

**UNIVERSIDAD NACIONAL DE INGENIERÍA**  
**FACULTAD DE INGENIERÍA ECONÓMICA, ESTADÍSTICA Y CC.SS.**  
**ESCUELA PROFESIONAL DE INGENIERÍA ECONÓMICA**



**TESIS**

**“Análisis de correlación entre los factores socioeconómicos y el tiempo de operación de las MYPE en Lima hasta el año 2013”**

**Para obtener el Título Profesional de Ingeniero Economista**

**ELABORADO POR**

Bach. Carlos Miguel Yalta Vargas

**ASESOR**

Dr. Alfonso López-Chau Nava

**Lima – Perú**

**2017**

## **Dedicatoria**

*A mis padres Sara y Carlos en agradecimiento por motivarme de forma constante a lograr mis metas y objetivos*

*A mi hermano Favio, con quien he aprendido que la creatividad es la mejor solución a los problemas que se nos presentan día a día.*

*A mi familia que siempre ha sido de apoyo durante mis años de estudio.  
A mis bellas abuelitas Rita y Aurora, así como a mi gran abuelo Miguel.*

## **Agradecimientos**

*Un excelso reconocimiento a mi alma mater la Universidad Nacional de Ingeniería y la facultad de ingeniería económica, estadística y CC.SS que me acogió y preparó para afrontar como un profesional íntegro los desafíos venideros.*

*Mi más sincero agradecimiento al Dr. Alfonso López Chau-Nava, asesor de tesis, por su valiosa orientación, recomendaciones y comentarios de los borradores realizados para el que el presente trabajo de investigación pueda brindar nuevas luces y enfoques a los problemas de nuestra sociedad moderna.*

*Agradezco a mis amigos y colegas por motivarme de forma competitiva a seguir preparándome y avanzando en mi carrera profesional, a todos ellos muchas gracias.*

# Tabla de contenido

<b>RESUMEN EJECUTIVO</b> .....	6
<b>INTRODUCCIÓN</b> .....	7
<b>CAPÍTULO I: PLANTEAMIENTO DE LA INVESTIGACIÓN</b> .....	10
1.1. Planteamiento del problema .....	11
1.2. Formulación del problema.....	16
1.3. Objetivos.....	16
1.3.1. Objetivo general .....	16
1.3.2. Objetivos específicos .....	17
1.4. Justificación .....	17
<b>CAPÍTULO II: MARCO TEÓRICO</b> .....	19
2.1. Antecedentes.....	20
2.2. Marco Conceptual .....	23
2.3. Teoría general.....	31
2.3.1. Modelo para datos de conteo .....	31
2.4. Teorías específicas .....	35
2.4.1. Regresión de Poisson.....	35
2.4.1.1. Estimación y Cuasi estimación de máxima verosimilitud para la Regresión de Poisson 36	
2.4.1.2. Interpretación de los coeficientes de regresión .....	38
2.4.2. Sobredispersión .....	39
2.4.3. Modelo Binomial Negativo .....	41
2.5. Hipótesis .....	44
<b>CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN</b> .....	46
3.1. Procedimiento de investigación .....	47
3.2. Secuencia metodología del análisis econométrico.....	47
3.2.1. Visión macro económica del problema .....	48
3.2.2. Enfoque micro econométrico del problema.....	49
3.2.3. Predicción de escenarios .....	50
3.3. Diseño de la investigación .....	51
3.4. Población y muestra .....	52
3.4.1. Población .....	52
3.4.2. Muestra .....	52
3.5. Operacionalización de las variables.....	53
3.6. Instrumentos de recolección de datos .....	55

3.7. Procesamiento y análisis.....	55
<b>CAPÍTULO IV: DESARROLLO DE LA METODOLOGÍA.....</b>	<b>56</b>
4.1. Análisis macro económico .....	57
4.1.6. PBI vs Régimen MYPE .....	57
4.1.6. Términos de intercambio vs Régimen MYPE .....	68
4.2. Análisis micro econométrico .....	75
4.2.6. Descripción de las variables del modelo.....	75
4.2.2. Análisis estadístico de las variables. ....	78
4.2.3. Estimación inicial y pruebas de ajuste .....	81
4.2.4. Test de sobredispersión.....	82
4.2.5. Estimación final.....	82
4.2.6. Test de variables omitidas .....	83
4.3. Predicción de escenarios .....	85
<b>CAPÍTULO V: ANÁLISIS DE RESULTADOS.....</b>	<b>86</b>
5.1. Factores socioeconómicos analizados y su nivel de correlación. ....	87
5.2. Gráficos de los resultados obtenidos en la regresión.....	89
<b>CAPÍTULO VI: DISCUSIÓN .....</b>	<b>94</b>
6.1. Discusiones .....	95
6.2. Posibilidades y limitaciones del estudio .....	95
6.2.1. Posibilidades .....	95
6.2.2. Limitaciones.....	95
6.3. Conclusiones.....	96
6.4. Recomendaciones.....	98
Bibliografía	99
ANEXOS	101

## RESUMEN EJECUTIVO

**Palabras Claves:** Microeconomía, MYPES, Regresión de Poisson, Modelo Binomial Negativo, Test de sobredispersión, predicción de escenarios.

Existe actualmente un horizonte de oportunidades en innovación y emprendimiento, ya sea netamente tecnológico o de bienes y servicios para atender una demanda específica, o sencillamente conocer cómo desarrollar operativamente determinado negocio, en ambos casos el producto final mediato es la creación de una Micro empresa que luego seguirá creciendo hasta avanzar a la categoría de Pequeña, donde continuarán afrontando diversas dificultades, así como la necesidad de implementar nuevas estrategias de costos y ventas, en su batalla por avanzar al siguiente nivel.

El presente trabajo tiene como objetivo analizar la correlación existente entre diversos factores y su influencia directa en el tiempo de operación de las MYPES, a partir de la realización de regresiones discretas para posteriormente evaluar escenarios futuros donde estas empresas puedan seguir operando por un periodo mayor al promedio que es de dos años en promedio, ya que debido al potencial y versatilidad de las Micro y Pequeñas empresas para atender la demanda de Grandes negocios es fundamental encontrar la combinación de factores sociales, económicos y ahora tecnológicos también que den como resultado un avance gradual de las empresas de este sector a una consolidación empresarial a nivel nacional. Esta contribución espera permitir que posteriormente se puedan realizar reformas y normativas tanto a nivel de gobiernos locales como el central que permita al micro y pequeño empresario continuar desarrollando su modelo de negocio atendiendo a diversos sectores de la economía, generando además una alternativa atractiva para los jóvenes la creación y desarrollo de negocios propios con un escalamiento gradual garantizado durante los años, que termine brindado mejor calidad de vida e ingresos.

## INTRODUCCIÓN

El desempeño durante los últimos quince años de las Micro y Pequeñas empresas (MYPES) es sin lugar a dudas uno de los pilares más importantes en el desarrollo de la economía local de forma inmediata, sino también generando empleos contribuyendo a la disminución gradual del nivel de desempleo. Según el enfoque adoptado en esta investigación son los pequeños negocios los que dinamizan y acentúan el crecimiento económico, la reducción del desempleo y en el mediano plazo una reducción de los niveles de pobreza en países en desarrollo como el nuestro. Desde un punto de vista económico, dentro del proceso de crecimiento de una empresa MYPE, está demanda mayor fuerza laboral. Adicionalmente, tiene como consecuencia inmediata el incremento de sus ventas, que impacta en los beneficios sin mencionar la necesidad de seguir formalizándose generando una cultura de pago de tributos.

No obstante, las MYPES en su largo y arduo camino por la consolidación de su modelo de negocio descubren diversos obstáculos que restringe o limita su avance durante los primeros dos años de funcionamiento formal. Diversos estudios enfocados en los factores que llevarían al fracaso de este tipo de empresas concluyen en que la tasa de mortalidad de las micro y pequeña empresas es mayor en países en desarrollo que en los países desarrollados. Una de las causas que llevaría a poseer estos niveles elevados de cierre de empresas es debido a que se ven obligadas a implementar estrategias específicas de corto y largo plazo para reducir la tasa de mortalidad a la que se encuentran expuestas por ser nuevas, ya que siempre existe un nivel de riesgo elevado durante los primeros años de inversión y puesta en marcha.

Dentro del territorio nacional, la información estadística que se maneja de la base de datos del Instituto Nacional de Estadística e Informática (INEI), referidas a las micro y pequeñas empresas, reflejan que la participación de las MYPES en el Producto Bruto Interno

hasta el 2013 ha sido del 55% y han representado el 90% del universo de empresas en el Perú y el 85% de la población económicamente activa.

Es cierto que las MYPES se encuentran funcionando dentro de un marco económico y normativo adverso, contrario a la promoción de este sector importante y creciente de la economía nacional, ya que los índices de supervivencia son inferiores a los que poseen este tipo de empresas en otras regiones de Latinoamérica, sin mencionar que la consolidación del modelo de negocio se ve afectado por el contexto de criminalidad y corrupción de instituciones en muchos de los casos.

Según el estudio del Mirador Global de Emprendimiento (Global Entrepreneurship Monitor) realizado en el Perú, se estimó que 9.2 millones de ciudadanos estaban involucrados de forma directa o indirecta en una actividad emprendedora, de los cuales las tres cuartas partes tuvieron un periodo de operación de entre 0 y 42 meses, mientras que la parte restante llegó a sobrepasar este periodo, brindándoseles la clasificación de consolidadas. Dentro de este último grupo, solamente el 3% llevó a cabo proyectos de desarrollo de nuevos productos o incursionó en nuevos mercados.

Estos datos reflejan la realidad diaria operativa de las MYPES, además de incrementar la importancia de estudios dedicados a la investigación de los factores que limitan el crecimiento y desarrollo de estas empresas, de manera que se puedan consolidar en sus respectivas áreas de negocio una mayor proporción de negocios e ideas de emprendimiento. No obstante, la literatura actual refleja que no se han hecho estudios que identifiquen los factores que limitan el crecimiento y desarrollo de las MYPES a pesar de su gran importancia en la economía. En este sentido, el presente estudio tiene por objetivo seleccionar diversos factores socioeconómicos evaluados previamente y asumiendo que son una causa importante en la limitación del tiempo de operación para, de esta manera, realizar un proceso de correlación entre los años de operatividad desde el año de fundación hasta el 2013. Además se utilizan métodos econométricos y de probabilidades para poder aproximar una predicción sobre que posibles combinaciones de factores incrementan esta posibilidad de seguir



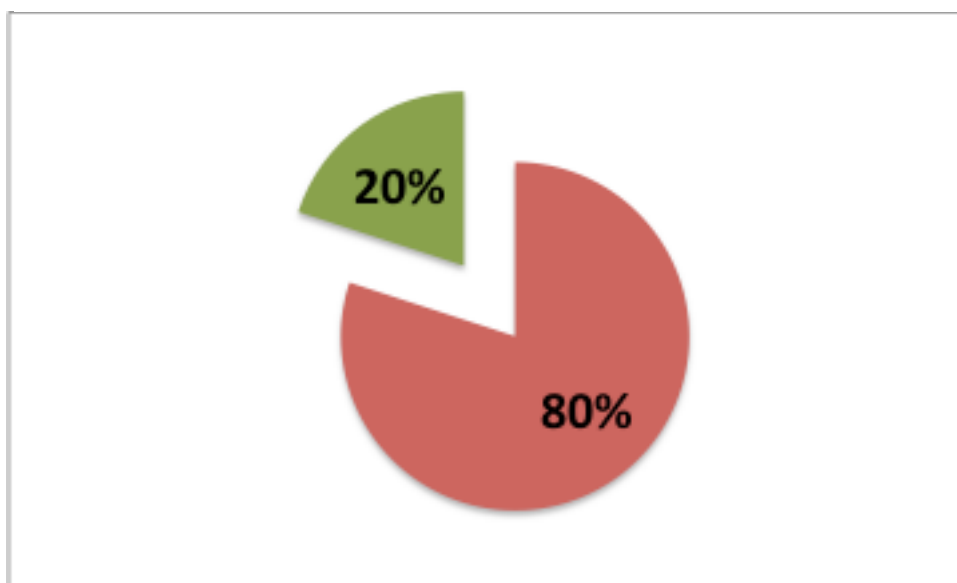
funcionando más de tres o cuatro años, permitiendo su consolidación y reduciendo la tasa de mortalidad que es una característica en este tipo de negocios.

# CAPÍTULO I: PLANTEAMIENTO DE LA INVESTIGACIÓN

## 1.1. Planteamiento del problema

El tiempo de operación de las MYPES o años de funcionamiento luego de registrarse dentro de dicho régimen se encuentra influenciado por diversos factores sociales y económicos tanto internos como de fuera de la empresa, siendo en muchos casos el factor endógeno el cual determinara si continúan produciendo dentro de los diversos sectores de la economía. Para poder comprender las causas inherentes a este particular comportamiento de los agentes económicos es necesario realizar un análisis de regresión micro econométrico para determinar la relación entre cada una de las causales con el objetivo de poder realizar una simulación que permita diseñar mejores estrategias que tengan como fin el incremento en los años de operación.

**Gráfico 1 Porcentaje de empresas con tiempo de vida menor a 3 años**



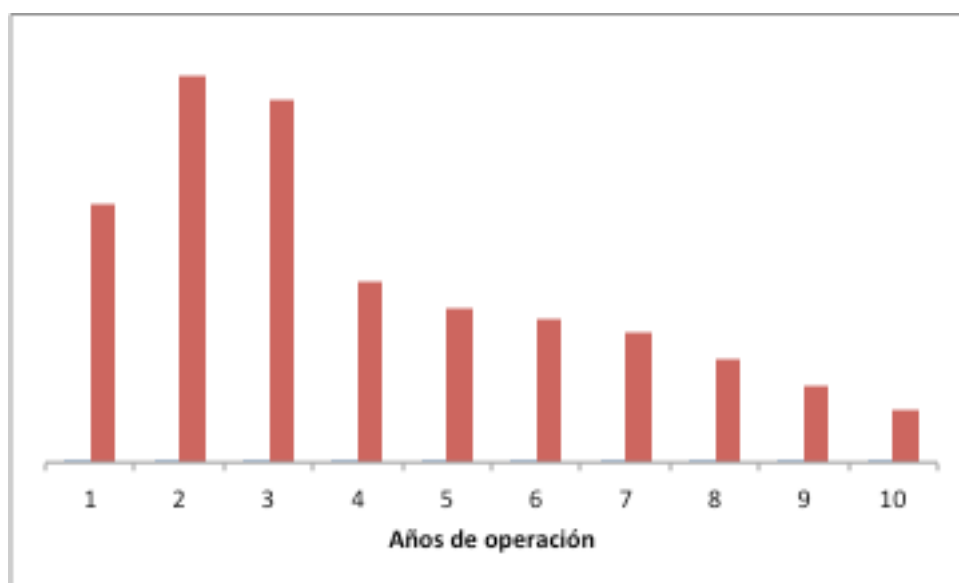
**Fuente: Encuesta de la Micro y Pequeña empresa (EMYPE)**

**Elaboración: propia**

En nuestro país el 80% de las MYPES no dura más de 3 años, es decir para el año 2012 de 100 empresas creadas, conformadas o constituidas en el año 2010, solamente 20 de estas podrán sobrellevar adelante su gestión y el resto decidirá dejar de funcionar. Una característica importante de las MYPES es que en su mayoría son negocios familiares, es decir, los empleados y empleadores poseen lazos de parentesco haciendo de las decisiones de la empresa, decisiones emocionales que involucran otros problemas fuera del negocio.

Asimismo diversos estudios sobre el problema y los factores que llevan a estas empresas a funcionar por un corto periodo de tiempo buscan el problema analizando las grandes inversiones desde la perspectiva macroeconómica, que según estos estudios, son la principal fuente de generación de empleo para las MYPES. Sin embargo al considerar solo variables macro en el tiempo no es posible conocer las causas internas o externas para dejar de operar, en consecuencia se pierde conocimiento de más de 2000 empresas<sup>1</sup> de dicho sector que son encuestadas. Indicando la deficiencia metodológica existente para el desarrollo de políticas públicas, ya sean de gobierno local o central, para fomentar el desarrollo, crecimiento y consolidación del mayor sector empresarial del país.

### **Gráfico 2 Clasificación de micro y pequeña empresa por años de operación**



**Fuente: Encuesta de la Micro y Pequeña empresa (EMYPE)**

**Elaboración: propia**

El inadecuado empleo de las herramientas metodológicas como la selección del modelo que se ajuste al comportamiento de la información puede ocasionar que al momento de realizar la predicción de escenarios tengamos un comportamiento simulado que no nos permita entender con exactitud el problema, generando además que las instituciones públicas tengan

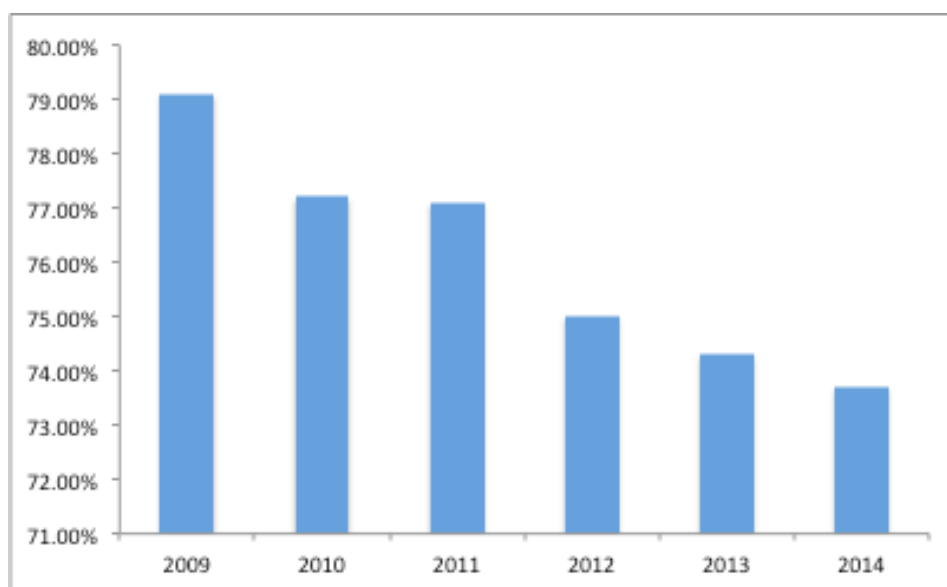
<sup>1</sup> Cada dos años el Instituto Nacional de Estadística e Informática realiza la encuesta EDMYPE donde se registran las características de ese grupo. A la fecha la última encuesta se realizó el 2013.

información errónea con la cual diseñar las políticas. Por lo tanto es importante conocer la distribución estadística que no proveerá el mejor entendimiento aproximando la realidad a la simulación de escenarios donde se podrán evaluar diversos campos de acción que permitan incrementar el tiempo de funcionamiento.

La dificultad para realizar una regresión micro econométrica específicamente se inicia a partir de la selección del modelo con el cual se traza la ruta para relacionar las variables exógenas y endógenas, ya que cada modelo sigue supuestos o restricciones que permitan su mejor aprovechamiento. Otro punto importante recae sobre los estadísticos del modelo, los cuales permiten entender la forma y ajuste de nuestra regresión, la cual puede tener defectos que impacten la predicción siendo por tanto imprescindible encontrar los métodos de reducción para cada uno.

En referencia a la predicción de escenarios para el modelo de datos de conteo esta ayuda a asegurar y garantizar una adecuada aplicación de las modificaciones o políticas públicas que se adecuen como objetivo de las MYPES para mejorar su capacidad de administración, y por ende lograr un crecimiento sostenido superior a los 3 años. Entonces es clara la limitación en el campo del diseño de políticas nacionales de impacto a este sector cuando no se analiza adecuadamente con información estadística histórica y por el contrario se aplican sanciones o

### ***Gráfico 3 Porcentaje de trabajadores en MYPES laborando en el sector informal***



Fuente: Produce – Sunat - Encuesta de la Micro y Pequeña empresa (EMYPE)

Elaboración: propia

penalizaciones que empujan a muchos de los empresarios al sector informal nacional<sup>2</sup>.

Es evidente que cada operador político desea involucrarse en las reformas públicas que mejoren su apreciación para poder seguir postulando a cargos de mayor relevancia. Por ende busca impactar a una mayor cantidad de personas realizando los menores cambios posibles que perjudiquen sus aspiraciones. De esta forma muchos diseñan leyes o normas en base a estudios cualitativos de percepción que no analizan información histórica además de no involucrar análisis econométrico para casos donde se podrían simular escenarios bajo cualquier tipo de condición que implique una política pública nueva.

La metodología de análisis cualitativo es la de mayor aceptación para estos casos de estudios ya que pretenden explicar en base a cuatro factores el desempeño de las Micro y Pequeñas Empresas tales como el desarrollo profesional, la motivación por el crecimiento, habilidades gerenciales para motivar dicho crecimiento y acceso a los recursos y la demanda del mercado (Perren, 1999). Cabe resaltar que dentro de las características seleccionadas para el análisis cualitativo, un factor común dentro del enfoque micro econométrico se encuentra en las habilidades gerenciales para motivar el crecimiento y el desarrollo profesional. Para el primer caso dentro de las habilidades de gerencia se tiene la información de la Encuesta de la Micro y Pequeña Empresa (EMYPE)<sup>3</sup> con lo cual podemos conocer de cerca los principales métodos a los cuales recurren los empresarios para poder desarrollar sus modelos de negocio que pueden ser desde asociaciones o implementación de mecanismos informáticos que reducen los costos e incrementan la exposición de sus productos.

Y en el caso del desarrollo profesional ambos enfoques resaltan el efecto crucial que tiene en la micro o pequeña empresa la formación profesional de sus dirigentes, con lo cual pueden impulsar un desarrollo a mayor escala empleando herramientas que para un gerente sin formación académica pueda resultar complicado. Afectando además dentro del trato a los

---

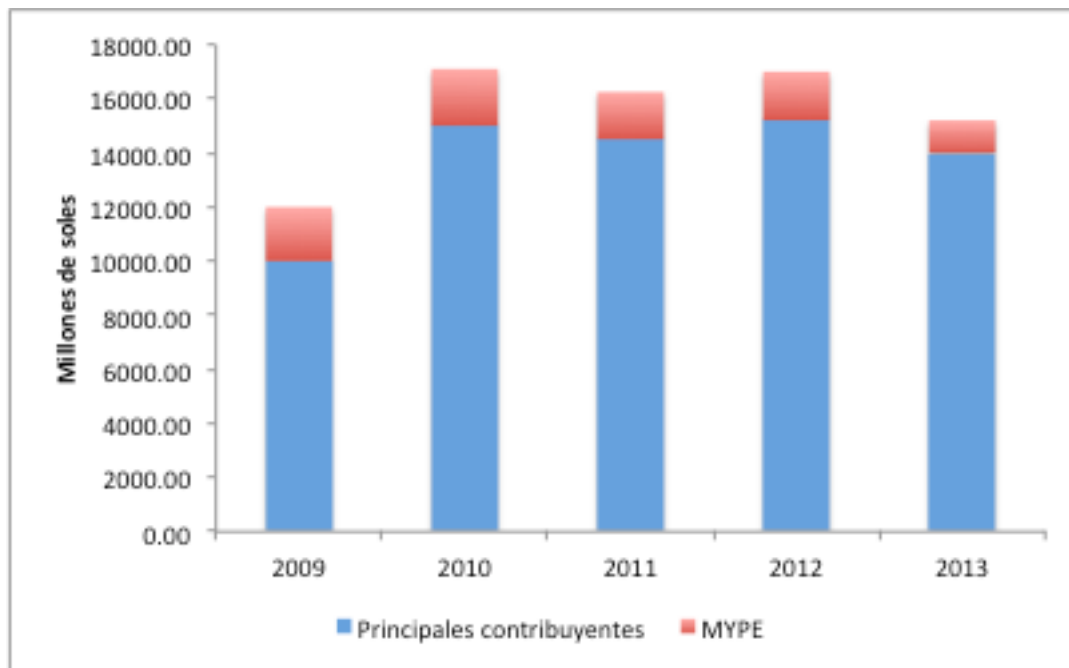
<sup>2</sup> En el año 2015 dentro del área urbana la población ocupada con empleo informal representó el 65.9% de la PEA. Según la Encuesta Nacional de Hogares ENAHO del INEI.

<sup>3</sup> En el marco del presupuesto por resultados, el Ministerio de la Producción (Produce) formuló el programa estratégico "Incremento de la Productividad de las MYPE". El programa propuso la construcción de una línea base para evaluar la intervención de los servicios de PRODUCE en la micro y pequeña empresa. El INEI es la institución encargada de obtener la información estadística cuantitativa y cualitativa para calcular y estimar los valores de los indicadores que permitan caracterizar los aspectos más sustantivos de las MYPE ubicadas en las capitales de cada departamento. (Dirección Nacional de Censos y Encuestas del Instituto Nacional de Estadística e Informática, 2011)

colaboradores del negocio, los cuales en la mayoría de los casos poseen una relación amical o consanguínea con los dueños dificultando de esa manera el desarrollo del proyecto o idea de negocio.

De esta forma en ambos casos el análisis involucra factores internos que afectan directamente a las MYPE y que para el caso de estudio desarrollado en este trabajo son tan importantes como los factores externos los cuales en su mayoría provienen de las políticas diseñadas por las municipalidades distritales o gobiernos locales, las cuales plantean una visión sancionadora y fiscalizadora de los diversos comercios que operan en su jurisdicción dejando de lado muchas veces la capacitación a estos negocios con el objetivo de reducir la evasión a las normas. Además la Superintendencia Nacional de Administración Tributaria (Sunat) posee una política, hasta el 2015, de recaudación y fiscalización masiva con el objetivo de recaudar la mayor cantidad de tributos evadidos para poder financiar los programas sociales. Sin embargo las medidas adoptadas contribuyeron a un mayor incremento de la informalidad y evasión tributaria cuando se debió optar por entender las causas que motivan a un empresario de la micro o pequeña empresa a defraudar a la entidad recaudadora de tributos.

**Gráfico 4 Impacto en la recaudación por parte de las MYPES**



Fuente: Sunat – BCRP - Encuesta de la Micro y Pequeña empresa (EMYPE)

Elaboración: propia

En conclusión el problema de estudio del presente trabajo de investigación consiste en analizar la relación existente entre diversos factores sociales y económicos de las micro y pequeña empresas para explicar el corto periodo de funcionamiento o duración en el mercado nacional.

## 1.2. Formulación del problema

Las principales cuestiones que se plantean para la presente investigación son las siguientes:

- ¿En qué medida se relacionan los factores socioeconómicos y la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región Lima hasta el año 2013?
- ¿En qué medida se relacionan el nivel de educación de los dueños y la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región Lima hasta el año 2013?
- ¿En qué medida se relacionan el nivel de formalidad del negocio y la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región Lima hasta el año 2013?
- ¿En qué medida se relacionan el tipo de organización bajo la cual realizan sus operaciones de negocio y la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región Lima hasta el año 2013?
- ¿Cuáles son los factores sustantivos en la predicción de escenarios que permitan impulsar la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región Lima hasta el año 2013?

## 1.3. Objetivos

### 1.3.1. Objetivo general

- Establecer el nivel de relación entre los factores socioeconómicos y la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región de Lima hasta el año 2013.



### 1.3.2. Objetivos específicos

- Establecer el nivel de relación entre el nivel de educación de los dueños y la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región Lima hasta el año 2013.
- Establecer el nivel de relación entre el nivel de formalidad del negocio y la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región Lima hasta el año 2013.
- Establecer el nivel de relación entre el tipo de organización bajo la cual realizan sus operaciones de negocio y la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región Lima hasta el año 2013.
- Establecer los factores sustantivos y cruciales en la predicción de escenarios que permitan impulsar la sostenibilidad de las Micro y Pequeñas empresas (MYPE) en la región Lima hasta el año 2013.

### 1.4. Justificación

Se proponen tres ideas principales que justifican la investigación del problema a tratar.

- Comprender las formas de administrar de las MYPE, las cuales conforman más del 40% de las empresas siendo un factor decisivo en el funcionamiento de diversos sectores económicos así como de asignación de salarios. La importancia radica en que algunos sectores solo funcionan de forma legal en promedio tres años, entonces se necesita evaluar los determinantes de años de operación que permitan hacer los cambios necesarios como consecuencia funcionar por más tiempo.
- La desaceleración económica como afecta el desempeño y tiempo de funcionamiento de las MYPE, el crecimiento para este año 2017 de la economía se calcula en menos de 2.5%, significa que de acuerdo a los índices macroeconómicos de inversión, las grandes empresas están retrasando sus proyectos para evaluar las próximas medidas o políticas a tomar por el presente gobierno. Pero el efecto es mayor cuando tienes a micro, pequeñas y medianas empresas con una gran dependencia de estos sectores

empresariales grandes. Por tal motivo es necesario ver el efecto respecto al último año de nuestro caso de estudio (2013) sobre el tiempo de operación y plantear alternativas que disminuyan el impacto económico y social.

- La inversión pública no está dirigida a estos sectores, es importante resaltar el hecho que muchos negocios MYPE ven una oportunidad de crecimiento con el sector público a través de las compras del estado. Por lo tanto se necesita cuantificar el efecto de excluir a este sector empresarial y plantear soluciones que no perjudiquen el tiempo de funcionamiento.

En resumen, el propósito de este trabajo de investigación es brindar un mayor entendimiento sobre los factores que afectan al sector con mayor cantidad de empresas en cuanto al tiempo de funcionamiento. Generando modelos econométricos precisos para simular las diferentes propuestas que puedan plantear las instituciones públicas, además de otorgar una visión cuantitativa mostrando el impacto en los costos de oportunidad por cada factor seleccionado en el estudio.

# CAPÍTULO II: MARCO TEÓRICO

## 2.1. Antecedentes

Dentro del territorio peruano existen escasos estudios y trabajos de investigación relevantes respecto al tiempo de funcionamiento y operación de las Micro y Pequeña empresas (MYPES). Las estadísticas generales para América Latina muestran un promedio anual, que evidencia la realidad donde 8 de 10 empresas fracasan antes de los cinco años y solamente 1 de 10 logra alcanzar la madurez corporativa con más de 10 años de funcionamiento (Soriano, 2005)<sup>4</sup>

Para los dueños de estas micro y pequeñas unidades de negocio muchas de las razones del fracaso o lentitud al momento de desarrollarse proviene de factores externos como la demora en los trámites burocráticos dentro de las municipalidades distritales, exceso de papeleo para regularizar con entes fiscalizadores, y sobre todo el tema tributario donde las MYPES dependiendo del régimen en el cual se encuentran deben mantener costos mensuales altos de contabilidad con el objetivo de cumplir lo estipulado por la Superintendencia Nacional de Administración Tributaria (SUNAT). Por lo tanto desde el enfoque de los micro y pequeño empresarios existe una sensación de marginación por parte del Gobierno que no les permite progresar mediante las actividades que desarrollan.

Sin embargo los analistas empresariales desean conocer las causas o factores que propician el fin del funcionamiento de un elevado porcentaje de negocios que conforma en su totalidad alrededor del 90% de empresas registradas<sup>5</sup>, empezando por un análisis de los problemas internos y externos que experimentan. Por ende se establece que la causa principal estaría dentro de las propias empresas identificando a los responsables de la dirección y la forma en que gestionan los procesos diarios que involucran especialización en varios temas como informática o administración básica, propiciando que resalte un indicador importante como es el de la educación adquirida por el líder, la cual permitiría según estudios (Avolio, Mesones, & Roca, 2011) realizar una mejor planificación empresarial teniendo en cuenta las decisiones claves sobre las cuales deben dirigir los negocios.

---

<sup>4</sup> El artículo de Soriano muestra un análisis sobre el impacto de los factores internos y externos en el funcionamiento del negocio.

<sup>5</sup> Fuente: Instituto Nacional de Estadística e Informática – Notas de prensa Setiembre 2013

Dentro de los estudios realizados para la región se abordan diversos casos donde los factores involucrados en el desarrollo o caída de las MYPES son sociales, educacionales o gubernamentales, sin embargo el tema de la financiación y fondos de financiamiento (Lira, 2009) dirigidos a estos sectores solamente son investigados bajo la perspectiva de la contabilidad, por ende se indaga para este trabajo el enfoque financiero que permite diseñar planes estructurados y soportar la expansión de dichas empresas.

Desarrollando la idea del impacto de las finanzas en este sector, se sabe que el sistema financiero en nuestro país se encuentra diseñado para brindar diversos servicios a costos competitivos direccionados por la política monetaria estable, que permite arriesgar a los bancos un mayor nivel de activos. Adicionalmente, diseñan planes exclusivos de micro préstamos con tasas considerablemente bajas comparadas con los préstamos regulares. Toda esta nueva oferta permite al sector MYPE seguir evolucionado hasta alcanzar una madurez que le permita en un futuro funcionar sin la presencia del dueño fundador.

Dentro de la investigación sobre financiamiento se desarrollan los productos y proveedores de fondos para este sector importante, siendo las operaciones crediticias de mayor relevancia las cuales se clasifican en corto y largo plazo. Dentro de este esquema por convención se determina que las de corto plazo sean cuyo plazo de pago es menor a un año, en tanto que las operaciones cuyo plazo sea mayor a un año caerán bajo la categoría de largo plazo. Asimismo se señala que por el principio de temporalidad, las solicitudes de apoyo financiero para capital de trabajo serán básicamente operaciones de corto plazo mientras que las destinadas a financiar compras de activos pertenecerán al largo plazo (Lira, 2009)<sup>6</sup>.

Otro punto importante señalado dentro de la investigación la cual sirve de antecedente es donde se involucra a la Corporación Financiera de Desarrollo (COFIDE) el cual es un banco que presta a otros bancos o instituciones financieras supervisadas por la Superintendencia de Banca Seguros y AFP (SBS). Los recursos de COFIDE que provienen de organismos multilaterales, agencias de gobiernos, bancos comerciales y del mercado de capitales local, se caracterizan por financiar a la mayoría de actividades y segmentos empresariales en el

---

<sup>6</sup> El estudio se enfoca principalmente en diseñar un manual práctico para el manejo de los micro y pequeño empresarios donde puedan comparar diversos servicios financieros.

Perú, en condiciones blandas de costo y plazo donde los plazos son más amplios o flexibles y las tasas de interés menores a la promedio.

El financiamiento por parte de COFIDE está dirigido a proyectos de inversión, capital de trabajo y adquisición de activos fijos de empresas que ya están en marcha. De esto un porcentaje considerable de los fondos administrados se destinan a financiar y apoyar las actividades económicas del sector MYPE local, básicamente de dos maneras: Prestando dinero a los empresarios de las MYPE a través de las IFI, y brindando productos y servicios no financieros tanto de asesoría en temas administrativos, financieros y tributarios. (Lira, 2009)

Existen diversos fondos de inversión especializados enfocados en desarrollar el sector MYPE a través de emprendimientos o nuevas ideas de negocio. Sin embargo algunos requieren de un nivel mayor de apalancamiento y por ende un mejor sustento financiero al momento de postular a los mismos reduciendo la posibilidad de aplicar.

Por último el documento elaborado sobre el análisis de la competitividad para las regiones de Europa y América Latina en cooperación con las cámaras de comercio europeas y los principales organismos de promoción de la inversión (CEPAL, 2013).

El informe señala que las microempresas, pequeñas y medianas empresas son agentes económicos clave, ya que buena parte de la población y de la economía dependen de las actividades desempeñadas. En los países que conforman parte de la investigación, las MYPES representan aproximadamente el 99% del total de empresas, generan una gran parte del empleo (67%) y actúan en una amplia variedad de ámbitos de la producción y de los servicios. En general el hecho de que este sector tenga un mayor peso en el nivel de empleo que en la producción indica que sus niveles de productividad son inferiores a los de las grandes empresas, ya que necesita de un mayor grupo de trabajadores para generar los bienes y servicios que brinda. Por otra parte, su limitada participación en las exportaciones habla de una clara orientación al mercado interno.

Se señala además que en ambas regiones los diferentes gobiernos han puesto en marcha instrumentos y programas de apoyo a las PYMES para mejorar su desempeño, fomentar su desarrollo productivo y generar un ambiente de negocios favorable. Sin embargo se

evidencias que en muchos casos los resultados no han llegado a cumplir las expectativas y no se han observado mejoras significativas en la productividad ni mucho menos en la competitividad de este tipo de empresas dentro de la región (CEPAL, 2013).<sup>7</sup>

Otro punto importante señalado en el informe es donde se señala la influencia del proceso de globalización el cual está ocasionando cambios notables dentro de la tecnología, la organización industrial y la división internacional del trabajo. Como resultado de esto se está produciendo una fragmentación de los sistemas productivos, ya que muchas actividades se trasladan a diferentes localizaciones por todo el mundo y las grandes empresas externalizan procesos productivos a MYPES especializadas. En este contexto, las MYPES latinoamericanas y Europas enfrentan una mayor competencia, pero también surgen para ellas posibilidades antes insospechadas. De esta manera el desempeño de las MYPES se encuentra sustancialmente diferenciado entre ambos bloques siendo marcada esta brecha debido también a las políticas de estado enfocadas en promocionar estos sectores.

## 2.2. Marco Conceptual

Las variables de conteo o recuento se definen como el número de sucesos o eventos que ocurren en una misma unidad de observación en un intervalo espacial o temporal definido. Así, por ejemplo, el número de artículos adquiridos por una tienda deportiva durante un año es un conteo. Es importante considerar dentro del análisis de datos de conteo que, muchas muestras o bases de datos poseen el valor numérico cero (0). A partir de esta definición propuesta (James, 1995), se derivan dos características principales de una variable de recuento, lo cual la diferencia de una variable cuantitativa continua, que son su naturaleza discreta y no negativa.

El modelo de Regresión de Poisson (MRP) es el modelo de referencia en estudios de variables de recuento (Cameron & Trivedi, 1998). Consiste en ser un modelo que resulta especialmente adecuado para modelar valores enteros no negativos, generalmente cuando la frecuencia de ocurrencia es baja.

---

<sup>7</sup> Según el índice de competitividad global elaborado por el Fondo Económico Mundial (World Economic Forum – WEF) del 27 de Setiembre del 2016 el Perú se ubica en el ranking 67 de un total de 138 países observados.

- **Tipo de distribución de probabilidad**

La distribución de Poisson es la que corresponde a datos de conteo de la misma forma que la distribución Normal lo es para los datos continuos. Dentro de la distribución de Poisson se tiene un único parámetro que es la media el cual debe ser siempre positivo. De esta manera este único parámetro determina la distribución en su totalidad. Por otra parte, en la Distribución Normal existen dos parámetros que son media y la varianza, las cuales caracterizan la distribución de probabilidades.<sup>8</sup>

A diferencia de la distribución multinomial, se asume una distribución de Poisson cuando el tamaño de muestra  $n$  es aleatorio, lo cual lleva a considerar que todas las celdas de una tabla de contingencia son variables aleatorias independientes con distribución Poisson (Córdova, 2013).

- **Distribución de probabilidad de las variables aleatorias discretas**

Las variables aleatorias discretas se denominan de esa forma si es posible contar con su conjunto de resultados posibles definido. Las variables aleatorias discretas son variables cuyo intervalo es un campo finito donde podemos conocer la cantidad exacta.

Cuando es necesario conocer un comportamiento estadístico para este grupo de variables discretas es cuando se emplean las distribuciones de Probabilidades discretas, denominados así por ser una lista de los resultados de un experimento con la probabilidad que se estarían esperando asociándose a dichos cálculos resultantes.

Por lo tanto si  $x$  es una variable aleatoria discreta, la función dada por  $f(x)$  para cada  $x$  contenida en dicho intervalo se denomina función de probabilidad o distribución de probabilidad. Una función puede fungir como la distribución de probabilidad de una variable aleatoria discreta  $x$ , si y solo si sus valores,  $f(x)$  cumple las condiciones siguientes:  $f(x) \geq 0$  para cada valor contenido en su dominio y  $\sum f(x) = 1$ , donde la sumatoria se extiende sobre todos los valores contenidos en su dominio.

---

<sup>8</sup> El cálculo de la media y varianza en las distribuciones de probabilidad se realiza mediante el método de los momentos.



- **Función de distribución acumulativa**

La distribución acumulada  $F(x)$  de una variable aleatoria discreta  $X$ , cuya distribución de probabilidad es  $f(x)$ , es:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), -\infty \leq x \leq \infty$$

- **Esperanza matemática**

Sea  $X$  una variable aleatoria con distribución de probabilidad  $f(x)$ . La media o valor esperado de  $X$  es:

$$\mu = E(X) = \sum_x xf(x)$$

El resultado muestra un valor medio teórico de todos los valores que puede tomar la variable de estudio representando de esta manera una medida de centralización.

- **Varianza**

Definido como la medida del cuadrado de la distancia promedio entre la media y cada elemento de la población. Entonces si se determina que  $x$  es una variable aleatoria con una distribución de probabilidad,  $f(x)$  y media  $\mu$ . La varianza de  $X$  estaría calculada por medio de la siguiente expresión:

$$\sigma^2 = E[(X - \mu)^2] = \sum_x (x - \mu)^2 f(x)$$

- **Desviación estándar**

Definido como una medida de dispersión de la misma dimensión física de la variable. Su expresión matemática se encuentra determinada por la raíz cuadrada positiva de la varianza; una medida de la dispersión, expresada en las mismas unidades que los datos o valores originales más no en las unidades cuadradas de la varianza, permitiendo que podamos analizar y comparar con otras medidas de tendencia central.

- **Regresión lineal**

En estadística la regresión lineal o ajuste lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente  $Y$ , las variables independientes  $X_i$  y un término aleatorio  $\varepsilon$ . Este modelo puede ser expresado como:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Donde:

$Y_t$ : Variable dependiente, explicada o regresando.

$X_1, X_2, \dots, X_p$ : Variables explicativas, independientes o regresores.

$\beta_0, \beta_1, \beta_2, \dots, \beta_p$ : Parámetros, miden la influencia que las variables explicativas tienen sobre el regresor.

- **Estimación de Máxima Verosimilitud**

En estadística la estimación por máxima verosimilitud<sup>9</sup> considera una variable aleatoria  $x$  descrita por la función de probabilidad  $f(x)$ , donde el espacio de muestras está constituido por todos los posibles valores de  $x$ , por lo tanto un conjunto de  $n$  observaciones independientes de  $x$  se llama una muestra de tamaño  $n$ . Seguido es posible definir un nuevo espacio de muestras constituido por todos los posibles valores del vector  $x = (x_1, \dots, x_n)$ . Es decir, la muestra se considera formada por una sola medida aleatoria, caracterizada por las cantidades  $(x_1, \dots, x_n)$ .

Luego procedemos a considerar la situación donde se han realizado  $n$  medidas de una variable aleatoria cuya función de probabilidad se desconoce, ya que uno de los problemas centrales de la estadística es inferir las propiedades de  $f(x)$  basándose en las observaciones  $(x_1, \dots, x_n)$ . Específicamente hablando deseamos construir funciones de los  $x_i$  para estimar las propiedades de  $f(x)$ . Porque a menudo se tiene una hipótesis para la función de probabilidad  $f(x, \theta)$  de un parámetro desconocido (o de forma mas general de un vector de parámetros

---

<sup>9</sup> Este método fue introducido por Fisher en la década de 1920.

$\theta = (\theta_1, \dots, \theta_n)$ ). Por ende el objetivo primordial es construir funciones de los  $x_i$  que permitan estimar los parámetros  $\theta$ . Definiendo de esta forma una estadística asociada a la muestra para estimar las propiedades de la función de probabilidad tales como la media y varianza.

La estimación empleando este método consiste en estimar los valores de  $\theta$  dada una muestra finita de datos. La suposición principal consiste en que existen  $n$  medidas de  $x, x_1, \dots, x_n$ . Puesto que las medidas son independientes, la probabilidad de que  $x_1$  esté en  $[x_1, x_1 + dx_1]$ ,  $x_2$  en  $[x_2, x_2 + dx_2]$ , se encuentra determinada por la siguiente expresión para cualquier valor de  $i > 1$ :

$$\prod_{i=1}^n f(x_i; \theta) dx_i$$

De esta manera mediante la medición de la función de probabilidad y los parámetros se esperaría una alta probabilidad para los datos medidos.

- **Distribución de Poisson**

La distribución de Poisson fue derivada por Simeon Denis Poisson, quien en 1837 publicó un trabajo de investigación en el que se presentaba una nueva distribución para el cálculo de probabilidades aplicado al ámbito penal. Poisson encontró que cuando el tamaño de una muestra es grande y la probabilidad de ocurrencia de un evento es pequeña, el valor esperado tiende a una constante. (Córdova, 2013)

Los experimentos que resultan en valores numéricos de una variable aleatoria  $X$ , misma que representa el número de resultados durante un intervalo de tiempo dado o en una región específica, se llaman experimentos de Poisson. Además el intervalo de tiempo determinado puede ser de diferentes duraciones. Representando además gráficamente dicha región por un segmento de línea, área, volumen, etc.

El proceso de Poisson se caracteriza por las siguientes propiedades:

1. El número de resultados que ocurren en un intervalo de tiempo o región específicos es independiente del número que ocurre en cualquier otro intervalo disjunto de tiempo o región del espacio disjunto. De esta manera se dice que el proceso de Poisson no tiene memoria.

2. La probabilidad de que un resultado sencillo ocurra en un intervalo de tiempo muy corto o en una región pequeña es proporcional a la longitud del intervalo de tiempo o al tamaño de la región y no depende del número de resultados que ocurren fuera de este intervalo o región.
3. La probabilidad de que más de un resultado ocurra en ese intervalo de tiempo tan corto o en esa región tan pequeña es despreciable.

Además el número  $X$  de resultados que ocurren en un experimento de Poisson se llama variable aleatoria de Poisson y su distribución de probabilidad recibe el nombre de distribución de Poisson.

La distribución de Poisson de la variable aleatoria  $X$ , que representa el número de resultados que ocurren en un intervalo de tiempo dado o en una región específica se encuentra denominado por la siguiente expresión:

$$p(x, \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, 3, \dots, N$$

Dentro de la ecuación el parámetro  $\lambda$  equivale al valor de la esperanza matemática de  $x$ .

Los principales estadísticos se encuentran definidos de la siguiente forma:

1. Media o esperanza matemática :  $E(x) = \mu = \lambda$
2. Varianza :  $\sigma^2 = \lambda$
3. Desviación estándar :  $\sigma = \sqrt{\lambda}$

Existe el flujo elemental de sucesos dentro de la distribución de probabilidad Poisson que posee las siguientes propiedades:

1. Calidad de estacionario: Esta propiedad consiste en que la probabilidad de que ocurran  $x$  sucesos en cada intervalo de tiempo depende solamente del número  $x$  y de la duración  $t$  del intervalo de tiempo y no depende del comienzo de su cuenta. En otras palabras, la probabilidad de aparición de  $x$  sucesos en un intervalo de tiempo de duración  $t$  depende solo de  $x$  y de  $t$ .
2. La propiedad de ausencia de efecto posteriori: Esta propiedad se caracteriza porque la probabilidad de que ocurran  $x$  sucesos en cualquier intervalo de tiempo no depende

de que hayan ocurrido o no los sucesos en los instantes de tiempo que preceden al comienzo del intervalo considerado. En otras palabras, la prehistoria del flujo no influye en la probabilidad de que los sucesos ocurran en un futuro próximo.

3. Propiedad de ordinario: Define que la aparición de dos o más sucesos en un intervalo pequeño de tiempo es prácticamente imposible. En otras palabras, la probabilidad de que ocurra más de un suceso en un pequeño intervalo de tiempo es despreciable en comparación con la probabilidad de que ocurra solamente un suceso.

En conclusión el promedio de sucesos que ocurren en una unidad de tiempo se denomina intensidad de flujo  $\lambda$ . En ese contexto si la intensidad constante del flujo  $\lambda$  es conocida, la probabilidad de que ocurran  $k$  sucesos de un flujo elemental en el tiempo  $t$  se determina por la fórmula de Poisson:

$$p(x; \lambda t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!}$$

Además el flujo puede caracterizarse por ser estacionario o no estacionario.

- **Distribución Binomial Negativa**

Esta distribución puede considerarse como una extensión o ampliación de la distribución geométrica. La distribución binomial negativa es un modelo adecuado para tratar aquello proceso en los que se repite un determinado ensayo o prueba hasta conseguir un número determinado de resultados favorables. Es por lo tanto de gran utilidad para aquellos muestreos que proceden de esta manera. Si el número de resultados favorables buscados fuera la unidad entonces estaríamos en el caso de la distribución geométrica. Está implicada por tanto la existencia de una dicotomía de resultados posibles en cada prueba y la independencia de cada prueba o ensayo, o la reposición de los individuos muestreados.

La forma de derivación de esta distribución proviene de un proceso experimental puro o de Bernoulli en el que se presentan las siguientes condiciones:

1. El proceso consta de un número no definido de pruebas separadas o separables. El proceso concluirá cuando se obtenga un determinado número de resultados favorables.

2. Cada prueba pueda dar dos resultados posibles mutuamente excluyentes denominados como éxito y fracaso o un suceso A y no A.
3. La probabilidad de obtener un resultado A en cada una de las pruebas es  $p$  siendo la probabilidad de no A representado por la letra  $q$ . Lo que define una relación complementaria expresada matemáticamente de la siguiente forma  $p + q = 1$ .
4. Las probabilidades  $p$  y  $q$  son constantes en todas las pruebas. Todas las pruebas son independientes. Si se trata de un experimento de extracción este se llevará a cabo con devolución del individuo extraído, a no ser que se trate de una población en la que el número de individuos tenga carácter infinito.
5. Derivando la distribución se tiene que la variable aleatoria  $x$  sea el número de pruebas necesarias para conseguir  $k$  éxitos o resultados A entonces la variable aleatoria  $x$  seguirá una distribución binomial negativa con parámetros  $p$  y  $k$ .

$$x \rightarrow BN(p, k)$$

Donde la variable aleatoria  $x$  podrá tomar solo valores superiores a  $k$  y donde  $x \in \{k, k + 1, k + 2, \dots, k + n\}$ . Además dado que las prueba son independientes y conocemos que  $P(A) = p$  y  $P(\text{no } A) = q$  como se muestra en la siguiente expresión:

$$\underbrace{q \cdot q \cdot q \dots q}_{x \text{ veces}} \cdot \underbrace{p \cdot p \cdot p \dots p}_{k \text{ veces}} = q^{x-k} p^k$$

La cual sería la probabilidad de  $x$  si el suceso fuera precisamente con los resultados en ese orden. Dado que pueden darse otras órdenes, en concreto se tendrían  $\binom{x-1}{x-k}$  formas u órdenes distintos. La función de cuantía de la distribución binomial negativa quedará como:

$$P(x) = \binom{x-1}{x-k} q^{x-k} \cdot p^k$$

La función generatriz de momentos nos permitirá calcular los principales estadísticos de la distribución por ende se tendría:

$$\varphi(t) = E[e^{tx}] = \sum_{\forall i} e^{tx_i} \binom{x_i-1}{x_i-k} p^k q^{x_i-k} = p^k \sum_{\forall i} e^{tx_i} \binom{x_i-1}{x_i-k} q^{x_i-k} = p^k (e^{-t} - q)^{-k}$$

Aplicando el teorema de los momentos hallamos la media y varianza que resultan ser:

1. Media:  $\mu = \frac{k}{p}$
2. Varianza:  $\sigma^2 = \frac{kq}{p^2}$

## 2.3. Teoría general

### 2.3.1. Modelo para datos de conteo

En diferentes contextos económicos la variable dependiente o de respuesta es un valor entero positivo o una cuenta no negativa que nosotros deseamos explicar y analizar en términos de un conjunto de regresores. Distinto del modelo de regresión clásico, la variable respuesta en este caso es discreta, con una distribución que ubica la masa probabilística en solo un valor entero no negativo. Varios modelos han tratado de explicar dicho fenómeno como el modelo de elección binaria y el modelo de duración, los cuales pueden ser mostrados como muy cercanos a los modelos de regresión para datos de conteo. Modelos de regresión de recuento, como otros de variable dependiente discreta o limitada, por ejemplo Logit y Probit, son no lineales con muchas propiedades y características especiales íntimamente conectadas a discreción y no linealidad. (Cameron & Trivedi, 1998)

Analizando algunos ejemplos de microeconomía, tomando como punto de partida una muestra de datos que son observaciones de corte transversal independientes. Tenemos el caso de los estudios de fertilidad que regularmente modelan el número de nacimientos de bebés comparado a un intervalo específico de edad de la madre, con el objetivo de analizar su variación en términos del nivel de escolaridad de la madre, edad, e ingresos del hogar (Winkelmann, 1995). En algunos modelos de decisiones para familias, el número de hijos podría aparecer como una variable explicativa con el reconocimiento de que es endógena. En otro campo de estudio se tienen análisis de accidentes que modelan la seguridad en las aerolíneas, utilizando como datos para la medición el número de accidentes experimentados por una aerolínea dentro de algún periodo, y determinando su relación con la rentabilidad de la compañía aérea, así mismo como otras medidas relacionadas a la salud financiera de la aerolínea. Otro espectro de análisis para estos modelos son los estudios de demanda recreacional, los cuales buscan colocar un valor sobre los recursos naturales, tales como

reservas nacionales. Modelando el número de viajes al sitio recreacional (Gurmu & Trivedi, 1996). En el sector salud, se analiza la demanda de servicios empleando datos relacionados al número de veces que personas particulares consumen o acceden a los servicios de salud, tales como visitas al doctor o días en el hospital durante los años previos al cierre de la investigación (Cameron A. , Trivedi, Milne, & Piggott, 1988). De esta forma se presentan los modelos de regresión para datos de conteo donde son relevantes las características de la información y relaciones entre variables para poder desarrollar cada caso.

La presentación de los casos de estudio permitió establecer que cuando la variable de interés únicamente toma valores enteros no negativos, que representan el número de sucesos ocurridos en un periodo determinado, los modelos apropiados, dada la naturaleza de la misma son los denominados modelos para datos de recuento (o de conteo). Estos pueden encontrarse como se señaló anteriormente en diversos campos y áreas de la economía como son: el número de patentes registradas, el número de huelgas en un determinado sector, los días de estancia en una zona turística, etc. La especificación básica en datos de recuento la constituye el modelo de regresión de Poisson. En este se supone que la probabilidad de cada recuento ( $y_i$ ) se determina a partir de una distribución de Poisson, cuya media  $\lambda_i$ , es función de un conjunto de variables explicativas, definidas de la siguiente manera:

$$p(y_i, \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

La interpretación de los parámetros en este modelo puede obtenerse a partir de:

$$\frac{\partial E(y_i, x)}{\partial x_j} = E(y_i, x) \cdot \beta_j$$

De forma general estos pueden ser interpretados como semi-elasticidades, esto es, el cambio proporcional en la media condicional ante un cambio unitario en la variable explicativa correspondiente. Sin embargo, para variables explicativas cualitativas, esta interpretación supondría una buena aproximación para valores pequeños de los parámetros, en otro caso, esta se mide a través de  $e^{\beta_j}$  interpretado como el número de veces en que la media condicionada es más grande, al pasar de la categoría de referencia a cualquiera de las categorías de cada uno de los factores.



Existe también una posibilidad sobre la interpretación en relación al efecto marginal sobre la probabilidad de un resultado. En este caso, se tiene la siguiente expresión:

$$\frac{\partial P(Y_i = j)}{\partial x_m} = P(Y_i = j)[j - e^{x_i' \beta}] \beta_m$$

Así, este efecto marginal sigue un patrón específico que viene impuesto por los supuestos del modelo. Como puede observarse, la dirección del cambio en la probabilidad depende del signo del parámetro y de la diferencia de  $[j - e^{x_i' \beta}]$ .

La estimación de los parámetros del modelo puede llevarse a cabo mediante el método de Máxima Verosimilitud (MV). Para el modelo de Poisson, la función de log-verosimilitud y las condiciones de primer orden vienen dadas de la forma:

$$\ln L = \sum_{i=1}^N (-e^{x_i' \beta} + y_i x_i' \beta - \ln(y_i!))$$

$$\sum_{i=1}^N (y_i - e^{x_i' \beta}) x_i = 0$$

Dada la no linealidad de las funciones en  $\beta$ , es necesario emplear métodos de optimización numérica para la obtención de las estimaciones.

Existen limitaciones dentro del modelo de Poisson las cuales son el supuesto de equidispersión y la dificultad de recoger la abundancia de ceros habitualmente presente.

Un modelo más general para este tipo de datos es el denominado Binomial Negativo. Este puede ser utilizado en presencia de sobredispersión. Su representación es obtenida mediante la incorporación de un término de perturbaciones en la media del modelo de Poisson, dando lugar a diferentes formulaciones. Su función de probabilidad puede expresarse de la siguiente forma:

$$P\left(Y_i = \frac{y_i}{x_i}\right) = \frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i}\right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\alpha^{-1} + \lambda_i}\right)^{y_i}$$

Donde se ha supuesto que

$$\lambda_i^* = E\left[\frac{y_i}{x_i, v_i}\right] = \lambda_i v_i = e^{x_i' \beta} e^{\varepsilon_i}$$

Por otro lado existen diversos modelos que permiten manejar la presencia de sobredispersión y el exceso de ceros en la muestra y es el denominado modelo en dos partes (Hurdle). En este se supone que el proceso que siguen las observaciones cero es diferente del que sigue las positivas. Para ello, se combina un modelo binario y otro truncado. De forma general, se tiene que:

$$P(Y_i = 0) = f_1(0)$$

$$P(Y_i = y_i) = (1 - f_1(0)) \frac{f_2(y_i)}{1 - f_2(0)}$$

Donde las funciones  $f_1$  y  $f_2$  representan distribuciones de probabilidad, que pueden ser o no del mismo tipo,  $f_1(0)$  recoge la probabilidad de cruzar el primer estadio de decisión y  $f_2(0)$  se aplica para truncar en el cero la segunda distribución. Los diferentes supuestos que se realizan sobre estas distribuciones conducen a las versiones Poisson y Binomial negativa.

Finalmente se tiene los modelos con exceso de ceros que suponen otra alternativa para poder recoger de forma adecuada la abundancia de valores, nulos, en los que se considera que estos valores pueden haber sido generados por dos situaciones diferentes, lo que lleva a una probabilidad suplementaria para estos valores a la obtenida por una distribución estándar. De forma similar al modelo en dos partes, en este se combina un proceso binario y una distribución para los recuentos ceros y positivos. Formalmente,

$$P(Y_i = 0) = f_1(0) + (1 - f_1(0))f_2(0)$$

$$P(Y_i = y_i) = (1 - f_1(0))f_2(y_i)$$

Donde  $f_1$  representa la distribución de probabilidad del proceso binario, habitualmente recogida mediante un modelo Logit y  $f_2$  la distribución de los recuentos, que será Poisson o Binomial Negativa. Este modelo puede considerarse también un modelo mixto de dos componentes, donde uno de ellos sigue una distribución degenerada en cero y el otro, es un modelo de regresión estándar para datos de recuento. Dependiendo de los supuestos que se realicen sobre la media del proceso y las probabilidades, se obtendrán especificaciones alternativas del mismo.

## 2.4. Teorías específicas

En algunos casos, tales como el número de nacimientos, el conteo es la variable de último interés. En otros casos, como la demanda médica y resultados de los gastos en investigación y desarrollo, la variable de interés final es continua, la mayoría de veces gastos o recibos medidos en dólares, sin embargo la mejor información disponible en su lugar es un conteo o recuento. En muchos de los casos, la muestra está concentrada en un pequeño puñado de valores discretos, sean estos, 0, 1 y 2. De esta forma la data o conjunto de información puede ser sesgada a la derecha. Y por último poseer intrincada heteroscedasticidad con una varianza aumentando con la media. (Cameron & Trivedi, 1998)

### 2.4.1. Regresión de Poisson

La regresión de Poisson es el punto de partida para el análisis de los datos de conteo, a pesar de ser en la mayoría de los casos inadecuada para su modelamiento. Por ende existe una calificación: En algunos casos una alta concentración de ceros en la muestra podría coexistir con una gran cantidad de valores de conteo, creando un desafío para modelar difícil y complicado. Este desafío es seleccionar una forma funcional que pueda capturar adecuadamente una gran media y la mayor proporción de ceros. En muchos otros ejemplos como el número de nacimientos, virtualmente toda la información está restringida a simples dígitos, y el número promedio de eventos es casi diminuto.

Estas características motivan a la aplicación de métodos especiales y modelos para regresión de conteo. Existen por tanto dos caminos para proceder.

La primera aproximación es un método completamente paramétrico que especifica completamente la distribución de los datos, respetando siempre la restricción de  $y$  para valores enteros no negativos. Esta aproximación fue tomada de aplicaciones tempranas, la mayoría de bioestadística, donde regresiones de conteo fueron vistas como una extensión y generalización de una vasta literatura en la distribución de recuentos independientes e idénticamente distribuidos.

La segunda aproximación es una aproximación de media-varianza, la cual especifica que la media condicional sea no negativa y además que la varianza condicional sea una función de la media condicional. Estos modelos mejoran la no negatividad y heteroscedasticidad pero no dirigen la discrecionalidad de la data. Esta aproximación, en un caso no limitado solo para datos de conteo, fue introducida por Nelder y Wedderburn, direccionando a la aproximación del modelo lineal generalizado ampliamente usado en estadística (Cameron & Trivedi, 1998).

### 2.4.1.1. Estimación y Cuasi estimación de máxima

#### verosimilitud para la Regresión de Poisson

El modelo estocástico natural para recuentos es un proceso de Poisson para la ocurrencia del evento de interés. Esto implica un distribución de Poisson para el numero de ocurrencias del evento, con densidad, o más formal la función de densidad probabilística.

$$\Pr[Y = y] = \frac{e^{-\mu} \mu^y}{y!}, y = 0, 1, 2, \dots,$$

Donde  $\mu$  es la intensidad o ratio del parámetro. Nos referimos a la distribución como  $\mathcal{P}[\mu]$ . Los primeros dos momentos son:

$$E[Y] = \mu$$

$$V[Y] = \mu$$

Esto muestra la bien conocida equidispersión (igualdad de media y varianza) propiedad de la distribución de Poisson.

Introduciendo la observación con el subíndice  $i$ , junto a las variables  $y$  y  $\mu$ , el marco de referencia independiente e idénticamente distribuido es extendido para el caso de regresión. El modelo de regresión de Poisson esta derivado de la distribución de Poisson si parametrizamos la relación entre el parámetro media  $\mu$  y coviamos (regresores)  $x$ . El supuesto estándar es usar el exponencial de la media como parámetro, de la siguiente forma

$$\mu_i = \exp(x_i' \beta), i = 1, \dots, N$$

Donde siguiendo el supuesto existen  $K$  covariables linealmente independientes, usualmente incluyendo una constante. Porque  $V[y_i|x_i] = \exp(x_i'\beta)$ , además por las expresiones anteriores la regresión de Poisson es heteroscedástica intrínsecamente.

Dadas las relaciones anteriores y el supuesto de que las observaciones  $(y_i|x_i)$  son independientes, el estimador más adecuado sería el de máxima verosimilitud. La función log verosimilitud está denominada de la siguiente forma:

$$\ln L(\beta) = \sum_{i=1}^N \{y_i x_i' \beta - \exp(x_i' \beta) - \ln y_i!\}$$

La estimación de máxima verosimilitud de Poisson, denotada por  $\widehat{\beta}_p$ , es la solución para  $K$  ecuaciones no lineales correspondientes a la condición de primer orden para máxima verosimilitud,

$$\sum_{i=1}^N (y_i - \exp(x_i' \beta)) x_i = 0$$

Si  $x_i$  incluye un término constante entonces los residuos de  $y_i - \exp(x_i' \beta)$  suman cero según la demostración previa. La función log verosimilitud es globalmente cóncava; por lo tanto resolviendo estas ecuaciones utilizando un Gauss-Newton o Newton-Raphson produce parámetros únicos estimados.

En la literatura econométrica las estimaciones pseudo máxima verosimilitud (PMV) o cuasi máxima verosimilitud (CMV) se refieren a estimaciones por el método de máxima verosimilitud, bajo falta de especificación de la densidad especificada. Los términos PMV o CMV son usados muchas veces de forma intercambiable. La distribución del estimador es obtenido bajo supuestos más débiles acerca del proceso de generación de la data o información que los utilizados para especificar de forma adecuada la función de verosimilitud. En la literatura estadística CMV usualmente se refiere a la estimación por mínimos cuadrados generalizados no lineales.

Para la regresión de Poisson, CML en el último sentido es equivalente a la máxima verosimilitud estándar.

El estimador de PMV de Poisson  $\widehat{\beta}_p$ , posee condiciones de primer orden  $\sum_{i=1}^N (y_i - \exp(\mathbf{x}'_i \beta)) \mathbf{x}_i = 0$ . Como ya se definió, la sumatoria en la lado izquierdo tiene esperanza cero si  $E[y_i | \mathbf{x}_i] = \exp(\mathbf{x}'_i \beta)$ . Por lo tanto la PMV de Poisson es consistente bajo el supuesto débil de especificaciones correctas de la media condicional; esto es, la información no necesita ser Poisson distribuida. Usando los resultados previos, la matriz de varianza es de la forma sándwich, con lo siguiente:

$$V_{PMV}[\widehat{\beta}_p] = \left( \sum_{i=1}^N \mu_i \mathbf{x}_i \mathbf{x}'_i \right)^{-1} \left( \sum_{i=1}^N \omega_i \mathbf{x}_i \mathbf{x}'_i \right) \left( \sum_{i=1}^N \mu_i \mathbf{x}_i \mathbf{x}'_i \right)^{-1}$$

Y  $\omega_i = V[y_i | \mathbf{x}_i]$  es la varianza condicional de  $y_i$ .

De forma estándar para la teoría de máxima verosimilitud ML si el supuesto más fuerte ha hecho que la regresión de Poisson sea especificada paramétricamente correcta, entonces esa  $\omega_i = \mu_i$ , el estimador  $\widehat{\beta}_p$  es consistente para  $\beta$  y normal asintóticamente con la matriz de covarianza muestral

$$V[\widehat{\beta}_p] = \left( \sum_{i=1}^N \mu_i \mathbf{x}_i \mathbf{x}'_i \right)^{-1}$$

En el caso donde  $\mu_i$  es de la forma exponencial.

La estimación de máxima verosimilitud de Poisson y la Pseudo estimación son idénticas pero poseen varianzas diferentes.

### 2.4.1.2. Interpretación de los coeficientes de regresión

Para modelos lineales, con  $E[y | \mathbf{x}] = \mathbf{x}' \beta$ , los coeficientes  $\beta$  son fácilmente interpretados como el efecto de un cambio en una unidad en regresores  $\mathbf{x}_i \mathbf{x}'_i$  de la media condicional. Para modelos no lineales esta interpretación necesita ser modificada. Para cualquier modelo con media condicional exponencial, la diferencial produce

$$\frac{\partial E[y | \mathbf{x}]}{\partial x_j} = \beta_j \exp(\mathbf{x}' \beta)$$

Donde el escalar  $x_j$  denota el  $j$ -ésimo regresor. Por ejemplo, si  $\widehat{\beta}_p = 0.25$  y  $\exp(x'_i\widehat{\beta}) = 3$ , entonces un cambio en una unidad en el  $j$ -ésimo regresor incrementa la esperanza de  $y$  por 0.75 unidades. Esta respuesta parcial depende de  $\exp(x'_i\widehat{\beta})$ , el cual se espera que varíe a través de los individuales. Es muy sencillo observar que  $\beta_j$  mide el cambio relativo en  $E[y|x]$  inducido por el cambio en unidad de  $x_j$ . Si  $x_j$  esta medido en una escala logarítmica,  $\beta_j$  es una elasticidad.

Para fines de presentación de informes una variable única de respuesta, un buen candidato es una estimación del promedio de respuesta  $N^{-1} \sum_i \frac{\partial E[y_i|x_i]}{\partial x_{ij}} = \widehat{\beta}_j \times N^{-1} \sum_i \exp(x'_i\widehat{\beta})$ . Para modelos de regresión de Poisson con intercepto incluido, esto puede ser mostrado para simplificar  $\widehat{\beta}_j \bar{y}$ .

## 2.4.2. Sobredispersión

El modelo de regresión de Poisson es usualmente muy restrictivo para datos de conteo o recuento, dirigiendo hacia modelos alternativos diferentes al presentado. El problema fundamental es que la distribución está parametrizada en términos de un único parámetro escalar ( $\mu$ ) entonces todos esos momentos de  $y$  son una función de  $\mu$ . Por contraste la distribución normal tiene parámetros separados para posición ( $\mu$ ) y escala ( $\sigma^2$ ). Por la misma razón el exponencial de un solo parámetro es muy restrictivo para datos de duración y distribuciones de dos parámetros más generales como la Weibull son superiores. Notar que esta complicación no aparece con datos binarios. Luego la distribución es claramente la Bernoulli de un parámetro, porque si la probabilidad de éxito es  $p$  entonces la probabilidad de fracaso debe ser  $1 - p$ . Para datos binarios el problema por el contrario recae en como parametrizar  $p$  en función de los regresores.

Una forma en la que esta restricción se manifiesta por sí misma es que en muchas aplicaciones una densidad de Poisson predice la probabilidad de que un cero de la cuenta sea considerablemente menos de los que es en realidad observado dentro de la muestra. Este está denominado como el problema del exceso de ceros, tanto como existan más ceros en la data que en las predicciones de Poisson.

Una segunda y más obvia deficiencia del modelo de Poisson es que para datos de conteo la varianza usualmente excede la media, una característica llamada sobredispersión. El modelo de Poisson en cambio implica igualdad entra la varianza y la media, una propiedad denominada equidispersión.

Sobredispersión tiene cualitativamente similares consecuencias para el fallo de los supuestos de homocedasticidad en el modelo de regresión lineal. Provisto de la media condicional está correctamente especificado, eso es, la estimación de máxima verosimilitud de Poisson es todavía consistente. Esto es claro de la inspección de las condiciones de primer orden, desde el lado izquierdo en las formulas previamente establecidas las cuales hubieran esperado un valor de cero si  $E[y_i|x_i] = \exp(x_i'\beta)$ . Esta propiedad consistente aplica de forma más general a la cuasi estimación de máxima verosimilitud cuando la densidad especificada se encuentra dentro de la Familia de Exponenciales Lineales (FEL). Ambos Poisson y normal son miembros de FEL. Esto sin embargo es importante para controlar en sobredispersión. Primero, en configuraciones más complicadas tales como truncar o censurar, la sobredispersión conduce al problema fundamental de inconsistencia. Segundo, inclusive en la más sencilla de las configuraciones una gran sobredispersión dirige a un error estándar bruto disminuido y un estadístico  $t$  abruptamente aumentado en la salida de una MV, y por lo tanto es importante usar previamente el estimador de varianza robusta entregado. Tercero, si uno quiere estimar probabilidades de un numero de eventos, en vez de solamente la media condicional, estos dependen de parámetros adicionales del proceso de generación de datos.

Sobredispersión podría señalar una presencia de un error de especificación mucho más básico, especialmente en configuración de datos que involucran truncar y censurar si ellos son ignorados en la estimación. En ese caso la media condicional esta incorrectamente especificada y la presencia simultánea de sobredispersión entonces dirige a inconsistencia, no solo ineficiencia, de la estimación por máxima verosimilitud.

Un test estadístico de sobredispersión es por lo tanto altamente deseado luego de correr una regresión de Poisson. La mayoría de modelos de conteo con sobredispersión especifican sobredispersión de la siguiente forma

$$V[y_i|x_i] = \mu_i + \alpha g(\mu_i)$$



Donde  $\alpha$  es un parámetro desconocido y  $g(\cdot)$  Es una función conocida, comúnmente  $g(\mu) = \mu^2$  o  $g(\mu) = \mu$ . Es asumido que bajo ambas hipótesis nula y alternativa la media está correctamente especificada como, por ejemplo,  $\exp(\mathbf{x}'_i\beta)$  mientras que bajo la hipótesis nula  $\alpha = 0$  y entonces  $V[y_i|\mathbf{x}_i] = \mu_i$ . Un simple test estadístico de sobredispersión para  $H_0: \alpha = 0$  contra  $H_1: \alpha \neq 0$  o  $H_1: \alpha > 0$  puede ser calculado estimando el modelo de Poisson, construyendo valores precisos  $\hat{\mu}_i = \exp(\mathbf{x}'_i\hat{\beta})$ , y corriendo la regresión auxiliar de mínimos cuadrados ordinarios.

$$\frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i} = \alpha \frac{g(\hat{\mu}_i)}{\hat{\mu}_i} + u_i$$

Donde  $u_i$  es un término de error. El  $t$  estadístico reportado para  $\alpha$  es asintóticamente normal bajo la hipótesis nula de la no sobredispersión (Cameron & Trivedi, 1998) aunque aquí regresores generados es usado. Este test puede ser usado también para baja dispersión,  $\alpha < 0$  en tal caso la varianza condicional es menor que la media condicional.

### 2.4.3. Modelo Binomial Negativo

El modelo binomial negativo, un ejemplo específico de un modelo mixto continuo, puede ser obtenido de diferentes maneras. La siguiente justificación usando una distribución mixta es una de las más clásicas y tiene amplia aceptación.

Supongamos que la distribución de una variable de conteo aleatorio  $y$  es Poisson, condicional en el parámetro  $\lambda$ , además  $f(y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}$ . Supongamos ahora que el parámetro  $\lambda$  es aleatorio, en vez de ser una función de regresores  $\mathbf{x}$  completamente determinística. En particular, sea  $\lambda = \mu v$ , donde  $\mu$  es una función determinística de  $\mathbf{x}$ , por ejemplo  $\exp(\mathbf{x}'\beta)$ , y  $v > 0$  es idéntica e independientemente distribuida (iid) con densidad  $g(v|\alpha)$ . Este es un ejemplo de heterogeneidad inobservada, como diferentes observaciones podrían tener diferentes  $\lambda$  (heterogeneidad) pero parte de esta diferencia es debido al componente (inobservado) aleatorio  $v$ . Notar que  $E[\lambda|\mu] = \mu$  si  $E[v] = 1$ , entonces la interpretación de los parámetros de la pendiente permanecen tal cual en el modelo de Poisson.

La densidad marginal de  $y$ , incondicional en el parámetro aleatorio  $v$  pero condicional en los parámetros determinísticos  $\mu$  y  $\alpha$ , son obtenidos por integración de  $v$ . Esto resulta

$$h(y|\mu, \alpha) = \int f(y|\mu, v)g(v|\alpha)dv$$

Donde  $g(v|\alpha)$  es llamado distribución mixta y  $\alpha$  denominado parámetro desconocido de la distribución mixta. La integración define una distribución "promedio". Para algunas elecciones específicas de  $f(\cdot)$  y  $g(\cdot)$ , la integral tendrá una solución de forma cerrada o explícita.

Si  $f(y|\alpha)$  es la densidad de Poisson y  $g(v) = v^{\delta-1}e^{-v\delta}\delta^\delta / \Gamma(\delta)$ ,  $v, \delta > 0$ , es la densidad gamma con  $E[v] = 1$  y  $V[v] = 1/\delta$ , obtenemos la binomial negativa como una función de densidad mixta como se muestra a continuación:

$$\begin{aligned} h[y|\mu, \delta] &= \int_0^\infty \frac{e^{-\mu v}(\mu v)^y}{y!} \frac{v^{\delta-1}e^{-v\delta}\delta^\delta}{\Gamma(\delta)} dv \\ &= \int_0^\infty \frac{e^{-(\mu+\delta)v}(\mu)^y v^{y+\delta-1}\delta^\delta}{y! \Gamma(\delta)} dv \\ &= \frac{\mu^y \delta^\delta}{\Gamma(\delta)y!} \int_0^\infty e^{-(\mu+\delta)v} v^{y+\delta-1} dv \\ &= \frac{\mu^y \delta^\delta \Gamma(y+\delta)}{\Gamma(\delta)y! (\mu+\delta)^{y+\delta}} \\ &= \frac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\mu+\alpha^{-1}}\right)^y \end{aligned}$$

Donde  $\alpha = 1/\delta$ ,  $\Gamma(\cdot)$  denota la integral de gamma la cual se especializa en un factorial para los argumentos enteros, y la cuarta línea sigue luego algunos métodos de algebra y usos de la definición de la función gamma. Casos especiales de la binomial negativa incluye la Poisson ( $\alpha = 0$ ), y la geométrica ( $\alpha = 1$ ).

Como es el caso de diversas distribuciones mixtas, la binomial negativa también posee justificación independiente. Puede aparecer en diferentes maneras y uno necesita no siempre pensar en ella como una distribución mixta.

La derivación algebraica de la binomial negativa como una mixtura entre Poisson y gamma puede ser dada una interpretación Bayesiana. La distribución previa de  $\mu$  es gamma, dado  $\alpha$ , esta es esperada que la próxima distribución tenga una forma cerrada. Por lo tanto, la estimación de máxima verosimilitud y la media posterior Bayesiana, bajo el siguiente supuesto de un anterior valor vago en  $\alpha$ , coincidirán.

Los primeros dos momentos de la distribución binomial negativa son

$$E[y|\mu, \alpha] = \mu$$

$$V[y|\mu, \alpha] = \mu(1 + \alpha\mu)$$

La varianza por lo tanto excede la media, desde  $\alpha > 0$  y  $\mu > 0$ . En efecto, puede ser mostrado fácilmente que la sobredispersión siempre aumenta si  $y|\lambda$  es Poisson y la heterogeneidad inobservada es de la forma multiplicativa  $\lambda = \mu v$ , donde  $E[v] = 1$ . Notar también que la sobredispersión puede contar con diversas formas combinadas de expresiones presentadas anteriormente que cumplan los supuestos.

Dos variantes estándar de la binomial negativa son usadas en aplicaciones de regresión. Ambas variantes especifican  $\mu_i = \exp(x_i'\beta)$ . La variante más común deja a  $\alpha$  ser un parámetro para ser estimado, en tal caso la función de varianza condicional,  $\mu + \alpha\mu^2$  es cuadrática en la media.

La otra variante del modelo binomial negativo tiene una función de varianza lineal,  $V[y|\mu, \alpha] = (1 + \gamma)\mu$  obtenida mediante el reemplazo de  $\alpha$  por  $\gamma/\mu$ . Estimación por máxima verosimilitud es de nuevo sencilla. Algunas veces esta variante es llamada binomial negativa 1 (BN1) en contraste con la variante que posee una función de varianza cuadrática la cual ha sido denominada modelo binomial negativo 2 (BN2) (Cameron & Trivedi, 1998). Ambas variantes del modelo son sencillamente estimadas por máxima verosimilitud. En ambas variantes los coeficientes tienen la misma interpretación desde  $E[y|x] = \exp(x'\beta)$ , haciendo de la BN2 la variante mayormente empleada en las aplicaciones de investigación.

El modelo BN2 ha sido escogido debido a que es muy útil en trabajos aplicados. Aparenta tener la flexibilidad necesaria para proveer un buen ajuste para diferentes cantidades de modelos de recuento. Lo hace en parte debido a que la especificación de varianza cuadrática

es una buena aproximación en muchas situaciones empíricas. Una consecuencia infortunada del hecho que BN2 casi siempre provea un buen ajuste es que si el supuesto de Poisson falla existe una tendencia a saltar a la alternativa binomial negativa, ignorando otras posibilidades. Tales como una aproximación mecánica debería ser evitada debido al bajo rendimiento de la Poisson puede también ser por las especificaciones de bajo nivel de la función de media condicional, y observar que usando el modelo binomial negativo mantienes la misma media condicional.

El modelo binomial negativo es menos robusto cuando no se especifica la distribución que la Poisson. Inclusive si la media condicional está correctamente especificada la estimación por máxima verosimilitud en modelos binomiales negativos es inconsistente, excepto por el caso especial del modelo BN2, mientras que la estimación por máxima verosimilitud para  $\beta$  es todavía consistente.

## 2.5. Hipótesis

### **Hipótesis General**

- Los factores sociales y económicos seleccionados afectan e influyen de manera directamente proporcional a los años de operación de las MYPE, incrementando las probabilidades de seguir en funcionamiento.

### **Hipótesis específicas**

- Niveles de educación mayor a 15 años o formación profesional especializada por parte de los gerentes generales o administradores a cargo generan un impacto positivo frente a los años de operación de las MYPE, incrementando las probabilidades de seguir en funcionamiento.
- Niveles bajos de informalidad por parte de los negocios, evidenciados por la tenencia de una personería jurídica, la posesión de licencias de funcionamiento o el hecho de contar con locales alquilados, genera una relación directamente proporcional frente a

los años de operación de las MYPE, incrementando las probabilidades de seguir en funcionamiento.

# CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN

### 3.1. Procedimiento de investigación

El presente trabajo de investigación tiene como primera fase el análisis de documentos a partir de la búsqueda bibliográfica y selección de investigación documental general y específica, consolidándolo a través de fichas o registros electrónicos que sea de utilidad para la descripción y análisis del contenido para la construcción del Marco teórico.

Luego y en segundo lugar se realizará un análisis macroeconómico del problema de investigación relacionando el producto bruto interno y los términos de intercambio con el objetivo de plantear una visión general del contexto en el cual se desarrolla el análisis micro econométrico planteado en esta investigación.

De forma paralela se presentará la descripción y análisis de las variables a utilizar en el modelo de regresión como parte del contenido para la aplicación de las fases de la metodología del análisis de regresiones econométricas. Por último se emplearán métodos de predicción y simulación de escenarios para las diversas posibilidades dentro de las variables seleccionadas.

### 3.2. Secuencia metodología del análisis econométrico

Para cumplir con los objetivos del presente trabajos se presenta la metodología del análisis de correlación en los modelos econométricos, que tiene por objetivo la aplicación de procedimientos estadísticos y matemáticos a problemas sociales o donde se necesite el análisis económico, los cuales se caracterizan por ser complejos, únicos y con gran cantidad de información de por medio.

Para tratar con los problemas asociados al comportamiento de los agentes económicos dentro del mercado, el análisis de correlación econométrica consiste en realizar tres fases: Visión macro económica, Enfoque micro econométrico y por último la Predicción de escenarios.

### 3.2.1. Visión macro económica del problema

En la primera fase se plantea el problema socioeconómico para su análisis de manera que se puedan identificar las variables macro económicas que permitan modelar el comportamiento del fenómeno, para luego ser analizados posteriormente.

Para tal efecto se realizan la búsqueda de la información en base a las variables relevantes siempre y cuando se tenga datos de instituciones públicas que permitan más adelante contrastar los resultados obtenidos con la fuente. Es fundamental delimitar el estudio mediante los objetivos para que no se trabaje o procese información innecesaria que a pesar de influir en menor medida puede modificar las relaciones entre variables.

Primero por lo tanto se escoge la variable dependiente del modelo econométrico, siendo está conformada por una serie de tiempo. Luego se procede a definir los regresores o variables independientes los cuales explicaran el comportamiento del término endógeno, por ende es necesario contar con diversas alternativas para que luego se puedan ir descartando conforme se construye el modelo econométrico.

Algo muy importante de clarificar consiste en tener como consideración que se están evaluando indicadores macro económicos, donde se deja de lado los impactos reflejados por la cantidad de agentes económicos que simbolizan dicho indicador. Mientras que solo se enfoca en un análisis histórico siempre relacionado con el producto bruto interno, con el objetivo de contar con una guía sobre el desempeño de cada sector económico.

Después de haber seleccionado el modelo que explicará el fenómeno económico mediante las variables independientes y dependiente. Procedemos a realizar un análisis individual de cada una de las series de tiempo escogidas, mostrando sus principales estadísticos junto a una gráfica que proporcione información sobre la estacionalidad. De contar con dicha estacionalidad se procede a eliminar la misma mediante un proceso econométrico, quitando los ciclos que afectan la línea de tiempo. Finalizando dicha etapa para cada variable seleccionada, se realiza el análisis grupal de variables que consiste en la visualización de graficas de dispersión entre las variables exógenas y la endógena con el objetivo de vislumbrar la existencia de relaciones directamente proporcionales que justifiquen la regresión



econométrica. Ya que si no se cuenta con ninguna relación grafica se estarían agregando datos espurios a la muestra perjudicando el modelo. Al final se calcula la matriz de correlación entre todas las variables con lo cual se estaría observando de forma primaria si las variables seleccionadas poseen un nivel de correlación significativo que se deba considerar al momento de la regresión econométrica.

Por último se realiza la regresión econométrica mediante mínimos cuadrados, obteniendo de esta manera los primeros coeficientes de regresión para cada variable independiente. Considerando los estadísticos y probabilidades, descartamos o aumentamos el número de variables en el modelo, siempre y cuando no se vea perjudicado el estadístico  $R^2$ , que mide el nivel de ajuste de nuestra ecuación estimada. Para finalizar se analizan los residuos con el objetivo de corregir algún problema de auto correlación agregando el termino respectivo a la ecuación econométrica.

### 3.2.2. Enfoque micro econométrico del problema

Luego de la fase de exploración con visión macro del caso de estudio seleccionado se procede a delimitar de forma micro las variables que impactaran al término de interés. Para esto se debe seleccionar el conjunto de datos que brindara toda la información, siendo el principal aporte las características para cada individuo. De esta forma analizamos de forma transversal diversos atributos para un grupo considerable de muestra. Por ende seleccionamos las variables que expliquen para un determinado año un fenómeno social o económico recurrente.

Luego de seleccionar las variables explicativas del modelo micro econométrico, se procede a crear la variable explicada o dependiente. Esta tendrá que ser el resultado de los cambios y relaciones entre cada uno de los regresores, de forma que permita explicar el comportamiento en modelo ajustado. Después se procede a la selección de la distribución que explica dicho comportamiento, siendo para este caso particular un modelo de datos de conteo por tener como variable explicada valores positivos enteros. Para esto empleamos las regresiones de Poisson y Binomial negativa, las cuales están diseñadas para seguir el comportamiento de variables categorizadas incluyendo la probabilidad en cada iteración de regresión.

Finalizando la selección de variables y con la creación del termino dependiente a ser explicado se procede con la descripción de las variables del modelo determinado. De esta forma se crea una nomenclatura específica para cada una de ellas que servirá al momento de programar en los programas estadísticas. Cada una de las variables se presenta con sus principales indicadores y características, luego se procede con un análisis de las variables donde analizamos el comportamiento de cada muestra junto a los estadísticos más resaltantes.

Posteriormente se procede con la estimación inicial y las pruebas de ajuste, las cuales consisten en realizar el cálculo de la regresión econométrica considerando el tipo de distribución que explique de forma acertada el comportamiento del modelo. Los estadísticos obtenidos dentro del primer computo nos mostraran si es el adecuado por lo tanto se compara con las hipótesis iniciales de cada uno para descartar cualquier imperfección. La prueba de ajuste permite acercar la regresión realizada aún más al modelo propuesto, de manera que se verifique la predisposición matemática a explicar dicho fenómeno socioeconómico.

Para el caso de la regresión micro econométrica de datos de conteo se necesita realizar un test de sobredispersión al momento de finalizar la primera estimación, ya que al tener características particulares entre la media y la varianza se debe descartar que exista para que no perjudique el nivel de ajuste del modelo. En el caso que la hipótesis inicial sea rechazada para el estadística que determina la sobredispersión se procede con la realización de una regresión del modelo binomial negativo, el cual elimina esta característica asegurando una mejor explicación al momento de calcular la regresión.

### 3.2.3. Predicción de escenarios

Como resultado de las dos etapas previas y terminando con la regresión del modelo micro econométrico binomial negativo se realiza el cálculo de los impactos marginales que tendrán el aumento o disminución de cada una de las variables explicativas para nuestro caso de estudio. Además se utiliza la probabilidad estadística para simular escenarios donde se consideren las diversas opciones de cada una de las variables categorizadas. Este proceso consiste en emplear el término independiente creado al inicio de la fase de enfoque micro econométrico para observar su evolución para cada valor empezando desde la unidad hasta

un límite determinado, que muestre el cambio en la probabilidad condicional asociada a cada uno. Para la predicción de escenarios por lo tanto es necesario contar con el modelo final y que no posea sobredispersión, ya que el objetivo principal es diseñar casos que permitan entender las diversas posibilidades dentro del caso de estudio.

### 3.3. Diseño de la investigación

El presente documento de acuerdo al propósito corresponde a un diseño de investigación no experimental, ya que no existe manipulación activa de alguna variable. Toda la información se encuentra disponible en las bases de datos del Instituto Nacional de Estadística e Informática (INEI).

Por la clase de medios utilizados para obtener la información relevante se considera como una investigación documental ya que está apoyado en fuentes de carácter documental, es decir tiene revisiones bibliográficas relacionadas al tema de análisis econométrico y de factores cualitativos sociales a través de modelos de regresión micro y macro. Dicha revisión bibliográfica busca documentación general y específica de fuentes documentadas en bibliotecas, sitios web, revistas especializadas y estudios de investigación sobre factores que afectan el desarrollo de las MYPE.

Por el nivel de conocimiento se considera que la investigación es correlacional debido a la búsqueda por determinar el tipo de relación entre los factores socioeconómicos que se relacionan con el tiempo de operación de las micro y pequeña empresas.

El método de análisis, se ha tomado como referencia a los modelos econométricos con la teoría de distribución de probabilidades para muestras de conteo la que se encuentra vinculada con la metodología del análisis correlacional donde se plantean ciertas condiciones para asegurar que un modelo es el mejor para explicar el caso de estudio.

El método estadístico consiste en recopilar la información de una base de datos existente y trabajarla en base a los datos que se necesitan para el modelo, luego del procesamiento se analiza con el programa Stata, obteniéndose como resultado los coeficientes de regresión para cada ecuación econométrica que nos permita contrastar con las hipótesis planteadas e interpretadas desde la visión económica y social.

El trabajo se enmarca en el desarrollo de una tesis con un enfoque desde el punto de vista de aplicación del análisis econométrico para el estudio de las MYPE que muchas instituciones públicas no consideran al momento de estudiar los factores determinantes en el funcionamiento de este sector empresarial.

Este trabajo de investigación es de enfoque cuantitativo ya que realiza mediciones estadísticas y computo de regresiones para la explicación de fenómenos socioeconómicos cualitativos que poseen características medibles. Por lo tanto se realizan correlaciones entre las principales variables causantes, las cuales fueron seleccionadas previamente para determinar el tipo de relación con el tiempo de operación de las MYPE, empleando metodología de análisis correlacional econométrico.

### 3.4. Población y muestra

#### 3.4.1. Población

Para determinar la población de micro y pequeñas empresas, se presenta la unidad de análisis

Unidad de análisis: Empresas nacionales o con accionistas peruanos que facturen anualmente entre 300 UIT y 1000 UIT (Unidad Impositiva Tributaria)

Por lo tanto nuestra población consiste en el 96% de negocios registrados dentro del Ministerio de la Producción (Produce) bajo el régimen MYPE.

#### 3.4.2. Muestra

La muestra se determinará según la formula general expuesta en todo libro de estadística o artículo de investigación, siendo la fórmula para poblaciones menores a 10,000 unidades de análisis la siguiente:

$$n = \frac{N\sigma^2Z^2}{(N - 1)e^2 + \sigma^2Z^2}$$

Donde,  $N$  es el tamaño de la población,  $n$  el tamaño de la muestra,  $\sigma$  es la desviación estándar de la población a valor constante de 0.5,  $Z$  que es el valor a base de la función normal dado el nivel de confianza se emplea 95% y su valor de  $Z$  sería de 1.96 y por último  $e$  como el error muestral que se asocia a la diferencia de 1-nivel de confianza.

En el caso que la población sea mayor a 10,000 unidades de análisis la fórmula es la siguiente:

$$n \frac{\sigma^2 Z^2}{e^2}$$

Por lo tanto luego de realizar los cálculos se obtiene como valor para la muestra la cantidad de 2352 unidades de análisis, las cuales estarían conformadas por empresas del sector MYPE. Todo esto teniendo un nivel de confianza del 95%, siendo estos resultados empleados en la fase de enfoque micro econométrico.

### 3.5. Operacionalización de las variables

Se parte del concepto de la variable, definiéndola como cualquier característica o cualidad de la realidad que es susceptible de asumir diferentes valores, es decir puede variar, aunque para un determinado objeto se considere que puede tomar un valor fijo.

De acuerdo al sujeto de investigación las variables se han clasificado en categóricas y discretas.

Las variables categóricas, clasifican a los sujetos distribuyéndolos en grupos, de acuerdo a algún atributo previamente establecido y están muy ligados con las variables discretas las cuales se definen como aquellas que no admiten posiciones intermedias entre dos números.

Este tipo de variables se ha subdividido en dos:

Variabes dicotómicas que poseen dos categorías, como por ejemplo la característica de un inmueble empleado por las empresas bajo las opciones de que solo puede ser propio o alquilado.

Variabes policotómicas que establecen tres o más categorías, como por ejemplo el nivel de educación del gerente de la empresa MYPE.

Asimismo la investigación presenta variables discretas cuando se mide la edad de los encargados del negocio y para el cálculo del tiempo de operación de las empresas.

Además las variables se dividen en dos grupos siendo las independientes y las dependientes.

La principal variable dependiente es el tiempo de operación de las MYPE, la cual se mide en años completos por lo que no acepta términos decimales. Mientras que como variables independientes se puede mencionar la edad del gerente general, el nivel de educación o formación profesional del mismo y el régimen del negocio al cual pertenece. El cuadro muestra las variables empleadas en este modelo de regresión micro econométrico.

**Tabla 1 definición de variables**

<b>Variables</b>	<b>Indicador</b>	<b>Tipo de indicador</b>	<b>Método de Cálculo</b>	<b>Unidad de análisis</b>	<b>Base de datos</b>	<b>Fuente</b>
<b>Operación</b>	Años de funcionamiento de la MyPE	Autoidentificación numérica	2013 – Año de inicio de operaciones	Empresas	Encuesta de Micro y Pequeña Empresa (EMYPE 2013)	Beatrice Avolio (2011)
<b>Edad</b>	Años cumplidos del encargado de la MyPE al 2013	Numérica	Identificación mediante encuesta	Individuos	Encuesta de Micro y Pequeña Empresa (EMYPE 2013)	Beatrice Avolio (2011)
<b>Nivel educativo</b>	Grado de educación del encargado de la MyPE	Categoría	1 : Primaria 2 : Secundaria 3 : Superior	Individuos	Encuesta de Micro y Pequeña Empresa (EMYPE 2013)	Beatrice Avolio (2011)
<b>Organización</b>	Tipo de empresa constituida	Dicotómica	0 : Natural 1 : Jurídica	Empresas	Encuesta de Micro y Pequeña Empresa (EMYPE 2013)	Cepal
<b>Licencia de Funcionamiento</b>	Licencia municipal emitida y válida hasta el año 2013	Dicotómica	0 : No cuenta 1 : Si cuenta	Empresas	Encuesta de Micro y Pequeña Empresa (EMYPE 2013)	Paul Lira Briceño (2013)
<b>Local</b>	Tipo de posesión del local, propio o alquilado	Dicotómica	0 : Alquilado 1 : Propio	Empresas	Encuesta de Micro y Pequeña Empresa (EMYPE 2013)	Cepal

Fuente: Encuesta de la Micro y Pequeña empresa (EMYPE)

Elaboración: propia

### 3.6. Instrumentos de recolección de datos

Recolección de datos secundarios, el registro de observación documental o recopilación bibliográfica ha partido de la búsqueda de páginas educativas en Internet y ubicación de datos concretos necesarios de información general y específica para el tema de las micro y pequeña empresas, siendo las características de estas el principal objeto de búsqueda.

Todas las variables se basan en fuentes secundarias como bases de datos estadísticas ya que no se realizarán encuestas.

No se utilizó ningún experimento, porque se estableció como límite el proceso de información estadística nacional la cual se encontraba disponible de acceso público.

### 3.7. Procesamiento y análisis

A partir de los datos secundarios recopilados, del uso de las herramientas econométricas existentes en cada una de las fases de la metodología del análisis econométrico, y de la aplicación del software Stata, así como complementos de Excel se obtendrá información relevante la cual se mostrará, interpretará, discutirá y que conducirá a las conclusiones y recomendaciones del presente trabajo de investigación.

# CAPÍTULO IV: DESARROLLO DE LA METODOLOGÍA



## 4.1. Análisis macro económico

### 4.1.6. PBI vs Régimen MYPE

El análisis principal relacionado con el tiempo de vida de las empresas del régimen micro y pequeña, debe estar vinculado a las variables macroeconómicas donde se desenvuelven y operan dichos negocios de variados rubros. Por tal motivo se analizará el problema de investigación buscando alguna correlación existente entre el Producto Bruto Interno (PBI) y la cantidad de empresas en los diversos regímenes de interés (Régimen Especial a la Renta, Régimen Único Simplificado e Independiente) con la hipótesis inicial de que a un incremento constante del PBI durante el periodo de estudio se tendría también un aumento significativo del registro de empresas en el régimen MYPE.

#### Análisis individual: descripción de variables

Las variables empleadas para esta parte del análisis serán el Producto Bruto interno como variable dependiente (Y), y las variables independientes Cantidad de contribuyentes en los tres tipos de régimen: RER, RUS e independiente.

- **Producto Bruto Interno (PBI)**

La variable dependiente del PBI es la cantidad de bienes y servicios finales producidos durante un periodo determinado de tiempo dentro de un espacio geográfico establecido.

Variable dependiente: PBI

Periodo de estudio: 2000 – 2013

Unidades: Millones de soles

Fuente de información: INEI

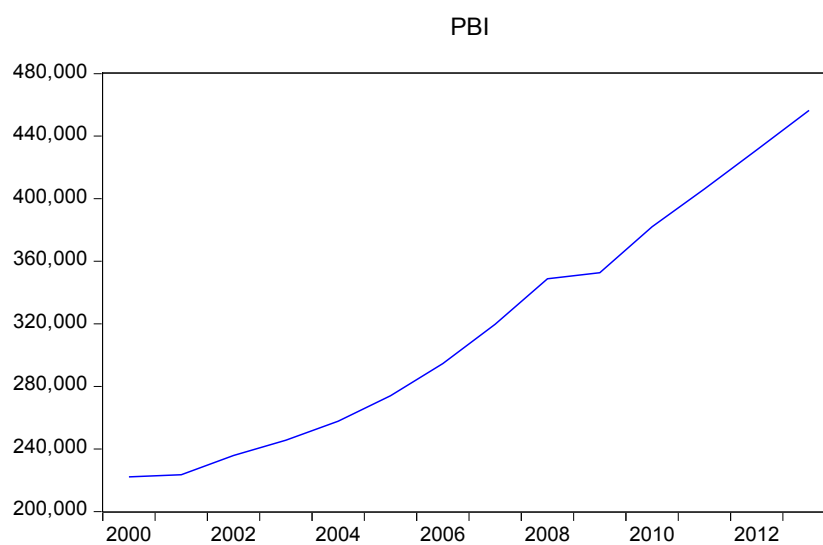
**Tabla 2 : Estadísticas descriptivas de la variable PBI**

PBI	
Media	317908.5
Mediana	307145.5
Máximo	456435.0
Mínimo	222207.0
Desviación Estándar	79538.29
Asimetría	0.345023
Kurtosis	1.793002
Jarque-Bera	1.127588
Probabilidad	0.569046
Sum	4450719.
Sum Desv. Cuadr.	8.22E+10
Observaciones	14

Fuente: INEI

Elaboración: propia

**Gráfico 5 Evolución del PBI**



Fuente: INEI

Elaboración: propia

- **Cantidad de contribuyentes en el Régimen Especial de impuesto a la Renta (RER)**

Dentro del régimen de contribuyentes a la Sunat existe la categoría de Régimen Especial a la Renta que está dirigido a micro empresarios o personas naturales con negocio que solo generan ingresos netos anuales de hasta 525,000 soles. Por lo tanto el pago de tributos al cual se encuentra afecto son el Impuesto General a las Ventas (IGV) y el 1.5% sobre el

ingreso neto mensual. Otro beneficio para la micro y pequeña empresa está por el lado de la simplificación contable ya que solamente deben manejar dos libros contables durante la operación de la empresa.

Periodo de estudio: 2000 – 2013

Variable independiente: RER

Unidades: Miles de empresas

Fuente de información: Sunat, MTPE, INEI

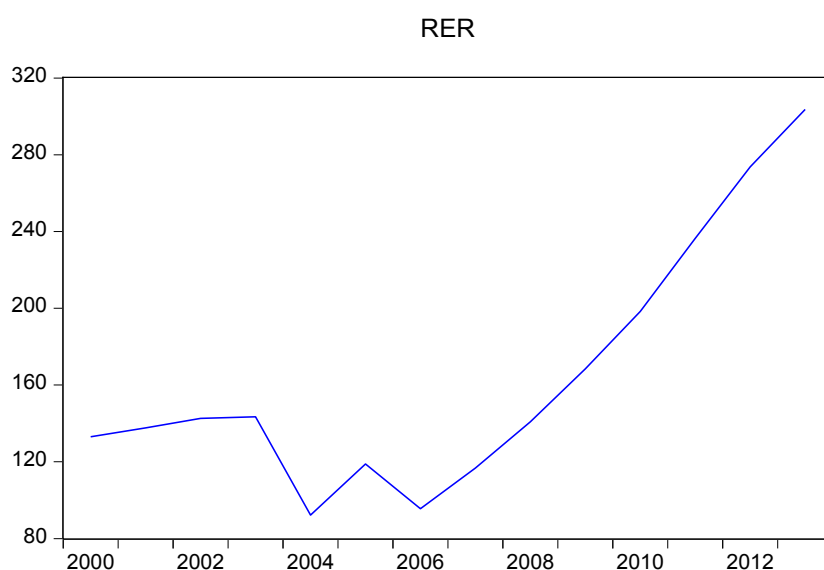
***Tabla 3 Estadísticas descriptivas de la variable RER***

<b>RER</b>	
<b>Media</b>	164.4000
<b>Mediana</b>	141.7000
<b>Máximo</b>	303.6000
<b>Mínimo</b>	92.30000
<b>Desviación estándar</b>	65.11016
<b>Asimetría</b>	0.991886
<b>Kurtosis</b>	2.800013
<b>Jarque-Bera</b>	2.318954
<b>Probabilidad</b>	0.313650
<b>Sum</b>	2301.600
<b>Sum Desv. Cuadr.</b>	55111.32
<b>Observaciones</b>	14

Fuente: Sunat - MTPE - INEI

Elaboración: propia

## Gráfico 6 Serie de tiempo del Regimen Especial a la Renta



Fuente: Sunat - MTPE - INEI

Elaboración: propia

- **Cantidad de contribuyentes en el Régimen Único Simplificado (RUS)**

Este régimen especial dentro de la categoría de contribuyentes de la Sunat ha experimentado diversas modificaciones y cambios a lo largo de los distintos periodos de gobierno hasta ahora quedando con el Nuevo RUS, donde se especifican las responsabilidades tributarias especiales así como los requisitos para poder registrarse dentro de dicho régimen. La característica principal consiste en realizar pagos mensuales de acuerdo a una tabla dividida en categoría con montos relacionados a las compras y/o ingresos, de esta forma se eliminan diversos pagos tributarios existentes dentro del régimen general, simplificando la operación y manejo de las micro empresas o pequeños productores.

Periodo de estudio: 2000 – 2013

Variable independiente: RUS

Unidades: Miles de empresas

Fuente de información: Sunat, MTPE, INEI

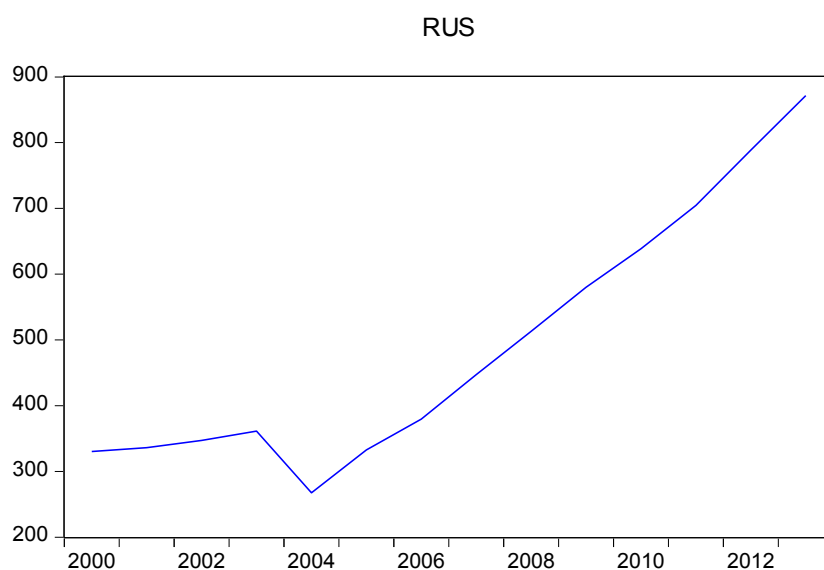
**Tabla 4 Estadísticas descriptivas de la variable RUS**

RUS	
Media	492.7429
Mediana	413.4500
Máximo	871.4000
Mínimo	267.6000
Desviación Estándar	193.3329
Asimetría	0.695559
Kurtosis	2.147096
Jarque-Bera	1.553214
Probabilidad	0.459964
Sum	6898.400
Sum Desv. Cuadr.	485909.0
Observaciones	14

Fuente: Sunat - MTPE - INEI

Elaboración: propia

**Gráfico 7 Serie de tiempo del Regimen Unico Simplificado (RUS)**



Fuente: Sunat - MTPE - INEI

Elaboración: propia

- **Cantidad de contribuyentes registrados como trabajadores independientes.**

La categoría de trabajadores independientes agrupa a diversos tipos de contribuyentes ya sean personas naturales con negocio o una Empresa Independiente de Responsabilidad Limitada, estas se caracterizan por tener un flujo de ingresos y trabajo variable, es decir

pueden emplear recibos por honorarios para justificar la generación de renta en contraparte a un servicio brindado. De esta forma se encuentran afectos a la renta de tercera categoría pagando un porcentaje determinado por Sunat sobre los ingresos anuales registrados.

Periodo de estudio: 2000 – 2013

Variable independiente: INDEP

Unidades: Miles de empresas

Fuente de información: Sunat, MTPE, INEI

***Tabla 5 Estadísticas descriptivas de la variable trabajadores independientes***

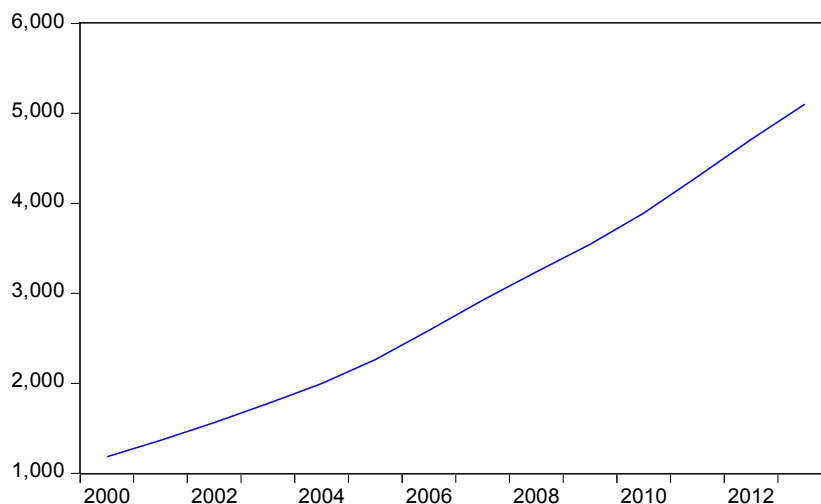
<b>INDEP</b>	
<b>Media</b>	2888.593
<b>Mediana</b>	2755.200
<b>Máximo</b>	5102.400
<b>Mínimo</b>	1183.600
<b>Desviación Estándar</b>	1278.729
<b>Asimetría</b>	0.299520
<b>Kurtosis</b>	1.840476
<b>Jarque-Bera</b>	0.993618
<b>Probabilidad</b>	0.608469
<b>Sum</b>	40440.30
<b>Sum Desv. Cuadr.</b>	21256929
<b>Observaciones</b>	14

Fuente: Sunat - MTPE - INEI

Elaboración: propia

### **Gráfico 8 Serie de tiempo de los trabajadores independientes (INDEP)**

INDEP



Fuente: Sunat - MTPE - INEI

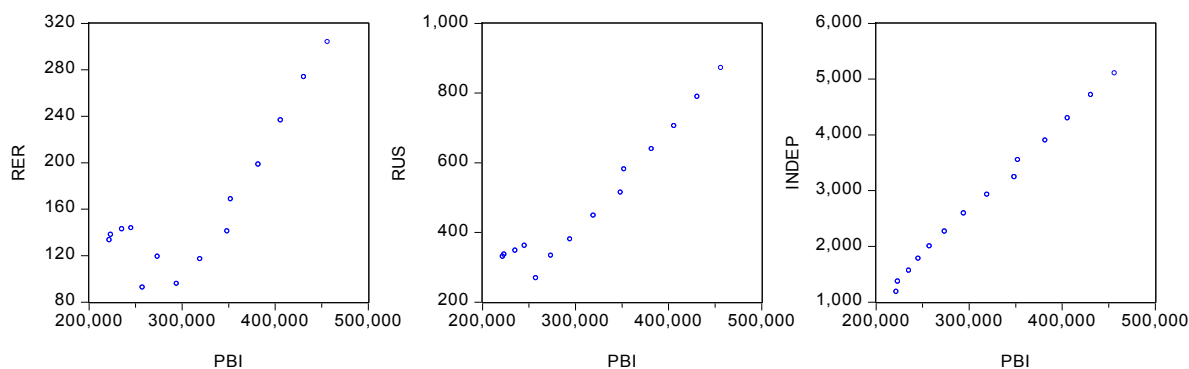
Elaboración: propia

### **Análisis grupal de variables**

La base de datos contiene 14 observaciones para analizar la relación existente entre el nivel del Producto Bruto Interno (PBI), medido a partir de la participación y aporte de las empresas registradas en los siguientes regímenes tributarios, Régimen Especial de Impuesto a la Renta (RER) , Régimen Único Simplificado (RUS) y Trabajadores independientes (INDEP). De esta forma se infiere que existe una relación entre la cantidad de empresas que se registran en categorías especiales de contribución frente al incremento o disminución del PBI.

Se elabora la primera inspección gráfica y numérica de las series para obtener un primer análisis del grado y la forma de relación existente entre las variables.

**Gráfico 9** *Dispersión del PBI vs RER, RUS, INDEP*



Fuente: Sunat - MTPE - INEI

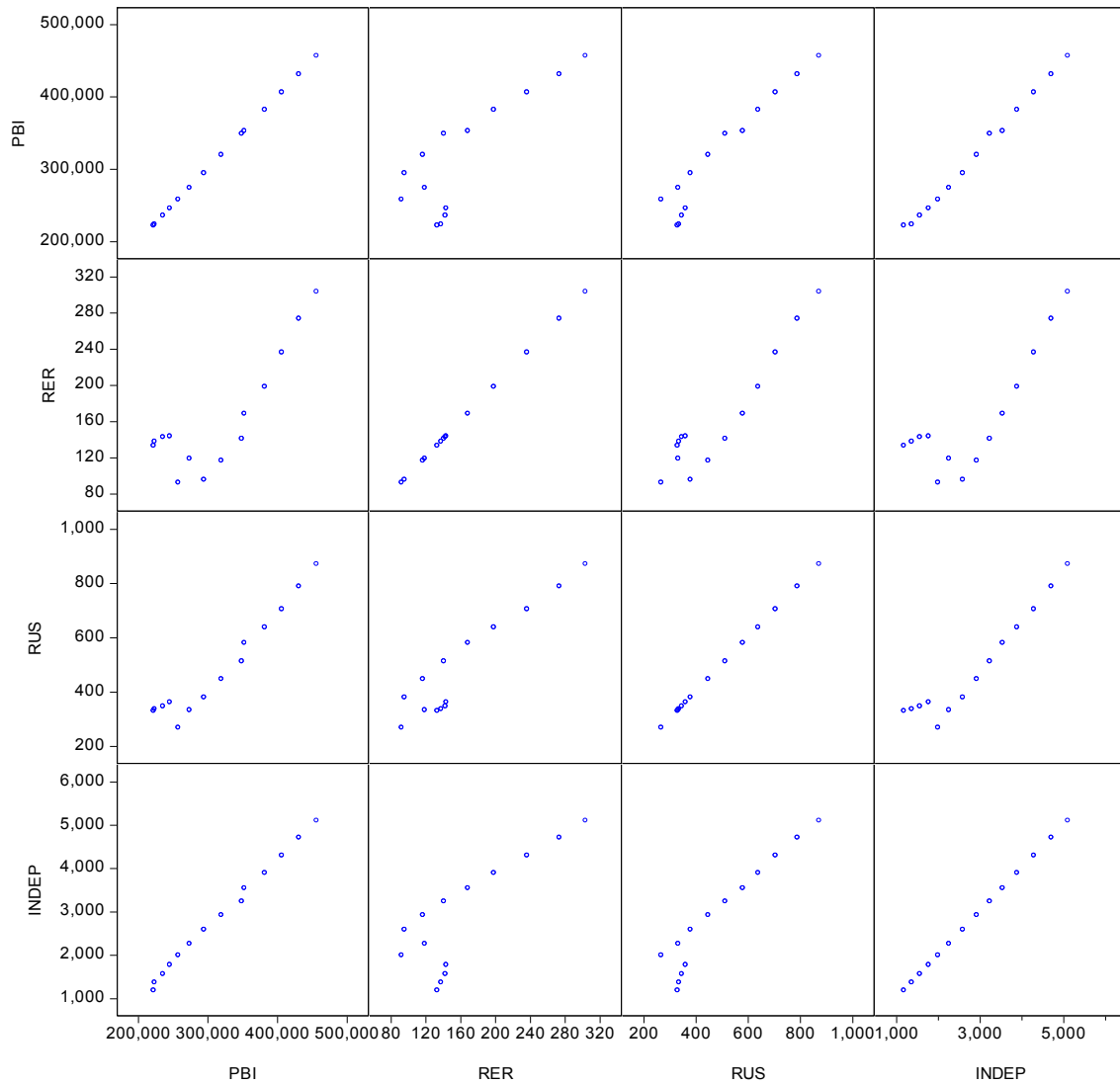
Elaboración: propia

Para el caso de estudio, los gráficos muestran una clara relación positiva lineal entre el nivel de Producto Bruto Interno (PBI) y cada una de las variables consideradas, si bien parece que la relación es más fuerte con la cantidad de trabajadores independientes (INDEP), se tiene un comportamiento claro para las otras variables de estudio.

Para obtener una visión más general de la relación entre el conjunto de variables consideradas está, se obtendría mediante la comparación dentro de una matriz de cada variable incluida las explicativas y representándolas en una gráfica para cada par de variables.



**Gráfico 10 Matriz de gráficas de dispersión**



**Fuente:** Sunat - MTPE - INEI

**Elaboración:** propia

Para estudiar el nivel de dependencia en números entre las variables de estudio se procede a obtener la matriz de correlaciones. De donde se concluye que las variables RUS e INDEP son las que presentan mayor correlación lineal con el nivel del PBI. Además existe cierta correlación lineal entre RUS y RER, así como entre INDEP y RUS.

**Tabla 6 Resultados de la matriz de correlación**

	PBI	RER	RUS	INDEP
PBI	1.000000	0.835964	0.963911	0.998117
RER	0.835964	1.000000	0.939814	0.826901
RUS	0.963911	0.939814	1.000000	0.956853
INDEP	0.998117	0.826901	0.956853	1.000000

Fuente: Sunat - MTPE - INEI

Elaboración: propia

### Estimación por mínimos cuadrados ordinarios

Se procede a realizar la primera regresión lineal en base a la ecuación:

$$PBI = \beta_1 * RER + \beta_2 * RUS + \beta_3 * INDEP + c$$

La tabla N°7 muestra los coeficientes obtenidos para cada variable independiente y su nivel de significancia con el cual se determina si es prescindible o no para el modelo. Para un nivel de confianza del 95% se tiene que la variable RER no cumple con el parámetro. Por la tanto se procede a realizar la segunda regresión configurando otra ecuación de regresión para obtener valores más ajustados.

**Tabla 7 Coeficientes de regresión**

Variable	Coeficiente	Estadístico-t	Prob.
RER	-110.9895	-1.377620	0.1984
RUS	108.9523	2.074880	0.0648
INDEP	50.99514	10.57148	0.0000
C	135165.5	39.39127	0.0000

Fuente: Sunat - MTPE - INEI

Elaboración: propia

Se procede a realizar la segunda regresión lineal en base a la ecuación:

$$PBI = \beta_1 * RUS + \beta_2 * INDEP + c$$

**Tabla 8 Coeficientes de la segunda regresión**

Variable	Coeficiente	Estadístico-t	Prob.
RUS	43.17045	1.900156	0.0839
INDEP	55.83855	16.25585	0.0000
C	135341.7	37.95159	0.0000

Fuente: Sunat - MTPE - INEI

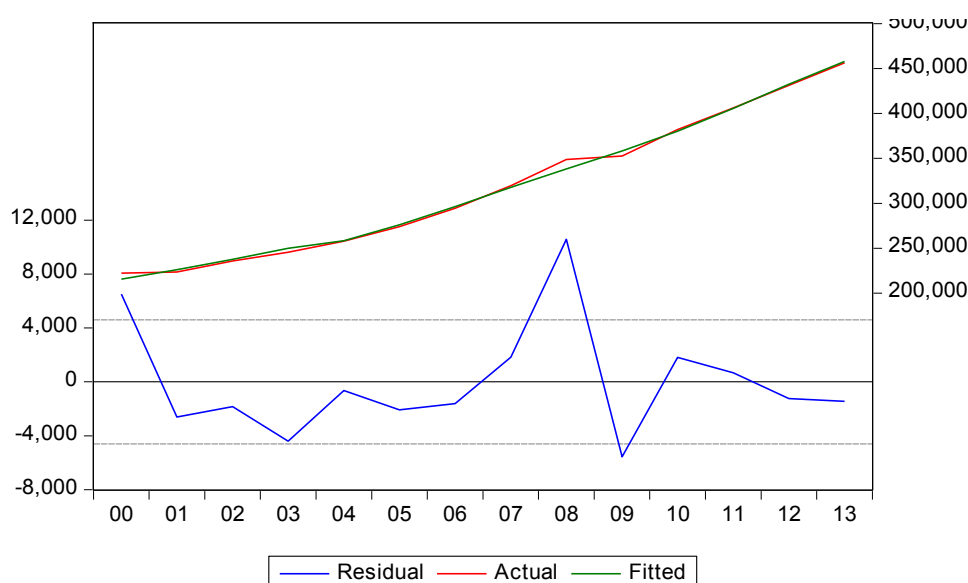
Elaboración: propia

Se observa que todas las variables cuentan con el signo esperado y además son significativas al 10%. El estadístico R2 indica un valor del 99,7% lo cual implica un alto grado de ajuste al modelo lineal. Las variables independientes RUS e INDEP poseen una dependencia lineal positiva, de esta manera se tendría un incremento de 43.17 millones en el PBI cuando se incrementa en 1 unidad la cantidad de empresas en el régimen RUS o de 55.83 millones en el caso de los trabajadores independientes (INDEP).

- Análisis de auto correlación

El estadístico Durbin-Watson indica un valor de 2.21, el cual es cercano a 2 por lo tanto no existe presencia de auto correlación. Además en la gráfica N°7 se observa que el residuo no sobrepasa las bandas siendo este otro indicador de estabilidad.

**Gráfico 11 Test de Auto correlación**



Fuente: Sunat - MTPE - INEI

Elaboración: propia

## Análisis de resultados y conclusiones

Tomando como punto de partida la hipótesis inicial donde se asegura una relación directamente proporcional entre el crecimiento de la economía y la cantidad de empresas en los regímenes MYPE, además comparando con los resultados obtenidos de la regresión econométrica se puede contemplar que para las empresas MYPE del RUS se tiene una relación positiva, evidenciada por el coeficiente de regresión (43.17). Por ende para cada mil empresas registradas en el RUS se espera un incremento del PBI de 43.17 millones de soles. En el caso de los trabajadores independientes de igual forma se observa una relación directamente proporcional con un coeficiente de regresión positivo (55.83) demostrando el impacto que generaría en el PBI (incremento de 55.83 millones de soles) una adición de 1000 personas que realicen labores independientes. Por lo tanto en conclusión se observa que para lograr un crecimiento importante y sostenido del PBI es necesario tener en consideración a los regímenes MYPE, brindado a ese sector numerosas políticas locales o nacionales que les permita continuar aportando durante un periodo mayor a los 3 años.

### 4.1.6. Términos de intercambio vs Régimen MYPE

El enfoque macroeconómico establecido en el análisis previo permitió vislumbrar una nueva relación entre las actividades económicas de exportación e importación con la evolución anual del sector de las micro y pequeña empresas, evidenciado mediante el incremento del PBI dentro del periodo de análisis. De esta forma se pretende encontrar una correlación entre el índice de términos de intercambio con el Producto Bruto Interno y la cantidad de empresas en los diversos regímenes de interés (Régimen Especial a la Renta, Régimen Único Simplificado e Independientes), con la hipótesis inicial de que los términos de intercambio, los cuales relacionan precios de exportación con precios de importación, se ven impactados positivamente ante un incremento constante en el número de empresas de los regímenes MYPE e independiente causando una mayor participación en actividades de comercio internacional.

- Análisis individual: descripción de la variable dependiente

Las variables empleadas para esta sección serán el índice de Términos de Intercambio como variable dependiente (Y), y las variables independientes Producto Bruto Interno (PBI) y la cantidad de contribuyentes en los regímenes RER, RUS e independiente.

a. Índice de Términos de Intercambio

Índice que relaciona un índice de precios de exportación con un índice de precios de importación. Refleja el poder adquisitivo de nuestras exportaciones respecto de los productos que importamos del exterior.

Variable dependiente: TOT

Periodo de estudio: 2000-2013

Unidad: índice con base 2007

Fuente de información: BCRP

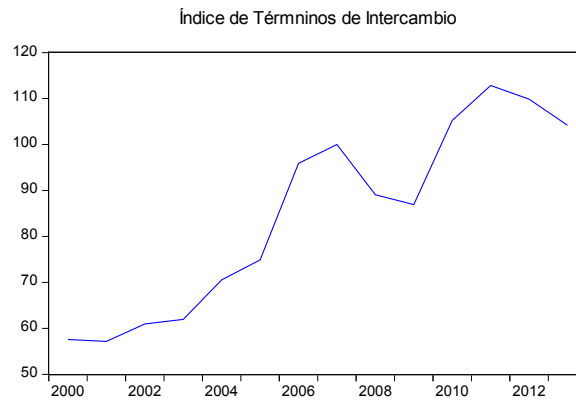
***Tabla 9 Estadísticas descriptivas de la variable TOT***

TOT	
<b>Media</b>	84.78904
<b>Mediana</b>	87.99536
<b>Máximo</b>	112.8408
<b>Mínimo</b>	57.14284
<b>Desviación Estándar</b>	20.53411
<b>Asimetría</b>	-0.110234
<b>Kurtosis</b>	1.480454
<b>Jarque-Bera</b>	1.375281
<b>Probabilidad</b>	0.502761
<b>Sum</b>	1187.047
<b>Sum Desv. Cuadr.</b>	5481.445
<b>Observaciones</b>	14

Fuente: BCRP

Elaboración: propia

## Gráfico 12 Serie de tiempo del índice de términos de intercambio



Fuente: BCRP

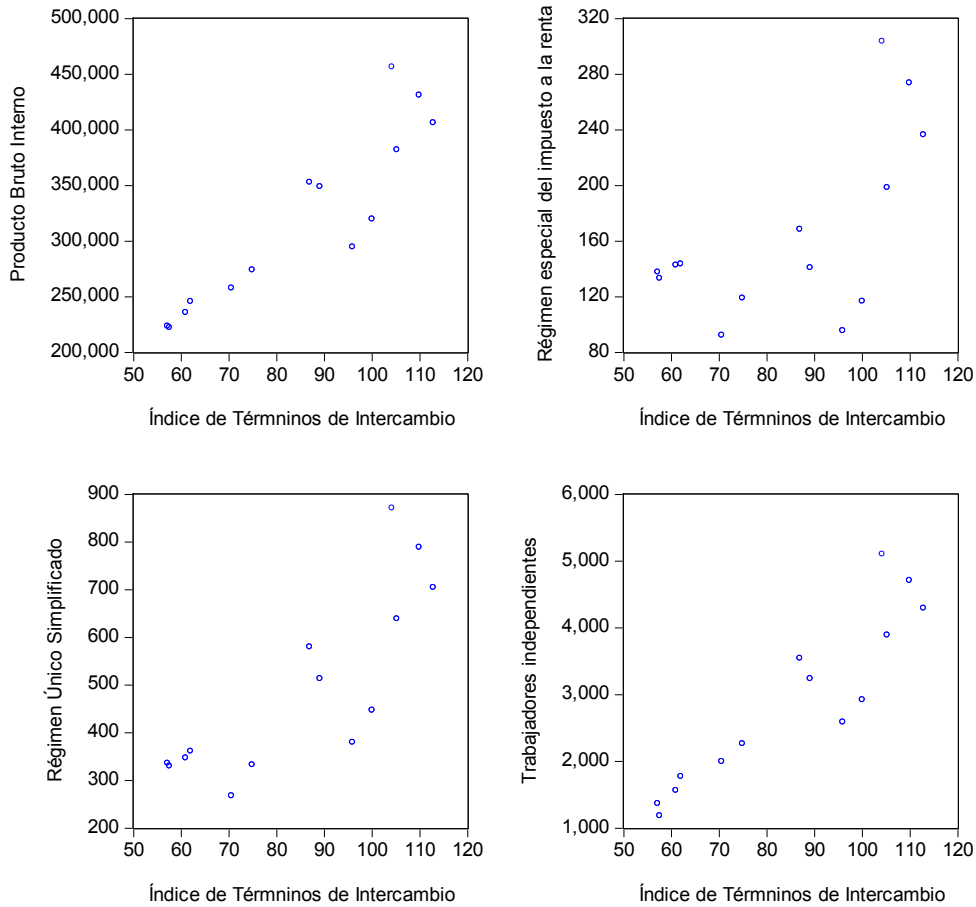
Elaboración: propia

- Análisis grupal de variables

La base de datos contiene 14 observaciones con el objetivo de analizar una relación existente entre el Índice de términos de intercambio (TOT), y el incremento del producto bruto interno (PBI) en conjunto con un aumento gradual de la cantidad de empresas en los regímenes MYPE.

Se elabora la primera inspección gráfica y numérica de las series para obtener un primer alcance del grado y forma de relación existente entre las variables.

**Gráfico 13 Dispersión TOT vs PBI, RUS, INDEP**



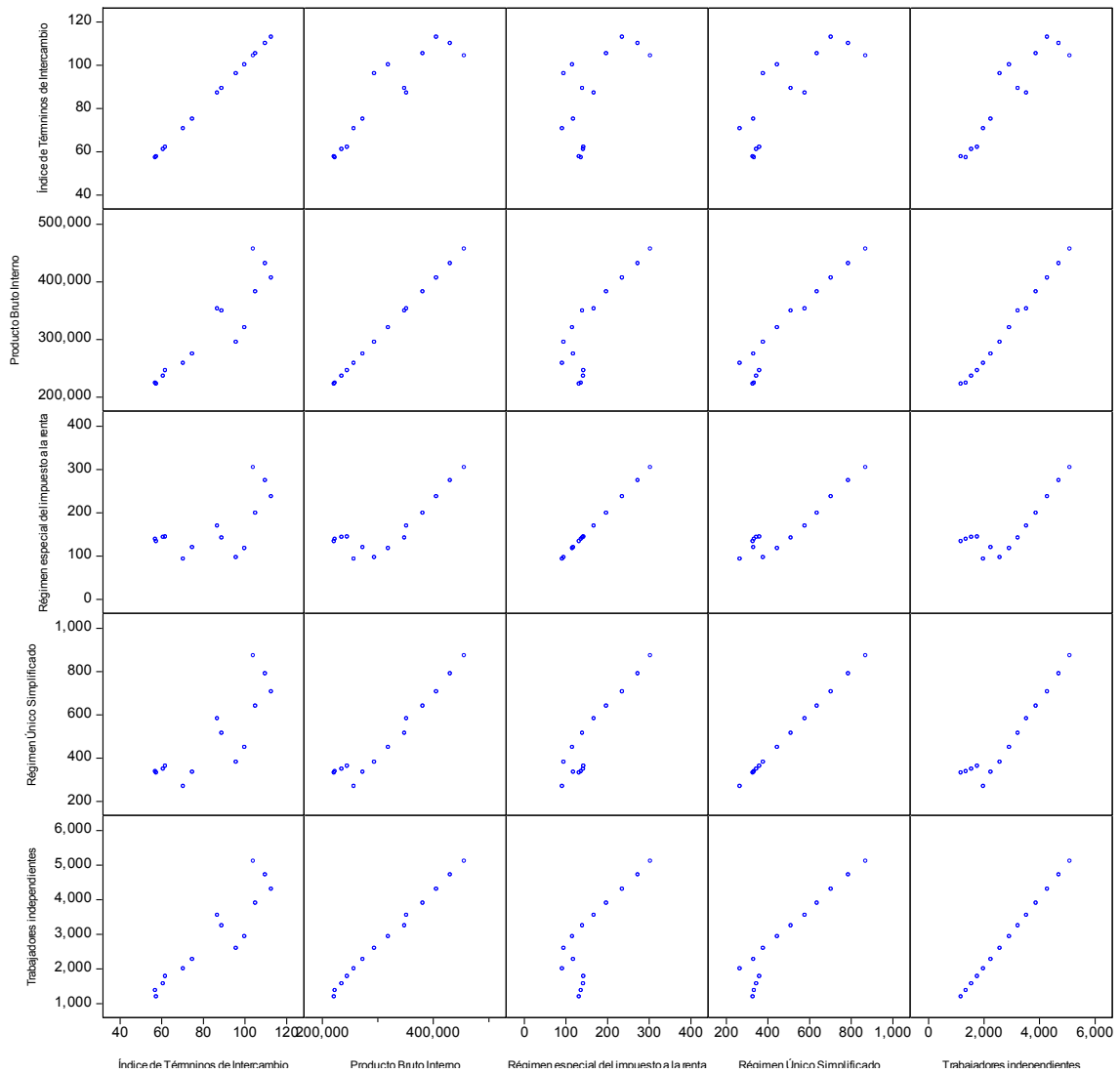
**Fuente:** BCRP - Sunat - MTPE - INEI

**Elaboración:** propia

Para el caso de estudio, los gráficos muestran una clara relación positiva lineal entre el índice de términos de intercambio (TOT) con las variables independientes producto bruto interno (PBI) y cantidad de trabajadores independientes (INDEP). En el caso de las otras variables explicativas no se puede determinar la existencia de alguna relación de forma gráfica.

Para obtener una visión general de la relación entre el conjunto de variables consideradas, se procede a una comparación variable a variable dentro de la matriz de dispersión.

**Gráfico 14 Matriz de dispersión**



**Fuente:** BCRP - Sunat - MTPE - INEI

**Elaboración:** propia

La matriz de correlaciones calcula el nivel de dependencia entre variables mostrado en la gráfica N°14. De donde se concluye que las variables INDEP y PBI son las que presentan mayor correlación lineal, además de la variable RUS con un 81.4% de correlación siendo este considerable.



**Tabla 10 Correlación de variables**

	TOT	PBI	RER	RUS	INDEP
TOT	1.000000	0.913837	0.611806	0.814529	0.917035
PBI	0.913837	1.000000	0.835964	0.963911	0.998117
RER	0.611806	0.835964	1.000000	0.939814	0.826901
RUS	0.814529	0.963911	0.939814	1.000000	0.956853
INDEP	0.917035	0.998117	0.826901	0.956853	1.000000

Fuente: BCRP - Sunat - MTPE - INEI

Elaboración: propia

- Estimación por mínimos cuadrados ordinarios

Se procede a realizar la primera regresión lineal en base a la ecuación:

$$TOT = \beta_1 * PBI + \beta_2 * RER + \beta_3 * RUS + \beta_4 * INDEP + c$$

La tabla N°3 muestra los coeficientes obtenidos para cada variable y su nivel de significancia con lo cual concluimos que no son significativas para el modelo ya que superan la probabilidad del 5%. Según el análisis previo priorizamos las variables relacionadas a las MYPE.

**Tabla 11 Coeficientes de primera regresión**

Variable	Coeficiente	Estadístico-t	Prob.
C	20.19312	0.285739	0.7815
PBI	0.000205	0.392634	0.7037
RER	-0.183158	-1.264638	0.2378
RUS	0.020549	0.198531	0.8470
INDEP	0.006761	0.243720	0.8129

Fuente: BCRP - Sunat - MTPE - INEI

Elaboración: propia

Se procede a realizar la segunda regresión lineal en base a la ecuación:

$$TOT = \beta_1 * RER + \beta_3 * RUS + c$$

**Tabla 12 Coeficientes de la segunda regresión**

Variable	Coeficiente	Estadístico-t	Prob.
C	45.66624	7.318842	0.0000
RER	-0.415189	-4.072577	0.0018
RUS	0.217923	6.347213	0.0001

Fuente: BCRP - Sunat - MTPE - INEI

Elaboración: propia

Se observa que todas las variables son significativas para un 95% nivel de confianza. Donde el estadístico R2 es de 86.58%, lo cual implica un grado importante de ajuste del modelo lineal. La variable RER posee una dependencia lineal negativa por ende se tendría una disminución de 0.415 unidades en el índice de términos de intercambio cuando se incremente en una unidad la cantidad de empresas en el régimen RER y en el caso del RUS ocurriría lo contrario con un incremento de 0.217 unidades en el índice TOT por cada empresa nueva registrada en el RUS.

### Análisis de resultados y conclusiones

De esta forma se encuentra que existe una relación mayor para el caso de las empresas del régimen MYPE que se encuentran registradas en el RUS en contraparte con las empresas del Régimen Especial a la Renta (RER), esto debido a que en el último caso ya se encuentra definidas las actividades económicas que se encuentran afectas a dicho régimen donde la mayoría se encuentra enfocada al mercado nacional sin ningún intercambio internacional ya sea exportación e importación. Mientras que el Régimen Único Simplificado permite registrar a micro y pequeña empresas que puedan participar de diversas transacciones ya sean nacionales o internacionales incrementando el índice de precios de exportación incrementando de forma directa al índice de términos de intercambio.

Empero tenemos una situación no evaluada al momento de realizar el análisis macro econométrico, y es el impacto de los choques externos ocasionados por crisis económicas internacionales o caída en los precios de los metales, los cuales pueden generar un descenso importante en los términos de intercambio (TOT) no contemplado en el modelo desarrollado. Por lo tanto se recomienda considerar los factores externos al momento de realizar una futura

regresión con el objetivo de conocer en qué grado modificaría la relación entre las variables explicativas y explicada el ingreso de una variable no controlada por los cambios en la economía nacional.

## 4.2. Análisis micro econométrico

El problema existente dentro del funcionamiento de los mercados emergentes, es la existencia de una gran cantidad de micro y pequeñas empresas que son capaces de manejar grandes actividades productivas o propiciar un ambiente de crecimiento del PBI.

En el Perú el 80% de las MYPES no dura más de 3 años, es decir para el año 2012 de 100 empresas creadas, conformadas o constituidas en el año 2010, solamente 20 de estas podrán sobrellevar adelante su gestión y el resto decidirá dejar de funcionar. Una característica importante de las MYPES es que en su mayoría son negocios familiares, es decir, los empleados y empleadores poseen lasos de parentela haciendo de las decisiones de la empresa, decisiones emocionales que involucran otros problemas personales.

Por lo tanto, identificamos el problema del corto tiempo de operación de muchas MYPES, relacionado con diversos factores estructurales de este tipo de negocios.

### 4.2.6. Descripción de las variables del modelo.

#### **Variable endógena**

- **Operación (oper)**

Años de operación de la empresa, se calcula restando el año de inicio de operaciones menos el años de estudio obteniendo de esta forma la cantidad de periodos anuales en funcionamiento.

$oper = 2013 - (\text{Año de inicio de operaciones})$

Unidades: Años

Periodo: Anual

Fuente: EMYPE 2013

## **Variables exógenas**

- Edad (edad)

Nivel en años del representante o encargado del manejo del negocio.

Unidades: Años

Periodo: Anual

Fuente: EMYPE 2013

- Nivel educativo (edu)

Variable cualitativa que especifica el nivel de instrucción educativa de la persona encargada del negocio. Se considera desde primaria, secundaria hasta superior.

Elaboramos variables dummy que nos permitan clasificar la formación educativa como una variable dicotómica:

educ1: representa formación primaria

educ2: representa formación secundaria

educ3: representa formación superior

Unidades: Números

Valores: 0,1

Fuente: EMYPE 2013

- Organización (org)

Tipo de asociación jurídica o de persona natural, en este campo se contempla las diversas sociedades anónimas además de las empresas con responsabilidad limitada.

Elaboramos una variable dicotómica que permite conocer qué porcentaje se inscribió en registros públicos como persona jurídica, entonces:

Persona natural: 0

Persona jurídica: 1

Unidades: Números

Valores: 0,1

Fuente: EMYPE 2013

- Licencia de funcionamiento (licen)

La licencia municipal para ejercer alguna actividad económica dentro de un establecimiento, es una variable importante que nos ayuda a medir el grado de fiscalización y cercanía de los sectores públicos. Además de indicar el nivel de burocracia dentro de las entidades municipales.

No cuenta con licencia de funcionamiento: 0

Si cuenta con licencia de funcionamiento: 1

Unidades: Números

Valores: 0,1

Fuente: EMYPE 2013

- Local (propio)

Régimen al cual está sometido el representante del negocio, de acuerdo al establecimiento donde labora y realiza las actividades comerciales. Esta variable presenta dos categorías donde se especifica si la propiedad es alquilada o propia.

Propiedad alquilada: 0

Local propio: 1

Unidades: Números

Valor: 0,1

Fuente: EMYPE 2013

## 4.2.2. Análisis estadístico de las variables.

- **Operación**

El número de años de operación muestra una distribución que con un sesgo hacia la derecha por el coeficiente de asimetría. Además según Kurtosis sería una distribución leptocúrtica, con los datos más concentrados cerca del valor modal. Además muestra que existen empresas con una antigüedad en sus operaciones de 68 años y empresas con un año de funcionamiento.

**Tabla 13 Estadísticas de la variable operación**

oper	
Media	10.52593
Desviación estándar	7.41736
Varianza	55.01724
Asimetría	1.89567
Kurtosis	8.63772
Mínimo	1
Máximo	68
Observaciones	2835

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

- **Edad**

La edad de la persona encargada posee una distribución con una media de 45 años, es decir cumple con el perfil de las micro y pequeñas empresas conformadas por ex trabajadores dependientes que deciden convertirse en independientes. El coeficiente de asimetría muestra que no existe sesgo por ser positivo y cercano a cero, esto debido a la cantidad de valores distribuidos alrededor del promedio que indican una alta propensión por conformar MYPES en edad adulta avanzada y jóvenes.

El coeficiente de Kurtosis muestra que se trataría de una distribución mesocúrtica, similar a la distribución normal con un coeficiente cercano a tres.

**Tabla 14 Estadísticas de la variable edad**

edad	
Media	45.43112
Desviación estándar	10.99269
Varianza	120.8392
Asimetría	0.2692479
Kurtosis	2.762237
Mínimo	20
Máximo	83
Observaciones	2352

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

- **Nivel de educación**

El nivel de educación en promedio para los encargados de las MYPES es de un nivel superior, según los coeficientes se concluye que hay más valores en el nivel superior que en el nivel educativo primaria. Esta característica puede ser un factor decisivo al momento de tomar decisiones empresariales dentro de estas organizaciones.

**Tabla 15 Estadísticas de la variable nivel de educación**

educ	
Media	2.617935
Desviación estándar	0.5727378
Varianza	0.3280286
Asimetría	-1.202962
Kurtosis	3.450666
Mínimo	1
Máximo	3
Observaciones	2353

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

- **Organización jurídica de la empresa**

Las empresas pueden optar por distintas opciones a la hora de conformar una sociedad, entonces según los datos en promedio las personas o grupo de socios conformas sociedades anónimas, es decir una personería jurídica, esto es un claro ejemplo de las empresas familiares que constituyen sociedades anónimas cerradas con el objetivo proteger su patrimonio y no negociar fusiones o adquisiciones.

**Tabla 16 Estadísticas de la variable organización jurídica**

org	
Media	0.4927721
Desviación estándar	0.5000541
Varianza	0.2500541
Asimetría	0.0289146
Kurtosis	1.000836
Mínimo	0
Máximo	1
Observaciones	2352

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

- **Licencia municipal**

Las personas prefieren en mayor medida la licencia municipal ya que les permite ahorrar costos al momento de ejercer la actividad comercial. En este caso la encuesta muestra que menos del 25% de encuestados no cuenta con una licencia que les permite ejercer su actividad económica.

**Tabla 17 Estadísticas de la variable licencia municipal**

licen	
Media	0.7019558
Desviación estándar	0.4574963
Varianza	0.2093020
Asimetría	0.0289146
Kurtosis	-0.8830617
Mínimo	0
Máximo	1
Observaciones	2352

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

- **Régimen de propiedad del local**

La mayoría de locales donde se realizan diversas actividades económicas constan de alquiler, es decir se agencian un establecimiento que no les pertenece y realizan un contrato de locación para tener los beneficios de un espacio cerrado. Pero además puede ser un factor crucial en el desarrollo de la micro y pequeña empresa ya que si se mantienen en



establecimientos de alquiler podría estar incurriendo en costos muy altos que afecten el desempeño de su labor productiva.

**Tabla 18 Estadísticas de la variable tipo de local**

local	
Media	0.4247449
Desviación estándar	0.4944093
Varianza	0.2444406
Asimetría	0.304489
Kurtosis	1.092714
Mínimo	0
Máximo	1
Observaciones	2352

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

### 4.2.3. Estimación inicial y pruebas de ajuste

- **Regresión de Poisson**

Realizamos la estimación del modelo considerando la variable endógena años de operación y las variables exógenas edad, organización, licencia municipal, tipo de local, educación primaria y educación superior. Obteniendo los siguientes resultados:

La primera regresión muestra que nuestras variables son significativas para nuestro modelo de conteo, Poisson. Tenemos un Pseudo R<sup>2</sup> de 0.0892 que se encuentra entre 0 y 1, dentro de los parámetros tenemos relaciones directas con la edad y el nivel de educación tanto primaria como superior. Mientras que existe una relación inversa con el tipo de organización que se constituye.

En las bondades de ajuste observamos una medida basada en el Log - Likelihood con los parámetros diferentes de cero es mayor al estimado con el intercepto. La hipótesis nula de que todos los coeficientes son cero menos el intercepto se rechaza ya que LR(6)= 1800.296 por lo tanto no son ceros.

**Tabla 19 Regresión de Poisson**

oper	Coeficiente	z	P> z
edad	0.0186796	32.22	0.0000
orga	-0.1246155	-9.28	0.0000
licen	0.2500621	16.44	0.0000
local	0.1195457	9.31	0.0000
educ1	0.1124202	3.84	0.0000
educ3	0.1436036	9.57	0.0000
_cons	1.235242	39.60	0.0000

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

#### 4.2.4. Test de sobredispersión

El modelo de regresión Binomial Negativa es adecuado cuando los datos cumplen con el modelo de Poisson pero presentan sobredispersión. Entonces al realizar la estimación *nbreg*, obtenemos un parámetro alfa (*alpha*) que representa la medida de la dispersión, entonces si el parámetro es estadísticamente igual a cero, el modelo se reduce al modelo de Poisson, pero si el parámetro alfa es mayor a cero revelaría la existencia de sobredispersión. En el caso del *p-value*, como es menor a 0.001 se rechaza el modelo de Poisson en favor del modelo Binomial Negativo.

Concluimos por lo tanto que tenemos que realizar la regresión mediante el modelo Binomial Negativo.

#### 4.2.5. Estimación final

Realizamos la estimación del modelo considerando nuestra variable endógena años de operación y las variables exógenas edad, organización, licencia municipal, tipo de local, educación primaria y educación superior. Obteniendo los siguientes resultados:

**Tabla 20 Regresión binomial negativa**

oper	Coficiente	z	P> z
edad	0.0179657	15.99	0.0000
orga	-0.1240375	-4.67	0.0000
licen	0.2461704	8.64	0.0000
local	0.1243858	4.98	0.0000
educ1	0.1082978	1.78	0.0750
educ3	0.1511377	5.20	0.0000
_cons	1.264121	21.62	0.0000
lnalfa	-1.385122		
alfa	0.2502932		

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

Los resultados muestran que nuestras variables son significativas excepto por la variable educ1, que representa si las personas poseen educación primaria. Por lo tanto tenemos relación directa con los años de edad del representante o encargado de la MYPE, también con la condición de la licencia de funcionamiento, si posee un local alquilado o propio puede influir en los años de operación de manera directa.

Existe un pseudo R2 con valor igual a 0.0281.

En las bondades de ajuste observamos una medida basada en el Log - Likelihood con los parámetros diferentes de cero y mayores al estimado con el intercepto. La hipótesis nula de que todos los coeficientes son cero menos el intercepto se rechaza ya que LR (6)= 426.656 por lo tanto no son ceros. Además el criterio de la información de AIC = 6.272 es menor que en el modelo de regresión de Poisson, confirmado la elección del modelo Binomial Negativo.

#### 4.2.6. Test de variables omitidas

Al realizar el test de Ramsey para variables omitidas es necesario volver a realizar una regresión de tipo lineal con el objetivo de obtener un estadístico que pueda indicar si se está dejando de lado algún término crucial. Este test nos permitirá conocer desde otra perspectiva

distinta a la metodología desarrollada si el modelo contiene las variables independientes necesarias.

**Tabla 21 Regresión lineal**

oper	Coficiente	z	P> z
edad	0.2095922	15.17	0.0000
orga	-1.424461	-4.51	0.0000
licen	2.555779	7.65	0.0000
local	1.321376	4.39	0.0000
educ1	1.554296	2.08	0.0380
educ3	1.526156	4.44	0.0000
_cons	1.409281	-2.02	0.0440

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

Luego de realizar la regresión de tipo lineal para variables categóricas y dicotómicas se obtienen los siguientes coeficientes significativos para el modelo.

**Tabla 22 Resultado del test de variables omitidas**

Ho: El modelo no posee variables omitidas	
F(3, 2342)	12.74
Prob > F	0.0000

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

Finalmente se realiza la medición del test de Ramsey para variables omitidas donde indica que en este caso el modelo no contaría con datos relevantes que hayan sido retirados de la formulación. Sin embargo el enfoque metodológico de modelos de regresión para datos de conteo adoptado en la investigación permite confirmar este resultado a pesar de no ser un test principal.

### 4.3. Predicción de escenarios

Los escenarios enseñados consisten en dos casos de comparación elementales:

Primer caso, cuando no se cuenta con licencia, el local donde funciona la empresa es alquilado y el tipo de organización es una personería jurídica es decir puede ser una sociedad anónima.

Segundo caso, cuando si cuenta con licencia de funcionamiento otorgada por parte del municipio, el local es propio y la organización es una persona natural la cual consiste en ser un trabajador independiente que emite recibo por honorarios o una persona con negocio.

Por lo tanto para ambos casos se procederá a comparar un horizonte de vida de 9 años de operación bajo los dos grupos de condiciones por separado y luego se obtendrá una visión probabilística sobre que alternativa es la mejor.

**Tabla 23 Predicción de escenarios para el tiempo de operación**

Probabilidad de operar t-años dadas las condiciones	Caso 1: Sin licencia, local alquilado, persona jurídica.	Caso 2: Con licencia, local propio, persona natural
P(t=0/x)	0.0130	0.0032
P(t=1/x)	0.0344	0.0096
P(t=2/x)	0.0569	0.0184
P(t=3/x)	0.0754	0.0280
P(t=4/x)	0.0874	0.0374
P(t=5/x)	0.0926	0.0456
P(t=6/x)	0.0921	0.0522

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

Según los resultados obtenidos podemos observar que la probabilidad de que una micro o pequeña empresa dure 5 años es de 9% para el primer caso mientras que en el segundo caso simulado la probabilidad es de la mitad con 4.56%. Todos los escenarios se realizaron bajo un nivel de confianza del 95%.

# CAPÍTULO V: ANÁLISIS DE RESULTADOS

## 5.1. Factores socioeconómicos analizados y su nivel de correlación.

La regresión binomial negativa nos permite conocer el nivel y tipo de correlación existente entre las variables independientes presentadas al inicio de la investigación y la variable dependiente. Por lo tanto se presenta un cuadro resumen con información relevante de los resultados obtenidos para luego ser discutidos en el siguiente capítulo de esta investigación.

**Tabla 24 Cuadro resumen regresión binomial**

<b>Factores socioeconómicos</b>	<b>Nivel de correlación</b>	<b>Tipo de correlación</b>	<b>Hipótesis inicial</b>
<b>Edad del responsable de la empresa MYPE.</b>	<b>0.179</b>	<b>Positiva</b>	Correlación positiva de la edad con el tiempo de operación de una MYPE
<b>Tipo de organización, si es una sociedad anónima cerrada o cualquier otra personería jurídica.</b>	<b>0.124</b>	<b>Negativa</b>	Correlación positiva de registrarse como una persona jurídica en el régimen MYPE con el tiempo de operación o funcionamiento.
<b>Licencia de funcionamiento del establecimiento empleado para las actividades empresariales.</b>	<b>0.246</b>	<b>Positiva</b>	Correlación positiva de no contar con una licencia de funcionamiento para la MYPE con el tiempo de

			operación.
<b>Local empleado para el desarrollo de sus actividades es alquilado.</b>	<b>0.124</b>	<b>Positiva</b>	Correlación positiva de alquilar el establecimiento de funcionamiento de la MYPE con el tiempo de operación.
<b>Nivel educativo primario completo o incompleto del responsable de administrar la MYPE.</b>	<b>0.108</b>	<b>Positiva</b>	Correlación positiva de poseer una formación educativa menor o igual a los 6 años con el tiempo de operación de la MYPE.
<b>Nivel educativo superior (técnico o universitario) completo o incompleto del responsable de administrar la MYPE.</b>	<b>0.151</b>	<b>Positiva</b>	Correlación positiva de haber cursado una formación educativa menor a los 16 años con el tiempo de operación de la MYPE.

Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

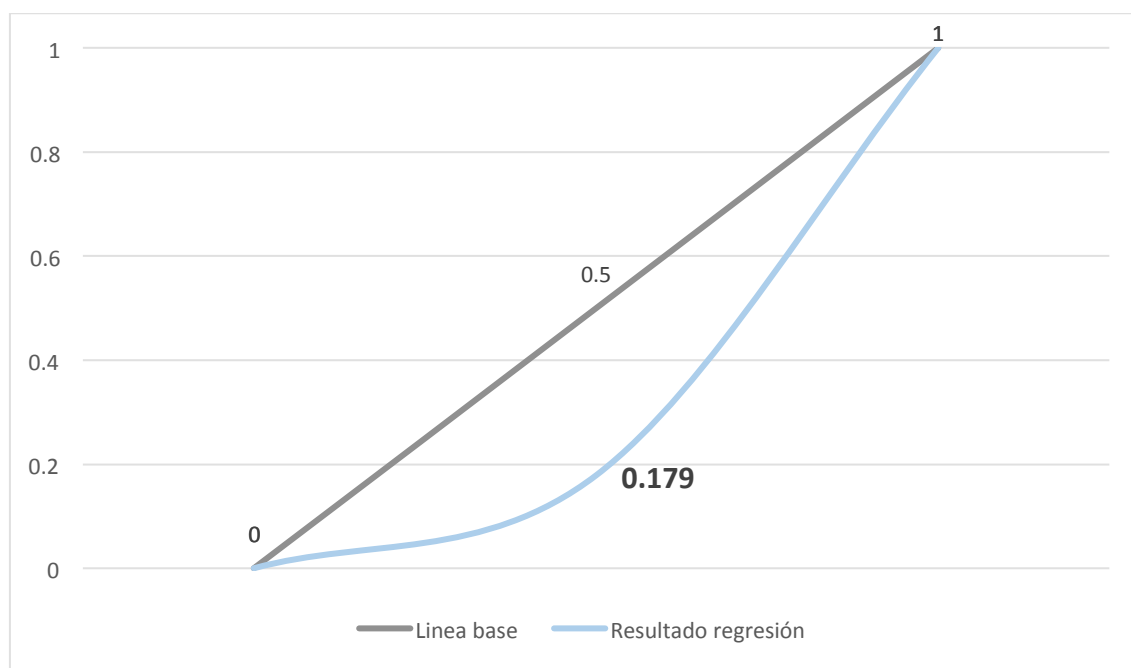


## 5.2. Gráficos de los resultados obtenidos en la regresión.

Las siguientes representaciones muestran el impacto cuantitativo del incremento en una unidad de cualquiera de las variables analizadas sobre los años de operación que pueda tener una empresa en el régimen MYPE.

La primera variable a estudiar es la edad del responsable sobre la administración de la empresa, la cual se encuentra en promedio dentro de los 37 años. La línea base muestra que para un incremento en una unidad de la edad promedio de los encargados beneficiará de forma positiva en 6 meses más de operación de la empresa. En el caso del resultado de la regresión, se tiene que para nuestro caso de estudio el impacto sería de 2.148 meses por cada año adicional que posea el responsable.

**Gráfico 15 Edad del responsable de la empresa MYPE**



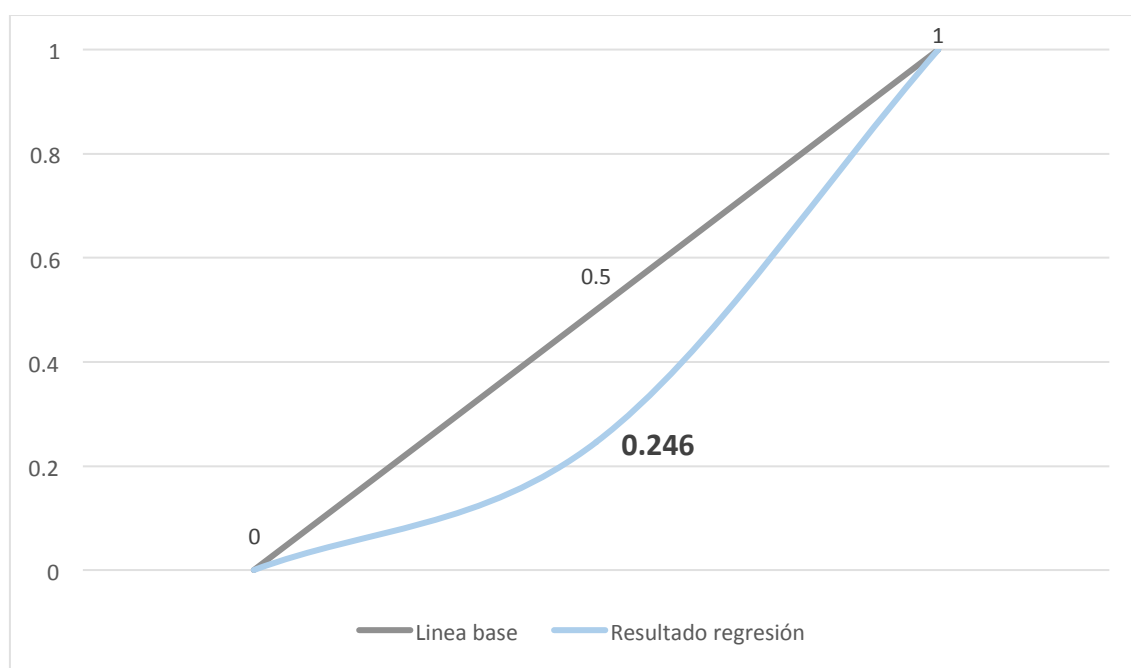
Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

Continuando con el análisis de la siguiente variable explicativa, se tiene el tipo de organización con el cual se crea una empresa, para este estudio se dividieron en dos categorías, personas jurídicas y naturales, siendo la primera considerada para la regresión. Por lo tanto el resultado muestra que al registrarse como una sociedad anónima cerrada se tiene un impacto negativo reduciendo el tiempo de operación de una MYPE en 1.488 meses.

La tercera variable de análisis es la licencia de funcionamiento, donde se empleó para la regresión el caso donde no contaban con una licencia para operar dentro de su localidad. El resultado de la regresión evidencia que para el hecho de operar dentro de la ilegalidad beneficia al negocio incrementando el tiempo de funcionamiento en 2.952 meses respecto de las MYPE que cuentan con los permisos correspondientes.

**Gráfico 16 Sin Licencia de funcionamiento**

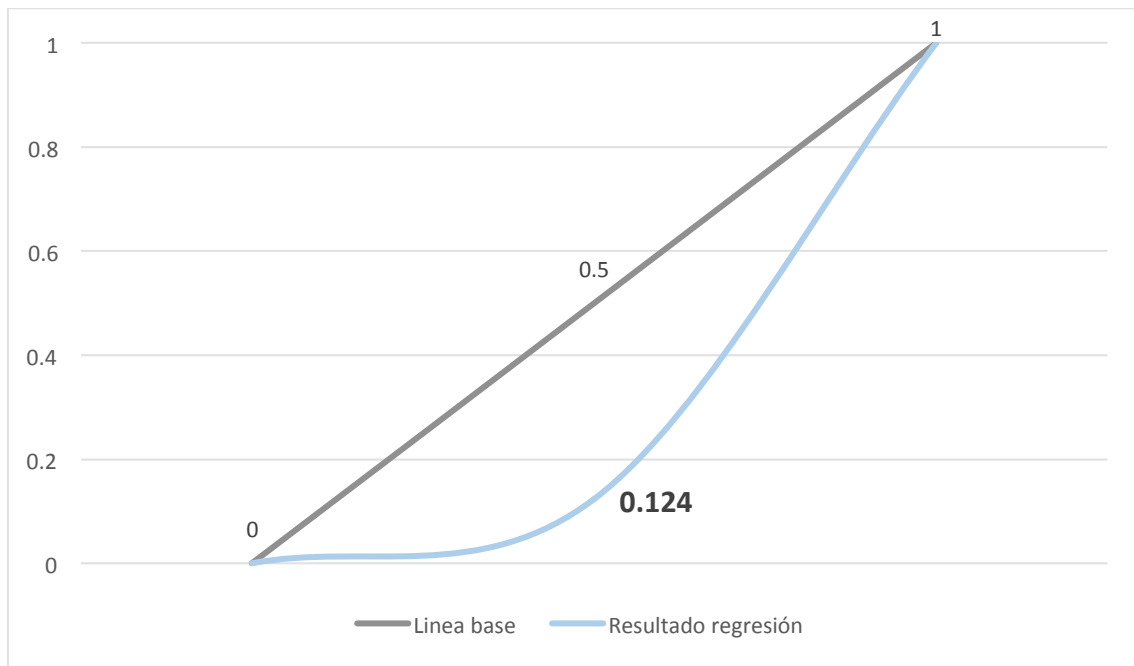


Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

Los activos inmuebles utilizados por los empresarios MYPE para desarrollar sus negocios pueden ser propios o alquilados dependiendo del rubro y la inversión necesaria. El estudio señala que en el caso de que una empresa de este régimen opte por establecer sus operaciones dentro de un local alquilado le permite generar un impacto positivo incrementando en 1.488 meses su tiempo de vida.

**Gráfico 17 Local alquilado**



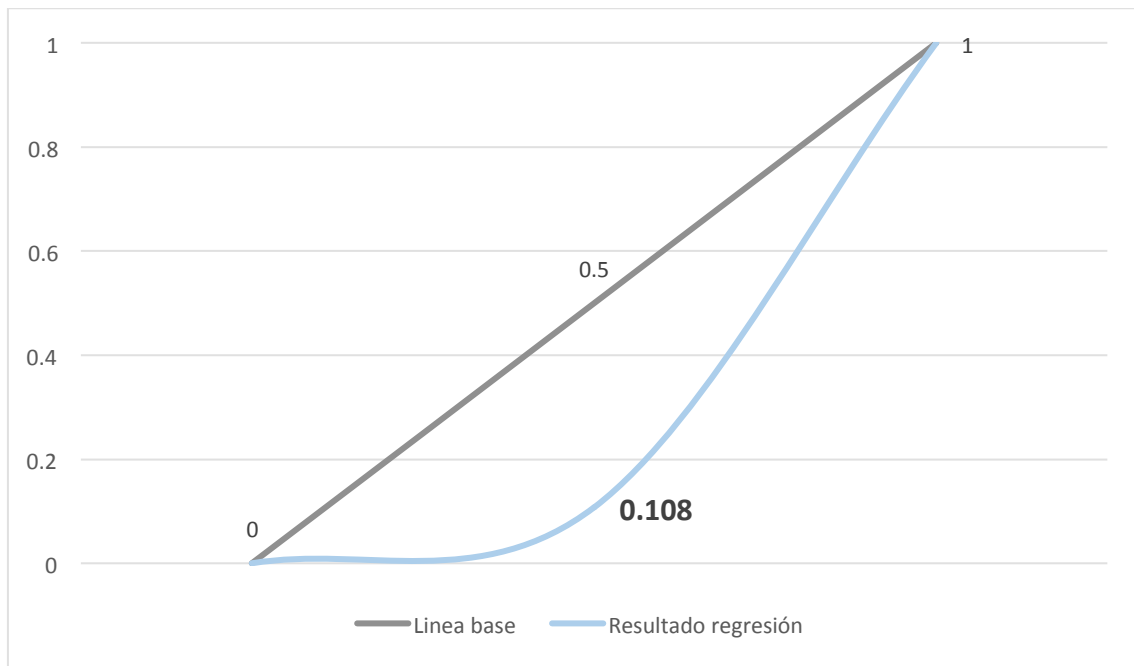
Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

La variable educación se utilizó dentro del análisis de dos casos relevantes, cuando se tenga un nivel educativo de primaria y en el caso de especializaciones a nivel superior universitario o técnico. Se consideran solo estos dos casos de tres debido al gran porcentaje que representa para la muestra seleccionada.

En el primer caso, se tiene un impacto positivo de 1.296 meses en el tiempo de operación de la MYPE.

**Gráfico 18 Nivel educativo primario (Menor o igual a 6 años)**

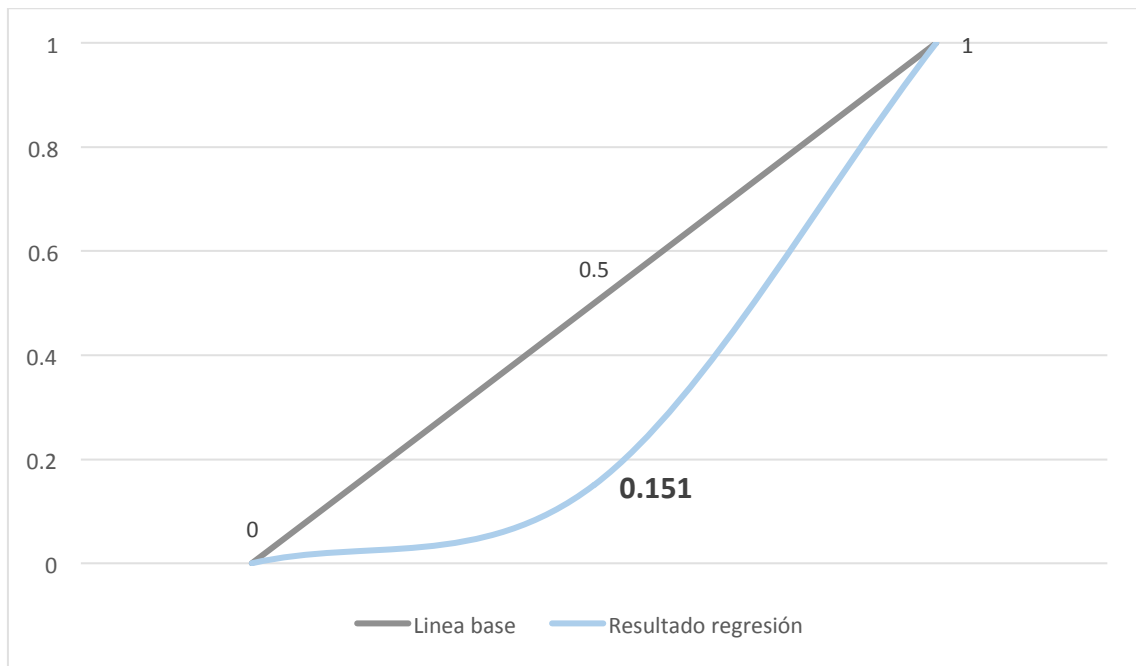


**Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013**

**Elaboración: propia**

El segundo caso, muestra un impacto positivo mayor, ya que si el responsable de la empresa cuenta con estudios superiores completos o incompletos puede aportar al periodo de vida 1.812 meses. Los cuales pueden permitir realizar cambios, mejoras u obtener más ingresos que permitan seguir operando.

**Gráfico 19 Nivel educativo superior (Menor o igual a 16 años)**



Fuente: Encuesta de la Micro y Pequeña Empresa – EMYPE 2013

Elaboración: propia

# CAPÍTULO VI: DISCUSIÓN

## 6.1. Discusiones

Los resultados de la tesis elaborada llevan a la siguiente discusión, dentro de la fase de análisis micro econométrico se encontró una relación inversamente proporcional sobre el tipo de organización de la empresa MYPE y el tiempo de operación que dura. Lo cual podría afectar el desarrollo de un mayor nivel de informalidad al tener un impacto inverso. Mientras que en la fase de predicción de escenarios se discutieron los dos casos simulados para un horizonte de vida de 5 años donde se tuvo que los negocios sin licencia y funcionando en locales alquilados tendrían mayores probabilidades de superar esa brecha que los negocios con licencias de funcionamiento y habiendo adquirido el local donde desarrollan sus actividades empresariales.

## 6.2. Posibilidades y limitaciones del estudio

### 6.2.1. Posibilidades

El desarrollo del presente trabajo de investigación constituye el primer análisis de los factores que determinan el tiempo de operación de las MYPE bajo un enfoque de regresiones micro econométricas utilizando modelos para datos de conteo. Ampliando el rango de posibilidades para incluir otras variables explicativas que puedan mostrar de forma más exacta las causas del horizonte de vida tan corto para este sector.

En general los estudios sobre este tipo de casos, se enfocan en el análisis cualitativo de factores internos mediante la realización de encuestas a un pequeño grupo de empresas seleccionadas, por lo que el hecho de usar técnicas econométricas nos brinda una muestra mucho más amplia acercándonos de forma cuasi precisa al verdadero comportamiento de las MYPE.

### 6.2.2. Limitaciones

Luego de realizar cada procedimiento dentro de las fases establecidas se puede decir que los resultados podrían haber reflejado mejor la realidad si se hubieran contado con datos actualizados al presente año o con 2 años de antigüedad como máximo, sin embargo como

se observa actualmente la realidad hace 4 años era otra distinta con niveles altos de crecimiento del PBI que podían motivar el ingreso de mayores empresas a los diferentes sectores económicos.

La encuesta en la cual se basa este estudio para desarrollar los cálculos estadísticos contiene información relacionada en su mayoría con factores tecnológicos que no son relevantes para nuestros objetivos específicos.

El tamaño de la muestra es ínfimo comparado con la cantidad real de empresas en el régimen MYPE según Produce las cuales superan el millón, por lo tanto una muestra más cercana a la población permitirá obtener resultados más precisos de su comportamiento bajo diversas condiciones de estudio.

### 6.3. Conclusiones

- El análisis dentro de la fase de visión macro económica del problema de investigación sugiere que, para el primer análisis realizado frente a los niveles de Producto Bruto Interno (PBI) contra la cantidad de empresas en el régimen MYPE, al observar un panorama de crecimiento continuo durante al menos 10 años las empresas tienden a experimentar una sensación de positivismo respecto a los negocios. Siendo así la motivación principal para los empresarios del sector MYPE la creación de empresas que se articulen dentro del proceso de producción de bienes y servicios de las grandes empresas. Generando así un incremento continuo durante los años de la cantidad de empresas MYPE que se registran evidenciado por el registro de Produce.
- El análisis macro económico del problema de investigación donde se relaciona además los términos de intercambio frente a la cantidad de empresas en el régimen MYPE muestra como conclusión una relación inversa con un sector en especial denominado Régimen Especial a la Renta (RER), el cual se encuentra restringido a actividades de comercio internacional. De manera que el incremento en las exportaciones reduciría la cantidad de empresas en el régimen RER, lo cual se demuestra por los datos históricos en la regresión realizada.



- Para realizar el modelo se presentó inicialmente un enfoque de regresión de Poisson basado en el análisis de datos de conteo. Sin embargo se concluye que, en base al test de sobredispersión para el caso particular de estudio, la regresión binomial negativa es la más adecuada para el estudio de tiempos de duración con unidades discretas donde se tengan como variables explicativas variables categóricas.
- Dentro de la fase de análisis micro econométrico se concluye que las principales variables que impactan el modelo de datos de conteo binomial negativo son la licencia de funcionamiento y el nivel de educación del encargado del negocio.
- Dentro de la fase de análisis micro econométrico se obtuvo una relación inversamente proporcional reflejada por la variable organización que se encuentra conformada por dos valores, persona natural y jurídica, por ende se concluye que al ser un negocio registrado como persona natural disminuirán las posibilidades de seguir operando conforme pasen los años.
- Dentro de la fase de predicción de escenarios donde se mostraron dos casos simulados, se concluye que el primer caso, el cual consiste en una empresa MYPE sin licencia de funcionamiento y con un local alquilado presenta mayores probabilidades de seguir operando a los 5 años de haber iniciado. Por lo tanto el modelo empleado refleja la realidad actual de muchos negocios registrados en Sunat y operando en diversos distritos de la capital los cuales se encuentran al borde de la informalidad. Debido a que incumplir con alguna de las medidas impuestas por los gobiernos locales corre el riesgo de ser sancionados deteniendo la producción diaria o mensual de sus bienes o servicios.
- Como conclusión final, se aborda el tema de los regímenes en los cuales las empresas pueden registrarse al momento de considerar desarrollar una idea. Sin embargo se observa en la muestra extraída de la encuesta EMYPE del INEI que muchos negocios optan por llevar sus operaciones al borde de las normas desde el punto de vista del investigador. Mientras que desde la visión de las instituciones públicas no se realizan procedimientos de verificación de los documentos necesarios y exigidos por cada uno, sino por el contrario se realizan procedimientos sancionadores que incrementan el costo

de operación de las MYPE dejando de lado el sector con mayor crecimiento anual y que brinda empleo directo e indirecto a más de un millón de personas.

#### 6.4. Recomendaciones

- Se recomienda a las instituciones públicas considerar modificar sus procesos al momento de fiscalizar los negocios locales, ya que según el estudio presentado se estaría reduciendo la probabilidad de existir por más de tres años superando la expectativa promedio de la región.
- Se recomienda a las personas que deseen abrir un negocio por primera vez iniciar como personas naturales o independientes y conseguir locales alquilados. Además de conseguir un nivel considerable de especialización educativa para el sector económico al cual deseen ingresar. Esto para que las probabilidades de funcionar por al menos tres años sean más favorables.
- Tomar en cuenta las variables, licencia de funcionamiento, tipo de organización y propiedad del local donde operan, ya que son variables con un alto índice de correlación.
- Se recomienda considerar otros factores internos asociados como el uso de la tecnología dentro de las MYPE para conocer el impacto de las herramientas informáticas dentro del tiempo de operaciones. En muchos casos de forma empírica se sabe que estas empresas no invierten en publicidad por internet o sitios web ya que lo perciben como un gasto innecesario.
- Dentro del análisis econométrico se recomienda considerar la variable ingresos percibidos por la empresa para conocer la capacidad de influencia que tienen sobre los años de funcionamiento.

## Bibliografía

Avolio, B., Mesones, A., & Roca, E. (2011). Factores que Limitan el Crecimiento de las Micro y Pequeñas Empresas en el Perú (MYPES). *Strategia* , 70-80.

Soriano, C. (26 de Noviembre de 2005). *www.rgpymes.com*. Retrieved 7 de Setiembre de 2016 from <http://www.prweb.com/releases/2005/11/prweb314871.htm>

Perren, L. (1999). Factors in the growth of micro-enterprises (Part 1): developing a frame work. *Journal of Small Business and Enterprise Development* , VI (4), 366-385.

Lira, P. (2009). *Finanzas y Financiamiento Las herramientas de gestión que toda pequeña empresa debe conocer* (Primera ed.). Lima: Nathan Associates Inc.

CEPAL. (2013). *Como mejorar la competitividad de las PYMES en la Unión Europea y América Latina y El Caribe*. Santiago de Chile: Naciones Unidas.

James, L. (1995). *Modelling Frequency and Count Data* (Primera ed.). Oxford: Clarendon Press.

Cameron, A. C., & Trivedi, P. K. (1998). *Regression Analysis of Count Data* (Primera ed.). Cambridge: Cambridge University Press.

Córdova, M. (2013). *Estadística Descriptiva e Inferencial* (Quinta ed.). Lima: Moshera S.R.L.

Winkelmann, R. (1995). Duration Dependence and Dispersion in Count-Data Models. *Journal of Business & Economic Statistics* , XIII (4), 467-474.

Gurmu, S., & Trivedi, P. (1996). Excess Zeros in Count Models for Recreational Trips. *Journal of Business & Economic Statistics* , XIV (14), 469-477.

Cameron, A., Trivedi, P., Milne, F., & Piggott, J. (1988). A Microeconomic Model of the Demand for Health Care and Health Insurance in Australia. *The Review of Economic Studies* , 55 (1), 85-106.

Dirección Nacional de Censos y Encuestas del Instituto Nacional de Estadística e Informática. (1 de 06 de 2011). *Publicaciones digitales INEI*. Recuperado el 05 de 08 de 2017, de [https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib0980/Libro.pdf](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib0980/Libro.pdf)

Pérez de Vargas Luque, A., & V, A. (1996). *Métodos multivariantes en bioestadística* (Primera ed.). Madrid: Centro de Estudios Ramón Areces.

Buse, A. (1982). The likelihood ratio, Wald and Lagrange multiplier tests: an expository note. *American Statistician* , I (36), 153-157.

Cameron, A., & Trivedi, P. (2001). Essentials of count data regression. En A. Cameron, & P. Trivedi (Edits.), *A companion to Theoretical Econometrics* (págs. 331-348). Cambridge: Badi H. Baltadi.

Cameron, A., & Windmeijer, F. (1996). R-squared measures for count data regression models with applications to health care utilization. *Journal of Business and Econometric Statistics* , XIV (2), 209-220.

Cameron, A., & F, W. (1997). An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics* , I (10), 329-342.

Charnes, A., Frome, E., & Yu, P. (1976). The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *Journal of the American Statistical Association* , LXXII (353), 214-222.

Gauss Cordeiro, M. (2000). *Modelos Lineales Generalizados* (Primera ed.). Lima: Instituto de Matemáticas y Ciencias Afines IMCA.

Frome, E., Kutner, R., & Beauchamp, D. (1973). Regression analysis of Poisson-Distributed Data. *Journal of the American Statistical Association* , LXVII (344), 935-940.

Instituto Nacional de Estadística e Informática. (2000). *INEI Perú*. Recuperado el 7 de Agosto de 2017, de <https://www.inei.gob.pe/>

Banco Central de Reserva del Perú. (1993). *Entidad autónoma que preserva la estabilidad monetaria del país, regula la moneda y el crédito, administra las RIN e informa sobre las finanzas nacionales*. . Recuperado el 7 de Agosto de 2017, de <http://www.bcrp.gob.pe/>

Ministerio de la Producción. (2000). *Produce*. Recuperado el 7 de Agosto de 2017, de <http://www.produce.gob.pe/>

# ANEXOS

- **Líneas de programación en Stata**

\*\*Modelos de conteo

\*=====

\*Análisis de la data

\*=====

\*usar data

use data, clear

describe

gen oper = 2013 - P2\_4

rename P2\_11\_1 lice

rename P2\_10 org

rename P2\_12 loca

recode local (3 4 = 2)

tab local

rename P2\_20A cargo

rename P2\_21 sexo

rename P2\_22 edad

rename P2\_23 edu

recode edu (1 2 3 4=1 "primaria") (5 6 =2 "secundaria") ///

(7 8 =3 "tecnic") (9 10 =4 "universidad"), gen(educ)

tab educ

rename P5\_1 tecgest

forvalue n=2/6 {

replace P10\_6\_1=`n' if (P10\_6\_`n'==1)

}

rename P10\_6\_1 finan

label define finan 1 "Banco" 2 "Caja Municipal" 3 "Caja Rural" ///

4 "EDPYME" 5 "ONG's" 6 "Otro", modify

tab finan

label list finan

rename P6\_1 tegtic

```

keep ncuest ccdd nombredd ccpp nombrepp ccdi nombredi org lice local ///
cargo sexo edad educ oper tecgest tegtic finan factor

recode educ (1=1 "Primaria") (2=2 "Secundaria") (3 4=3 "Superior") , gen (educa)

gen educ1=1 if educa==1
recode educ1 .=0
gen educ2=1 if educa==2
recode educ2 .=0
gen educ3=1 if educa==3
recode educ3 .=0

recode org (2 3 4 5 =1 "Persona jur'dica") (1=0 "Persona natural") , gen(orga)

recode lice (2=0 "no") (1=1 "si"), gen(licen)

recode loca (2=0 "Alquilado") (1=1 "Propio"), gen(local)

saveold data,replace

poisson oper, nolog

*=====
*Regresión modelo de Poisson
*=====

poisson oper edad orga licen local educ1 educ3, nolog

*Análisis de ratios odds
listcoef edad educ, help
listcoef edad educ, percent help

poisson oper edad orga licen local educ1 educ3, nolog
outreg using 07prmnbrm, replace
nbreg oper edad orga licen local educ1 educ3, nolog
outreg using 07prmnbrm, merge

```

```
quietly poisson oper edad orga licen local educ1 educ3
prvalue
quietly nbreg oper edad orga licen local educ1 educ3
prvalue
```

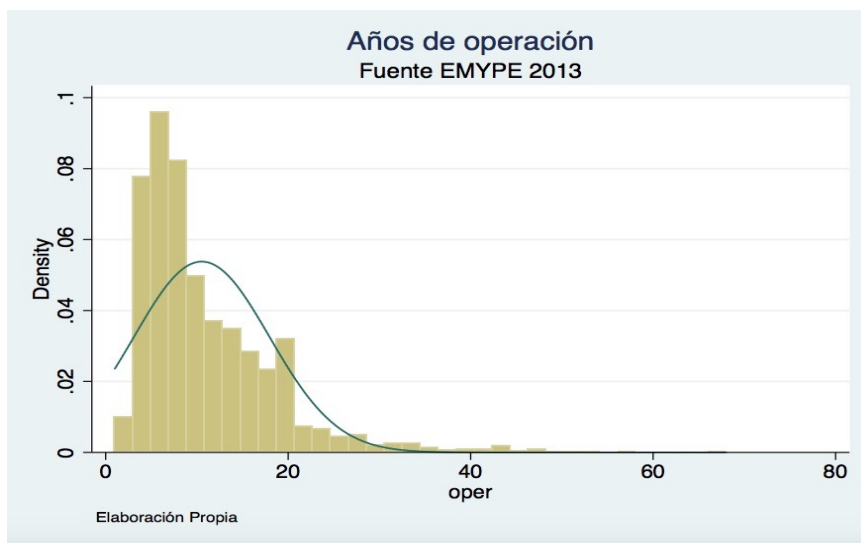
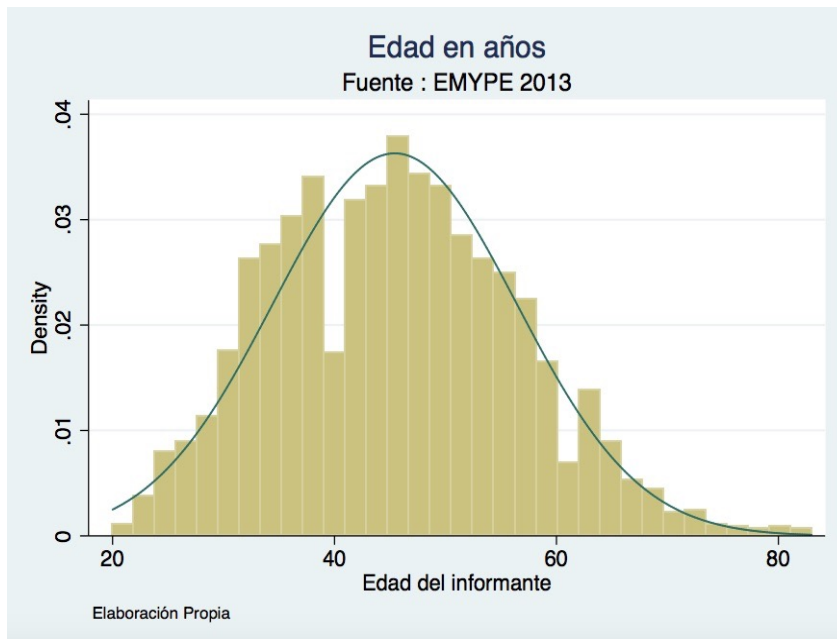
```
listcoef edad licen orga local educ1 educ3,help
prchange edad licen, rest(mean)
mfx compute
prvalue, x(licen=1 local=1 orga=0) rest(mean) brief
prvalue, x(licen=0 local=0 orga=1) rest(mean) brief
```

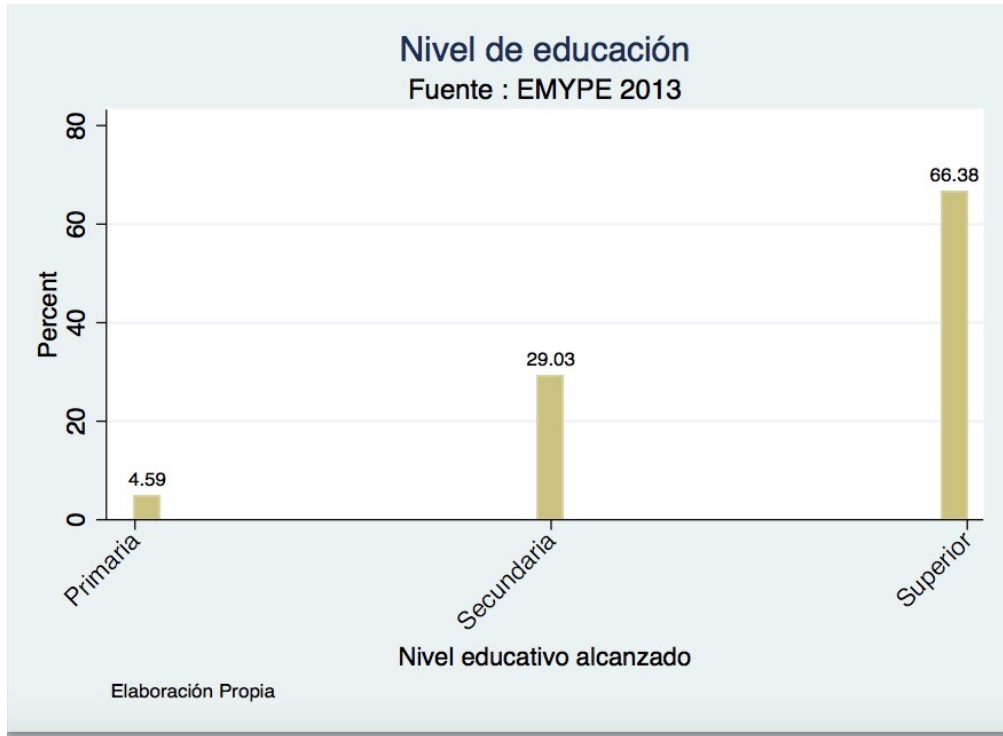
```
hist educa, percent addlab title("Nivel de educaci—n") ///
subtitle("Fuente : EMYPE 2013") note("Elaboraci—n Propia") ///
xlabel(1 "Primaria" 2 "Secundaria" 3 "Superior", angle(45))
```

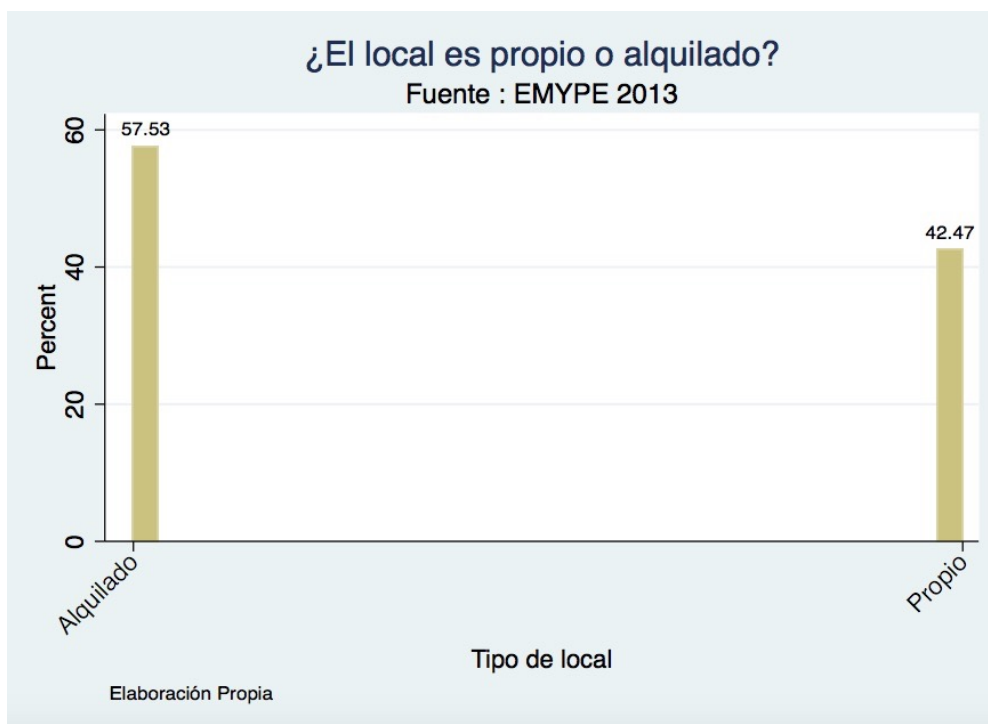
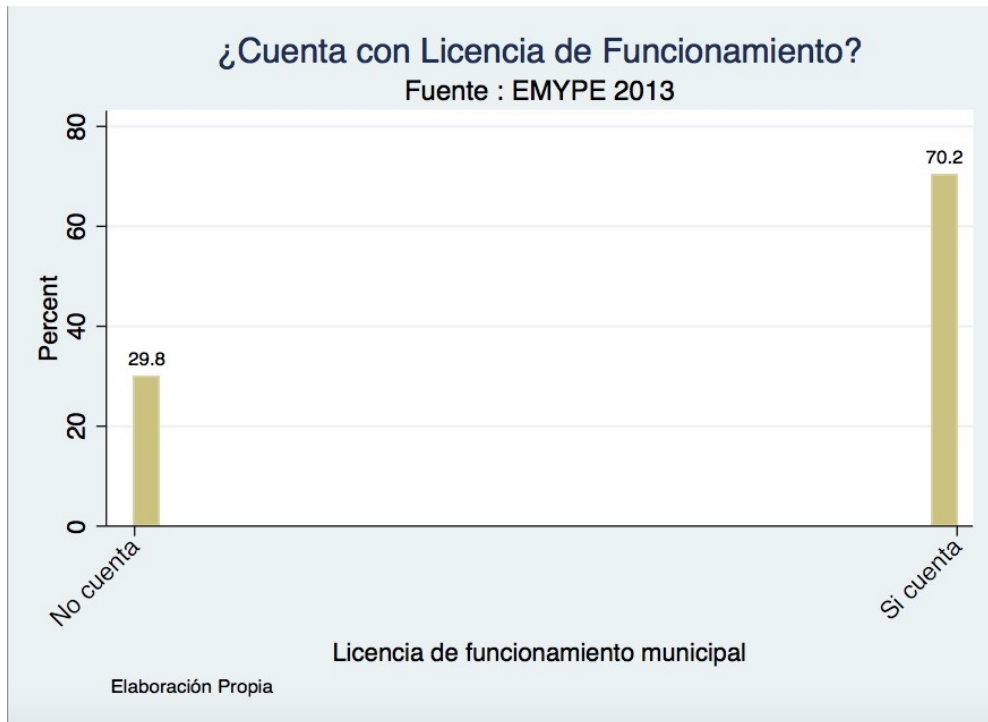
```
hist orga, percent addlab title("Tipo de organizaci—n") ///
subtitle("Fuente : EMYPE 2013") note("Elaboraci—n Propia") ///
xlabel(0 "Persona Natural" 1 "Persona Jur'dica", angle(45))
```



- Gráficos obtenidos para las variables del modelo micro econométrico.







**Gráficos y tablas obtenidos para las variables del modelo macro econométrico usando el software estadístico STATA.**

- **Primera regresión modelo macro econométrico PBI**

Variable Dependiente: PBI  
Método: Mínimos Cuadrados  
Muestra: 2000 2013  
Observaciones incluidas: 14

Variable	Coefficiente	Error Est.	Estadístico-t	Prob.
RER	-110.9895	80.56608	-1.377620	0.1984
RUS	108.9523	52.51014	2.074880	0.0648
INDEP	50.99514	4.823843	10.57148	0.0000
C	135165.5	3431.358	39.39127	0.0000
R-cuadrado	0.997619	Media var dependiente		317908.5
R-cuadrado ajustado	0.996905	D.E. var dependiente		79538.29
S.E. de la regresión	4424.803	Criterio información Akaike		19.86280
Sum residuos cuadrados	1.96E+08	Criterio Schwarz		20.04538
Log verosimilitud	-135.0396	Criterio Hannan-Quinn		19.84589
Estadístico-F	1396.856	Estad. Durbin-Watson		2.715435
Prob(estadístico-F)	0.000000			

- **Segunda regresión modelo macro econométrico PBI**

Variable dependiente: PBI  
Método: Mínimos Cuadrados  
Muestra: 2000 2013  
Observaciones incluidas: 14

Variable	Coefficiente	Error Est.	Estadístico-t	Prob.
RUS	43.17045	22.71943	1.900156	0.0839
INDEP	55.83855	3.434983	16.25585	0.0000
C	135341.7	3566.168	37.95159	0.0000
R-cuadrado	0.997168	Media var dependiente		317908.5
R-cuadrado Ajustado	0.996653	D.E. var dependiente		79538.29
S.E. de regresión	4601.841	Criterio información Akaike		19.89371
Sum residuos cuadrados	2.33E+08	Criterio Schwarz		20.03065
Log verosimilitud	-136.2560	Criterio Hannan-Quinn		19.88103
Estadístico-F	1936.292	Estad. Durbin-Watson		2.212836
Prob(Estadístico-F)	0.000000			

- **Primera regresión modelo macro econométrico TOT**

Variable Dependiente: TOT  
Método: Mínimos Cuadrados  
Muestra: 2000 2013  
Observaciones incluidas: 14

Variable	Coefficiente	Error Est.	Estadístico-t	Prob.
C	20.19312	70.66980	0.285739	0.7815
PBI	0.000205	0.000521	0.392634	0.7037
RER	-0.183158	0.144830	-1.264638	0.2378
RUS	0.020549	0.103505	0.198531	0.8470
INDEP	0.006761	0.027740	0.243720	0.8129
R-cuadrado	0.912687	Media var dependiente		84.78904
R-cuadrado Ajustado	0.873881	S.D. var dependiente		20.53411
S.E. de regresión	7.292339	Criterio información Akaike		7.083979
Sum residuos cuadrados	478.6039	Criterio Schwarz		7.312214
Log verosimilitud	-44.58785	Criterio Hannan-Quinn		7.062852
Estadístico-F	23.51922	Estadístico Durbin-Watson		1.525949
Prob(Estadístico-F)	0.000088			

- **Segunda regresión del modelo TOT**

Variable Dependiente: TOT  
Método: Mínimos Cuadrados  
Muestra: 2000 2013  
Observaciones incluidas: 14

Variable	Coefficiente	Error Est.	Estadístico-t	Prob.
C	45.66624	6.239545	7.318842	0.0000
RER	-0.415189	0.101947	-4.072577	0.0018
RUS	0.217923	0.034334	6.347213	0.0001
R-cuadrado	0.865802	Media var dependiente		84.78904
R-cuadrado ajustado	0.841402	S.D. var dependiente		20.53411
S.E. de regresión	8.177570	Criterio información Akaike		7.228077
Sum residuos cuadrados	735.5992	Criterio Schwarz		7.365017
Log verosimilitud	-47.59654	Criterio Hannan-Quinn		7.215400
Estadístico-F	35.48420	Estadístico Durbin-Watson		1.234458
Prob(Estadístico-F)	0.000016			

- Pantallas de tablas estadísticas de las variables

oper				
	Percentiles	Smallest		
1%	2	1		
5%	3	1		
10%	4	1	Obs	2835
25%	5	1	Sum of Wgt.	2835
50%	8		Mean	10.52593
		Largest	Std. Dev.	7.41736
75%	14	52		
90%	20	54	Variance	55.01724
95%	24	57	Skewness	1.89567
99%	39	68	Kurtosis	8.63772

Edad del informante				
	Percentiles	Smallest		
1%	24	20		
5%	28	21		
10%	32	21	Obs	2352
25%	37	21	Sum of Wgt.	2352
50%	45		Mean	45.43112
		Largest	Std. Dev.	10.99269
75%	53	80		
90%	60	82	Variance	120.8392
95%	64	83	Skewness	.2692479
99%	72	83	Kurtosis	2.762237

Nivel educativo alcanzado				
	Percentiles	Smallest		
1%	1	1		
5%	2	1		
10%	2	1	Obs	2353
25%	2	1	Sum of Wgt.	2353
50%	3		Mean	2.617935
		Largest	Std. Dev.	.5727378
75%	3	3		
90%	3	3	Variance	.3280286
95%	3	3	Skewness	-1.202962
99%	3	3	Kurtosis	3.450666

Organización jurídica de la empresa				
	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	2352
25%	0	0	Sum of Wgt.	2352
50%	0		Mean	.4927721
		Largest	Std. Dev.	.5000541
75%	1	1		
90%	1	1	Variance	.2500541
95%	1	1	Skewness	.0289146
99%	1	1	Kurtosis	1.000836

Licencia de funcionamiento municipal				
	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	2352
25%	0	0	Sum of Wgt.	2352
50%	1		Mean	.7019558
		Largest	Std. Dev.	.4574963
75%	1	1		
90%	1	1	Variance	.2093029
95%	1	1	Skewness	-.8830617
99%	1	1	Kurtosis	1.779798

Tipo de local				
	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	2352
25%	0	0	Sum of Wgt.	2352
50%	0		Mean	.4247449
		Largest	Std. Dev.	.4944093
75%	1	1		
90%	1	1	Variance	.2444406
95%	1	1	Skewness	.304489
99%	1	1	Kurtosis	1.092714

- Regresión de Poisson

Poisson regression		Number of obs	=	2352
Log likelihood = -9195.5577		LR chi2(6)	=	1800.30
		Prob > chi2	=	0.0000
		Pseudo R2	=	0.0892

oper	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
edad	.0186796	.0005798	32.22	0.000	.0175431 .019816
orga	-.1246155	.0134323	-9.28	0.000	-.1509423 -.0982888
licen	.2500621	.0152147	16.44	0.000	.2202418 .2798824
local	.1195457	.0128345	9.31	0.000	.0943905 .1447009
educ1	.1124202	.0292964	3.84	0.000	.0550003 .16984
educ3	.1436036	.0150032	9.57	0.000	.1141978 .1730094
_cons	1.235242	.0311917	39.60	0.000	1.174108 1.296377

```
. fitstat
```

Measures of Fit for poisson of oper

Log-Lik Intercept Only:	-10095.706	Log-Lik Full Model:	-9195.558
D(2345):	18391.115	LR(6):	1800.296
		Prob > LR:	0.000
McFadden's R2:	0.089	McFadden's Adj R2:	0.088
ML (Cox-Snell) R2:	0.535	Cragg-Uhler(Nagelkerke) R2:	0.535
AIC:	7.825	AIC*n:	18405.115
BIC:	186.830	BIC':	-1753.718
BIC used by Stata:	18445.457	AIC used by Stata:	18405.115



- Regression final: Binomial negativa

Negative binomial regression		Number of obs =		2352		
Dispersion = mean		LR chi2(6) =		426.66		
Log likelihood = -7367.8649		Prob > chi2 =		0.0000		
		Pseudo R2 =		0.0281		
oper	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	.0179657	.0011233	15.99	0.000	.015764	.0201674
orga	-.1240375	.0265429	-4.67	0.000	-.1760606	-.0720145
licen	.2461704	.028502	8.64	0.000	.1903074	.3020333
local	.1243858	.0249731	4.98	0.000	.0754393	.1733323
educ1	.1082978	.0608915	1.78	0.075	-.0110473	.2276429
educ3	.1511377	.0290558	5.20	0.000	.0941894	.2080861
_cons	1.264121	.0584823	21.62	0.000	1.149498	1.378744
/lnalpha	-1.385122	.0391573			-1.461869	-1.308375
alpha	.2502932	.0098008			.2318026	.2702587

. fitstat

Measures of Fit for nbreg of oper

Log-Lik Intercept Only:	-7581.193	Log-Lik Full Model:	-7367.865
D(2344):	14735.730	LR(6):	426.656
		Prob > LR:	0.000
McFadden's R2:	0.028	McFadden's Adj R2:	0.027
ML (Cox-Snell) R2:	0.166	Cragg-Uhler(Nagelkerke) R2:	0.166
AIC:	6.272	AIC*n:	14751.730
BIC:	-3460.792	BIC':	-380.078
BIC used by Stata:	14797.834	AIC used by Stata:	14751.730

- Test de variables omitidas

```
. regress oper edad orga licen local educ1 educ3
```

Source	SS	df	MS			
Model	19788.5348	6	3298.08913	Number of obs =	2352	
Residual	117448.785	2345	50.0847697	F( 6, 2345) =	65.85	
				Prob > F =	0.0000	
				R-squared =	0.1442	
				Adj R-squared =	0.1420	
Total	137237.32	2351	58.3740195	Root MSE =	7.0771	

oper	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
edad	.2095922	.0138128	15.17	0.000	.1825057	.2366787
orga	-1.424461	.3157186	-4.51	0.000	-2.043577	-.8053438
licen	2.555779	.3341006	7.65	0.000	1.900616	3.210942
local	1.321376	.3011542	4.39	0.000	.7308202	1.911933
educ1	1.554296	.7475164	2.08	0.038	.0884342	3.020158
educ3	1.526156	.34357	4.44	0.000	.8524236	2.199889
_cons	-1.409281	.6981687	-2.02	0.044	-2.778373	-.0401891

```
. estat ovtest
```

Ramsey RESET test using powers of the fitted values of oper  
Ho: model has no omitted variables  
F(3, 2342) = 12.74  
Prob > F = 0.0000

- Predicción de escenarios

```
. prvalue, x(licen=0 local=0 orga=1) rest(mean) brie
```

```
nbreg: Predictions for oper
```

		95% Conf. Interval	
Rate:	7.8584	[ 7.3908,	8.326]
Pr(y=0 x):	0.0130	[ 0.0109,	0.0150]
Pr(y=1 x):	0.0344	[ 0.0296,	0.0391]
Pr(y=2 x):	0.0569	[ 0.0502,	0.0636]
Pr(y=3 x):	0.0754	[ 0.0680,	0.0827]
Pr(y=4 x):	0.0874	[ 0.0806,	0.0941]
Pr(y=5 x):	0.0926	[ 0.0873,	0.0979]
Pr(y=6 x):	0.0921	[ 0.0886,	0.0955]
Pr(y=7 x):	0.0872	[ 0.0857,	0.0887]
Pr(y=8 x):	0.0794	[ 0.0792,	0.0796]
Pr(y=9 x):	0.0702	[ 0.0686,	0.0718]

```
. prvalue, x(licen=1 local=1 orga=0) rest(mean) brie
```

```
nbreg: Predictions for oper
```

		95% Conf. Interval	
Rate:	12.886	[ 12.28,	13.493]
Pr(y=0 x):	0.0032	[ 0.0027,	0.0036]
Pr(y=1 x):	0.0096	[ 0.0084,	0.0109]
Pr(y=2 x):	0.0184	[ 0.0161,	0.0206]
Pr(y=3 x):	0.0280	[ 0.0249,	0.0311]
Pr(y=4 x):	0.0374	[ 0.0337,	0.0411]
Pr(y=5 x):	0.0456	[ 0.0416,	0.0497]
Pr(y=6 x):	0.0522	[ 0.0482,	0.0562]
Pr(y=7 x):	0.0569	[ 0.0532,	0.0607]
Pr(y=8 x):	0.0597	[ 0.0565,	0.0630]
Pr(y=9 x):	0.0608	[ 0.0581,	0.0634]