta volatilidad a la que están sujetos los valores y emplean técnicas más sofisticadas para su determinación, mientras los modelos tradicionales emplean solamente variables contables y microeconómicas son menos precisas para predecir la insolvencia, sin embargo se consideran útiles en un entorno estable y con montos no muy relevantes.

La aplicación de un modelo de riesgo permite: determinar el riesgo, limitar y tener certidumbre sobre el nivel de riesgo que puede impactar el capital de la institución y tomar decisiones sobre un umbral de riesgo para deshacer posiciones y evitar pérdidas.

Cuadro 1.3 Comparativo entre desventajas de modelos para la medición del riesgo de crédito

Desventajas de los modelos		
Tradicionales o estáticos	Sistemas expertos	<ul style="list-style-type: none"> • Subjetividad • Consistencia
	Sistemas de calificación	<ul style="list-style-type: none"> • Alta correlación entre las variables • Datos históricos • Técnicas estadísticas avanzadas • No están integrados a la gestión de riesgo • No separa por sectores • Las puntuaciones no se correlacionan con la probabilidad de impago
Modernos o dinámicos	KMV	<ul style="list-style-type: none"> • Los precios del mercado no representan el valor real de las empresas • Enfoque micro-económico • Se necesita una gran base de datos
	CyRCE	<ul style="list-style-type: none"> • Se requiere conocimiento de técnicas estadísticas avanzadas • La identificación de segmentos no siempre resulta sencillo

Fuente: Saavedra & Saavedra (2010)

1.1.4 Aplicación del análisis discriminante y regresión logística en el estudio de la morosidad en las entidades financieras

“Aplicación de dos técnicas estadística a una muestra de clientes de entidades financieras con el fin de valorar el riesgo de crédito. A su vez, se determina el método que permite discriminar mejor entre clientes morosos y no morosos” (Mures, García & Vallejo, 2005).

El objetivo fundamental del trabajo se centra en el análisis y evaluación de la morosidad, como forma de manifestación del riesgo de crédito, mediante la determinación de aquellos factores con mayor influencia en dicho fenómeno. Esto se realiza a través de un estudio comparativo de dos técnicas estadísticas adecuadas para este análisis, como son el análisis discriminante, tradicionalmente utilizado por las entidades financieras en sus *credit scoring*, y la regresión logística, que presenta ciertas ventajas respecto al anterior, aplicándose ambas a una muestra de clientes de tres tipos de entidades financieras.

Respecto a la información que las entidades recogen a la hora de estudiar la concesión de un crédito, destacan tres elementos que resultan en dicha concesión: por un lado, la personalidad del solicitante, referente al carácter, integridad y naturaleza del individuo, que determina si pagará, o, por el contrario, existirá riesgo de crédito; por otro, la capacidad y el capital, relativos a la disponibilidad de ingresos para pagar y a la situación financiera y patrimonial del cliente, respectivamente, que establecen el límite de crédito que puede concederse, esto es, si el cliente podrá pagar y cuánto.

Variables relativas a la operación de crédito o préstamo, tales como el importe solicitado, la duración o plazo de devolución del crédito, el destino de la inversión, la procedencia de la operación y las garantías aportadas.

Variables relativas al perfil del solicitante, que incluyen aquéllas referidas a sus características personales, como edad, estado civil o cargas familiares; las que reflejan su situación profesional, como categoría laboral, ingresos mensuales o antigüedad en el trabajo, junto a otras que definen su relación con la entidad, como compensaciones ofrecidas, número de operaciones activas e importe total de las mismas o si tiene otras operaciones en las que se ha producido impago.

De acuerdo con todo lo expuesto, siendo el objetivo del estudio encontrar un modelo con el menor número posible de variables y la máxima capacidad predictiva, se ha considerado, tanto en la aplicación del análisis discriminante como de la regresión logística, un procedimiento de selección por pasos sucesivos o *stepwise selection*, cuyos resultados se muestran a continuación.

Los dos modelos estimados mediante análisis discriminante y regresión logística son significativos, como lo muestran los estadísticos calculados para valorar la significación de cada modelo. Además, poseen una elevada capacidad predictiva, sobre todo entre los clientes no morosos, ya que la primera técnica clasifica correctamente a la totalidad de tales individuos, alcanzando, la segunda, una tasa de aciertos del 98'1%. Dichos porcentajes son menores en cuanto a los individuos morosos, siendo los porcentajes de clasificación correcta del 88'9% para el análisis discriminante y del 94'4% para la regresión logit, lo que resulta en una tasa de aciertos del 97'1% para ambos modelos.

La conclusión global que se extrae es que tanto la técnica de análisis discriminante como la de regresión logística son adecuadas para el estudio y predicción de la morosidad, consiguiéndose mediante ambas técnicas una elevada eficacia predictiva. Esto es así porque las condiciones exigidas para la aplicación del análisis discriminante, en cuanto a la igualdad de varianzas en los dos grupos definidos por la variable dependiente y la normalidad de las variables independientes en la población de la que procede la muestra, se verifican en la muestra de clientes objeto del estudio. Por ello, los resultados obtenidos mediante esa técnica son similares a los conseguidos mediante regresión logística, que es el método alternativo aplicado para el análisis de variables cualitativas cuando se incumplen tales restricciones.

1.1.5 Predicción del fracaso empresarial: una revisión

Continuamente se evidencia el problema de la crisis financiera que afecta a un sin número de personas y empresas que se ven en la obligación de cerrar o buscar soluciones arriesgadas generando así un colapso en las actividades que realiza con el objetivo de obtener fondos para seguir realizando sus operaciones.

Es este el punto de partida debido a que las empresas para salir de la crisis requieren de créditos para solventarse y seguir con sus operaciones, es allí donde se presenta el riesgo de crédito y el aumento de los derechos de cobro impagados.

Una empresa fracasa o es fallida cuando:

- Deja de pagar una o varias de sus deudas, aunque sea de forma temporal.
- Reúne las condiciones previstas en la normativa vigente sobre quiebra (o procedimientos similares)
- Su situación patrimonial muestra un valor reducido en los activos o una escasez de tesorería que pueden desencadenar el fracaso.

“Equivocarse en la previsión del fracaso empresarial supone un riesgo y, por tanto, un coste. Este coste vendrá normalmente de clasificar una empresa con dificultades como sana (error tipo I), aunque también puede tener coste clasificar una empresa sana como empresa en dificultades (error tipo II), eso dependerá del decisor y su modelo de negocio (Jones, 1987).

Por ejemplo, para el caso de la banca, según un estudio empírico realizado por Altman, Haldeman y Narayanan (1977), cometer el error tipo I supondrá que el banco pierda el equivalente al 70% del valor del préstamo, a causa de la insolvencia del deudor; mientras que la pérdida de cometer un error tipo II es, únicamente, el coste de oportunidad. Esto es, la posible menor rentabilidad que se obtenga por no prestar a la empresa sana considerada fallida. Para este caso los autores estiman un coste entre el 2% y el 4% del valor del préstamo.

Para una empresa de calificación del riesgo que ofrezca cobertura para casos de insolvencia, el error tipo I supone tener que indemnizar a la empresa asegurada (aquí no hay garantías del insolvente) por lo que su pérdida es del 100%, en cambio un error tipo II sólo supone una pérdida igual a la prima que ha dejado de percibir al no asegurar a una empresa sana”. (Tascón & Castaño, 1998).

Las capacidades necesarias de una persona o empresa para enfrentarse al fracaso y combatirlo deben ser de: Conocimientos, Habilidades y Actitudes:

Cuadro 1.4 Capacidades para enfrentar el fracaso

Conocimientos Tecnología Estrategia Marketing Finanzas Operaciones	Habilidades Gestión de Equipos Negociación Comunicación Creatividad etc.	Actitudes Liderazgo Tolerancia al riesgo Compromiso Responsabilidad Adaptación Cooperación
---	---	--

Fuente: Adaptado por el autor de Tascón (1998).

1.1.6 Un Mapa de Riesgo de Crédito para el Sistema Financiero Colombiano.

Los mapas de riesgos pueden representarse con gráficos o datos correspondientes a la calificación de los riesgos con sus respectivas variables y a su evaluación de acuerdo con el método utilizado en cada empresa, los cuales pueden ser agrupados en tablas, con información referente a los riesgos; a su calificación, evaluación, controles y los demás datos que se requieran para contextualizar la situación de la empresa y sus procesos, con respecto a los riesgos que la pueden afectar y a las medidas de tratamiento implementadas. Los mapas de riesgo tienen como beneficio los siguientes puntos (Mejía, 2011):

- Permite un mejor entendimiento en relación con la situación de los riesgos de la empresa en conjunto y de sus procesos o sus proyectos, al proporcionar información en forma global o discriminada.
- En los casos en los cuales la gerencia no tiene conciencia de la necesidad de invertir en las medidas de control o financiamiento de los riesgos, o en el entrenamiento y sensibilización del personal, la información contenida en los mapas de riesgos puede servir de motivación para apoyar al desarrollo de los programas de administración de riesgos, orientar efectivamente las acciones al definir prioridades para su manejo y al disponer de propuestas sobre las medidas de tratamiento.
- Con el diseño e implementación de los mapas de riesgos se promueve el trabajo en equipo, lo cual incrementa el entendimiento de los participantes sobre los procesos analizados y crea un mayor nivel de responsabilidad y colaboración entre las

dependencias, porque con ellos se logra entender las relaciones que tienen los procesos entre sí y sus implicaciones en la generación y administración de riesgos.

- El mapa de riesgos permite también monitorear el desempeño de la organización en la administración de sus riesgos, con el establecimiento de comparativos anuales a partir de las evaluaciones de los diferentes riesgos y el análisis de la efectividad de las medidas de control implementadas.

Según Capera et. Al (2012), las herramientas con las que más se identifican las situaciones de riesgo y vulnerabilidades del sistema, es el mapa de riesgo, que tiene como ventaja de expresar gráficamente dos dimensiones de importancia en la administración de contingencias: la probabilidad de ocurrencia asociada a un evento y su daño potencial. El seguimiento de los mapas de riesgo en el tiempo permite identificar las situaciones de riesgo más probables y sus efectos sobre el sistema. En el Gráfico 1 se presenta un esquema general de esta herramienta; en el eje vertical se encuentra una medida de probabilidad de ocurrencia, y en el eje horizontal se coloca una medida de pérdidas del sistema analizado. Con base en los valores obtenidos para las medidas descritas anteriormente, es posible identificar los eventos que representan el mayor riesgo para el sistema.

Es de vital importancia que la empresa cuente con un comité de riesgos que se encargue de identificar, evaluar y valorar los riesgos de cada uno de los procesos y elaborar un mapa de riesgos que contribuya a la fácil administración de los mismos.

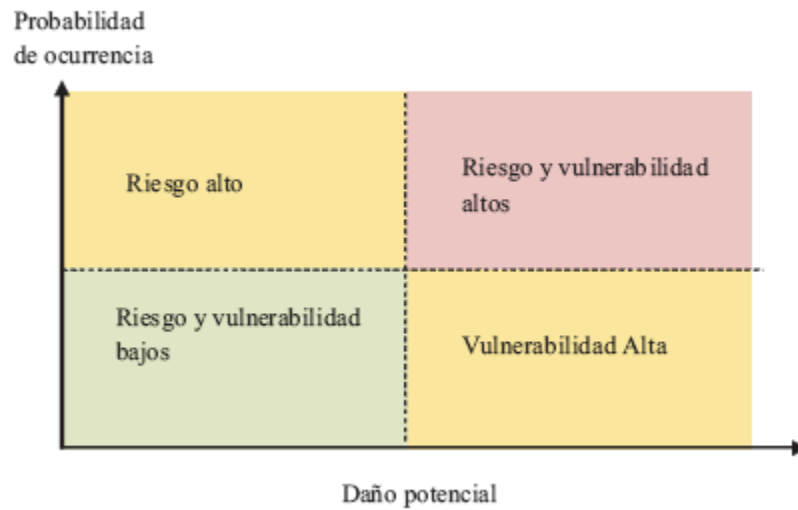
1.1.7 Determinantes de fragilidad en las empresas Colombianas

Una de las mayores preocupaciones para toda empresa es caer en un estado de insolvencia: el no poder cumplir con sus obligaciones financieras y declararse en quiebra o bancarrota. Este tipo de amenaza a la estabilidad financiera de las empresas es relevante no solo para inversionistas y empleados, sino también para prestamistas del sector financiero, auditores y autoridades reguladoras entre otros.

Por medio de la técnica de Regresión Probit se identifican las variables relevantes para pronosticar el estrés o fragilidad financiera de las empresas de manera que se puedan tomar acciones preventivas, correctivas o simplemente para monitorear el riesgo de crédito del sector corporativo privado de la economía (Martínez, 2003).

Se puede calcular por medio de indicadores de rentabilidad, endeudamiento, liquidez y eficiencia frecuentemente utilizados al realizar análisis financieros, identificando como relevantes las razones financieras utilidad antes de impuestos/activos, obligaciones financieras/activos y disponible/activos.

Cuadro 1.5 Esquema general del mapa de riesgo



Fuente: Capera, et. Al, 2012

El modelo Probit presupone que se conoce la distribución de la v.a. X que determina el nivel de estrés. Más concretamente, lo que el modelo presupone es que el analista será capaz de ajustar, mediante una distribución teórica, la distribución que sigue la v.a. Y , la cual es una transformación lineal de X . Habitualmente, dicha variable Y seguirá una distribución normal, logística, o de valores extremos. En ocasiones, sin embargo, será mejor ajustar la distribución de $Z = \log(Y)$ mediante una log-normal, una log-logística, o una Weibull. La distribución que se elija dependerá de los valores observados para la variable de estrés X , siendo necesario realizar un test de bondad de ajuste para comprobar que el modelo es válido.

El objetivo del análisis será estudiar el efecto causado por los cambios en la variable de estrés X sobre la probabilidad de éxito π (es decir, sobre la probabilidad de que el dispositivo falle).

La expresión genérica de estos modelos, suponiendo que no hay variables factores, sería la siguiente:

$$\pi = c + (1 - c) * F(\beta_0 + X * \beta)$$

Donde:

π = probabilidad de éxito (i.e., probabilidad de que el dispositivo falle),

$F(Y)$ = función de distribución de la v.a. $Y = \beta_0 + \beta \cdot X$,

β_0 = término constante,

X = variable de estrés,

β = coeficiente (desconocido) asociado a la variable de estrés,

c = tasa de fallo natural (proporción de dispositivos que fallan por causas naturales).

1.2 JUSTIFICACIÓN

Los ingenieros financieros deben comprender que el quehacer de su profesión radica en el diseño de modelos que, apoyados en técnicas y herramientas estadísticas e informáticas, brinden soluciones tentativas a problemas financieros de hoy en día, que cada vez son más complejos debido a la dinámica acelerada, generada en los mercados por la globalización de la economía y de la información.

En la región se pueden encontrar empresas en busca de subsistencia y empresas en busca de rentabilidad; las empresas que están en busca de la subsistencia, carecen de la ambición de obtener mayor beneficio de sus inversiones en las actividades económicas que desarrollan. En cambio, las empresas que andan en busca de la rentabilidad, quisieran tener a la mano herramientas que les permitan tomar decisiones acertadas en su planificación de estrategias para operar en el mercado. El presente estudio permite hacer un aporte descriptivo de la situación al que se enfrentan las empresas PYME en Colombia de acuerdo al reflejo de sus actividades económicas, en los estados financieros; lo que se pretende es brindar a los empresarios la base para determinar porque deben tener ciertos valores en sus cuentas de los estados financieros; lo que se traduce en formular, por ejemplo, estrategias para disminuir o aumentar los niveles de cartera de acuerdo a los estándares de riesgo de insolvencia que genera el análisis de regresión logística binaria.

Los modelos de regresión logística aplicados a la solución de problemas financieros normalmente utilizan variables de carácter cuantitativo para explicar algún suceso dicotómico. Esta investigación hace su aporte en la posible categorización de variables cuantitativas a variables cualitativas con múltiples estados, con el objetivo de generar probabilidades de un suceso de dos estados por medio de la estandarización¹ de estados múltiples para las posibles variables explicadoras de tal suceso. Es decir, se ha pretendido, por ejemplo, definir si una empresa tiene posibilidades de caer en estado de insolvencia si su estructura financiera está compuesta en mayor proporción por sus deudas o por su capital.

1.3 OBJETIVOS DE LA INVESTIGACIÓN

1.3.1 Objetivo general

Diseñar un modelo para la medición del riesgo de insolvencia empresarial: un caso de estudio en las empresas PYME de Colombia.

1.3.2 Objetivos específicos

- Construir una base de datos estándar con la información financiera empresarial de las PYMES incluyendo los periodos fiscales desde 2003 hasta el año 2011.

¹ No confundir con estandarización estadística.

- Caracterizar la base de datos empresarial de la Superintendencia de Sociedades de Colombia 2003-2011.
- Describir la caracterización cuantitativa de las empresas PYME a estudiar, por medio de un posicionamiento promedio y un modelo de regresión múltiple.
- Describir el comportamiento histórico de las variables de las empresas PYME objeto de estudio por medio del ajuste de distribución de variables y el valor de prueba K-S
- Diseñar un modelo de regresión logística binaria por medio de un grupo de entrenamiento de empresas PYME de las actividades económicas más representativas del sector real y con información de los estados financieros como variables explicadoras del estado de insolvencia de estas empresas.

2. MARCO REFERENCIAL

2.1 MARCO TEÓRICO

2.1.1 Regresión Logística

Molinero (2001) comenta que el objetivo de la Regresión Logística consiste en modelar como influye en la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de diversos factores y el valor o nivel de los mismos. También puede ser usada para estimar la probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías (politómico).

Si se utiliza como variable dependiente la probabilidad P de que ocurra el suceso y construimos la siguiente función:

$$Zn = \frac{P}{1 - P}$$

Ahora si tenemos una variable que puede tomar cualquier valor, por lo que podemos plantearlos el buscar para ella una ecuación de regresión tradicional:

$$Zn = \frac{P}{1 - P} = a + b [\textit{variable explicativa}]$$

Que se puede convertir con una pequeña manipulación algebraica en:

$$(P)\textit{Probabilidad suceso} = \frac{1}{1 + e^{-a-b[\textit{variable explicativa}]}}$$

El número de factores puede ser más de uno (Molinero, 2001).

Regresión Logística Múltiple

Botella, Martínez & Alacreu (2003) explican que para el caso de una Regresión Logística con múltiples estados de respuesta, los coeficientes se estiman y los contrastes de hipótesis se realizan del mismo modo que en el modelo simple, aunque con el modelo múltiple (igual que en regresión lineal) se pueden hacer contrastes no solo sobre cada coeficiente sino también sobre el modelo completo o para comparar modelos.

$$\text{Log} \left[\frac{P}{1 - P} \right] = a + bX$$

$$\frac{P}{1 - P} = e^{a+bX}$$

$$\frac{P}{(1-P)} \times (1-P) = e^{a+bX}(1-P)$$

$$P = e^{a+bX} - e^{a+bX}P$$

$$P + e^{a+bX}P = e^{a+bX}$$

$$P(1 + e^{a+bX}) = e^{a+bX}$$

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}}$$

$$\frac{1}{(1 + e^{a+bX})(e^{-a-bX})} = \frac{1}{e^{-a-bX} + e^{a+bX-a-bX}} = \frac{1}{e^{-(a+bX)} + 1}$$

$$P = \frac{1}{1 + e^{-(a+bX)}}$$

Con lo que Beltrán (2010) indica que la técnica de regresión logística es una técnica ampliamente utilizada cuando se tiene por objetivo identificar el grupo al cual pertenece una unidad experimental pero que, además, genera la probabilidad de tal clasificación y es menos estricta en el cumplimiento de supuestos para la utilización del modelo. Por ejemplo, uno de los supuestos más difíciles de cumplir y que se requiere en la aplicación del Análisis Discriminante, es el supuesto de normalidad; que en el caso de la Regresión Logística no se requiere cumplir el supuesto de normalidad multivariado del conjunto de variables regresivas; lo que se traduce en que se puede trabajar con las variables originales sin necesidad de transformarlas (Nava & Surendra, 2007).

En cuanto a la caracterización de cada una de las variables explicadoras, se busca encontrar, a través del método de Máxima Verosimilitud, los coeficientes que maximicen la función logística. Como explica Nava & Surendra (2007): "... Mediante la interpretación de los coeficientes del modelo estimado se busca hallar la características considerándolas simultáneamente a todas ellas... que son más discriminatorias... si alguna de las variables independientes es una variable discreta con k niveles, se debe incluir en el modelo como un conjunto de k-1 variables de diseño o "variables dummies". El cociente de las probabilidades correspondientes a los dos niveles de la variable respuesta se denomina ODDS y su expresión es:

$$\frac{P\left(Y = \frac{1}{X}\right)}{1 - P\left(Y = \frac{1}{X}\right)} = e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}$$

Si se aplica el logaritmo natural se obtiene el modelo de regresión logística:

$$\text{Log} \left[\frac{P\left(Y = \frac{1}{X}\right)}{1 - P\left(Y = \frac{1}{X}\right)} \right] = \text{Log}\left(e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Donde:

Log = Logaritmo de la razón de proporciones de los niveles de la variable respuesta.

Los β_j estimados representan la tasa de cambio de una función de la variable dependiente "Y" por unidad de cambio de la variable independiente "X".

El coeficiente β_j expresa el cambio resultante en la escala de medida de la variable "Y" y para un cambio unitario de la variable "X", por ejemplo, para la variable X_1 , $\beta_1 = g(x_1+1) - g(x_1)$ representa el cambio en el logit frente a un incremento de una unidad en la variable X_1 . La interpretación se hace en términos de la razón de odds (OR).

$$OR = \frac{\frac{P\left(Y = 1/X_j + 1\right)}{P\left(Y = 0/X_j + 1\right)}}{\frac{P\left(Y = 1/X_j\right)}{P\left(Y = 0/X_j\right)}} = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_j (X_j + 1) + \dots + \beta_p X_p}}{e^{\beta_0 + \beta_1 X_1 + \beta_j X_j + \dots + \beta_p X_p}}$$

Una cuestión en este tipo de análisis es determinar si todas las variables consideradas en la función discriminante contienen información útil y si solamente algunas de ellas son suficientes para diferenciar los grupos. Cuando las variables independientes tienen mucha relación entre sí, el modelo no puede distinguir que parte de la variable dependiente es explicada por una u otra variable → se conoce como multicolinealidad (Villagarcía, 2004). Por lo que es imperativo evaluar la correlación entre las variables explicadores del modelo.

Función de distribución logística

Botella, Martínez & Alacreu (2003), enuncian la explicación de la función de la distribución logística, planteando la siguiente cuestión: ¿Qué significado tienen los coeficientes del modelo α_0 y α_1 ?

$$\frac{P}{1 - P} = \exp(\alpha_0 + \alpha_1 X)$$

$$\frac{P}{1-P} = \exp^{\alpha_0} + \exp^{\alpha_1 X}$$

Respondiendo al interrogante dicen que $\exp(\alpha_0)$ representaría el valor del ODDS cuando la variable explicativa toma el valor cero, es decir, cuanto más probable es el éxito que el fracaso cuando la variable explicativa vale cero:

$$\frac{P}{1-P} = \exp(\alpha_0 + \alpha_1 X) \text{ y } \frac{P'}{1-P'} = \exp(\alpha_0 + \alpha_1(X+1)) = \exp(\alpha_1) = \frac{\frac{P'}{1-P'}}{\frac{P}{1-P}} = OR$$

Por lo que $\exp(\alpha_1)$ representa el OR por unidad de incremento de la variable explicativa X.

OR = Medida de asociación

$\exp(\alpha_0)$ es el OR por el aumento de una unidad en la variable Xi manteniendo constantes las otras (controlado por ellas).

Las variables cualitativas en el modelo logístico

Es incorrecto que en el modelo intervengan variables cualitativas, ya sean nominales u ordinales. La solución a este problema es crear tantas variables dicotómicas como respuestas menos dos. Estas son las dummy (variables internas, indicadores de diseño).

El coeficiente de la ecuación para cada variable dummy; se corresponde al Odds Ratio de esa categoría con respecto al nivel de referencia (la primera respuesta), cuantifica como cambia el riesgo respecto a primera respuesta. El hecho de que la constante en el modelo logístico no sea significativa significa que cuando las variables independientes toman el valor de cero, el logaritmo toma un valor de cero también.

2.1.2 Riesgo de crédito

“El riesgo de crédito se define como la pérdida potencial que se registra con motivo del incumplimiento de una contraparte en una transacción financiera (o en alguno de los términos y condiciones de la transacción). También se concibe como un deterioro en la calidad crediticia de la contraparte o en la garantía o colateral pactada originalmente”. (De Lara, 2005).

“El riesgo crédito... surge cuando las contrapartes están indispuestas o son totalmente incapaces de cumplir sus obligaciones contractuales”. (Jorion, 2000).

Knop (2004), comenta: "... el riesgo de crédito es la posibilidad, con mayor o menor incertidumbre, de incurrir en una pérdida cuando la contrapartida incumple total o parcialmente sus compromisos financieros según lo estipulado en el contrato que los rige".

Según estos autores y demás literatura revisada acerca del Riesgo de Crédito, las pérdidas y el incumplimiento son consecuencias que hacen del crédito, a través de los diferentes instrumentos financieros, un producto que requiere un detallado estudio por parte de las entidades que lo ofrecen para garantizar las ganancias futuras de los dueños de la empresa.

El Riesgo de Crédito, en perspectiva, es definido en 5 tipos de riesgo (Knop, 2004):

- Riesgo de contrapartida.
- Riesgo emisor.
- Riesgo por país.
- Riesgo de liquidación.
- Riesgos relacionados, como el deterioro de la calidad de cartera.

Cada uno de estos grupos expresa que el estudio del riesgo crediticio es realizado de acuerdo al interés del gestor de riesgo. La perspectiva del presente estudio de investigación es determinar la probabilidad de que una entidad deudora sea incapaz de cumplir con los términos de pago y fecha pactados en un contrato de crédito con otra entidad. Esto es lo que se denomina Riesgo de Contrapartida.

Cálculo del Riesgo de Crédito

En la actualidad hay desarrollados, entre otros, cuatro tipos de modelos para determinar el riesgo de crédito de las empresas (De Lara, 2005):

- Modelos tradicionales, entre los que se encuentra el de las 5 Ces.
- Modelos econométricos.
- Modelo KMV y Moody's.
- Redes neuronales artificiales.

Estos modelos tienen por objetivo determinar la probabilidad de que una entidad incumpla en sus compromisos crediticios, exceptuando a los modelos tradicionales, los cuales son de carácter cualitativos. Los tomadores de decisiones deben interactuar armoniosamente tanto con los estudios cualitativos como con los cuantitativos ya que ambos no deben ser percibidos de forma mutuamente excluyentes sino que, por el contrario, deben complementarse con el fin de generar resultados más fundamentados y certeros. A continuación hay una descripción de los modelos econométricos utilizados en el presente estudio para la medición del riesgo crediticio.

El análisis discriminante múltiple (ADM)

Es una técnica estadística que, partiendo de un conjunto de observaciones (universo), separa cada una de las observaciones en dos categorías. Para el caso del riesgo de crédito, esta técnica es empleada, generalmente, para diferenciar las empresas que están en quiebra de las que no lo están. Uno de los modelos de análisis discriminante más representativo, a modo de ejemplo, utilizado para determinar el riesgo de crédito es el Z-score de Altman (1968), el cual es una combinación lineal con unos pesos específicos de cinco razones financieras que clasifica en: las empresas que incumplen en sus deudas, de aquellas que no lo hacen (De Lara, 2005).

El enfoque del análisis discriminante aplicado en esta investigación es similar al cálculo de las ecuaciones de regresión lineal. Pero es de resaltar que la regresión discierne del análisis discriminante en que:

- La variable dependiente es de carácter "no métrica" en el análisis discriminante y "métrica" en el análisis de regresión lineal.
- El análisis discriminante presenta pasos adicionales para su desarrollo.

Modelo de regresión lineal múltiple

Los modelos de regresión tienen como fin predecir los futuros valores de una variable dependiente numérica a través de la relación con al menos una variable independiente numérica (Berenson, Levine y Krehbiel, 2001). Para el caso del modelo lineal múltiple, existen al menos dos variables independientes relacionadas linealmente y que explican el comportamiento de una sola variable dependiente. El modelo, para n variables explicadoras, es representado formalmente como:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_n X_{ni} + \epsilon_i$$

Donde,

β_0 = Ordenada en Y

β_1 = Pendiente de Y con la variable X_1 , y las variables X_2, \dots, X_n constantes

β_2 = Pendiente de Y con la variable X_2 , y las variables X_1, \dots, X_n constantes

β_n = Pendiente de Y con la variable X_n , y las variables X_1, \dots, X_{n-1} constantes

ϵ_i = Error aleatorio en Y para la observación i

Una explicación más práctica radica en que el valor de las β son los coeficientes de la ecuación de regresión que deben ser hallados y después reemplazar en las X los valores medidos para cada observación i . En otras palabras: "La regresión múltiple y el análisis de correlación

implican un proceso de tres pasos como el que usamos en la regresión simple. En este proceso: Describimos la ecuación de regresión múltiple; examinamos el error estándar de regresión múltiple de la estimación, y utilizamos el análisis de correlación múltiple para determinar qué tan bien la ecuación de regresión describe los datos observados (Levin & Rubin, 2004).

Supuestos en un Análisis de Regresión Lineal.

- Parámetros lineales. Esto significa que es posible escribir el modelo como una ecuación lineal. Por ejemplo,

$$Y_{Estimada} = a + bX_1 + cX_2 + \dots + nX_n + e$$

- Muestra aleatoria. Hay una muestra aleatoria del universo que quiere ser representada en el modelo.
- Media condicional cero. El valor esperado del error e es cero.
- Multicolinealidad. Ninguna de las variables independientes debe estar linealmente relacionada.
- Homoscedasticidad. La varianza del error es la misma para todas las combinaciones de las variables independientes.
- Normalidad. El error e se distribuye de acuerdo con una distribución normal.
- Autocorrelación. Los términos de error e son independientes uno del otro.

El coeficiente de correlación múltiple ρ .

Esta es una medida que indica el grado de relación existente entre dos o más variables. Los valores generados en un análisis de correlación van desde -1 (los valores negativos representan una relación inversa) hasta 1 (los valores positivos representan una relación directa).

A una correlación de -1 , se le llama correlación inversa perfecta; un resultado de 1 se denomina correlación directa perfecta. Si el resultado llegase a ser muy próximo a cero, se dice que no existe relación entre las variables estudiadas.

Para el cálculo del coeficiente de correlación es necesario determinar cuál es el objetivo del estudio. Esto debido a que la regresión y la correlación son dos técnicas separadas: la regresión tiene como fin la predicción y la correlación la asociación. "Para situaciones en las que el interés

principal es el análisis de regresión, el coeficiente de correlación de la muestra (r) se obtiene a partir del coeficiente de determinación (r^2) (Berenson, Levine y Krehbiel, 2001):

$$r^2 = \frac{\text{suma de cuadrados de regresión}}{\text{suma de cuadrados total}} = \frac{SCR}{SCT}$$

$$\text{Coeficiente de correlación} = r = \sqrt{r^2}$$

Evaluación del modelo de Regresión Lineal Múltiple

Cuando el modelo de regresión múltiple está construido, inicialmente debe ser evaluada la bondad de ajuste. Esta es medida con el error estándar y el coeficiente de determinación múltiple. Finalmente es llevada a cabo la prueba de significancia para el modelo la cual consta del coeficiente de regresión estandarizado (prueba t) y el cociente F.

La bondad de ajuste

Medir la bondad de ajuste significa que tan aproximados fueron los resultados obtenidos en la regresión, a los datos originales. Si una regresión no genera un ajuste considerable, entonces el modelo será rechazado inmediatamente sin necesidad de hacer más pruebas. El ajuste es medido con los siguientes indicadores.

Error estándar de estimación (Se): Este mide los grados de dispersión alrededor del plano de la regresión. Entre más pequeño sea Se, más ajustado y preciso será el modelo de regresión.

$$Se = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n - k - 1}}$$

En donde $n - k - 1$ es el número de grados de libertad y k es el número de variables a la derecha; el numerador bajo el radical de la fórmula es la suma de los errores elevada al cuadrado y se minimizará de acuerdo con el concepto de mínimos cuadrados ordinarios.

Coeficiente de determinación múltiple (R^2): Expresa la fuerza de relación entre la variable dependiente y las variables explicadoras. Entre más alto sea el valor de R^2 mayor poder explicativo tendrá el modelo. El valor de esta medida se encuentra siempre entre 0 y 1, es decir: $0 \leq R^2 \leq 1$

El coeficiente de determinación ajustado ($\overline{R^2}$): El coeficiente de determinación tiene la desventaja de que es vulnerable a manipular sus resultados incluyendo nuevas variables al modelo así no tengan significancia, esto debido a algún tipo de correlación casual que se pueda presentar, pero que en el fondo no tiene nada que ver con el objetivo del estudio. El coeficiente

de determinación ajustado se "ajusta a la medida del poder explicativo para el número de grados de libertad" (Webster, 2000).

$$\overline{R^2} = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}$$

Pruebas de significancia del modelo de regresión

Estas pruebas miden si el modelo de regresión construido presenta algún valor explicativo, es decir, si el modelo realmente presenta una relación entre la variable dependiente y las variables independientes o si existe alguna significancia de cada variable en el modelo. Para estas pruebas se construyen una Hipótesis nula y una Hipótesis alternativa que refute a la primera.

Prueba del cociente F: Esta prueba consiste en analizar si todas las variables independientes carecen de una relación lineal con la variable dependiente (H_0) o si por el contrario al menos una de las variables independientes tiene alguna relación lineal con la variable dependiente (H_A). Si alguna de las variables independientes no tiene relación lineal con la variable explicada entonces su coeficiente (en la ecuación de regresión) debería ser cero (Webster, 2000). Formalmente se tiene que:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_n = 0$$

$$H_A: \text{Al menos un } \beta \text{ no es cero}$$

Teniendo planteadas las hipótesis a hallar el valor de F y construir la regla de decisión la cual se logra a través de los grados de libertad de la parte de la regresión y de los errores.

$$F = \frac{\text{Cuadrado medio de la variación entre muestras}}{\text{Cuadrado medio de la variación dentro de las muestras}}$$

$$F = \frac{CMR}{CME}$$

Regla de decisión: El valor de F, con un nivel de confianza del 1%, 5%, etc., debe ser comparado con el valor de α que se halle en la tabla de distribución F de acuerdo a los grados de libertad de la parte explicada y el error. Si $F \leq \alpha$ no se rechaza la hipótesis nula; si $F > \alpha$ se rechaza la hipótesis nula.

Coefficiente de regresión estandarizado (prueba t): La prueba anterior mide en conjunto la significancia de las variables en el modelo. La prueba t mide la relación de cada una de las variables independientes con la variable dependiente. Si la variable independiente X_i no tiene relación alguna con la variable dependiente (H_0), tal variable independiente no es significativa en el modelo. Esto es:

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

El valor de t , con un nivel de confianza del 1%, 5%, etc., debe ser comparado con el valor de α que se halle en la tabla de distribución t de acuerdo a los grados de libertad de la parte del error. Si $-\alpha \leq t \leq \alpha$ (en la prueba de dos colas) no se rechaza la hipótesis nula; si $t > \alpha$ se rechaza la hipótesis nula.

2.1.3 Empresas PYME en Colombia

Para la descripción de los respectivos subsectores objeto de trabajo en la investigación se basó en la Resolución No. 00432 del 19 de noviembre de 2008, que describe a cada uno de los mismos así:

Cuadro 2.1 Actividades económicas estudiadas

CÓDIGO CIU	CLASIFICACIÓN
A0121	Cría especializada de ganado vacuno
F4521	Construcción de edificaciones para uso residencial
F4530	Construcción de obras de ingeniería civil
G5030	Comercio de partes, piezas (autopartes) y accesorios (lujos) para vehículos automotores
G5141	Comercio al por mayor de materiales de construcción, vidrio, artículos de ferretería y equipo y materiales de fontanería y calefacción
G5190	Comercio al por mayor de productos diversos npc
J6599	Intermediación financiera, excepto los seguros y los fondos de pensiones y cesantías
K7010	Actividades inmobiliarias realizadas con bienes propios o arrendados

Cría especializada de ganado vacuno.

Actividades que realiza. Explotación de recursos naturales vegetales y animales. Comprende las actividades de cultivo, cría de animales, explotación (aprovechamiento) de madera, cosecha de plantas, cuidado de animales y captura de los mismos en su hábitat natural. En cuanto a la cría especializada:

- La cría y/o levante de ganado vacuno.
- La producción de leche cruda de vaca.
- La producción de semen bovino.
- Servicios de engorde del ganado en corrales.

- La producción de mantequilla, queso y otros productos lácteos, como actividad secundaria no modifica la clasificación de la unidad.

Construcción de edificaciones para uso residencial y construcción de obras de Ingeniería Civil

Actividades que realizan. Preparación de terreno, construcción corriente y construcción especial; construcción de obras de ingeniería civil; el acondicionamiento y la terminación de edificios y obras; acondicionamiento de edificaciones y de obras civiles y terminación y acabado de edificaciones y de obras civiles. También se incluyen las obras nuevas, reparaciones, ampliaciones y reformas, la erección de edificios y estructuras prefabricadas in situ y también la construcción de obras de índole temporal.

Comercio de partes, piezas (autopartes) y accesorios (lujos) para vehículos a nivel nacional; comercio al por mayor de materiales de construcción ferretería y vidrio; y venta al por mayor de productos diversos

Actividades que realiza. Venta al por mayor y al por menor (venta sin transformación) de cualquier tipo de productos y la prestación de servicios relacionados con la venta de mercancía; reparación de vehículos automotores y la instalación y reparación de efectos personales y enseres domésticos.

Intermediación financiera, excepto los seguros y los fondos de pensiones y cesantías

- Distribuir fondos por medios distintos del otorgamiento de préstamos:
 - o Inversiones en valores mobiliarios, por ejemplo, acciones, obligaciones, títulos, lotes de valores mobiliarios de sociedades o fondos comunes de inversión, etc.
 - o Transacciones por cuenta propia de corredores de bolsa.
 - o Inversiones en bienes inmuebles efectuadas primordialmente por cuenta de otros intermediarios financieros (por ejemplo, las sociedades de inversión inmobiliaria).
 - o Suscripción de créditos recíprocos, opciones y otras operaciones financieras de cobertura.

Actividades inmobiliarias realizadas con bienes propios o arrendados

Actividades que realiza:

- La compra, venta, alquiler y explotación de bienes inmuebles propios o arrendados tales como edificios de apartamentos, viviendas y edificios no residenciales, incluso salas de exposiciones, terrenos; la urbanización y el fraccionamiento de terrenos en lotes, etc.
- El acondicionamiento y la venta de terrenos, la explotación de apartamentos amoblados y de zonas residenciales para viviendas móviles.

2.2 MARCO CONCEPTUAL

Test de Hosmer – Lemeshow

Es una evaluación que permite medir la bondad de ajuste de un modelo de regresión logística. Es un agrupamiento basado en las probabilidades estimadas por el modelo. Agrupa los n sujetos en g patrones según criterios estadísticos. Se obtiene calculando el estadístico chi cuadrado de Pearson que ampara las frecuencias observadas y las estimadas, en cada grupo y categoría de la variable respuesta, esto es, el estadístico χ^2 calculado a partir de una tabla $2 \times g$ la conservada significación de la misma indica un buen ajuste del modelo” (Beltrán, 2010).

Variables indicadoras (dummy)

“En los modelos de regresión logística... con la modelización se asume que las variables son “cuantitativas”, si una variable categórica se introduce en el análisis con las categóricas “0,1,2,3” y la interpretación del coeficiente que acompañara a la misma, indicará que para la categoría “2” el efecto es el doble que para la categoría codificada como “1”, cuando esto puede no tener ningún sentido (pensemos, por ejemplo, que “0” representa 0 gramos; “1” representa 100 gramos; “2” representa 165 y “3” representa 235 gramos).

...La solución es crear tantas variables como categorías menos 1, denominadas Variables Dummy... Estas variables no tienen ningún sentido por si solas y por ende deben figurar en los modelos y se debe contrastar su inclusión siempre en bloque (Botella, Martínez & Alacreu, 2003).

Confusión

Existe confusión cuando la asociación entre dos variables difiere significativamente según se considere o no, otra variable. A esta última variable se le denomina variable de confusión para la asociación en

$$\text{Log} \frac{P}{1-P} = \alpha_0 + \alpha_1 X_1$$

donde α_1 cuantifica la asociación. Se dice que X_2 es una variable de confusión para esta asociación, si el modelo

$$\text{Log} \left(\frac{P}{1-P} \right) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2$$

produce una estimación para α_1 diferente del modelo anterior... Se dice que las variables X_2, \dots, X_k son variables de confusión si la estimación de α_1 obtenida por el modelo

$$\text{Log} \left(\frac{P}{1-P} \right) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k,$$

es diferente de la obtenida en el modelo simple. En ambos casos se dice que la estimación de α_1 obtenida en los modelos múltiples está controlada o ajustada por X_2 o por X_2, \dots, X_k .

Interacción

Existe interacción cuando la asociación entre dos variables varía según los diferentes niveles de otra u otras variables. El modelo más sencillo que hace explícita la interacción entre dos variables X_1 y X_2 es

$$\text{Log} \left(\frac{P}{1-P} \right) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1 X_2$$

Si α_3 es distinto de 0, el odds ratio por unidad de cambio en X_1 manteniendo fijo X_2 es diferente para cada valor que X_2 de X_2 . Del mismo modo el Odds Ratio por una unidad de cambio en X_2 manteniendo fijo X_1 es diferente para cada valor X_1 de X_1 . Es obvio que primero debe contrastarse la interacción y después, en caso de que no exista, la confusión (Botella, Martínez & Alacreu, 2003).

Los coeficientes del modelo logístico como calificadores de riesgo:

ODDS: el odds asociado a un suceso es el cociente entre la probabilidad de que ocurra frente a la probabilidad de que no ocurra:

$$\text{Odds} = \frac{P}{1-P}$$

donde:

P = probabilidad del suceso.

"... el modelo más preciso será el que muestre un error estándar en el coeficiente de regresión de X , o el que aporte un intervalo de confianza de la OR asociada a la exposición X más estrecho" (Aguayo, 2007).

Prueba Wald

$H_0: \beta = 0$, es decir que $OR = 1$ (que la variable independiente no tiene ningún efecto sobre la variable dependiente); nivel de significancia 5%.

Wald: valor obtenido para el estadístico de contraste de hipótesis $H_0: \beta = 0$. Esta prueba se llama test de Wald y su valor de halla mediante el cociente del coeficiente al cuadrado y su error estándar al cuadrado. Este estadístico sigue una distribución X^2 con 1 grado de libertad (Perez & Martínez, 2002).

Es posible que algunas de las supuestas variables explicativas no sean tales y no tengan ningún efecto sobre la variable respuesta; para poder identificar y eliminarlas del modelo, se recurre a la prueba Wald. La cual se limita a contrastar la hipótesis de nulidad del coeficiente β_i asociado a la variable X_j .

$$H_{0j}: X_j \text{ no influye sobre } Y: \beta_j = 0$$

$$H_{1j}: X_j \text{ influye sobre } Y: \beta_j \neq 0$$

... que se distribuye como una X^2 con 1 grado de libertad cuando la muestra es grande, siendo S_j^2 la varianza del estimador de β_j . El contraste se realiza a un nivel de significancia del 5%.

El estadístico de contraste para la j -ésima variable explicativa es:

$$W_j = \frac{\beta_j^2}{S_j^2}$$

Nota importante acerca del Wald: si la probabilidad asociada al estadístico de Wald es supuesto al 5% se acepta la H_0 de nulidad del coeficiente β_j . Si esta por debajo del 5% se rechaza la hipótesis nulidad H_0 y se acepta que el coeficiente $\beta_j \neq 0$.

Cuando la variable es numérica, como puede ser por ejemplo la edad, o el índice de masa corporal, es una medida que cuantifica el cambio en el riesgo cuando se pasa de un valor del factor a otro, permaneciendo constantes el resto de variables. Así Odds Ratio que supone pasar de la edad X_1 a la edad X_2 , siendo "b" el coeficiente correspondiente a la edad en el modelo logístico, es:

$$OR = \exp[b \cdot (X_2 - X_1)]$$

Según Calderón & De Los Godos (2009), cuando el coeficiente "b" de la variable es positivo obtendríamos un Odds Ratio mayor que 2 y corresponde por tanto a un factor de riesgo. Por el contrario, si "b" es negativo el Odds Ratio será menor que 1 y se trata de un factor de protección.

Chi² Wald: Permite contrastar si el coeficiente es significativamente diferente de cero.

Modelo Lineal Generalizado (MLG):

Es una flexible generalización de la regresión de mínimos cuadrados ordinaria. Relaciona la distribución aleatoria de la variable dependiente en el experimento (la función de distribución) con la parte sistemática (predictor lineal) a través de una función llamada "función de enlace". Formulados por Jhon Nelder y Robert Wedderburn, como una manera de unificar varios modelos estadísticos, incluyendo la regresión lineal, regresión logística y regresión de Poisson, bajo un solo marco teórico. Esto permitió desarrollar un algoritmo general para la estimación de máxima verosimilitud.

2.3 MARCO LEGAL Y NORMATIVO

Ley 1450 de 2011. Artículo 43: Hace referencia a las definiciones del tamaño empresarial en Colombia.

Ley 1116 de 2006. Régimen de insolvencia empresarial. Deroga título II de la Ley 222 de 1995 con el artículo 126; reemplaza Ley 550 de 1999 pero a partir de primero de julio de 2007, fecha en la cual culmina la prórroga dada a la Ley 550 de 1999.

Ley 905 de 2004. Estimula la promoción y formación de mercados altamente competitivos mediante el fomento a la permanente creación y funcionamiento de la mayor cantidad de Mipymes.

Ley 590 de 2000. Artículo 2: Parámetros para clasificar las empresas por su tamaño de acuerdo a nivel de activos y número de trabajadores.

Ley 550 de 1999. Título II: Acuerdos de reestructuración. Prorrogada hasta primero de julio de 2007.

Ley 222 de 1995. Título II: Régimen de procesos concursales. Concordato, liquidación obligatoria. Derogado por artículo 126 de Ley 1116 de 2006.

3. METODOLOGÍA

3.1 TIPO DE INVESTIGACIÓN

El trabajo de investigación desarrollado en este proyecto alcanzó dos niveles: un nivel descriptivo y un nivel exploratorio. El nivel descriptivo en la investigación se debió a la caracterización detallada que se hizo de las cuentas de los estados financieros de las empresas PYME en Colombia; de las actividades económicas más representativas, en unidades de producción, del sector real en Colombia; y del ajuste teórico de las cuentas de los estados financieros, elegidas para medir el riesgo de insolvencia.

El nivel exploratorio se dio en el diseño de un modelo de regresión logística aplicado a la ingeniería financiera en el sentido de trabajar con variables categóricas como posibles explicadoras del suceso de que una empresa PYME de Colombia tenga más probabilidad o no de caer en estado de insolvencia de acuerdo a su ubicación en los diferentes estados de las variables explicadoras.

3.2 FASES DE DESARROLLO

3.2.1 Fase I: Caracterización de la base de datos empresarial de la Superintendencia de Sociedades de Colombia 2003-2011

En este apartado es detallada la metodología seguida en el desarrollo de la primera fase de la investigación: *Medición del riesgo de crédito para empresas PYME en Colombia*. Esta fase estuvo dirigida a cumplir dos objetivos:

- Construir y afinar la base de datos que ya estaba funcionando en el complemento prototipo en *Excel*.
- Realizar un estudio para la determinación de las características de la base de datos descargada de la Superintendencia de Sociedades de Colombia.

Entrega de material a los auxiliares de investigación

Se entregaron cuatro archivos en formato *Excel 2007*, llamados:

- 2010 **(1)**
- EFECTIVO_2010 **(2)**
- REGIONES **(3)**
- CIIU **(4)**

El documento (1) contenía la lista de todas las empresas que presentaron su Balance General y Estado de Resultados, con sus respectivas cuentas, en el año 2010.

El documento (2) contenía la lista de las empresas que presentaron su Flujo de Efectivo, con sus respectivas cuentas, en el año 2010.

El documento (3) contenía una tabla donde se relacionan la división política de Colombia con sus respectivas regiones.

El documento (4) contenía una tabla que relaciona los códigos del CIIU con su respectiva descripción.

3.2.1.1 Construcción de la base de datos

En un estado inicial para la estructuración, diseño y registro de los datos en la base de datos, primero se descargó la información de la página *web* de la superintendencia de sociedades. Esta información estaba contenida en hojas de cálculo del programa *Microsoft Office Excel*; su estructura de consulta estaba organizada de acuerdo a las cuentas del plan único de cuentas de Colombia.

Después de tener la información descargada en el computador de trabajo, fueron estandarizados cada uno de los registros disponibles. Como esta información fue creada, seguramente, por diferentes colaboradores, en diferentes periodos de tiempo, las condiciones de presentación y de síntesis variaron en algunas de las cuentas. La necesidad de tener la información disponible de una manera uniforme fue indispensable para realizar las respectivas consultas y filtros, y para esto era necesario cumplir con las condiciones necesarias que exigía el manejador de bases de datos relacional.

Lo que se quería tener era una base de datos compuesta por unas tablas específicas que contuvieran individualmente la información del balance, estado de resultados y flujo de efectivo (desde el año de 1995 hasta el año 2010) de:

- Las empresas en general.
- Sectores económicos del sector real.
- Subsectores de los sectores reales.
- Subsectores para:
 - Microempresas.
 - Pequeñas empresas.
 - Medianas empresas.
 - Grandes empresas.
 - Subsectores en general.

Para llegar a tal punto en la estructuración de la base de datos solicitada, fue necesario planear y seguir una serie de pasos a través de las herramientas del programa informático *Microsoft Office Access*.

Contenido de la base de datos

La base de datos fue construida con el objetivo de manejar información masiva de manera eficiente y trivial. Para esto el diseño fue enfocado a la definición inicial de unos grupos contenedores de tablas específicas. Los grupos que fueron definidos *a posteriori* pero planificados a medida que se construía la base de datos, se ilustran en cuadro 3.1.

El cuadro 3.1 hace referencia la cantidad de tablas que fueron incluidas en cada uno de los grupos contenedores de la base de datos en *Access*. En total se definieron 22 grupos contenedores y 325 tablas con la información estandarizada de acuerdo a los objetivos de la investigación.

Pasos para la estructuración de la base de datos

Todos estos puntos detallados a continuación fueron desarrollados en *Access*.

- Punto de partida (Totalizando las empresas). Lo primero que se hizo para construir la base de datos fue importar, toda la información contenida en las hojas de cálculo, de *Excel* a *Access*. Para esto fue utilizada una herramienta de *Access* denominada *Datos Externos*, la cual contiene un comando denominado *Importar a una hoja de cálculo de Excel*. Después de realizar esto ya se contaba con toda la información referenciada.

Un problema encontrado fue que después de tener tablas con información de las empresas que registraron sus estados financieros en la superintendencia de sociedades fue que no todas las empresas registraban su información en años consecutivos, es decir, una empresa que presentó sus estados financieros en el año de 1995, 1997, 2000, sería un claro ejemplo de tal problema. Pero este problema se presentó más de manera técnica que de contenido para los objetivos de la investigación. Al no tener las empresas referenciadas en cada uno de los años registrados, fue necesario realizar consultas que permitieran filtrar a cada una de las empresas registradas al menos una vez en la superintendencia de sociedades para totalizar el número de empresas que habían realizado su registro.

Estas consultas iniciales fueron realizadas de acuerdo a las herramientas de consulta disponibles. Las consultas son objetos que permiten cargar información de tablas relacionadas o no, de acuerdo a los criterios de búsqueda del que esté consultando información en una base de datos relacional. La herramienta de *Access* para realizar consultas está referenciada como *Crear*, en la cual se encuentran una serie de comandos (ubicadas en la cinta de opciones de *Access*). Entre otras opciones se encuentran: el *Asistente para consultas* y *Diseño de consultas*. El primero es utilizado para realizar una consulta con la ayuda del asistente, cargada en *Access*. En la opción de diseño hay más libertad para realizar las consultas, pero se debe dominar en una buena proporción el manejo de *Access*. Esta consulta inicial fue realizada a través de la opción del asistente para consultas.

Cuadro 3.1 Lista grupos contenedores de la base de datos construida

GRUPOS CONTENEDORES	TABLAS CONTENIDAS
Información General	5
Balance sectores generales	16
Balance total empresas	16
Balance microempresas	16
Balance pequeñas empresas	16
Balance medianas empresas	16
Balance grandes empresas	16
Balance subsectores micro	16
Balance subsectores pequeña	16
Balance subsectores mediana	16
Balance subsectores grande	16
Efectivo sectores generales	16
Efectivo total empresas	16
Efectivo microempresas	16
Efectivo pequeñas empresas	16
Efectivo medianas empresas	16
Efectivo grandes empresas	16
Efectivo subsectores micro	16
Efectivo subsectores pequeña	16
Efectivo subsectores mediana	16
Efectivo subsectores grande	16
Total de tablas de base de datos	325

Luego de realizar las consultas para definir el número total de empresas que se habían registrado por los menos una vez, se encontró con que más de 40.000 empresas hicieron registros en uno de los años entre 1995 y 2010. Todo lo anterior permitió sentar las bases para la construcción de las demás tablas necesarias para los objetivos de la investigación.

- Estructuración de la base de datos. Como fue comentado anteriormente, la idea era tener una base de datos eficiente para su manejo y futuras consultas. La información conseguida inicialmente desde la superintendencia contenía la información de todas las cuentas del balance general, el estado de resultados y el flujo de efectivo. Entre otros, tenía referenciado el sector con su descripción y el código CIIU al que pertenecía, pero sin su respectiva descripción. Al no tener referenciado la descripción del CIIU, se hacía difícil distinguir a que subsector pertenecía cada una de las empresas. Para solucionar este problema fue necesario buscar la relación entre los códigos del CIIU y su respectiva descripción. La información fue encontrada en la página del DANE y en un documento denominado *Clasificación industrial uniforme de todas las actividades*

económicas revisión 3.1 adaptada para Colombia, creado por el DANE. Esta información estaba contenida en formato PDF; se realizó una transcripción de estos datos a *Excel* para poder relacionarlos e incluirlos en la base de datos de *Access*. La cantidad de subsectores encontrados para el sector real fueron 393.

En el año 2000 fue implementada la ley 590 del 2000. Entre otros aspectos, esta ley definió las características de las empresas del país de acuerdo a su tamaño dependiendo el nivel de activos y número de trabajadores. Conforme a esto se encontró la necesidad de agrupar las empresas a nivel de tamaño por activos. Esto es:

- Microempresas.
- Pequeñas empresas.
- Medianas empresas.
- Grandes empresas.

Una microempresa, por ejemplo, dedicada al comercio al por menor, tiene unas condiciones y características diferentes a otra empresa con la misma actividad pero de tamaño diferente de acuerdo a nivel de activos o número de trabajadores. Por tal motivo, estudiar o investigar un sector en general puede dar unos resultados muy generales, dependiendo del tipo de investigación, herramientas y metodologías, e imposibilitar una delimitación necesaria para resultados confiables y validación estadística, por ejemplo. De esta manera se procedió a filtrar las empresas según su tamaño por nivel de activos.

El criterio para la filtración de las empresas fue basado en la ley 905 del año 2004 en el que se aumenta el nivel de activos para denominar las empresas medianas y grandes. Las empresas medianas fueron definidas desde 5.000 salarios mínimos mensuales legales vigentes hasta 30.000 de estos mismos. Pero es de resaltar que hubo cuidado a la hora de implementar el criterio ya que se contaba con información desde el año de 1995, y que desde este año hasta el año 2000 no se tenía referenciado formalmente el tamaño de las diferentes empresas

- Relación de las empresas con sus respectivos sectores económicos. Teniendo el total de empresas que se habían registrado alguna vez en la superintendencia, y después de discriminarlas en micro, pequeñas, medianas y grandes, se procedió a promediar los sectores referenciados en cada año para cada una de las empresas. El promedio hallado fue *la media aritmética*. El procedimiento consistió en agrupar a las empresas por sectores y calcular el promedio para cada una de las cuentas estandarizadas de los estados financieros. Para tal fin se realizó un *diseño de consulta* (mencionado anteriormente), y el criterio fue filtrar las empresas de acuerdo al *promedio* por campos y *agrupación* por código de sector. *Access* cuenta con un comando llamado *Totales* el cual se encuentra ubicado en la ficha de la cinta de opciones *Diseño* que se despliega cuando es abierta tal ficha. El comando *Totales* permite totalizar, ya sea agrupando, sumando, contando, promediando, calculando volatilidad, entre otros, agrupando campos de texto como cadena de caracteres o números, que sean coincidentes como registros con los demás. Este procedimiento se hizo para cada uno de los

años desde 1995 hasta 2010. Fueron 66 sectores los que se promediaron en cada uno de los años.

Cada una de las empresas realiza actividades que pueden estar dentro de un mismo sector, pero en el que sus actividades económicas pueden diferir mucho de una a otra empresa. Por eso hay discriminación más detallada para clasificar las empresas de acuerdo a sus actividades económicas. Aquí es donde la clasificación industrial internacional uniforme (CIIU) define sus funciones como sistema normativo. En Colombia se pueden encontrar alrededor de 400 subsectores económicos pertenecientes al sector real, compuesto por los 66 sectores generales ya comentados. Para obtener esta discriminación, se transcribió la información contenida en un documento del DANE respecto a la clasificación industrial uniforme versión 3.1, en una hoja de cálculo de *Excel*. Después de esto se importó esta hoja de cálculo a la base de datos en *Access* con la que se estaba trabajando.

Teniendo esta información en la base datos como una tabla con los campos *ciiu* y *descripción ciiu*, se aplicó una herramienta de *Access* que relaciona tablas con campos en común, es decir, las tablas que contenían la información de cada una de las empresas tenían un campo denominado *ciiu*, pero no contaban con un campo que relacionara la descripción de este CIIU. La nueva tabla ingresada contenía el campo en común *ciiu* y además de esto tenía la descripción referente a cada registro de la tabla. En la cinta de opciones hay una ficha denominada *Herramientas de bases de datos*; ingresando a esta se despliegan varios grupos y galerías de comandos. Hay un comando denominado *Relaciones*. Al abrir este comando se despliega una ventana que permite escoger tablas para relacionarlas por campos coincidentes. Lo único que se hace es *clickar* y arrastrar el campo seleccionado con el coincidente en la otra tabla que se quiera relacionar. En este caso se relacionaron las tablas ya mencionadas, se realizó una consulta con estas tablas y con esto se logró construir una nueva tabla ya con la relación de cada una de las empresas con la descripción de sus respectivos subsectores.

- Referencia de la evolución histórica del salario mínimo. Como la clasificación de las empresas en Colombia de acuerdo a su nivel de activos fue determinada de acuerdo a la vigencia del salario mínimo mensual legal vigente, cada año el nivel, en pesos, de los topes de las categorías (micro, pequeñas, medianas, grandes) es diferente cada vez. Lo que se sintetiza en que deben ser recalculadas estas categorías de clasificación para cada año. En este sentido es primordial contar con la evolución histórica del salario mínimo. Para esto fue buscada la información en diferentes páginas de *internet* y se encontró gran cantidad de información y fue validada con las estadísticas históricas contenidas en la página *web* del DANE. Simplemente se tomó la información manualmente y se registró directamente en una nueva tabla de la base de datos. En el cuadro 3.2 es mostrada la información del salario mínimo conseguida.

Trabajo artesanal de la base de datos

En los párrafos anteriores se ha comentado brevemente cómo se empezó a trabajar con la información inicial de la que se tuvo disponibilidad. Ahora es momento de detallar todo el desarrollo mecánico de cada una de las 325 tablas. Es de recordar que hay grupos de tablas

con las mismas características de formato y contenido en donde sólo cambia el año de registro de datos; el detalle se resume entonces a los comentarios pertinentes para la construcción de las "tablas tipo" ubicadas en cada uno de los grupos contenedores.

El grupo *Información general* ya ha sido más que enunciado en los párrafos anteriores. Aquí están concentradas las tablas referentes al total de empresas registradas, sectores, subsectores descritos y la tabla del salario mínimo.

Balance sectores generales, contiene 16 tablas; cada una incluye, aproximadamente, 66 registros, compuesto cada uno por la información de los sectores generales del sector y los promedios de cada una de las cuentas pertenecientes al balance general y al estado de resultados. Bajo el modelo de diseño en que se creó este grupo también fue creado otro grupo con las mismas características de forma: *Efectivo sectores generales*. Este también contiene 16 tablas con las mismas características ya mencionadas en este párrafo; la única diferencia es el contenido de los campos, que para este grupo fueron las cuentas del flujo efectivo.

Balance total empresas y *Efectivo total empresas*, son los grupos que se conformaron en el punto de partida cuando fue realizada la totalización del número de empresas con sus respectivos estados financieros básicos. Estos dos grupos contienen, cada uno, 16 tablas con los registros de las empresas con sus respectivos estados financieros. Se recuerda que *Balance total empresas* tiene como campos las cuentas del balance general y del estado de resultados.

Balance microempresas, *Balance pequeñas empresas*, *Balance medianas empresas* y *Balance grandes empresas*, son 4 grupos contenedores los cuales en total sumarían 64 tablas. Cada uno de estos grupos contiene 16 tablas con la información del balance y el estado de resultados de cada una de las empresas de acuerdo a su tamaño en nivel de activos. Por ejemplo, *Balance microempresas*, tiene registradas cada una de las microempresas registradas en cada uno de los años desde 1995 hasta el año 2010 con su respectiva información referente al balance general y estado de resultados. Pero, de acuerdo a estas mismas características de grupos y tablas también se construyeron otros 4 grupos: *Efectivo microempresas*, *Efectivo pequeñas empresas*, *Efectivo medianas empresas* y *Efectivo grandes empresas*; cada grupo con las mismas características de los explicados en este párrafo, pero con la información de las cuentas del flujo de efectivo.

Balance subsectores micro, *Balance subsectores pequeña*, *Balance subsectores mediana* y *Balance subsectores grande*, son otros 4 grupos contenedores de 16 tablas cada uno. Para la construcción de estas tablas fue necesaria la estructuración de una gran cadena de relaciones entre la tabla *ciiu* con las tablas del balance del total de empresas:

- Primero fue creada una tabla temporal denominada *Temporal***Balance****. El primer grupo de asteriscos de esta denominación hace referencia al tamaño de empresas (micro, pequeñas, medianas o grandes); el último grupo de asteriscos hace referencia al año de la información. Para la definición de esta tabla se realizó una consulta donde se relacionara el *nit* de cada empresa con la cuenta de *activos totales*. Esto con el fin de

poder filtrar las empresas que pertenecían a un tamaño en particular de acuerdo al nivel de sus activos totales.

Cuadro 3.2 Historial del salario mínimo en Colombia desde 1995 hasta 2012

Año	Salario míni	Auxilio de ti
1990	\$ 41,025.00	\$ 3,798.00
1991	\$ 51,720.00	\$ 4,787.00
1992	\$ 65,190.00	\$ 6,033.00
1993	\$ 81,510.00	\$ 7,542.00
1994	\$ 98,700.00	\$ 8,705.00
1995	\$ 118,170.00	\$ 10,815.00
1996	\$ 142,125.00	\$ 13,567.00
1997	\$ 172,005.00	\$ 17,250.00
1998	\$ 203,825.00	\$ 20,700.00
1999	\$ 236,438.00	\$ 24,012.00
2000	\$ 260,100.00	\$ 26,413.00
2001	\$ 286,000.00	\$ 30,000.00
2002	\$ 309,000.00	\$ 34,000.00
2003	\$ 332,000.00	\$ 37,500.00
2004	\$ 358,000.00	\$ 41,600.00
2005	\$ 381,500.00	\$ 44,500.00
2006	\$ 408,000.00	\$ 47,700.00
2007	\$ 433,700.00	\$ 50,800.00
2008	\$ 461,500.00	\$ 55,000.00
2009	\$ 496,900.00	\$ 59,300.00
2010	\$ 515,000.00	\$ 61,500.00
2011	\$ 535,600.00	\$ 63,600.00
2012	\$ 566,700.00	\$ 67,800.00

Fuente: Adaptada por los autores (fuente DANE).

- Las tablas temporales se relacionaron posteriormente con las tablas de balance de un año en particular y se relacionaban las demás cuentas que dieran cuenta del balance general y del estado de resultados.
- Después de tener las empresas filtradas de acuerdo a su tamaño se procedió a hacer una consulta que generara los promedios de las cuentas, agrupando para esto a las empresas por subsectores. Esto es: Para la consulta se llamó a la cuenta *ciiu* y a las cuentas del balance y estado de resultados de las tablas *Balance año****, y se ordenó el comando de agrupar por *ciiu* y promediar (media aritmética) cada uno de los campos con cuentas numéricas. El resultado, finalmente, fue: las tablas del balance y estado de resultados por subsector y por tamaño de acuerdo al nivel de activos.

Igualmente este procedimiento fue realizado para estructurar los grupos contenedores: *Efectivo subsectores micro*, *Efectivo subsectores pequeña*, *Efectivo subsectores mediana* y *Efectivo subsectores Grande*. También quedaron contenidas 16 tablas por grupo.

3.2.1.2 Afinación base de datos

Lo pedido a continuación se aplicó a los documentos (1) y (2).

- Contar el número de empresas que presentaron Balance General y Estado de Resultados.
- Contar el número de empresas que presentaron Flujo de efectivo.
- Contar el número de cuentas (Caja, deudores, inventarios) que hay en el Balance general, Estado de Resultados y Flujo de Efectivo.

Esta descripción preliminar no requirió de gráficas ni tablas, solo un recuento: "*tantas empresas en este año...; hay tantas cuentas..., etc.*"

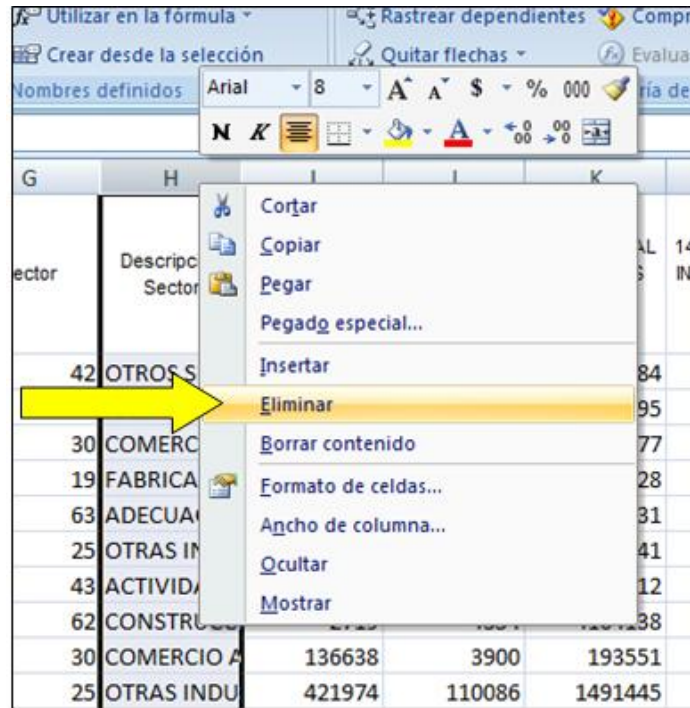
Consolidación de una única tabla y archivo

Como la base de datos requería de una afinación para hacer más eficiente la herramienta que se estaba desarrollando, fue necesario simplificar la redundancia de información; en este caso se sacrificó una insignificancia por esta consolidación. El procedimiento fue como sigue:

- Unir la tabla de Balance General y Estado de Resultados con la de Flujo de Efectivo solo cruzando la información de las empresas que se encontraban en ambos archivos (hubo empresas que registraron su Balance general y Estado de Resultados, pero que no registraron su flujo de efectivo. Por ende siempre hubo una cantidad menor de empresas en la tabla de Flujo de Efectivo). Para hacer esto se abrió el archivo (2); se seleccionó toda la tabla de Flujo de Efectivo incluyendo los rótulos y se copió en una hoja de cálculo del archivo (1) y se cerró el archivo (2). De esta manera solo quedó abierto el archivo (1) con su nueva tabla incluida.
- El siguiente paso consistió en eliminar las subcuentas de todos los estados financieros. Para esto solo se eliminaron las columnas: clic derecho en la cima de la columna y elegir *eliminar columna* (también se pueden seleccionar varias columnas al mismo tiempo y eliminarlas). Ver figura 3.1.

Nota: El nombre de las cuentas definitivas quedó tal cual se lista en el cuadro 3.3, sin tildes, ni minúsculas y sin espacios entre palabras. Lo anterior con el objetivo de filtrar la información sin errores de ejecución (esto aplica para todos los estados financieros).

Figura 3.1 Cómo eliminar columnas en Excel



Las cuentas que debieron quedar finalmente se listan en el cuadro 3.3:

Cuadro 3.3 Cuentas definitivas del Balance General

CUENTAS DEFINITIVAS DEL BALANCE GENERAL	
NIT	IMPUESTOS_GRAVAMENESYTASAS_CORTOPLAZO
RAZON_SOCIAL	OBLIGACIONES_LABORALES_CORTOPLAZO
CIUDAD	PASIVOS_ESTIMADOSYPROVISIONES_CORTOPLAZO
DEPARTAMENTO	DIFERIDOS_PAS_CORTOPLAZO
REGION	OTROS_PASIVOS_CORTOPLAZO
CIUU	BONOS_PAPELESCOMERCIALES_CORTOPLAZO
DESCRIPCION_CIIU	PASIVO_CORRIENTE
SECTOR	OBLIGACIONES_FINANCIERAS_LARGOPLAZO
DESCRIPCION_SECTOR	PROVEEDORES_LARGOPLAZO
DISPONIBLE	CUENTAS_PORPAGAR_LARGOPLAZO
INVERSIONES_CORTOPLAZO	IMPUESTOS_GRAVAMENESYTASAS_LARGOPLAZO
DEUDORES_CORTOPLAZO	OBLIGACIONES_LABORALES_LARGOPLAZO
INVENTARIOS_CORTOPLAZO	PASIVOS_ESTIMADOSYPROVISIONES_LARGOPLAZO
DIFERIDO_ACT_CORTOPLAZO	DIFERIDOS_PAS_LARGOPLAZO
ACTIVO_CORRIENTE	OTROS_PASIVOS_LARGOPLAZO
INVERSIONES_LARGOPLAZO	BONOS_PAPELESCOMERCIALES_LARGOPLAZO
DEUDORES_LARGOPLAZO	PASIVO_NOCORRIENTE

Cuadro 3.3 Cuentas definitivas del Balance General (continuación)

CUENTAS DEFINITIVAS DEL BALANCE GENERAL	
PROPIEDADES_PLANTAYEQUIPO	PASIVO_TOTAL
INTANGIBLES	CAPITAL_SOCIAL
DIFERIDO_ACT_LARGOPLAZO	SUPERAVIT_CAPITAL
OTROS_ACTIVOS	RESERVAS
VALORIZACIONES	REVALORIZACION_PATRIMONIO
ACTIVO_NOCORRIENTE	DIVIDENDOS_PARTICIPACION_ENACCIONES_CUOTAS
ACTIVO_TOTAL	RESULTADOS_EJERCICIO
OBLIGACIONES_FINANCIERAS_CORTOPLAZO	RESULTADOS_EJERCICIOS_ANTERIORES
PROVEEDORES_CORTOPLAZO	SUPERAVIT_PORVALORIZACIONES
CUENTAS_PORPAGAR_CORTOPLAZO	PATRIMONIO_TOTAL
	PASIVO_YPATRIMONIO

Las filas REGION y DESCRIPCION_CIIU, puede que no se encontraran en la tabla. En caso de que no estuvieran fue necesario incluirlas y que quedaran en ese mismo orden. Por lo que hubo que incluir columnas. Del estado de resultados, quedó (ver cuadro 3.4):

Cuadro 3.4 Cuentas definitivas del Estado de Resultados

CUENTAS DEFINITIVAS DEL ESTADO DE RESULTADOS
INGRESOS_OPERACIONALES
COSTO_VENTAS_YPRESTACION_SERVICIOS
UTILIDAD_BRUTA
GASTOS_OPERACIONALES_ADMINISTRACION
GASTOS_OPERACIONALES_VENTAS
UTILIDAD_OPERACIONAL
INGRESOS_NOOPERACIONALES
GASTOS_NOOPERACIONALES
UTILIDAD_ANTES_IMPUESTOS
IMPUESTO_RENTA_COMPLEMENTARIOS
GANANCIAS_PERDIDAS

Aquí (cuadro 3.4) la única cuenta que se eliminó era la de ajustes por inflación, si hubiere estado.

Del flujo de efectivo (ver cuadro 3.5):

Cuadro 3.5 Cuentas definitivas del Flujo de Efectivo

CUENTAS DEFINITIVAS DEL FLUJO DE EFECTIVO
NIT
RAZON_SOCIAL
CIUDAD
DEPARTAMENTO
CIU
SECTOR
DESCRIPCION_SECTOR
DEPRECIACIONES
AMORTIZACIONES
PROVISIONES
EFFECTIVO_GENERADO_OPERACION
FLUJO_EFFECTIVONETO_ACTIVIDADES_OPERACION
FLUJO_EFFECTIVONETO_ACTIVIDADES_INVERSION
FLUJO_EFFECTIVONETO_ACTIVIDADES_FINANCIACION
EFFECTIVO_PERIODOANTERIOR
EFFECTIVO_PERIODOACTUAL

3.2.1.3 Relación de la tabla Balance General y Estado de Resultados con la tabla Flujo de Efectivo

Hasta aquí ya habían quedado las cuentas definitivas de los tres estados financieros básicos. A continuación se detalla cómo se simplificó redundancia y espacio en memoria para tener una base de datos más eficiente en cuanto a la herramienta en desarrollo.

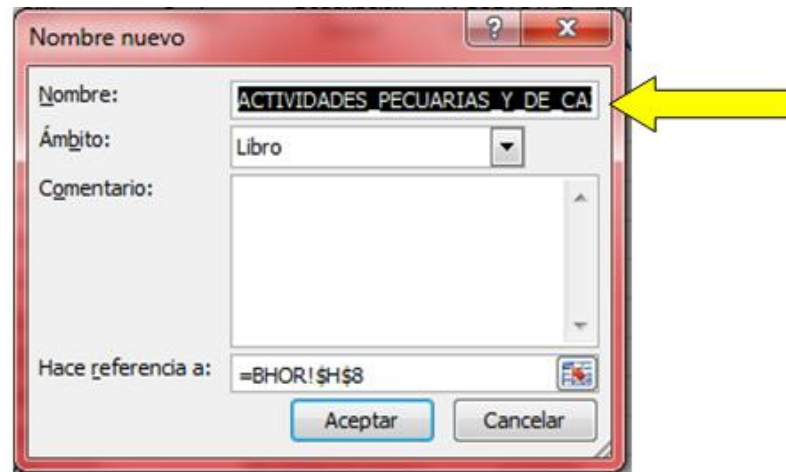
- Se seleccionó la tabla de Flujo de Efectivo y se le dio un nombre. Esto se hizo de la siguiente manera (ver figura 3.2 y figura 3.3):

Figura 3.2 Definición de nombre a un rango en una hoja de cálculo Excel



De ahí se le dio un nombre que identificara tal rango y después clic en *Aceptar*.

Figura 3.3 Definición de nombre a un rango en una hoja de cálculo Excel parámetros



- Lo siguiente fue seleccionar todos los campos (rótulos) de la tabla de Flujo de Efectivo, copiarlos y pegarlos a continuación de la tabla del Balance General y Estado de Resultados. Esto fue (ver figura 3.4 y figura 3.5):

Figura 3.4 Rótulos de la tabla definitiva Flujo de Efectivo

CD	CE	CF	CG	CH	CI	CJ	CK	CL
DEPRECIACIONES	AMORTIZACIONES	PROVISIONES	EFFECTIVO_GENERADO_OPERACION	FLUJO_EFFECTIVO_NETO_ACTIVIDADES_OPERACION	FLUJO_EFFECTIVO_NETO_ACTIVIDADES_INVERSION	FLUJO_EFFECTIVO_NETO_ACTIVIDADES_FINANCIACION	EFFECTIVO_PERIODOANTERIOR	EFFECTIVO_PERIODOACTUAL
31058	0	0	112015	82831	1480	17660	7482	81154

- Después se ubicó en la celda para llenar el primer registro de los nuevos campos de la tabla en la que estábamos ubicados. La idea era utilizar la fórmula BUSCARV contenida en *Excel* con el objetivo de buscar un registro, que en este caso fue el NIT de esta tabla; que lo buscara en la tabla Flujo de Efectivo con el nombre que le hubiéramos asignado y se pidió que trajera el valor de la columna DEPRECIACIONES. Esto se hizo para cada uno de los nuevos campos copiados a la tabla; el valor del registro buscado era el mismo para cada fila, lo que cambiaba era el número de la columna referencia que se quería traer a la celda en la que se estuviera.
- A modo de ejemplo, se llamó la tabla Flujo de Efectivo como EFECTIVO. La columna DEPRECIACIONES de tal tabla estaba ubicada en la posición 8. La condición FALSO de la fórmula hacía referencia a que la coincidencia debía ser exacta. La celda A2 hacía

referencia al registro que quería buscar en la tabla EFECTIVO; este registro lo representaba el NIT.

Figura 3.5 Cuentas del flujo de efectivo unidas a balance general y estado de resultados en la hoja de cálculo Excel

	BI	BJ	BK	BL	BM	BN	BO	BP
1	42 MAS: INGRESOS NO OPERACIONAL ES (ANEXO 5)	53 MENOS: GASTOS NO OPERACIONAL ES (ANEXO 5)	UTILIDAD ANTES DE IMPUESTOS Y AJUSTES POR INFLACION	54 MENOS: IMPUESTO DE RENTA Y COMPLEMENT ARIOS	59 GANANCIAS Y PERDIDAS			
2	900	18931	92847	32901	59946			
3	0	0	-907	3815	-4722			
4	7	11472	60918	24118	36800			
5	0	0	0	0	0			

Las fórmulas para un registro en específico hubieran sido, por ejemplo:

Para traer EFECTIVO_GENERADO_OPERACION

```
=+BUSCARV(A2;EFECTIVO;11;FALSO)
```

Para DEPRECIACIONES

```
=+BUSCARV(A2;EFECTIVO;8;FALSO)
```

Para AMORTIZACIONES

```
=+BUSCARV(A2;EFECTIVO;9;FALSO)
```

Así se hubiera hecho con todos los demás campos.

El resultado de todo el procedimiento anterior se reflejó en una nueva tabla con la información de los tres estados financieros: Balance General, Estado de Resultados y Flujo de Efectivo. Lo que quedaba por hacer en este punto era guardar el archivo como EstFin_ . La idea final era que este archivo solo contuviera esta tabla y en las demás pestañas no hubiera ningún tipo de información.

Nota: Para relacionar los campos REGIONES y/o DESCRIPCION_CIIU, se debieron relacionar los archivos (3) y (4).

Caracterización de la base de datos

La información es valiosa y poderosa, por eso fue indispensable conocer a fondo el contenido de esta; en este caso con el objetivo de generar estrategias de emprendimiento y negocios en el país. La caracterización consistió en ubicar la proporción de la cantidad de empresas que se encontraban en ciertas regiones y departamentos, calcular porcentajes y generar su respectiva representación tabulada y gráfica. Este fue uno de los puntos pilares en el que la rigurosidad metodológica y sistemática fueron los criterios fundamentales para desarrollar esta etapa de desarrollo de la investigación.

3.2.2 Fase II: Caracterización cuantitativa de las empresas a estudiar

Las actividades para la realización de esta segunda fase se resumen como sigue:

1. Determinar cuáles, de un listado de variables referentes a la situación financiera de las empresas, explicaban el comportamiento de las empresas de un subsector económico. Estas variables eran cuentas de los estados financieros y/o indicadores financieros internos de las empresas. Para tal fin se consultaron o calcularon las variables que eran necesarias.
2. Seguidamente se ordenaron las empresas de acuerdo al resultado en sus variables, para determinar una posición relativa entre las empresas, por variable, y así obtener una posición promedio o *ranking*.
3. Finalmente se realizó una regresión lineal múltiple y se simplificó si era necesario; esta simplificación se pudo hacer a través del *stepwise, backward* o *forward* (métodos de inclusión y exclusión de variables), utilizando para esto los test de significancia y bondad de ajuste estadísticos.

¿Cuáles variables debieron ser calculadas?

A continuación se presentan las variables que debieron ser calculadas y/o calculadas para cada una de las empresas del subsector que le correspondía, de acuerdo a los subsectores seleccionados para el estudio.

- a. Ventas.
- b. Crecimiento en ventas.
- c. Utilidad operativa / ventas.
- d. Utilidad neta / ventas.
- e. Razón corriente.
- f. Patrimonio neto (sin valorizaciones).
- g. Pasivo total / activos.

- h. Endeudamiento total / ventas.
- i. Endeudamiento financiero / ventas.
- j. Utilidad operativa / gastos financieros.
- k. Flujo de caja / gastos financieros.
- l. Utilidad neta.

A cada uno de los auxiliares le fue entregada una base de datos con la información de las empresas que pertenecían a un subsector en específico con dos modalidades: para pequeñas empresas y para medianas empresas. Es decir, se tuvieron que realizar los cálculos tanto para las pequeñas empresas de un subsector como para las medianas empresas de ese subsector.

Posición promedio (*ranking*) de las empresas de acuerdo a variables estudiadas

Después de haber calculado las variables del punto anterior se procedió a ordenar las empresas de primera a última de acuerdo a una variable en específico.

Por ejemplo, si era calculada la variable *razón corriente*, ocupó la primera posición frente a las demás, la empresa que obtuviera el nivel de razón corriente más alta; ocupando el último lugar, la empresa que tuviera la menor razón corriente. Otro ejemplo sería que la empresa que estuviera menos endeudada ocuparía el primer lugar y la empresa más endeudada, el último lugar.

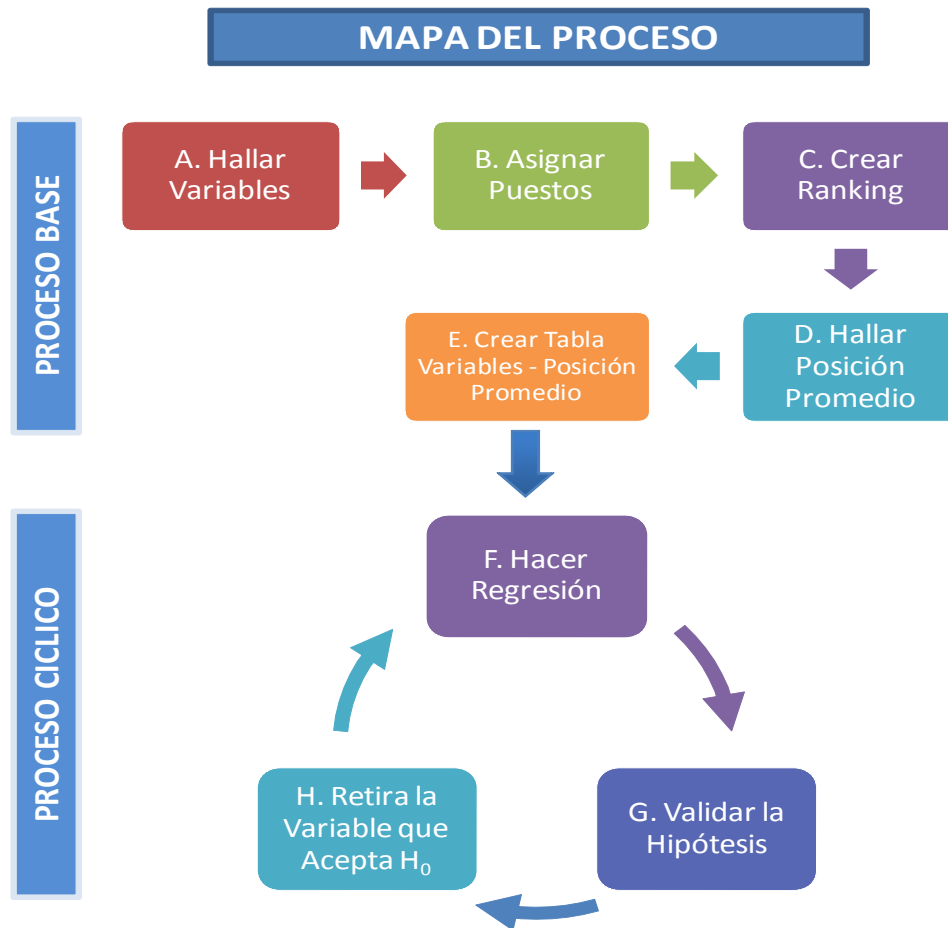
Al ir ordenando las empresas, se iba vislumbrando que una empresa no necesariamente ocupaba el primer o último lugar en todas las variables financieras, sin embargo, esta reordenación iba generando una posición general promedio, que fue igual a la media aritmética de la posición de cada una de las variables calculadas.

Medición de la bondad de ajuste y poder explicativo del modelo de regresión lineal

El objetivo intrínseco a las dos actividades anteriores, estuvo dirigido a establecer una relación lineal entre las variables financieras adoptadas y la posición promedio obtenida después de determinar una posición relativa por cada variable. La posición promedio era la variable *explicada*; el conjunto de variables financieras, las variables *explicadoras*. Lo que se debía hacer era realizar una regresión con las condiciones dadas. Para esto, primero se realizó la matriz de *covarianzas* o la matriz de *correlaciones*, dependiendo de si el conjunto de datos era de carácter homogéneo o heterogéneo. El objetivo de esta matriz era identificar si había multicolinealidad entre los datos, lo que representaba una redundancia de información; si era así se eliminaba una de las variables que estuviera casi perfectamente relacionada con otra variable explicadora (primer criterio), eliminando, efectivamente, la que menos estuviera correlacionada con la variable dependiente (segundo criterio). Este procedimiento dio como

resultado una función lineal compuesta por los coeficientes de las variables que tenían significancia en el modelo de regresión. La figura 3.6 ilustra la dinámica de trabajo de la fase II.

Figura 3.6 Mapa del proceso metodológico de la fase II



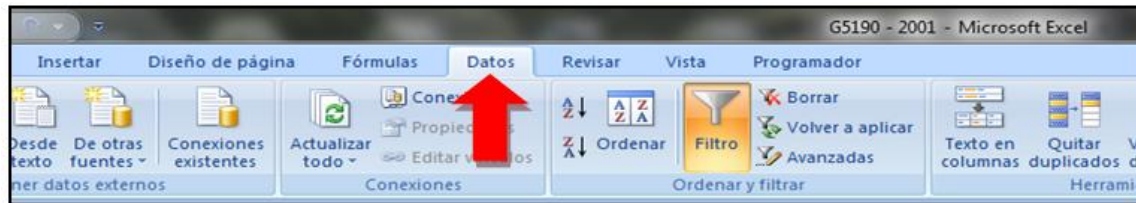
¿Cómo se hizo la regresión en la hoja de cálculo Excel?

Utilizando la función de *Excel* de análisis estadístico de Regresión, se aplicó a la tabla "Relación", tomando los datos así:

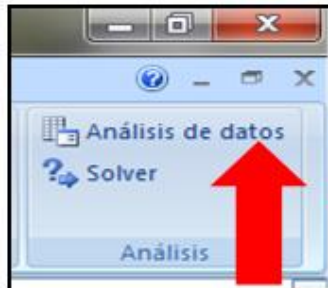
- Posición Promedio: Variables Dependiente (Y)
- Variables Calculadas: Variable Independiente (X)

Pasos:

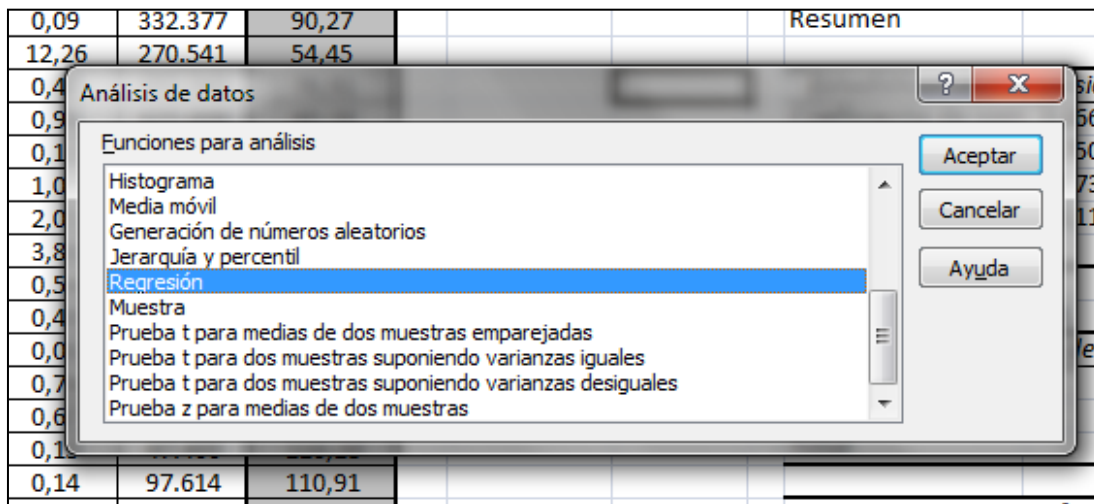
1. Elegir la pestaña DATOS.



2. Elegir la opción ANÁLISIS DE DATOS.



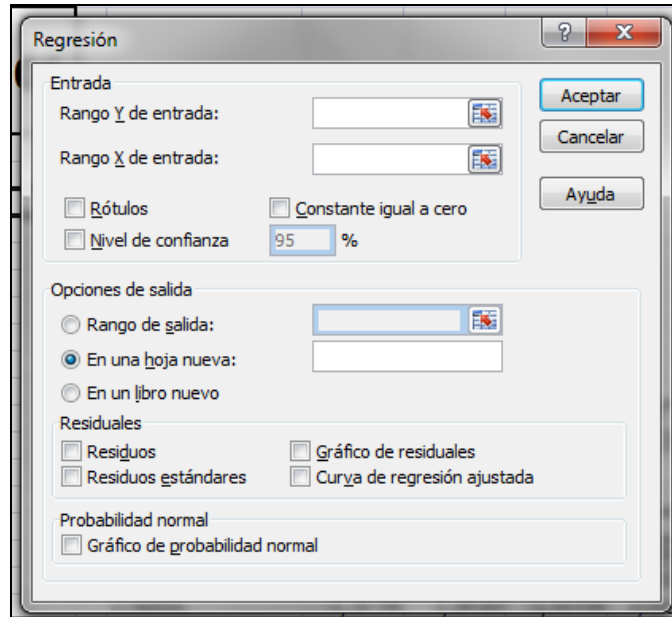
3. Elegir de la lista, la opción REGRESIÓN.



4. Definir el rango Y (posición promedio), el rango X (variables explicadoras) y el rango de salida.

5. Clic en ACEPTAR.

Cuando se terminó de hacer la regresión en la hoja de cálculo Excel, se generó la información del cuadro 3.6.



Cuadro 3.6 Tabla de regresión generada por el comando en la hoja de cálculo Excel

<i>Estadísticas de la regresión</i>									
Coefficiente de correlación	0,866448								
Coefficiente de determinación	0,750732								
R ² ajustado	0,73583								
Error típico	15,11504								
Observaciones	196								
ANÁLISIS DE VARIANZA									
	<i>Grados de libertad de cuadro de los cu</i>			<i>F</i>	<i>Valor crítico de F</i>				
Regresión	11	126606,3	11509,66	50,37835	1,3E-49				
Residuos	184	42037,46	228,4645						
Total	195	168643,8							
	<i>Coefficiente</i>	<i>Error típico</i>	<i>estadístico</i>	<i>probabilidad inferior 95%</i>	<i>superior 95%</i>	<i>inferior 95,0</i>	<i>superior 95,0</i>	<i>inferior 95,0</i>	<i>superior 95,0</i>
Intercepción	100,0362	5,447823	18,36259	0,000%	89,28793	110,7844	89,28793	110,7844	
A. VENTAS	-1,2E-06	2,2E-07	-5,35726	0,000%	-1,6E-06	-7,4E-07	-1,6E-06	-7,4E-07	
UTILIDAD OPERATIVA	6,975197	12,07705	0,577558	56,427%	-16,8521	30,80249	-16,8521	30,80249	
UTILIDAD NETA / VENTAS	-8,44936	4,445562	-1,90063	5,891%	-17,2202	0,321474	-17,2202	0,321474	
RAZÓN CORRIENTE	-10,3921	4,32562	-2,40245	1,728%	-18,9263	-1,85791	-18,9263	-1,85791	
PATRIMONIO NETO (SIN	-3,7E-06	1,54E-06	-2,40449	1,719%	-6,7E-06	-6,6E-07	-6,7E-06	-6,6E-07	
PASIVO TOTAL/ACTIVOS	32,25954	7,685448	4,197484	0,004%	17,09661	47,42247	17,09661	47,42247	
H. ENDEUDAMIENTO TOT	-0,37874	0,346127	-1,09422	27,529%	-1,06163	0,304149	-1,06163	0,304149	
I. ENDEUDAMIENTO FINAN	19,3408	4,810375	4,020643	0,008%	9,850215	28,83138	9,850215	28,83138	
J. UTILIDAD OPERATIVA /GA	-2,3672	0,258156	-9,16965	0,000%	-2,87653	-1,85788	-2,87653	-1,85788	
FLUJO DE CAJA /GASTOS	-0,10001	0,057522	-1,73859	8,378%	-0,21349	0,013481	-0,21349	0,013481	
L. UTILIDAD NETA	-6,4E-06	2,38E-06	-2,69247	0,775%	-1,1E-05	-1,7E-06	-1,1E-05	-1,7E-06	

3.2.3 Fase III: Ajuste de distribución de variables

Hasta este momento ya se contaba con los modelos de regresión definitivos que fueron simplificados para la explicación del comportamiento de las empresas de determinados subsectores económicos. Las variables fueron definidas, pero todavía no habían sido caracterizadas. La caracterización estadística requería de los siguientes pasos:

1. Graficar el histograma de frecuencias para cada una de las variables.
2. Comparar visualmente la distribución de los datos.
3. Aplicar *test* de normalidad *Kolmogorov-Smirnov.d*

El objetivo de los pasos anteriores era determinar a cuáles de las distribuciones de frecuencias parametrizada se ajustaban o no cada una de las variables estudiadas. Con la ayuda de los datos arrojados por el *Risk Simulator* (complemento numérico para Excel) se analizó la prueba estadística de P-Valor para determinar si se ajustaba o no.

Esta prueba estadística debía estar en un rango entre los 80% y 100% para que la variable estudiada, posiblemente, siguiera una distribución paramétrica, de lo contrario si el resultado de la prueba era menor que el 80% esta variable no se ajustaba a una distribución paramétrica. La línea azul continua, representaba los datos teóricos del test de normalidad *Kolmogorov-Smirnov*; las barras verticales representaban la distribución propia de los datos.

Procedimiento

1. Ubicándose en el resultado final de la regresión se seleccionó cada una de las variables que no se eliminaron (significativas en el modelo), y se procedió a desarrollar el ajuste de distribución mediante la aplicación SIMULADOR DE RIESGO en la hoja de cálculo Excel, de la siguiente manera:
 - a. Simulador de riesgo
 - b. Herramientas analíticas



- c. Ajuste de Distribución simple



3.2.4 Fase IV: Desarrollo del modelo de riesgo de crédito a través de la regresión logística

3.2.4.1 Definición preliminar de la muestra

Como población se eligieron las pequeñas y medianas empresas que presentaron sus estados financieros frente a la Superintendencia de Sociedades para el año 2011. Para tal año fueron 21.518 empresas las que se reportaron. Este tamaño de empresas según sus activos representó, aproximadamente, el 77% del total de empresas.

De todas las PYMES consideradas se encontraron 160 en estado de reestructuración, 105 en estado de reorganización, 27 en concordato y 9 empresas en liquidación. Para efectos del estudio, estos estados representan unas condiciones relacionadas en cuanto a la caracterización jurídica de estar en uno u otro estado, pero matemáticamente presentan una similitud. Por ejemplo, las empresas que están en reestructuración son empresas que se acogieron a la Ley 550 de 1999 para la reactivación empresarial, ley que fue modificada por la Ley 1116 de 2006, la cual dio una nueva denominación para las empresas que cayeran en estado de insolvencia el cual es "empresas en reorganización o liquidación". Con todo lo anterior se decidió considerar como parte de la muestra de estudio las empresas denominadas como en reestructuración, reorganización, concordato y liquidación; en total se tomaron 301 empresas con estas características.

La otra parte de la muestra estuvo compuesta por empresas que estuvieran "aparentemente sanas", es decir, por empresas que no tuvieran la denominación de estar en estado de reorganización o liquidación. Para tratar de tener homogeneidad en la comparación de empresas en reorganización y las "aparentemente sanas", se eligió un mismo número de empresas: 301 unidades "aparentemente sanas". También se tuvo cuidado de que la elección de estas empresas fuera proporcional en tamaño, igual que las empresas del primer grupo. Es decir, las 301 empresas en estado de insolvencia (reorganización, reestructuración, etc.), fueron

clasificadas en unos intervalos según el tamaño de los activos como se muestra en el cuadro 3.7.

Cuadro 3.7 Proporción de empresas en reorganización según tamaño por activos.

PROPORCIÓN DE EMPRESAS EN REORGANIZACIÓN SEGÚN TAMAÑO (Cifras en miles de pesos)		
LI	LS	FA
275.057	2.243.063	109
2.243.063	4.211.069	68
4.211.069	6.179.075	33
6.179.075	8.147.081	30
8.147.081	10.115.087	19
10.115.087	12.083.093	18
12.083.093	14.051.099	15
14.051.099	16.019.105	9
total		301

De acuerdo al cuadro 3.7, se procuró elegir casi el mismo tamaño de empresas según el número de intervalos, por ejemplo, elegir 109 empresas “aparentemente sanas” con activos entre, aproximadamente, 275 millones y 2.200 millones de pesos al igual que empresas en reorganización. En total se tuvo como muestra 602 empresas tanto en reorganización como aparentemente sanas.

3.2.4.2 Determinación del grupo de entrenamiento

En cualquier modelo de regresión se busca encontrar el valor de los coeficientes de las variables para realizar la descripción de un conjunto de datos, así como también predecir el comportamiento de una nueva observación aplicando la función de regresión hallada. Para lograr evaluar el poder estimativo y predictivo de un modelo de regresión es necesario aplicarlo en observaciones que no hayan sido consideradas para la determinación de los coeficientes del modelo; esto con la pretensión de no evaluar relajadamente el poder del modelo de regresión, ya que si se evalúa con observaciones que fueron utilizadas para hallar la ecuación de regresión, las características de esta observación permitirían una perfecta estimación y predicción para la observación.

El grupo de entrenamiento se compuso por casi el 50% de empresas que estuvieran dentro de cada uno de los intervalos definidos en el cuadro 3.7, tanto para empresas en reorganización como empresas aparentemente sanas. En el cuadro 3.8 se muestra la composición del grupo de entrenamiento.

Cuadro 3.8 Grupo de entrenamiento compuesto por empresas en reorganización y por empresas aparentemente sanas.

LI	LS	REORGANIZACIÓN	APARENTEMENTE SANAS
275.057	2.243.063	54	54
2.243.063	4.211.069	34	34
4.211.069	6.179.075	16	16
6.179.075	8.147.081	15	15
8.147.081	10.115.087	10	10
10.115.087	12.083.093	9	9
12.083.093	14.051.099	7	7
14.051.099	16.019.105	5	5
Total cada grupo		150	150
TOTAL		300	

Para el grupo de entrenamiento se definieron 300 empresas; la mitad compuesta por empresas en reorganización y la otra mitad por empresas aparentemente sanas. Por consiguiente las empresas restantes fueron utilizadas en aras de evaluar el poder estimativo y predictivo del modelo de regresión logística binaria.

Para garantizar la aleatoriedad en la selección de cada una de las unidades del grupo de entrenamiento, le fue asignado a cada uno de los elementos del marco muestral un número identificador, con lo que se aplicó la función de "generación de números aleatorios" de la hoja de cálculo *Excel* de *Microsoft®*, y así elegir las unidades identificadas con los números aleatorios generados por la función del *Excel*.

3.2.4.3 Hipótesis

¿Están relacionadas las variables independientes con el evento de que una empresa PYME caiga en estado de insolvencia?

3.2.4.4 Selección de variables

Para el modelo de regresión logística aplicado para evaluar el grado de insolvencia (variable dependiente) de las empresas PYME en Colombia, fueron definidas 8 variables independientes: 2 cuantitativas y 6 cualitativas.

Variable dependiente

La variable dependiente de respuesta dicotómica se definió como "Estado de insolvencia", la cual tomó como valores de respuesta "1" y "0"; significando "1" que la empresa era insolvente y "0" que la empresa estaba aparentemente sana o era solvente.

$Y_{\text{variable respuesta dicotómica}} = \text{"Estado de insolvencia"}$

$Y = 1, \rightarrow \text{"Insolvente (en reorganización)"}$

$Y = 0, \rightarrow \text{"Solvente (aparentemente sana)"}$

VARIABLES INDEPENDIENTES

Cada una de las variables independientes cualitativas elegidas para el modelo tiene estados dicotómicos representados por 1 (en caso de presencia de una característica) o 0 (en caso de ausencia de la característica). Por ejemplo, la variable independiente X_1 representó la razón corriente de una empresa, tomando el valor de "1" cuando la razón corriente era inferior a 1 y tomando el valor de "0" cuando la razón corriente estaba por encima de 1. La lógica de esta variable se puede entender en el sentido de que una empresa con una razón corriente por debajo de ciertos niveles indicaba que la empresa tal vez presentaría problemas para afrontar sus compromisos del corto plazo; no significando que este indicador mostrara insolvencia. El criterio de que la razón fuera mayor o igual a 1 parte del hecho de que la teoría del indicador y empíricamente hablando, tener una razón corriente igual a 1 significa que la empresa cuenta, exactamente, con los recursos para afrontar con los compromisos a corto plazo; por este motivo se tomó esta referencia.

Cualitativas

$X_1 = \text{"Razón Corriente"}$

$X_1 = 1, \rightarrow \text{"Razón corriente inferior a 1"}$

$X_1 = 0, \rightarrow \text{"Razón corriente superior o igual a 1"}$

$X_2 = \text{"Estructura Financiera"}$

$X_2 = 1, \rightarrow \text{"Pasivos} \geq \text{Patrimonio"}$

$X_2 = 0, \rightarrow \text{"Pasivos} < \text{Patrimonio"}$

$X_3 = \text{"Utilidad Neta"}$

$X_3 = 1, \rightarrow \text{"Utilidad neta Negativa"}$

$X_3 = 0, \rightarrow \text{"Utilidad neta 0 o Positiva"}$

$X_4 = \text{"Utilidad Operacional"}$

$X_4 = 1, \rightarrow \text{"Utilidad operacional Negativa"}$

$X_4 = 0, \rightarrow \text{"Utilidad operacional 0 o Positiva"}$

$X_5 = \text{"Resultado de Ejercicios Anteriores"}$

$X_5 = 1, \rightarrow \text{"Negativo"}$

$X_5 = 0, \rightarrow \text{"Positivo"}$

$X_6 = \text{"Cambio Flujo de Efectivo Neto"}$

$X_6 = 1, \rightarrow \text{"Efectivo neto disminuyó"}$

$X_6 = 0, \rightarrow \text{"Efectivo neto aumentó"}$

Cuantitativas

$X_7 = \text{"Grado de Apalancamiento"}$

$X_7 = 0 < x < 1, \text{entre más se acerca a 1 hay mayor apalancamiento}$

$X_8 = \text{"Concentración de Endeudamiento en Corto Plazo"}$

$X_8 = 0 < x < 1, \text{entre más se acerca a 1 hay mayor necesidad de efectivo}$

El cuadro 3.9 lista las variables con sus respectivas codificaciones

Cuadro 3.9 Codificación de variables para aplicar la Regresión Logística

CODIFICACIÓN DE VARIABLES	
NOMBRE	IDENTIFICADOR
Identificador Muestral	IDE_MUE
NIT	NIT
Razón Social	NOM_EMP
Estado de Insolvencia	EST_INS
Razón Corriente	RAZ_COR
Estructura Financiera	EST_FIN
Utilidad Neta	UTI_NET
Utilidad Operacional	UTI_OPE
Resultados de Ejercicios Anteriores	REE_ANT
Cambio en Flujo de Efectivo Neto	CFL_EFE
Grado de Apalancamiento	GRA_APA
Concentración Endeudamiento Corto Plazo	CON_ENC

RISARALDA	Farbicación de prendas de vestir	Cosntrucción de obras residenciales	Comercio de vehículos y actividades conexas	Comercio al por mayor	Comercio al por menor	Actividades inmobiliarias	Construcción de obras civiles	Otros	Actividades pecuarias y de caza
PROMEDIO	5,08%	6,01%	9,42%	21,23%	12,54%	8,92%	3,68%	37,49%	5,93%

EMPRESAS EJE CAFETERO	PROPORCIÓN SECTORIAL								
	Cosntrucción de obras residenciales	Comercio de vehículos y actividades conexas	Comercio al por mayor	Comercio al por menor	Actividades inmobiliarias	Construcción de obras civiles	Otros	Actividades pecuarias y de caza	Construcción de obras residenciales
PROMEDIO	6,78%	7,84%	16,51%	9,66%	8,03%	5,35%	50,14%	6,04%	5,55%

Cuadro 4.2 Participación de empresas por sectores en las regiones desde 2003 hasta 2011

EMPRESAS REGIÓN ANDINA	PROPORCIÓN SECTORIAL								
	Com x may	Com x men	Act Inve y Fina	Acti inmob	Cons obras civ	Otros	Act Pecua y caz	Cons obras res	Comr veh y act con
PROMEDIO	15,71%	8,67%	5,50%	7,72%	5,82%	57,13%	5,36%	4,47%	4,19%

EMPRESAS REGIÓN CARIBE	PROPORCIÓN SECTORIAL								
	Comercio al por mayor	Comercio al por menor	Actividades inmobiliarias	Construcción de obras civiles	Construcción de obras residenciales	Comercio de vehículos y actividades conexas	Otros	Actividades pecuarias y de caza	Productos alimenticios
PROMEDIO	16,31%	9,25%	9,58%	5,71%	4,32%	4,02%	55,42%	6,93%	5,74%

EMPRESAS REGIÓN PACÍFICA	PROPORCIÓN SECTORIAL									
	Comercio al por mayor	Comercio al por menor	Comercio de combustibles y lubricantes	Construcción de obras civiles	Comercio de vehículos y actividades conexas	Otros	Actividades pecuarias y de caza	Productos alimenticios	Agrícola con predominio exportador	Construcción de obras residenciales
PROMEDIO	15,61%	9,97%	4,44%	5,34%	4,88%	59,53%	5,35%	6,10%	4,13%	5,94%

EMPRESAS REGIÓN ORINOQUIA	PROPORCIÓN SECTORIAL							
	Comercio al por mayor	Comercio al por menor	Construcción de obras civiles	Otros	Actividades pecuarias y de caza	Productos alimenticios	Comercio de combustibles y lubricantes	Construcción de obras residenciales
PROMEDIO	22,79%	9,83%	11,73%	41,46%	8,15%	10,95%	5,59%	7,52%

Cuadro 4.3 Participación de empresas por sectores, regiones y departamentos desde 2003 hasta 2011

TOTAL EMPRESAS	PROPORCIÓN REGIONAL		
	Andina	Caribe	Otros
PROMEDIO	89,86%	8,13%	2,01%

TOTAL EMPRESAS	PROPORCIÓN SECTORIAL								
	Comercio al por mayor	Comercio al por menor	Actividades pecuarias y de caza	Actividades inmobiliarias	Construcción de obras civiles	Otros	Comercio de vehículos y actividades conexas	Construcción de obras residenciales	Actividades diversas de inversión y servicios financieros conexos
PROMEDIO	15,82%	8,76%	4,95%	7,73%	5,34%	60,39%	4,19%	4,39%	7,57%

TOTAL EMPRESAS	PROPORCIÓN DEPARTAMENTAL						
	Bogotá D.C.	Antioquia	Valle	Atlántico	Cundinamarca	Santander	Otros
PROMEDIO	52,68%	13,44%	10,59%	4,80%	3,16%	2,76%	13,28%

4.2 CARACTERIZACIÓN CUANTITATIVA DE LAS EMPRESAS ESTUDIADAS DENTRO DE SUS RESPECTIVOS SUBSECTORES

Los resultados que se buscaban eran conjuntos de variables para cada subsector, las cuales explicarían el comportamiento financiero de las empresas. Como la regresión permitió disminuir la redundancia de información innecesaria para explicar el comportamiento con un enfoque estadístico, a continuación se resumen las variables obtenidas para cada subsector con sus respectivos años desde 1995 hasta el 2010.

Cuadro 4.4 Número de variables que quedaron en la ecuación final de regresión para cada subsector desde el año 1995 hasta el año 2010

NÚMERO DE VECES VARIABLE EN LA ECUACIÓN DEL REGRESIÓN DESDE 1995-2010	CÓDIGO CIU	A0121	F4521	F4530	G5030	G5141	G5190	J6599	K7010
	VENTAS	12	13	16	9	12	16	7	15
	UTILIDAD OPERATIVA / VENTAS	13	4	8	9	6	13	7	14
	UTILIDAD NETA / VENTAS	5	4	7	8	4	12	7	5
	RAZÓN CORRIENTE	7	1	6	3	0	7	6	7

PATRIMONIO NETO	4	2	1	0	1	15	5	14
NIVEL DE ENDEUDAMIENTO	14	14	16	15	16	14	15	16
ENDEUDAMIENTO TOTAL / VENTAS	3	3	7	6	1	6	5	6
ENDEUDAMIENTO FINANCIERO / VENTAS	4	4	9	6	7	16	4	5
UTILIDAD OPERATIVA / GASTOS NO OPERATIVOS	4	14	11	7	8	14	1	9
FLUJO DE CAJA / GASTOS NO OPERATIVOS	6	1	7	1	3	10	1	9
UTILIDAD NETA	10	12	13	12	15	16	15	16

Las variables que más incidencia tuvieron, al hacer la regresión en cada uno de los subsectores con la línea de tiempo desde 1995 hasta 2010, se resumen como sigue:

- Cría especializada de ganado vacuno: *ventas, utilidad operativa / ventas y nivel de endeudamiento*. Cada una de estas quedaron en el modelo de regresión final en cada año: 12, 13 y 14 veces, respectivamente. Lo anterior se observa en cuadro 4.1.
- Construcción de edificaciones para uso residencial: Para este subsector se tuvo que: *ventas, 13 veces; nivel de endeudamiento, 14; utilidad neta, 12*. Se recuerda que estas son las variables más incidentes en los modelos de regresión finales para cada año.
- Construcción de obras de ingeniería civil: *Ventas, 16 veces; nivel de endeudamiento, 16; utilidad neta, 13 utilidad operativa / gastos no operacionales, 11*.
- Comercio de partes, piezas (autopartes) y accesorios (lujos) para vehículos automotores: *Pasivo total / activos, 15; Utilidad neta, 12 veces*.
- Comercio al por mayor de materiales de construcción, vidrio, artículos de ferretería y equipo y materiales de fontanería y calefacción: *Ventas, 12; Pasivo total / activos, 16; Utilidad neta, 15*.
- Comercio al por mayor de productos diversos npc: *Ventas, 16; patrimonio neto, 15; Endeudamiento financiero, 16; utilidad neta, 16*.
- Intermediación financiera, excepto los seguros y los fondos de pensiones y cesantías: *Nivel de endeudamiento, 15 veces; utilidad neta, 15*

- Actividades inmobiliarias realizadas con bienes propios o arrendados: *Ventas, 15; utilidad operativa / ventas, 14; patrimonio neto, 14; pasivo total / activos, 16; utilidad neta, 16.*

La información obtenida al reducir el número de variables que explicaban el comportamiento del sector, permitió establecer un número reducido y no redundante de indicadores financieros. De esta manera se pretende establecer unos valores de referencia para que las empresas de cada subsector puedan ser comparadas y se puedan generar estrategias de mejoramiento en los indicadores clave que estén afectando de manera directa a tales empresas.

4.3 AJUSTE DE DISTRIBUCIÓN DE VARIABLES

A continuación son mostrados los resultados de ajuste de cada una de las variables elegidas para estudiar los subsectores.

Cuadro 4.5 Ajuste del comportamiento de las variables de los sectores estudiados

VARIABLE	G5030		G5141		A0121	
	DISTRIBUCIÓN MÁS FRECUENTE	# DE VECES ENTRE 1995-2010	DISTRIBUCIÓN MÁS FRECUENTE	# DE VECES ENTRE 1995-2010	DISTRIBUCIÓN MÁS FRECUENTE	# DE VECES ENTRE 1995-2010
Ventas	Gamma y Erlang	4 veces cada una	Gamma	4	Logarítmica normal	6
Utilidad Operativa / Ventas	Laplace y Gumbel mínimo	2 veces cada una	Ninguna	0	Gumbel mínimo	6
Razón Corriente	Ninguna	-			Logarítmica normal	4
Patrimonio neto			Ninguna	-	Logarítmica normal	2
Pasivo total / Activo total	Normal	6	Normal y Gumbel mínimo	5 veces cada una	Erlang	4
Utilidad Neta	Laplace	2	Laplace	3	Laplace	4
Utilidad Neta / Ventas	Cauchy	2	Ninguna	0	Ninguna	0
Utilidad Operativa / Gastos No operativos	Cauchy	5	Cauchy	2	Cauchy	2
Flujo de caja / Gastos no operativos			Ninguna	0	Ninguna	0
Endeudamiento total / Ventas	Ninguna	-	Ninguna	-	F	2
Endeudamiento financiero / Ventas	Ninguna	-	Ninguna	-	Ninguna	0

VARIABLE	F4530		K7010		G5190	
	DISTRIBUCIÓN MÁS FRECUENTE	# DE VECES ENTRE 1995-2010	DISTRIBUCIÓN MÁS FRECUENTE	# DE VECES ENTRE 1995-2010	DISTRIBUCIÓN MÁS FRECUENTE	# DE VECES ENTRE 1995-2010
Ventas	Gamma	6	Logarítmica normal	3	Logarítmica normal	4
Utilidad Operativa / Ventas	Cauchy	2	Ninguna	0	Laplace	1
Razón Corriente	PearsonV	3	Logarítmica normal	3	Normal y logística	2 veces cada una
Patrimonio neto	Ninguna	0	Ninguna	0	Erlang	2
Pasivo total / Activo total	Normal	6	Gamma	3	Normal	7
Utilidad Neta	Ninguna	0	Logarítmica normal	1	Ninguna	0

Utilidad Neta / Ventas	Ninguna	0	Ninguna	0	Cauchy	6
Utilidad Operativa / Gastos No operativos	Cauchy	5	Ninguna	0	Cauchy	5
Flujo de caja / Gastos no operativos	Weibull desplazada	4	Logarítmica normal	7	Weibull desplazada	1
Endeudamiento total / Ventas	Logarítmica normal	2	Logarítmica normal	3	Ninguna	0
Endeudamiento financiero / Ventas	Ninguna	0	Ninguna	0	Ninguna	0

VARIABLE	J6399		F4521	
	DISTRIBUCIÓN MÁS FRECUENTE	# DE VECES ENTRE 1995-2010	DISTRIBUCIÓN MÁS FRECUENTE	# DE VECES ENTRE 1995-2010
Ventas	Logarítmica normal	4	Exponencial y Gamma	3 veces cada uno
Utilidad Operativa / Ventas	Ninguna	0	Ninguna	0
Razón Corriente	Logarítmica normal	3		
Patrimonio neto	Ninguna	0		
Pasivo total / Activo total	Gamma	4	Potencia 3	2
Utilidad Neta	Ninguna	0	Ninguna	0
Utilidad Neta / Ventas	Ninguna	0	Ninguna	0
Utilidad Operativa / Gastos No operativos	Ninguna	0	Ninguna	0
Flujo de caja / Gastos no operativos	Ninguna	0		
Endeudamiento total / Ventas	Weibull desplazada	2		
Endeudamiento financiero / Ventas	Ninguna	0	Ninguna	0

4.4 MODELO DE RIESGO DE INSOLVENCIA A TRAVÉS DE LA REGRESIÓN LOGÍSTICA BINARIA

Este primer modelo tiene el objetivo de ver la relación existente entre las variables independientes. Los cuadros presentados a continuación son los generados para la estructura completa del modelo de regresión logística.

4.4.1 Modelo de Regresión Logística 01

Cuadro 4.6 Resumen del procesamiento de los casos Regresión Logística modelo 01

Casos no ponderados		N	Porcentaje
Casos seleccionados	Incluidos en el análisis	300	100,0
	Casos perdidos	0	,0
	Total	300	100,0
Casos no seleccionados		0	,0
Total		300	100,0

El cuadro 4.6 indica que fueron seleccionados los 300 elementos del grupo de entrenamiento. Todos los casos fueron válidos y no se excluyó ninguna de las observaciones.

Cuadro 4.7 Codificación de la variable dependiente modelo Regresión Logística modelo 01

Valor original	Valor interno
Solvente	0
Insolvente	1

El cuadro 4.7 confirma los valores asignados por SPSS a la variable dependiente dicotómica cuando la empresa se presentó como Solvente "0" o como Insolvente "1".

Cuadro 4.8 Codificaciones de variables categóricas Regresión Logística modelo 01

CODIFICACIONES DE VARIABLES CATEGÓRICAS			
		Frecuencia	Codificación de parámetros
			1
Cambio en Flujo de Efectivo Neto	Efectivo neto disminuyó	160	1
	Efectivo neto aumentó	140	0
Estructura Financiera	Pasivos mayores o iguales que el patrimonio	106	1
	Pasivos menores que el patrimonio	194	0
Utilidad Neta	Utilidad neta negativa	198	1
	Utilidad neta positiva	102	0
Utilidad Operacional	Utilidad operacional Negativa	193	1
	Utilidad operacional positiva	107	0
Resultados de Ejercicios Anteriores	Negativo	150	1
	Positivo	150	0
Razón Corriente	Razón corriente inferior a 1	222	1
	Razón corriente superior o igual que 1	78	0

El modelo fue estructurado con 6 variables independientes de carácter cualitativo con dos estados. Según el cuadro 4.8, se presentó que:

- Para la variable CFL_EFE, 160 empresas disminuyeron su efectivo neto respecto al año anterior y 140 empresas aumentaron su efectivo neto respecto al año anterior.
- La variable EST_FIN muestra que 106 empresas tenían pasivos mayores o iguales que su patrimonio; 194 empresas tenían patrimonios menores o iguales a sus pasivos.
- UTI_NET: 198 empresas presentaron utilidad neta negativa; 102 empresas, utilidad neta positiva.
- UTI_OPE muestra que 193 empresas presentaron utilidad operacional negativa; 107 empresas, utilidad operacional positiva.
- REE_ANT: 150 empresas tenían resultados de ejercicios anteriores negativos; 150 empresas, con resultados de ejercicios anteriores positivos.
- RAZ_COR muestra que 222 empresas presentaron una razón corriente por debajo de 1; 78 empresas, una razón corriente igual o por encima de 1.

Se resalta que el cuadro 4.8 contiene en la tercera columna la codificación de cada uno de los estados de las variables independientes.

La categoría de referencia para cada una de las variables cualitativas es la que esté codificada con el valor de "0"; con lo cual se obtiene la probabilidad de ocurrencia de la variable dependiente teniendo el estado de la variable independiente con codificación "1" respecto a la misma variable con codificación "0".

Cuadro 4.9 Tabla de clasificación sin variables independientes Regresión Logística modelo 01

Observado			Pronosticado		
			Estado de Insolvencia		Porcentaje correcto
			Solvente	Insolvente	
Paso 0	Estado de Insolvencia	Solvente	0	150	,0
		Insolvente	0	150	100,0
		Porcentaje global			50,0

El cuadro 4.9 muestra un modelo preliminar en donde solo se ha tenido en cuenta una constante para el pronóstico de las observaciones. Solo con la constante, el modelo tuvo un poder de pronóstico del 50%. De las 150 empresas solventes, se pronosticó al total de estas como insolventes, lo que se tradujo en un 0% de porcentaje correcto de pronóstico para este grupo de observaciones. Para el grupo de las 150 empresas insolventes, se logró un pronóstico

del 100%, es decir, las 150 empresas insolventes fueron clasificadas como insolventes. Por lo que se tuvo un promedio de pronóstico correcto de 50%.

Cuadro 4.10 Constante en la ecuación Regresión Logística modelo 01

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 0 Constante	,000	,115	,000	1	1,000	1,000

El parámetro estimado en este modelo preliminar que es la constante β_0 , dio como resultado 0, lo que generó un $e^{\beta_0} = 1$. Esto significa que de acuerdo a las observaciones, por cada empresa insolvente hay una empresa solvente. Esto significa que según el grupo de entrenamiento elegido, por cada empresa solvente hay un 100% de probabilidades que haya una empresa insolvente. Lo que se traduce que es igual de probable elegir del grupo de entrenamiento una empresa que sea solvente o que sea insolvente.

Cuadro 4.11 Variables que no están en la ecuación Regresión Logística modelo 01

		Puntuación	gl	Sig.
Paso 0	Variables			
	RAZ_COR(1)	17,741	1	,000
	EST_FIN(1)	7,061	1	,008
	UTI_NET(1)	31,432	1	,000
	UTI_OPE(1)	13,961	1	,000
	REE_ANT(1)	85,333	1	,000
	CFL_EFE(1)	,054	1	,817
	GRA_APA	17,232	1	,000
	CON_ENC	27,891	1	,000
	Estadísticos globales	105,741	8	,000

La columna 2 del cuadro 4.11 contiene el valor X^2 resultado de la prueba de independencia entre la variable dependiente "Estado de insolvencia" (EST_INS) y cada una de las variables explicativas; esto con el objetivo de determinar si cada una de las variables independientes tenía alguna relación con la variable dependiente y así incluirlas en el modelo de regresión. Todas las variables, según el *test* de independencia de Chi-cuadrado, tuvieron relación con la variable dependiente, exceptuando la variable CFL_EFE, la cual presentó un p-valor por encima del nivel de significancia del 5% (8.17%), con lo que se aceptó la hipótesis nula de independencia entre la variable "Estado de insolvencia" y la variable "Cambio en el Flujo de Efectivo Neto".

En este punto se pudo determinar que la variable CFL_EFE no debería entrar en el modelo de regresión logística, sin embargo como se corrió este primer modelo con la inclusión de variables a través del método "introducir", todas las variables ingresaron para la evaluación del modelo, según las siguientes tablas de resultados.

Cuadro 4.12 Pruebas ómnibus sobre los coeficientes del modelo 01

		Chi cuadrado	gl	Sig.
Paso 1	Paso	129,846	8	,000
	Bloque	129,846	8	,000
	Modelo	129,846	8	,000

El resultado de la prueba Ómnibus (ver cuadro 4.12) muestra que la inclusión de las nuevas variables, al modelo, es significativa en comparación con el modelo que tiene en cuenta solo la constante. Esta prueba contrasta la hipótesis nula de que los coeficientes de todos los términos del modelo son iguales a cero, lo que indica, en este caso, que todo el grupo es significativo en el modelo.

Cuadro 4.13 Resumen del modelo 01

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	286,042 ^a	,351	,468

El cuadro 4.13 evalúa de forma global la validez del modelo. El estadístico "-2log de la verosimilitud" indica que la desviación del modelo es de 286,042, lo que no deja claro si el modelo se ajusta bien a los datos. R^2 de Cox y Snell para este caso indica que las variables independientes explican en un 35,1% la varianza de la variable dependiente; esto una proporción de explicación de varianza pobre. El R^2 de Nagelkerke, que es una corrección del estadístico anterior, indica un aumento en la explicación de la varianza, aumentándose hasta 46,8%; igualmente este resultado es pobre en explicación de la varianza de la variable dependiente. Pero se debe tener en cuenta que este estadístico es radical ya que incluso en un modelo de regresión perfecto, no se alcanza el valor de 1 (100%).

Cuadro 4.14 Tabla de clasificación Regresión Logística modelo 01

Observado			Pronosticado		
			Estado de Insolvencia		Porcentaje correcto
			Solvente	Insolvente	
Paso 1	Estado de Insolvencia	Solvente	117	33	78,0
		Insolvente	37	113	75,3
	Porcentaje global				76,7

El cuadro 4.14 muestra que de 150 empresas solventes, el modelo clasificó a 117 empresas efectivamente como solventes y a las 33 restantes como insolventes. Igualmente, de 150 empresas insolventes, clasificó a 37 empresas como solventes y a 113 empresas como insolventes. Todo lo anterior generó un poder de pronóstico del 76,7%, lo que significa que este primer modelo puede clasificar de forma correcta a 77 empresas de un total de 100 empresas.

Cuadro 4.15 Variables en la ecuación Regresión Logística del modelo 01

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	RAZ_COR(1)	-,665	,361	3,392	1	,066	,514
	EST_FIN(1)	1,219	,508	5,764	1	,016	3,383
	UTI_NET(1)	-,636	,404	2,470	1	,116	,530
	UTI_OPE(1)	,058	,392	,022	1	,883	1,060
	REE_ANT(1)	-1,705	,310	30,257	1	,000	,182
	CFL_EFE(1)	,222	,299	,552	1	,457	1,249
	GRA_APA	2,722	,889	9,381	1	,002	15,212
	CON_ENC	-1,486	,469	10,028	1	,002	,226
	Constante	,516	,906	,324	1	,569	1,676

Según muestra el cuadro 4.15 las variables UTI_NET, UTI_OPE, CFL_EFE y la constante no son significativas en este primer modelo de regresión logística.

4.4.2 Modelo de Regresión Logística 02

Se corrió el modelo automáticamente hacia adelante (forward method). Además de los parámetros incluidos en el modelo 01 para correr el SPSS, se indicó que se generara la bondad de ajuste de Hosmer-Lemeshow, el historial de iteraciones y las correlaciones de las estimaciones.

Cuadro 4.16 Historial de iteraciones con variables más significativas en el modelo 02

HISTORIAL DE ITERACIONES						
Iteración		-2 log de la verosimilitud	Coeficientes			
			Constant	REE_ANT(1)	CON_ENC	RAZ_COR(1)
Paso 1	1	326,792	1,067	-2,133		
	2	325,964	1,186	-2,372		
	3	325,964	1,19	-2,379		
	4	325,964	1,19	-2,379		
Paso 2	1	314,24	1,72	-1,955	-1,109	
	2	311,742	2,144	-2,241	-1,515	
	3	311,721	2,19	-2,267	-1,561	
	4	311,721	2,19	-2,268	-1,561	
Paso 3	1	307,188	2,258	-1,749	-1,309	-0,684
	2	303,823	2,887	-2,018	-1,776	-0,924
	3	303,786	2,963	-2,048	-1,834	-0,953
	4	303,786	2,964	-2,049	-1,835	-0,953
	5	303,786	2,964	-2,049	-1,835	-0,953

Cuadro 4.17 Prueba ómnibus sobre los coeficientes del modelo 02

PRUEBAS ÓMNIBUS SOBRE LOS COEFICIENTES DEL MODELO				
		Chi cuadrado	gl	Sig.
Paso 1	Paso	89,925	1	0
	Bloque	89,925	1	0
	Modelo	89,925	1	0
Paso 2	Paso	14,242	1	0
	Bloque	104,167	2	0
	Modelo	104,167	2	0
Paso 3	Paso	7,935	1	0,005
	Bloque	112,102	3	0
	Modelo	112,102	3	0

El resultado de la prueba Ómnibus (ver cuadro 4.17) muestra que la inclusión de las nuevas variables, al modelo, en los tres pasos, es significativa. Esta prueba contrasta la hipótesis nula de que los coeficientes de todos los términos del modelo son iguales a cero, lo que indica, en este caso, que todo el grupo es significativo en el modelo.

Cuadro 4.18 Bondad de ajuste de tres estadísticos al modelo 02

BONDAD DE AJUSTE CON TRES ESTADÍSTICOS			
Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	325,964 ^a	0,259	0,345
2	311,721 ^a	0,293	0,391
3	303,786 ^b	0,312	0,416
	DISMINUYÓ ↓	AUMENTÓ ↑	AUMENTÓ ↑

El estadístico “menos dos veces el logaritmo neperiano de la verosimilitud” muestra que al ir incluyendo las variables en paso 2 y 3, este va disminuyendo; lo que indica que el modelo efectivamente se fue ajustando más a medida que le eran incluidas más variables. El “R cuadrado de Cox y Snell”, así como el “R cuadrado de Nagelkerke” muestran que el poder explicativo de las variables en conjunto, aumenta a medida que se incluyen las demás variables significativas en el modelo (ver cuadro 4.18), de 25,9% a 31,2% y de 34,5% a 41,6%, respectivamente.

Cuadro 4.19 Tabla de contingencia para prueba de Hosmer y Lemeshow modelo 02

		Estado de Insolvencia = Solvente		Estado de Insolvencia = Insolvente		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	115	115	35	35	150
	2	35	35	115	115	150
Paso 2	1	51	52,751	12	10,249	63
	2	21	24,317	9	5,683	30
	3	24	21,886	6	8,114	30
	4	19	16,047	8	10,953	27
	5	18	13,56	21	25,44	39
	6	9	8,832	21	21,168	30
	7	4	6,073	26	23,927	30
	8	3	4,252	27	25,748	30
	9	1	2,283	20	18,717	21
Paso 3	1	45	45,943	8	7,057	53
	2	22	24,976	8	5,024	30
	3	22	22,22	8	7,78	30
	4	21	18,748	9	11,252	30
	5	17	13,7	13	16,3	30
	6	13	10,011	24	26,989	37
	7	6	6,456	24	23,544	30
	8	3	4,77	27	25,23	30
	9	1	3,177	29	26,823	30

En el cuadro 4.19 se compara la división en deciles de riesgo de las frecuencias observadas en cada uno de estos deciles con la frecuencia esperada en deciles del estado de insolvencia. Se observa que las frecuencias en deciles del paso 3 de las empresas solventes e insolventes tienen una proporción que es aproximada por las frecuencias de los deciles esperados. Lo anterior permite aseverar, al menos visualmente, que el modelo de regresión logística estructurado está bien ajustado a las observaciones reales del grupo de entrenamiento.

Cuadro 4.20 Estadísticos del modelo 02, factores de riesgo y/o protección

ESTADÍSTICOS DEL MODELO (FACTORES DE RIESGO Y/O DE PROTECCIÓN)									
		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95% para EXP(B)	
								Inferior	Superior
Paso 1a	REE_ANT(1)	2.379	.273	75.944	1	.000	10.796	6.322	18.435
	Constante	-1.190	.193	37.972	1	.000	.304		
Paso 2b	REE_ANT(1)	2.268	.280	65.626	1	.000	9.656	5.579	16.714
	CON_ENC	-1.561	.423	13.640	1	.000	.210	.092	.481
	Constante	-.077	.348	.049	1	.824	.926		
Paso 3c	RAZ_COR(1)	.953	.343	7.742	1	.005	2.594	1.325	5.078
	REE_ANT(1)	2.049	.289	50.110	1	.000	7.758	4.399	13.680
	CON_ENC	-1.835	.439	17.479	1	.000	.160	.068	.377
	Constante	-.038	.350	.012	1	.913	.963		

El cuadro 4.20 muestra que la variable REE_ANT (Resultados de ejercicios anteriores), para el paso 1 del modelo 02, es un factor de riesgo ya que está por encima de uno. En el paso 2, incluyendo la variable CON_ENC (Concentración del endeudamiento en el corto plazo), la segunda variable más significativa, la variable REE_ANT sigue como factor de riesgo y CON_ENC entra como factor de protección. En el último paso del modelo por el "método hacia adelante", quedan, finalmente, tres variables como explicadoras del comportamiento de la solvencia en una muestra representativa de las empresas PYME de Colombia. La variable REE_ANT queda como una factor de riesgo; CON_ENC queda como factor de protección y la variable RAZ_COR (Razón corriente) que como factor de riesgo.

En el cuadro 4.21 queda evidenciado el poder de pronóstico del modelo Regresión Logística con las variables mostradas en el cuadro 4.2. Para el paso 1 se tuvo que de 150 empresas solventes se clasificaron a 115 empresas efectivamente como solventes, y que de estas 150 empresas solventes, se clasificaron 35 empresas como insolventes, dando como resultado un porcentaje correcto de pronóstico del 76,7%. De 150 empresas insolventes (en el paso 1), se clasificaron 35 como solventes y a 115 empresas, efectivamente, como insolventes; con un

porcentaje de pronóstico correcto del 76,7%. Para el paso 2, con la inclusión de la variable CON_ENC, el porcentaje de pronóstico fue el mismo que en el paso 1. En el paso 3 (modelo con las tres variables: REE_ANT, CON_ENC y RAZ_COR), de 150 empresas solventes se clasificaron 114 empresas como solventes y 36 empresas como insolventes; de 150 empresas insolventes se clasificaron 33 como solventes y 117 efectivamente como insolventes, lo que generó un porcentaje promedio de pronóstico del 77%.

Cuadro 4.21 Tabla de pronóstico o clasificación para el modelo 02 de Regresión Logística

TABLA DE PRONÓSTICO					
Observado	Pronosticado				Porcentaje correcto
	Estado de Insolvencia		Solvente	Insolvente	
	Solvente	Insolvente			
Paso 1	Estado de	Solvente	115	35	76,7
	Insolvencia	Insolvente	35	115	76,7
	Porcentaje global				76,7
Paso 2	Estado de	Solvente	115	35	76,7
	Insolvencia	Insolvente	35	115	76,7
	Porcentaje global				76,7
Paso 3	Estado de	Solvente	114	36	76
	Insolvencia	Insolvente	33	117	78
	Porcentaje global				77

5. CONCLUSIONES

La fuerza productiva del país se puede sintetizar en el sector servicios ya que, en promedio, en el periodo 2003-2011, las actividades económicas están concentradas en el comercio al por menor y el comercio al por mayor; lo que sugiere que los modelos de desarrollo de todos los departamentos de Colombia están fundamentados en el comercio. Basta solo con ver la entrada masiva que han tenido las grandes superficies y centros comerciales extranjeros en el mercado colombiano. En Risaralda se observa un fuerte consumo de bienes importados. Todo lo anterior puede explicar el hecho de que en Risaralda, por ejemplo, no le están apuntando a la producción industrial o tecnológica, lo que podría generar beneficio a la sociedad ya que al trabajar con la producción, se puede crear valor agregado a los productos que se fueran a fabricar y generar una mayor rentabilidad. En promedio, el 15% de las empresas registradas ante la Superintendencia de Sociedades en el periodo 2003-2011, son empresas dedicadas al comercio al por menor de productos diversos. El 9%, en promedio, de las empresas que han presentado informes ante la Superintendencia de Sociedades, están dedicadas a la venta al por mayor; no es de extrañar ya que estas son las que surten a las compañías que se dedican al menudeo.

El comportamiento económico de todo el país está siendo reflejado en las políticas económicas de todos los departamentos ya que todos estos han seguido la misma tendencia en la composición de empresas; el departamento de Cundinamarca se destaca por una proporción de actividades económicas un tanto distanciada de la tendencia nacional; este departamento tiene una alta composición de empresas dedicadas a la producción de productos químicos. Para el departamento de Risaralda se observa, también, que la actividad de fabricación de prendas de vestir es representativa cerca de un 6% en la composición de la fuerza productiva del departamento de acuerdo a las pocas empresas que han registrado sus estados financieros ante la Superintendencia de Sociedades.

El comercio de vehículos, las actividades inmobiliarias, la construcción de obras residenciales y civiles y las actividades financieras, son las restantes actividades económicas que representan la dinámica de oferta de productos y servicios en Colombia. Pero se debe tener cuidado en el análisis ya que hay empresas de los demás subsectores que no son representativos en unidades de empresas, pero que sí lo son en nivel de activos, como por ejemplo, Ecopetrol y Claro.

Todo lo anterior deja al descubierto que Colombia es un país que no le está apuntando a la generación de valor agregado, lo cual es preocupante ya que esto genera una desaceleración de desarrollo, aunque haya crecimiento por el aumento de oferta de productos y servicios importados, e impide que las personas tengan iniciativas empresariales para aportar al desarrollo colectivo de la sociedad.

Los estados financieros son una buena herramienta para determinar variables de comportamiento de las empresas, ya que los registros de las cuentas del balance general, el

estados de resultados y el flujo de efectivo reflejan la situación estática y dinámica de estas en un momento determinado; a lo cual se le suma las aplicaciones de técnicas estadísticas y matemáticas para el análisis exhaustivo desde el punto de vista cuantitativo y pueden generar criterios para la toma de decisiones, análisis de correlaciones, pronósticos, optimización, etc.

En cuanto a la caracterización de las cuentas de los estados financieros que más peso tenían para la explicación del comportamiento de las empresas dentro de sus subsectores, se pudo apreciar que variables como nivel de ventas, razón corriente, utilidad operativa sobre ventas, nivel de endeudamiento y utilidad neta, son variables comunes a todos los subsectores, con una proporción de aparición para cada año de 80% en promedio; lo que se traduce en que estas variables son las que explican, en mayor medida, el comportamiento futuro de las empresas dentro del subsector al que pertenecen. Con lo que se podrían hacer pronósticos y determinar la tendencia de estos subsectores.

En el presente estudio de investigación se pudo determinar el comportamiento de ciertas cuentas de los estados financieros desde el año de 1995 hasta el año 2011 desde el punto de vista del ajuste estadístico. Lo que se encontró fue que la distribución de frecuencias de algunas de estas cuentas se ajustó en múltiples años a determinadas distribuciones teóricas continuas, como: gamma, erlang, logarítmica-normal, normal y exponencial. También se encontró que muchas de estas cuentas no se comportaron en ninguno de los años como alguna distribución teórica. Lo anterior permite estructurar modelos de simulación para los subsectores ya que si se tiene el comportamiento teórico de algunas de sus cuentas de los estados financieros, es más factible realizar simulaciones para la toma de decisiones, aproximadamente acertadas. Aunque no queda descartada la simulación para el comportamiento propio de las demás cuentas.

El riesgo de insolvencia puede ponerse en perspectiva de acuerdo a la definición de riesgo de crédito que es la expresión que se utiliza para analizar el riesgo, desde el punto de vista de las entidades financieras, de que sus clientes adeudados caigan en estado de insolvencia y no puedan cumplir con el pago de su capital e intereses. Entre el riesgo de insolvencia y el riesgo de crédito no se debe discernir sino en la perspectiva de los que utilizan la maquinaria para la estructuración, análisis e interpretación de los modelos de riesgo; para el caso de la utilización de la expresión "riesgo de insolvencia" es la que se utiliza para determinar el riesgo de que la empresa caiga en estado de insolvencia y no pueda pagar su deuda; el riesgo de crédito es lo mismo, pero es utilizado con el objetivo de determinar si una empresa está en la capacidad de endeudarse con la entidad; un valor agregado del riesgo de crédito es que se genera una puntuación en letras, aunque esto son puntuaciones que utilizan las calificadoras de riesgo en sus programas automatizados para el análisis de riesgo de crédito. En el presente estudio de investigación se buscó estructurar un modelo que permitiera determinar la probabilidad de que una empresa PYME cayera en estado de insolvencia de acuerdo a factores dicotómicos, representados por las cuentas de los estados financieros.

En medio del proceso para la determinación de un modelo para la medición del riesgo de crédito, resultó la idea de aplicar el modelo de regresión logística; inicialmente se tenía pensado

aplicar el análisis discriminante, pero se encontró con que de acuerdo a la información disponible en la elección de las cuentas de los estados financieros a estudiar, un muy bajo porcentaje permitió cumplir el supuesto de comportamiento normal de las variables explicadoras del estado de insolvencia. El único cambio que implicaba utilizar la regresión logística en vez del análisis discriminante, era que además de tener supuestos, para la estructuración del modelo, más flexibles, arrojaba la probabilidad de que las empresas objeto de estudio tenían de caer en estado de insolvencia de acuerdo a la determinación de factores de riesgo y de protección los cuales se generaron con el análisis de las variables elegidas como independientes para la explicación del estado de insolvencia.

El modelo de regresión logística para empresas PYME en Colombia, permitió establecer que las empresas que reflejaron en sus estados financieros una utilidad acumulada de ejercicios anteriores negativa, tendrían ocho veces más probabilidad de caer en estado de insolvencia; así como también, empresas con una razón corriente por debajo de uno tendrían tres veces más probabilidad de caer en estado de insolvencia que una empresa con una razón corriente por encima de uno.

El modelo de regresión logística para la medición del riesgo de insolvencia tuvo un poder de pronóstico del 77%, lo que indica que el modelo construido tuvo un poder de explicación relativamente alto, respecto al modelo de análisis discriminante que, normalmente, tiene un poder de pronóstico en las aplicaciones financieras del 75%. Encima de esto, la regresión logística supera al análisis discriminante ya que arroja una probabilidad de ocurrencia asociada.

6. REFERENCIAS BIBLIOGRÁFICAS

- Aguayo C., Mariano (2007). *Cómo hacer una Regresión Logística con SPSS® "paso a paso"*. [En línea], pp. 4. Disponible en: http://www.fabris.org/html/archivos/docuweb/Regres_log_1r.pdf
- Altman, E. I., R. Haldeman & P. Narayanan (1977). *ZETA Analysis: A new model to identify bankruptcy risk of corporations*. Journal of Banking and Finance 10, 29-54
- Ballón F., Raúl A. (2010). *Aplicación Modelo Altman a la Inversión de Reservas del Banco Central de Bolivia*. Gerencia de operaciones internacionales. Subgerencia de reservas. Departamento de negociaciones de Inversión. [En línea], [Citado el 2 de febrero de 2013]. Disponible en: <http://www.bcb.gob.bo/webdocs/ReservasInternacionales/DT-05-NotaTecnicaAplicacionModeloAltmanalaInversiondeReservaBdeBCB-08-10.pdf>
- Beltrán, Celina (2010). *Análisis de regresión logística aplicado a la clasificación textos académicos: Biometría y Filosofía*. [En línea]. pp. 2. Disponible en: <http://www.revistapistemologi.com.ar/biblioteca/05.Analisis%20de%20regresion%20logistica%20aplicado%20a%20la%20clasificacion%20textos%20academicos%20Biometria%20y%20Filosofia.pdf>
- Berenson, Mark L., Levine, David M. & Krehbiel, Timothy C. (2001). *Capítulo 9: Regresión lineal simple y correlación*. En: *Estadística para administración*. 2 ed. México: Editorial PEARSON Educación, pp. 504
- Botella R., P., Martínez B., M. A. & Alacreu G., M. (2003) *Regresión Logística*. Universidad Cardenal Herrera. [En línea], Págs. 10 y 11. Disponible en: <http://maphysco.ceuuch.es/~pbotella/IntroduccionalaRegresionLogisticaconR-Commander.pdf>
- Calderón S., July P., De Los Godos A., Luis (2008). *Regresión logística aplicada a la epidemiología*. [En línea], pp. 02. Disponible en: <http://www.inppares.org/revistasss/Revista%20IV%202009/10-%20Regres%20Logis.pdf>
- Capera, Laura; Cabrera, Wilmar; Morales, Miguel & Estrada, Darío (2012). *Un Mapa de Riesgo de Crédito para el Sistema Financiero Colombiano: Temas de Estabilidad Financiera*. [En línea], [Citado el 11 de febrero de 2013]. Disponible en: http://www.banrep.org/documentos/publicaciones/report_estab_finan/2012/ref_tema3_marr.pdf
- De Lara H., Alfonso (2005). *Capítulo 9: Modelos de riesgo de crédito*. En: *Medición y Control de Riesgos Financieros*. 3 ed. México: Editorial LIMUSA S.A, pp. 163
- Jones, F. L. (1987). *Current techniques in bankruptcy prediction*. Journal of accounting Literature, 6, pp. 131-164

- Jorion, Philippe (2004). *Capítulo 13: Riesgo Crédito. En: Valor en Riesgo: El nuevo paradigma para el control de riesgos con derivados*. México: Editorial LIMUSA S.A, pp. 267
- Knop, Roberto; Ordovás, Roland & Vidal, Joan (2004). Capítulo 3: Riesgo de Crédito. En: *Medición de Riesgos de Mercado y Crédito*. 1 ed. Barcelona: Editorial ARIEL S.A., pp. 187
- Levin, Richard I. & Rubin, David S. (2004). *Capítulo 13: Regresión múltiple y modelado. En: Estadística para administración y economía*. 7 ed. México: Editorial PEARSON Educación, pp. 566
- Martínez A., Oscar G. (2003). *Determinantes de fragilidad en las empresas colombianas: Banco de la República*. [En línea], [Citado el 8 de febrero de 2013]. Disponible en: <http://www.banrep.org/docum/ftp/borra259.pdf>
- Mejía Q., Rubí C. (2011). *Administración de Riesgos Un Enfoque Empresarial*. [En línea], [Citado el 10 de febrero de 2013]. Disponible en: <http://www.eafit.edu.co/escuelas/administracion/consultorio-con-tabla/Docu%20men%20de%20clase%2016%20Mapa%20de%20Riesgos.pdf>
- Moliner, Luis M. (2001). *La Regresión Logística*. [En línea], Disponible en: <http://www.sehelha.org/pdf/rlogis1.pdf>
- Mures Q., Jesús, García G., Ana & Vallejo P., M Eva (2005). *Aplicación del Análisis Discriminante y Regresión Logística en el estudio de la morosidad en las entidades financieras. Comparación de resultados*. [En línea], [Citado el 2 de febrero de 2013]. Disponible en: http://www3.unileon.es/pec/vnia/01/01_17_5_199.pdf
- Nava P., Luis & Surendra P., Sinha (2007). *Ajuste e interpretación de los modelos de regresión logística con variables categóricas y continuas*. [En línea], pp. 10. Disponible en: <http://med.javeriana.edu.co/publi/vniversitas/serial/v49n1/4-AJUS%20TE%20MODELO.pdf>
- Pérez F., Napoleón & Martínez H., David (2002). *Regresión logística SPSS*. [En línea], pp. 05. Disponible en: <http://pendiente de migracion.ucm.es/info/dosis/Preventiva/octoberado/TEMA13.pdf>
- Saavedra G., María L., Saavedra G., Máximo J. (2010). *Modelos para medir el riesgo de crédito de la banca*. [En línea], el artículo se recibió el 27 de abril de 2009 y se aprobó el 21 de mayo de. [Citado el 09 de febrero de 2013]. Disponible en: http://cuadernoadministracion.javeriana.edu.co/pdfs/Cnos_Admon_23-40_12_MSaavedra.pdf
- Tamara A., Armando L., Aristizábal V., Raúl E. & Velasquez C., Ermilson (2010). *Modelación de riesgo crediticio como elemento fundamental en el cálculo de la pérdida esperada en una institución financiera*. [En línea], [citado el 2 de febrero de 2013]. Disponible en:

http://www.eafit.edu.co/m_inisitios/cienciaa_plicada/Docum_ents/Model_uci%C3%B3n%20del_%20riesgo_20crediticio%20para%20estimar%20la%20p%C3%A9rdida%20esperada%20en%20una%20instituci%C3%B3n%20financiera.pdf

Tascón F., María T. & Castaño G., Francisco J. (1998). *Predicción del fracaso empresarial: una revisión*. [En línea]. [Citado el 11 de febrero de 2013]. Disponible en: http://personales.ya.com/aeca/pub/online/comunicaciónes_xvcongreso_aeca/cd/63b.pdf

Webster, Allen L. (2000). *Capítulo 12: Regresión múltiple y correlación*. En: *Estadística aplicada a los negocios y la economía*. 3 ed. Bogotá: Editorial Irwin Mc GRAW HILL, pp. 382