



**UNIVERSIDAD DE ORIENTE  
NÚCLEO DE SUCRE  
ESCUELA DE ADMINISTRACIÓN  
DEPARTAMENTO DE CONTADURÍA**

# **APLICACIONES ESTADÍSTICAS A LA ECONOMÍA (Econometría)**

Tutor Académico:  
Lcdo. **Miguel Romero**

Presentado por:  
**Pérez L., Daniel E.** C.I.: V-11.424.175.  
**Sánchez R., Jesús E.** C.I.: V-14.596.818

Trabajo de Curso Especial de Grado presentado como requisito parcial para  
optar al Título de LICENCIADO EN CONTADURÍA PÚBLICA

Cumaná, noviembre de 2008



**UNIVERSIDAD DE ORIENTE  
NÚCLEO DE SUCRE  
ESCUELA DE ADMINISTRACIÓN  
DEPARTAMENTO DE CONTADURÍA**

# **APLICACIONES ESTADÍSTICAS A LA ECONOMÍA (Econometría)**

Presentado por:

**Pérez L., Daniel E. C.I.: V-11.424.175.**

**Sánchez R., Jesús E. C.I.: V-14.596.818**

## **ACTA DE APROBACIÓN DEL JURADO**

Trabajo de Grado aprobado en nombre de la Universidad de Oriente, por el siguiente jurado calificador, en la Ciudad de Cumaná, a los 7 días del mes de noviembre de 2008

Lcdo. **Miguel Romero**

C.I.: V-8.879.006

Tutor

Cumaná, noviembre de 2008

# ÍNDICE

<b>DEDICATORIA</b>	<b>I</b>
<b>AGRADECIMIENTO</b>	<b>III</b>
<b>ÍNDICE DE TABLAS</b>	<b>V</b>
<b>ÍNDICE DE FIGURAS</b>	<b>VI</b>
<b>RESUMEN</b>	<b>VII</b>
<b>INTRODUCCIÓN</b>	<b>1</b>
<b>CAPÍTULO I</b>	<b>2</b>
<b>ASPECTOS GENERALES</b>	<b>2</b>
1.1. Planteamiento Del Problema	2
1.2. Objetivos	6
1.2.1. General	6
1.3. Justificación	8
1.4. Marco Referencial	9
1.4.1. Antecedentes de la Investigación	9
1.4.2. Bases Teóricas	10
1.4.3. Definición de Términos Básicos	12
1.5. Marco Metodológico	17
1.5.1. Nivel de la Investigación	17
1.5.2. Diseño de la Investigación	17
1.5.3. Fuentes de Información	18
1.5.4. Técnicas e Instrumentos de Recolección de Datos	18
1.5.5. Técnicas del Procesamiento y Análisis de Datos	19
<b>CAPÍTULO II</b>	<b>20</b>
<b>SERIES DE TIEMPO</b>	<b>20</b>
2.1. Series De Tiempo	20
2.1.1. El Doble Objetivo de las Series de Tiempo	21
2.1.2. Componentes para el Estudio de las Series de Tiempo	22
2.1.2.1. Tendencia Secular	22
2.1.2.2. El Componente Estacional	24

2.1.2.3. Variaciones Cíclicas	26
2.1.2.4. Variaciones Irregulares	28
2.2. Modelos De Series De Tiempo	29
2.3. Números Índice	30
2.3.1. El Uso de los Números Índice	31
2.3.2. Tipos de Números Índice	32
2.4. Técnicas De Suavizamiento	36
2.4.2. Promedios Móviles ó Medias Móviles	37
2.4.3. Suavizamiento Exponencial	39
<b>MÉTODOS DE ANÁLISIS DE SERIES TEMPORALES</b>	<b>41</b>
2.5.1. Análisis de Tendencia	41
2.5.2. Descomposición de las Series de Tiempo	45
<b>CAPÍTULO III</b>	<b>46</b>
<b>ANÁLISIS DE REGRESIONES</b>	<b>46</b>
3.1. Regresión	46
3.2. Objetivos Del Análisis De Regresión	47
3.3. Regresión Simple	47
3.4. Modelos De Regresión Simple	49
3.5. Determinación Del Modelo De Regresión	50
3.6. Método De Mínimos Cuadrados	55
3.7. Error Estándar En La Estimación	62
3.8. Análisis De Correlación	64
3.8.1. Correlación Lineal Simple	65
3.8.2. Coeficientes de Correlación	70
3.8.3. Pearson	71
3.8.4. Características del Coeficiente de Correlación	72
3.8.5. Tau-B de Kendall	76
3.8.6. Spearman	77
3.9. Limitaciones Del Análisis De Regresión	77
<b>CONCLUSIONES</b>	<b>80</b>
<b>RECOMENDACIONES</b>	<b>81</b>
<b>BIBLIOGRAFÍA</b>	<b>82</b>

## DEDICATORIA

- A **DIOS**, todopoderoso por darme la fuerza y voluntad necesaria para seguir adelante y no dejarme vencer por los obstáculos que encontramos en el andar de la vida, por escucharme en mis noches de oración, por estar siempre a mi lado en los momentos de tristeza y de soledad, por haberme guiado por el buen camino, ayudándome así a levantarme cuando he tropezado y caído, hasta cumplir con mis metas.
- A mi padre; **Jesús Sánchez** quien desde muy pequeño me enseñó a luchar por las cosas sin dejarme vencer, y que a pesar de todos los malos momentos que hemos vivido siempre nos ha sacado adelante con su esfuerzo y dedicación.
- A mi madre; **Nereida Rodríguez** que desde siempre se ha esforzado por que salga adelante, brindándome su cariño, amor, apoyo y comprensión, de no ser por ella tampoco hubiese llegado donde estoy, ya que, siempre tenía el consejo adecuado para levantarme los ánimos y las ganas de seguir adelante.
- A mis hermanos; **Mercedes y Carlos** porqué siempre han estado a mi lado cuando más los he necesitado.
- A mi novia **Esthelyn Mendoza** que ha sido una persona muy especial, porque desde que la conozco me ha brindado todo su cariño y apoyo y me ha ayudado en todo lo necesario.

*Jesús Sánchez*

## DEDICATORIA

- A DIOS todopoderoso, quien me ha dado la vida y me ha permitido ver su gloria puesta de manifiesto en mi vida.
- A mis pequeñas maravillas, mis hijas, Daniela y Patricia, por ser mi motivo.
- A mis abuelas, Mélida y María T., por su abnegación y entrega, en resaltar los valores de la familia.

*Daniel Pérez Loaiza*

## AGRADECIMIENTO

- Gracias a mis **Padres**, Nereida y Jesús por darme la oportunidad de tener una buena educación, para formarme no solo de manera intelectual, sino también espiritual. A mi Dios y a ustedes debo lo que soy, gracias por proporcionarme todas las herramientas necesarias para lograr este triunfo.
- A la **Universidad de Oriente Núcleo Sucre**, por darme la oportunidad de abrirme sus puertas y poder hacer realidad uno de mis sueños convertirme en un profesional.
- A mi compañero de clase y compañero de tesis **Daniel Pérez**, por todo el apoyo y la ayuda que me ha dado en la realización de este trabajo y por lo que a través de él he aprendido. Para ti mis mejores deseos de éxito y felicidad. **¡que dios te bendiga!**
- A mi Profesor y asesor **Licenciado Miguel Romero** por guiarme y ayudarme en todo lo necesario, para de esta forma cumplir mi meta.
- A mis **compañeros de clase**, por las veces que me ayudaron y en especial a **Luidimar** que siempre estuvo presente para aconsejarme y apoyarme en los malos momentos.

*Jesús Sánchez*



## AGRADECIMIENTO

- A Dios, por iluminarme el camino en todos los momentos de mi vida. Por enseñarme que con fe todo es posible.
- A mis padres, Luisa Loaiza y Rafael Pérez, por su entrega incondicional, quienes con su ejemplo, me impulsaron a seguir adelante.
- A mis hermanos, Alejandro y Rafael, a quienes quiero mucho.
- A mi esposa e hijo, Carmen Amato y José D., por haber estado siempre a mi lado.
- A mis compañeros de estudio, por su solidaridad, su sinceridad y por los gratos momentos compartidos, en especial a mi compañero de tesis Jesús, por su paciencia y dedicación.
- A todo el personal docente de la Universidad de Oriente, por su constante búsqueda de la excelencia académica.
- A todas aquellas personas que de una u otra forma contribuyeron al logro de este objetivo.

Les estaré a todos, eternamente agradecido.

*Daniel Pérez Loaiza*

## ÍNDICE DE TABLAS

<b>TABLA Nº 1. Producción de Calzado Anual .....</b>	<b>23</b>
<b>TABLA Nº 2. Recaudación de Impuesto Sobre la Renta en Venezuela (Millones de bolívares) datos irreales.....</b>	<b>25</b>
<b>TABLA Nº 3. Producción Petrolera.....</b>	<b>27</b>
<b>TABLA Nº 4. Compras (Unidades).....</b>	<b>37</b>
<b>TABLA Nº 5. Ventas de Autos.....</b>	<b>40</b>
<b>TABLA Nº 6. Pronostico de Clientes .....</b>	<b>42</b>

## ÍNDICE DE FIGURAS

<b>FIGURA Nº 1. Tendencia Secular</b> .....	24
<b>FIGURA Nº 2. Componente Estacional</b> .....	26
<b>FIGURA Nº 3. Variación Cíclica</b> .....	28
<b>FIGURA Nº 4. Medias o Promedios Móviles (Suavizamiento)</b> .....	38
<b>FIGURA Nº 5. Análisis de Tendencia</b> .....	44
<b>FIGURA Nº 6. Relación Estadística Lineal</b> .....	49
<b>FIGURA Nº 7. Relación Estadística No Lineal</b> .....	50
<b>FIGURA Nº 8. Línea Recta con Pendiente Positiva</b> .....	52
<b>FIGURA Nº 9A. Línea Recta con Pendiente Negativa</b> .....	52
<b>FIGURA Nº 9B. No Existe Ninguna Relación entre X e Y</b> .....	53
<b>FIGURA Nº 10. Valores Observados y Estimados</b> .....	62
<b>FIGURA Nº 11A. Relación Lineal Positiva Perfecta</b> .....	65
<b>FIGURA Nº 11B. Relación Lineal Negativa</b> .....	66
<b>FIGURA Nº 11C. Relación Exponencial o Curvilínea</b> .....	66
<b>FIGURA Nº 11D. No Existe Relación</b> .....	67
<b>FIGURA Nº 11E. Relación Lineal Positiva</b> .....	68
<b>FIGURA Nº 11F. Independencia Lineal</b> .....	69
<b>FIGURA 12. Una Posible Relación X –Y</b> .....	79



**UNIVERSIDAD DE ORIENTE  
NÚCLEO DE SUCRE  
ESCUELA DE ADMINISTRACIÓN  
DEPARTAMENTO DE CONTADURÍA**

## **APLICACIONES ESTADÍSTICAS A LA ECONOMÍA (Econometría)**

Asesor:  
Lcdo. **Miguel Romero**  
C.I.: V-8.879.006

Presentado por:  
**Pérez L., Daniel E.** C.I.: V-11.424.175.  
**Sánchez R., Jesús E.** C.I.: V-14.596.818

Cumaná, noviembre de 2008

### **RESUMEN**

La Econometría es una técnica de estudio de fenómenos económicos, aplicados por economistas con el fin de poder predecir y comprender futuros acontecimientos, a partir del análisis estadístico y matemático, de esta manera poder sugerir medidas de políticas económicas conforme a objetivos deseados. Para la puesta en práctica de esta técnica se utilizan varias herramientas estadísticas, como las series de tiempo, que es una sucesión de observaciones cuantitativa de un fenómeno ordenadas en el tiempo, esta metodología se basa en descomponer la series en varias partes: tendencia, variaciones estacionales, cíclicas e irregulares. Otra herramienta utilizada es el análisis de regresión y correlación, la idea básica del análisis de regresión es utilizar los datos de una variable cuantitativa independiente para predecir o explicar la variación de una variable cuantitativa dependiente, utilizando estas variables para desarrollar una ecuación de predicción, y el análisis de correlación, se refiere al grado de variación existente entre dos o más variables. En el desarrollo de este trabajo se aplicó una investigación de tipo descriptivo y el diseño fue de tipo documental, utilizando fuentes de información secundarias.

## INTRODUCCIÓN

El objetivo de la Economía –como ciencia social– es tratar de distribuir los escasos recursos entre la infinidad de necesidades de los pobladores de un territorio.

El comportamiento humano, ha sido uno de los grandes acertijos a través de todos los tiempos. Es por esta razón que los economistas han tratado, más que entenderlo, de asumirlo y explicarlo de algún modo. Las Matemáticas, han servido de apoyo para cuantificar los resultados de la puesta en práctica de las teorías económicas.

En la elaboración de la Econometría se unen las Matemáticas y la Estadística junto con la investigación social y la teoría económica.

El mayor problema con el que se enfrentan los económetras en su investigación es la escasez de datos, los sesgos que pueden causar los mismos y la ausencia o insuficiencia de una teoría económica adecuada. Aún así, la Econometría es la única aproximación científica al entendimiento de los fenómenos económicos.

La Estadística y la Teoría Económica, han servido de fundamento de la Econometría desde sus inicios. En la medida en que han ido incursionando los Económetras en la Economía se han visto en la obligación de hacer nuevas propuestas u observaciones a las teorías económicas existentes. Cabe destacar, la evolución que ha tenido la Economía como resultado de los estudios econométricos.

# CAPÍTULO I

## ASPECTOS GENERALES

### 1.1. Planteamiento Del Problema

La ciencia económica, como disciplina que estudia la conducta de seres humanos en relación mutua, debe considerarse como una ciencia social, es decir, no es una ciencia exacta, por lo tanto, no bastan las demostraciones lógicas o matemáticas para aceptar que sus postulados se verifican en la realidad. Hay en la teoría económica presencia de pensamientos complejos, nociones alternativas y disímiles, ideas sobre la articulación entre lo social, lo político y lo económico.

La economía nació como ciencia social, tratando de explicar la distribución de recursos escasos entre distintos fines posibles. El estudio de la Economía puede dividirse en dos grandes campos: La teoría de los precios, o microeconomía, y la macroeconomía.

La macroeconomía analiza parte de la ciencia social económica que enfoca el análisis del sistema económico, utilizando la abstracción de las principales variables del universo de relaciones económicas; construyendo modelos económicos que sirvan para predecir el comportamiento de las mencionadas variables. El estudio de temas macroeconómicos es motivado por problemas como caídas en el nivel de ingreso nacional y persistencia de elevados niveles de desempleo.

En las últimas décadas, se ha avanzado hacia modelos

macroeconómicos que parten de ecuaciones de comportamiento microeconómicos, lo que refuerza la idea de que ambas son sólo dos perspectivas diferentes de analizar un mismo problema. En el centro de la macroeconomía se encuentran las cuestiones de la creación, distribución y utilización del ingreso nacional y la ocupación de los factores de producción durante los ciclos coyunturales.

Los procesos macroeconómicos se forman de numerosas acciones y reacciones de agentes económicos individuales. Actualmente, los modelos macroeconómicos parten de ecuaciones de comportamiento macroeconómicas, para, mediante un proceso de agregación, deducir el comportamiento de agregados, lo que les otorga un mayor fundamento científico. Estos procesos son interdependientes entre sí. Esto se refleja actualmente en el hecho de que los modelos macroeconómicos tienen en cuenta las interdependencias existentes entre las diferentes magnitudes (producción, desempleo, inflación, etc.) y sectores macroeconómicos (familias, empresas, gobierno, sector externo, etc.).

La microeconomía es una rama de la economía que estudia el comportamiento de unidades económicas individuales, como pueden ser individuos, familias y empresas, y el funcionamiento de los mercados en los cuales ellos operan. La definición más clásica de microeconomía dice que la microeconomía es la parte de la economía que estudia la asignación de los recursos escasos entre finalidades alternativas. La teoría microeconómica utiliza modelos formales que intentan explicar y predecir, utilizando supuestos simplificadores, el comportamiento de los consumidores y productores, y la asignación de los recursos que surge como resultado de su interacción en el mercado. En general el análisis microeconómico se asocia con la teoría de precios y sus derivaciones. Se considera que el mayor

contribuyente al análisis microeconómico ha sido Alfred Marshall.

Es usual considerar a la microeconomía la rama opuesta a la macroeconomía, ya que mientras la microeconomía analiza el comportamiento de agentes individuales, como personas, familias, y firmas, la macroeconomía tiene como objeto de estudio a agregados, como el conjunto de familias de un país, o el conjunto de empresas de un país.

Actualmente, el estudio de la microeconomía se encuentra signado por los fundamentos de la escuela neoclásica, lo que se observa en la mayoría de los análisis microeconómicos. Por ejemplo, en los modelos utilizados en la microeconomía se parte de agentes racionales que tienen que resolver problemas de optimización condicionada. Para resolver estos problemas se utilizan técnicas de optimización.

Los economistas han tratado de concebir modelos que explicaran el comportamiento de sus semejantes en cuanto a la actividad económica. Por otra parte, ha habido un afán de tratar de corroborar esos modelos a través de su cuantificación matemática. Hay que considerar que tratan con uno de los fenómenos más complejos que conocemos, el comportamiento de los humanos.

Es entonces, como los conocimientos y las teorías micro y macro económicas se han vuelto insuficientes para explicar los fenómenos que acontecen en la economía, es por ello que estudiosos de las ciencias económica se han visto en la necesidad de crear una técnica que ayude a compensar las debilidades de estas dos vertientes (micro y macro economía) y que contribuya de manera más precisa (verificable) y eficaz al momento de tomar decisiones económicas u organizacionales.



De esta forma, surge la Econometría, la cual consiste en una combinación de Economía Matemática, Teoría de Probabilidad y Estadística, datos económicos y, claro está, teoría económica. Esta disciplina fortalece el carácter científico de la economía (o cuando menos lo intenta): compara los modelos económicos con lo que se observa en la realidad y, por lo mismo, da indicios respecto a cuales teorías resultan demasiado alejadas de lo observado. El que la Econometría difiera de la Estadística se debe esencialmente a que resulta virtualmente imposible, en cuestiones económicas, realizar experimentos bajo condiciones controladas. Lo anterior obliga a los econométricos a tomar los datos crudamente recolectados, sin conocer exactamente bajo que condiciones ocurrieron las cosas. Es necesario entonces crear métodos que puedan filtrar correctamente la información y tomar en cuenta esta carencia. Otro problema persistente en la disciplina estriba en la incertidumbre imperante en lo que refiere a las verdaderas relaciones entre variables económicas. Cuando los modelos omiten (por desconocimiento) relaciones importantes, la estimación de las restantes relaciones puede verse sumamente deteriorada. También destaca el hecho de que muchas series económicas presentan un componente de tendencia a través del tiempo; Ello implica que las variables no se mantienen alrededor de un mismo nivel sino que, van creciendo o disminuyendo conforme se hacen nuevas mediciones: se dice entonces que las series no son estacionarias. La no-estacionariedad resulta muy grave, ya que muchas técnicas estadísticas dejan de servir cuando está presente en los datos.

El mayor problema con el que se enfrentan los econométricos en su investigación es la escasez de datos, los sesgos que pueden causar los mismos y la ausencia o insuficiencia de una teoría económica adecuada. Aún así, la Econometría es la única aproximación científica al entendimiento de los fenómenos económicos.

Fundamentado en todo lo antes expuestos, se formularon las siguientes incógnitas:

- 1.- ¿Cuál es el uso de las herramientas estadísticas para la Economía?.
- 2.- ¿Qué beneficios proporciona la Econometría a la Economía?.
- 3.- ¿Puede la Econometría coadyuvar a aliviar el problema económico básico?.
- 4.- ¿Cómo contribuye las series de tiempo en la tomas de decisiones económicas?.
- 5.- ¿Cuál es el aporte del análisis de regresión en la toma de decisiones económicas?.

## **1.2. Objetivos**

### 1.2.1. General

Establecer cuales son las aplicaciones estadísticas a la Economía (Econometría).

### 1.2.2. Específicos

- Definir los elementos teóricos de Estadística y de Economía como base para la Econometría.
- Conocer la funcionalidad de las series de tiempo en la Econometría.
- Estudiar la determinación y el uso de los números índice.
- Identificar la operacionalización del análisis de regresión como herramienta estadística para la Economía.

- Medir, a través de ejemplos prácticos la necesidad del uso de las series de tiempo y del análisis de regresión para la Econometría.

### 1.3. Justificación

La Economía es una ciencia social muy compleja, ya que, estudia las necesidades humanas. A través del tiempo ésta ha venido evolucionando de acuerdo con las necesidades de su época, es así que mediante su avance han surgido varias ramas, que de algún modo facilitan el estudio de esta ciencia, de esta forma nace la Econometría, la cual hace valiosos aportes para la Economía, porque es capaz de medir, usando datos reales, los fenómenos económicos mediante la aplicación de métodos estadísticos, fortaleciendo así su carácter científico. Dando como resultado, una base que sustente la aplicación de las herramientas estadísticas, para ayudar a entender los problemas económicos.

El fin de esta investigación es dar a conocer los aspectos relevantes de la econometría y el uso que se le da dentro del campo económico y empresarial, aplicando series de tiempo y análisis de correlación. Esto es de gran importancia para nosotros como futuros profesionales, porque, nos permite tener herramientas con las cuales contar a la hora de tomar decisiones, ya que, mediante su aplicación podemos predecir situaciones futuras que puedan afectar las metas establecidas sea de una organización o de un gobernante. También, podemos hacer estimaciones sobre algunas variables que consideremos necesarias para la mejor toma de decisiones, como por ejemplo, la liquidez que puede tener una empresa para hacer frente a sus compromisos, la cantidad de materia prima que se va a utilizar para producir sus productos, el costo de producción, estimar el grado que puede alcanzar la inflación en un periodo, etc. Y de esta manera optimizar los recursos que se tienen. Por eso, tomar decisiones apoyadas en estudios econométricos, es garantía de un buen camino hacia la consecución de las metas propuestas.

La siguiente investigación busca primordialmente proporcionar, a las personas involucradas e interesadas en las ciencias administrativas y contables, una herramienta sencilla, de fácil comprensión y aplicación para la realización de estudios econométricos, con el objeto de poder prever el resultado de sus decisiones.

## **1.4. Marco Referencial**

### 1.4.1. Antecedentes de la Investigación

VENTOSA (2006), en su reporte ¿Qué es la Econometría? Expresa:

Los estudios empíricos en economía suelen ser realizados mediante el uso de técnicas econométricas. Estas se componen de diversos elementos entre los destacan la economía matemática, la probabilidad y la estadística así como el análisis de bases de datos económicos. El análisis econométrico apareció por primera vez a finales del siglo XIX. No obstante, no fue sino hasta la década de los 30 que su origen suele establecerse oficialmente.

Desde entonces a la fecha, esta disciplina ha sufrido un sinfín de revoluciones, algunas de ellas muy dramáticas. Los objetivos, empero, siguen siendo los mismos: proveer del herramental adecuado al economista para que éste pueda llevar a cabo estudios económicos con base en datos estadísticos.

MATA y DA COSTA (2004), en su trabajo presentan el siguiente resumen:

La complejidad económica, manifestada en los tres conceptos que definen el valor del dinero: capacidad adquisitiva, tasa de interés y tipo de cambio, se magnifica al involucrarse con la

complejidad política al momento de la previsión. Por ello la Ciencia Económica en su pretensión de proporcionar leyes explicativas para la previsión con un margen aceptable de incertidumbre, viene incorporando métodos de ajustes no lineales y delimitando los aportes de la Econometría como herramienta de uso común o tradicional. Al respecto, el presente ensayo hace una revisión sucinta de la presente etapa de la epistemología económica y da cuenta de una secuencia procedimental como aproximación preliminar a un método general para la evaluación de la complejidad, a partir de la Econometría.

FERNÁNDEZ y DÍAZ (2000):

La Econometría, término introducido por Ragnar Frisch en 1926, es una disciplina que siempre se ha movido entre las teorías económica y Estadística; sin embargo, la interacción de estos dos campos del conocimiento no ha sido ni directa ni sencilla. Desde sus orígenes, en las primeras décadas del siglo veinte, con Henry L. Moore, pasando por gente de la talla de Burns, Cowles, Haavelmo, Mitchell, Morgenstern y Tinbergen, hasta econométricos modernos como Box, Hendry, Jenkins, Granger. Koopmans y Spanos, esta ciencia ha mostrado un indudable avance que desde la perspectiva de la teoría estadística, todavía está en construcción.

#### 1.4.2. Bases Teóricas

ALLARD (1980)

Al tratar de explicar algunas de las ideas y técnicas básicas de la Econometría usando solamente un vocabulario matemático limitado, no quiere decir que aquellos que tengan mayor experiencia matemática no puedan lograr mejor y más rápida comprensión de la materia. Sin embargo, creo que hay buenas razones para lograr, incluso, una comprensión limitada de la econometría. En primer lugar, en toda investigación económica

es importante que el investigador esté a cargo personalmente del análisis de datos tanto como sea posible, ya que la interpretación de los resultados depende considerablemente de la comprensión de las decisiones que se tomen en cada paso de la elaboración de los modelos y del análisis de datos. Por lo que el investigador debe comprender las técnicas que se emplean.

En segundo lugar, el uso creciente de las técnicas de econometría en todos los aspectos de la economía aplicada dificulta al principiante la evaluación de una obra dada de investigación. Por último, está la necesidad de hacer inferencias estadísticas a partir de los datos empíricos con relación a muchos problemas en las ciencias sociales.

#### ALLARD (1980)

El estudio de la implicación de decisiones de política como por ejemplo la devaluación, los subsidios a las construcciones urbanas y las políticas monetarias, requieren determinado conocimiento o juicio acerca de las relaciones económicas causales involucradas. Una combinación de la lotería económica y el sentido común pueden indicar la naturaleza cualitativa de dichas relaciones, pero para toda evaluación cuantitativa generalmente se tiene que basar en experiencias anteriores. Gran parte de esta experiencia está contenida en los datos relativos a las variables económicas involucradas, y la mejor oportunidad de descubrir estas relaciones cuantitativas está en el análisis de los datos.

#### SALVATORE y REAGLE (2004)

La econometría se ha identificado mucho con el análisis de regresión. Este análisis permite relacionar una variable dependiente con una o mas variables independiente o explicativas. Puesto que las relaciones entre variables económicas suelen ser inexactas, es necesario incluir un término de error o de perturbación, con propiedades o características bien definidas.

### 1.4.3. Definición de Términos Básicos

#### **Correlación:**

- Es una medida sobre el grado de relación entre dos variables, sin importar cual es la causa y cual es el efecto. La dependencia de la que se habla en este sentido es la dependencia entre la varianza de las variables.
- Indica la fuerza y la dirección de una relación lineal entre dos variables aleatorias. Se considera que dos variables cuantitativas están correlacionadas cuando los valores de una de ellas varían sistemáticamente con respecto a los valores homónimos de la otra: si tenemos dos variables (A y B) existe correlación si al aumentar los valores de A lo hacen también los de B y viceversa. La correlación entre dos variables no implica, por sí misma, ninguna relación de causalidad.

#### **Econometría:**

- La Econometría que es el resultado de cierta posición sobre el papel de la Economía, consiste en la aplicación de la estadística matemática a datos económicos, para dar apoyo empírico a los modelos construidos por la economía matemática, y para obtener resultados matemáticos.
- La Econometría puede ser definida como el análisis cuantitativo de fenómenos económicos reales basados en los desarrollos simultáneos de la observación y la teoría, relacionados mediante métodos apropiados de inferencia.
- La Econometría es la integración de la teoría económica, de las matemáticas y de las técnicas estadísticas con el objetivo de contrastar



hipótesis sobre fenómenos económicos, estimar coeficientes sobre las relaciones económicas, y prever o predecir los valores futuros de las variables o fenómenos económicos.

### **Economía Aplicada:**

Analiza los problemas encontrados y las conclusiones en determinados campos de la Economía, como la Teoría de la Demanda, De la Producción, De la Inversión, Del Consumo y otros campos de la investigación económica aplicada.

### **Economía Teórica:**

Hace referencia a los métodos de medición de las relaciones económicas en general.

### **Espuria (o Correlación sin Sentido):**

- Esto consiste básicamente en encontrar relaciones estadísticas entre variables que de hecho no tienen ninguna conexión entre ellas.

### **Estadística Aplicada (Inferencia Estadística):**

- Se denomina estadística aplicada al área de la estadística que se ocupa de inferir resultados sobre la población a partir de una o varias muestras.
- La inferencia estadística es una parte de la Estadística que comprende los métodos y procedimientos para deducir propiedades (hacer inferencias) de una población, a partir de una pequeña parte de la misma

(muestra).

### **Estocástico:**

- Se denomina estocástico a aquel sistema que funciona, sobre todo, por el azar. La palabra proviene del griego: στοχαστικός, hábil en conjeturar. Significa "perteneciente o relativo al azar" según el DRAE. Las leyes conocidas de causa–efecto no explican cómo actúa el sistema (y de modo reducido el fenómeno) de manera determinista, sino en función de probabilidades.

### **Medidas de Dispersión:**

- Indican la mayor o menor concentración de los datos con respecto a las medidas de tendencia central.

### **Medidas de Tendencia Central:**

- Indican valores con respecto a los que los datos parecen agruparse.
- Nos dan un centro de la distribución de frecuencias, es un valor que se puede tomar como representativo de todos los datos.

### **Método de Mínimos Cuadrados:**

- Es una técnica de optimización matemática que, dada una serie de mediciones, intenta encontrar una función que se aproxime a los datos (un "mejor ajuste").
- El procedimiento más objetivo para ajustar una recta a un conjunto de

datos presentados en un diagrama de dispersión se conoce como "el método de los mínimos cuadrados".

### **Números Índice:**

- Es un número que expresa un cambio relativo de un lapso de tiempo a otro, en los conceptos de precio, cantidad o valor.
- Es una medida estadística diseñada para poner de relieve cambios en una variable o en un grupo de variables relacionadas con respecto al tiempo, situación geográfica, ingreso o cualquier otra característica. Una colección de números índice para diferentes años, lugares, etc.; se llama a veces serie de índices.

### **Regresión:**

- Consiste en modelizar la relación de dependencia entre las variables y predecir los valores de una de ellas (variable dependiente) en función de los valores de la otra (variable independiente o explicativa).

### **Regresión Lineal (Ajuste Lineal):**

- Es un método matemático que modeliza la relación entre una variable dependiente  $Y$ , las variables independientes  $X_i$  y un término aleatorio  $\epsilon$ .

### **Regresión Simple:**

- Cuando SOLO HAY UNA variable independiente o explicativa.

- Se utiliza para contrastar la hipótesis sobre la relación entre la variable independiente  $Y$  y una variable independiente o explicativa  $X$  y para realizar predicciones.

### **Series de Tiempo:**

- Sucesión estadística de los valores de una variable a lo largo del tiempo.
- Gráfica que traza una variable en el tiempo, por lo general colocando al tiempo sobre el eje horizontal.

### **Variable Dependiente:**

- Como su palabra lo dice, son características de la realidad que se ven determinadas o que dependen del valor que asuman otros fenómenos o variables independientes.
- Es el objeto o evento de estudio, sobre la cual se centra la investigación en general.

### **Variable Endógena (Dependiente):**

- Variable que se explica dentro del modelo económico a partir de sus relaciones con otras.
- "Y" es la variable endógena, cuyo valor es determinado por las exógenas,  $X_1$  hasta  $X_n$ . Cuales son las variables elegidas depende de la teoría económica que se tenga en mente, y también de análisis estadísticos y económicos previos.

**Variable Exógena, Independiente o Explicativa:**

- Los cambios en los valores de este tipo de variables determinan cambios en los valores de otra (variable dependiente).
- Es aquella propiedad de un fenómeno a la que se le va a evaluar su capacidad para influir, incidir o afectar a otras variables.

**1.5. Marco Metodológico****1.5.1. Nivel de la Investigación**

Para el desarrollo de este trabajo se aplicó una investigación de tipo descriptivo. Ésta procura brindar una buena percepción del funcionamiento de un fenómeno y de las maneras en que se comportan las variables, factores o elementos que lo componen.

Los estudios descriptivos llegan finalmente a conclusiones generales construidas por medio de abstracciones, que dan cuenta de los hechos observados y se llaman generalizaciones empíricas.

**1.5.2. Diseño de la Investigación**

Constituye el plan general del investigador para obtener respuestas a sus interrogantes o comprobar la hipótesis de investigación. El diseño de investigación desglosa las estrategias básicas que el investigador adopta para generar información exacta e interpretable. Los diseños son estrategias con las que intentamos obtener respuestas a preguntas como: contar, medir o describir.

El diseño de este trabajo fue de tipo documental.

### 1.5.3. Fuentes de Información

El contenido de esta investigación está respaldado por fuentes de tipo documental (o secundarias). El material consultado estuvo compuesto por: libros, revistas, trabajos de ascenso y tesis doctorales; las mismas se consultaron de forma impresa y digital.

### 1.5.4. Técnicas e Instrumentos de Recolección de Datos

Cuando hablamos de recolección de datos nos estamos refiriendo a información empírica abstraída en conceptos. La recolección de datos tiene que hacer con el concepto de medición, proceso mediante el cual se obtiene el dato, valor o respuesta para la variable que se investiga.

La medición, etimológicamente viene del verbo medir y significa comparar una cantidad con su respectiva unidad con el fin de averiguar cuantas veces la segunda esta contenida en la primera (Diccionario de la Real Academia Española).

En el proceso de recolección de datos la medición es una precondition para obtener el conocimiento científico.

El instrumento de recolección de datos está orientado a crear las condiciones para la medición. Adicionalmente, se recurrió a la encuesta oral.

### 1.5.5. Técnicas del Procesamiento y Análisis de Datos

Los datos son conceptos que expresan una abstracción del mundo real, de lo sensorial, susceptible de ser percibido por los sentidos de manera directa o indirecta. Todo lo empírico es medible. No existe ningún aspecto de la realidad que escape a esta posibilidad. Medición implica cuantificación.

Dentro de las técnicas empleadas para el procesamiento de los datos para el desarrollo de esta investigación, se propuso primeramente recolectar, clasificar y organizar, de una forma lógica los datos obtenidos, y presentar los resultados de acuerdo al nivel de conocimiento requerido para ir avanzando a los de mayor complejidad.

## **CAPÍTULO II**

### **SERIES DE TIEMPO**

#### **2.1. Series De Tiempo**

Una serie temporal es una colección de datos reunidos sobre la misma variable a lo largo del tiempo. Desde cuestiones macroeconómicas como el ingreso disponible, hasta sucesos microeconómicos tales como las ventas semanales de un producto particular en una tienda específica, las series de tiempo son datos fundamentales para los gerentes.

El objetivo del análisis de una serie temporal, de la que se dispone de datos en períodos regulares de tiempo, es el conocimiento de su patrón de comportamiento para prever la evolución futura, siempre bajo el supuesto de que las condiciones no cambiarán respecto a las actuales y pasadas.

Si al conocer la evolución de la serie en el pasado se pudiese predecir su comportamiento futuro sin ningún tipo de error, estaríamos frente a un fenómeno determinista cuyo estudio no tendría ningún interés especial.

Si los datos de series temporales presentan dificultades, también presentan oportunidades. Uno de los mejores predictores del comportamiento futuro de una variable es su comportamiento pasado. El análisis inteligente de los datos de series temporales puede producir a menudo intuiciones muy útiles para entender y predecir los valores futuros de la serie.



Una serie de tiempo es un conjunto de observaciones producidas en determinados momentos durante un periodo, semanal, mensual, trimestral o anual, generalmente a intervalos iguales.

Si bien el comportamiento de cualquier serie de tiempo puede observarse gráficamente, no en todos los casos es posible distinguir las particularidades que cada una puede contener. La experiencia basada en muchos ejemplos de series de tiempo, sin embargo, ha revelado que existen ciertos movimientos o variaciones características que pueden medirse y observarse por separado. Estos movimientos, llamados a menudo componentes, de una serie de tiempo y que se supone son causados por fenómenos distintos.

#### 2.1.1. El Doble Objetivo de las Series de Tiempo

Por un lado se busca explicar las variaciones observadas en la serie en el pasado, tratando de determinar si responden a un determinado patrón de comportamiento. Y por otro, si se consigue definir ese patrón o modelo, se intentará predecir el comportamiento futuro de la misma.

Para lograr alcanzar este doble objetivo se utiliza una metodología según la cual se admite que la serie temporal es una función del tiempo:

$$y_t = f(t)$$

Bajo este esquema, la serie sería una variable dependiente y el tiempo una independiente o explicativa. Sin embargo, es necesario dejar bien claro que el tiempo, en sí, no es una variable explicativa, es simplemente el “soporte” o escenario en el que se realiza o tiene lugar la serie temporal. El

tiempo no sirve para explicar el comportamiento de la serie. A esta forma de abordar el estudio de una serie temporal se le conoce como enfoque clásico, frente al causal, según el cual, cualquier serie, como variable que es, puede ser explicada por otra u otras series.

### 2.1.2. Componentes para el Estudio de las Series de Tiempo

Cualquier serie temporal se supone que es el resultado de cuatro componentes: tendencia, variaciones estacionales, variaciones cíclicas y variaciones residuales o accidentales. Pero esta descomposición de la serie no deja de ser un procedimiento diseñado para que el estudio de la misma resulte más fácil, pues esas componentes no siempre existen. Así cuando se trabaja con datos anuales la serie no puede presentar estacionalidad. A su vez las variaciones cíclicas son una componente ligada especialmente a las variables de tipo económico, pero que en variables de otra naturaleza puede que no esté presente. Estas componentes se definen en la forma siguiente:

#### 2.1.2.1. Tendencia Secular

Las tendencias a largo plazo (sin alteraciones de una serie de tiempo) de las ventas, el empleo, los precios de las acciones, y otras series económicas y comerciales.

De forma amplia, podemos definir la tendencia como aquella componente que recoge el comportamiento de la serie a largo plazo. Para poder detectarla es necesario que la serie conste de un número de observaciones elevado, a lo largo de muchos años, para que se pueda determinar si la serie muestra un movimiento a largo plazo que responda a una determinada ley de crecimiento, decrecimiento o estabilidad.

Muchas variables macroeconómicas, como el Producto Nacional Bruto (PNB), el empleo y la producción industrial están dominadas por una fuerte tendencia.

La tendencia de una serie de tiempo es el componente de largo plazo que representa el crecimiento o disminución en la serie sobre un periodo amplio. Las fuerzas básicas que ayudan a explicar la tendencia de una serie son el crecimiento de la población, la inflación de precios, el cambio tecnológico y los incrementos en la productividad.

Es decir, Movimientos seculares contienen los movimientos suaves de largo plazo, los cuales están dominados fundamentalmente por factores de tipo económico.

La siguiente tabla muestra la producción de zapatos en miles unidades para los años 1998 a 2007 de la empresa LAS TRENZAS, C.A.

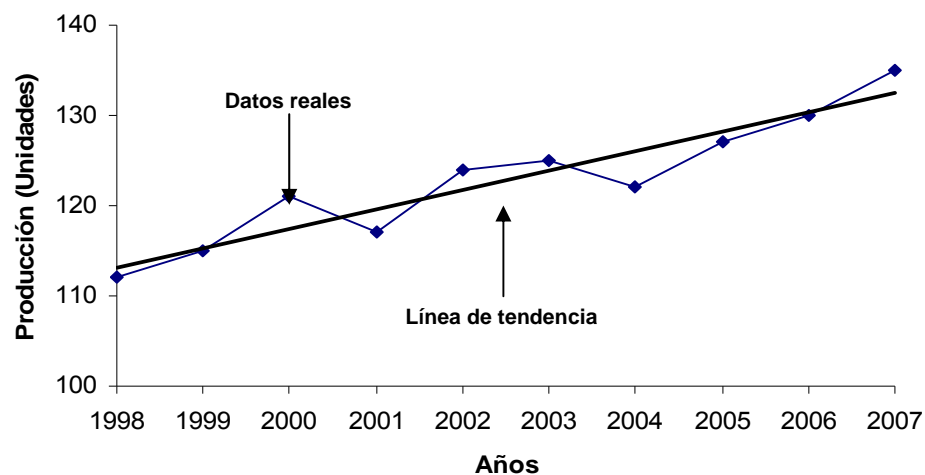
**TABLA Nº 1. Producción de Calzado Anual**

<b>Años</b>	<b>Producción (Unidades)</b>
1998	112
1999	115
2000	121
2001	117
2002	124
2003	125
2004	122
2005	127
2006	130
2007	135

**Fuente:** Los autores.

Veamos en la Figura N° 1 el comportamiento tendencial de estos valores para los períodos en estudio.

**FIGURA N° 1. Tendencia Secular**



**Fuente:** Los autores.

#### 2.1.2.2. El Componente Estacional

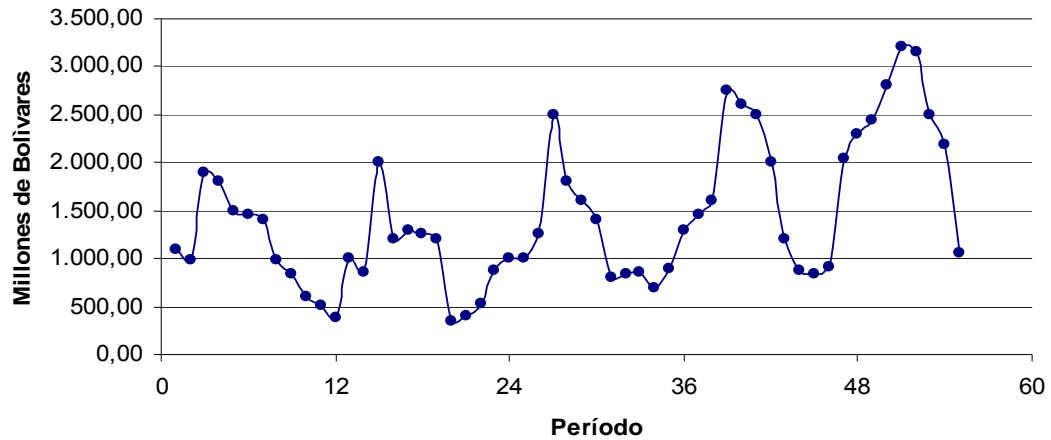
Son movimientos de la serie que se repiten de forma periódica. La razón de estas variaciones se basa en causas de tipo climatológico (producción, turismo, etc.) o de ordenación del tiempo (los días de la semana condiciona el comportamiento de ciertas series). La periodicidad generalmente es el año, aunque puede ser el mes, la semana o incluso el día.

**TABLA N° 2. Recaudación de Impuesto Sobre la Renta en Venezuela  
(Millones de bolívares) datos irreales**

	Años				
	2004	2005	2006	2007	2008
<b>Ene</b>	1.100,00	1.000,00	1.000,00	1.450,00	2.450,00
<b>Feb</b>	990,00	850,00	1.250,00	1.600,00	2.800,00
<b>Mar</b>	1.900,00	2.000,00	2.500,00	2.750,00	3.200,00
<b>Abr</b>	1.800,00	1.200,00	1.800,00	2.600,00	3.150,00
<b>May</b>	1.500,00	1.300,00	1.600,00	2.500,00	2.500,00
<b>Jun</b>	1.450,00	1.250,00	1.400,00	2.000,00	2.180,00
<b>Jul</b>	1.400,00	1.200,00	800,00	1.200,00	1.050,00
<b>Ago</b>	980,00	350,00	830,00	870,00	
<b>Sep</b>	830,00	400,00	850,00	840,00	
<b>Oct</b>	600,00	530,00	700,00	920,00	
<b>Nov</b>	510,00	870,00	900,00	2.050,00	
<b>Dic</b>	380,00	1.000,00	1.300,00	2.300,00	

**Fuente:** Los autores.

La Figura N° 2, muestra que cada año, durante el primer trimestre, se registra la mayor recaudación de Impuesto Sobre la Renta en Venezuela. Esto es motivado a que la fecha tope de declaración y pago de este tributo es el 31 de marzo de cada año. Cabe destacar, que no existe ninguna tendencia evidente en la recaudación de este tributo.

**FIGURA N° 2. Componente Estacional**

**Fuente:** Los autores.

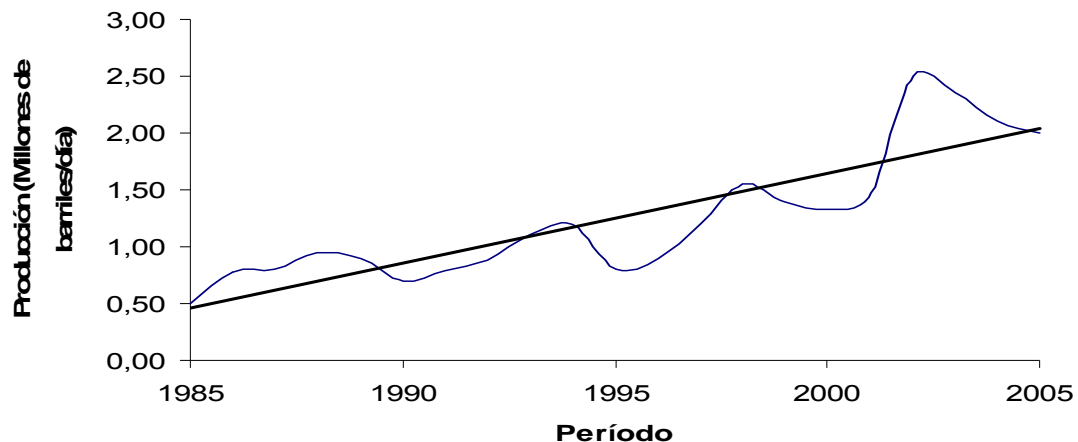
### 2.1.2.3. Variaciones Cíclicas

Esta componente tiene un marcado carácter económico, pues suele ser el resultado de la sucesión de las fases expansivas y recesivas de la economía. Son movimientos a plazo medio, periodos superiores al año, que se repiten de forma casi periódica, aunque no son tan regulares como las variaciones estacionales. Esta componente resulta difícil de aislar, pues ocurre, con frecuencia, que se pueden superponer ciclos de distintos periodos o amplitudes. La amplitud es el número de años que dura un ciclo completo. Para entender mejor la variación cíclica, presentamos el siguiente ejemplo en la Tabla N° 3 y la Figura N° 3.

**TABLA N° 3. Producción Petrolera**

<b>Años</b>	<b>Producción</b> (Millones de barriles/mes)
1985	0,500
1986	0,780
1987	0,800
1988	0,950
1989	0,900
1990	0,700
1991	0,790
1992	0,885
1993	1,100
1994	1,200
1995	0,800
1996	0,900
1997	1,200
1998	1,550
1999	1,400
2000	1,325
2001	1,430
2002	2,500
2003	2,350
2004	2,100
2005	2,000

**Fuente:** Los autores.

**FIGURA Nº 3. Variación Cíclica**

**Fuente:** Los autores.

#### 2.1.2.4. Variaciones Irregulares

Las series de tiempo también contienen variaciones irregulares o aleatorias, producidas por sucesos inusuales o impredecibles, que produce movimientos sin un patrón discernible. Son el resultado de factores fortuitos o aleatorios que inciden de forma aislada y no permanente en una serie. Tales movimientos son, como las huellas en los copos de nieve, únicos y probablemente no ocurrirán de nuevo de la misma manera. Pueden ser producidos por eventos tales como guerras, inundaciones, terremotos, elecciones políticas o embargos del petróleo.

Al analizar una serie de tiempo es necesario, entonces, tener en consideración el comportamiento de cada uno de estos componentes. Para ello el criterio más lógico a seguir es aislarlos secuencialmente partiendo de la serie original para luego analizarlos de manera individual. Si bien esto supone la utilización de métodos estadísticos adecuados, la mejor forma de



apreciarlos es a través de su observación visual.

## 2.2. Modelos De Series De Tiempo

La interacción de estas cuatro componentes genera la serie temporal. La forma en que se combinen puede ser muy variada, pero tradicionalmente se ha optado por dos modelos distintos. El aditivo y el multiplicativo, aunque en algunas ocasiones se mezclan ambos. Según el modelo que se adopte, la serie temporal será:  $Y_t = T_t + VE_t + C_t + I_t$  en el caso del modelo aditivo, y  $Y_t = (T_t).(VE_t).(C_t).(I_t)$  si se admite un esquema multiplicativo.

$Y_t$ : Es el valor de la serie de tiempo para el período  $t$

$T_t$ : Componente secular

$VE_t$ : Variación estacional

$C_t$ : Variación Cíclica

$I_t$ : Variación irregular

El **Modelo Aditivo** se puede entender en unidades absolutas en oposición a los porcentajes. Por ejemplo, un índice estacional para el consumo diario de agua en una ciudad puede tener un valor de 35 millones de litros por día en enero. Esto significa que la serie de consumo de agua para enero debería tender a superar en 35 millones de litros el valor pronosticado por la tendencia y el ciclo. Un modelo multiplicativo se puede convertir en un modelo aditivo tomando logaritmos (en cualquier base):

$$\log y_t = \log T_t + \log VE_t + \log C_t + \log I_t$$

El **Modelo multiplicativo** se puede entender en término de cambios porcentuales. Por ejemplo, un índice estacional para un mes cualquiera igual a 1,17 significa que la serie tiende a exceder en un 17% el valor pronosticado únicamente con la tendencia y el ciclo.

Se recomienda el modelo aditivo cuando la ecuación de tendencia es lineal, mientras que se recomienda el modelo multiplicativo cuando la ecuación de tendencia sea exponencial.

### **2.3. Números Índice**

Al paso de los años los Números Índice han llegado a ser cada vez más importantes para la administración como indicadores de la cambiante actividad económica o de negocios; de hecho, su uso se ha convertido en el procedimiento de más amplia aceptación.

Los números índices, constituyen un sencillo artificio para comparar los términos de una o varias series cronológicas; considerando ésta última como una sucesión de observaciones de una variable tomada en instantes sucesivos.

En muchos problemas de Economía interesa combinar, mediante un promedio adecuadamente definido varios índices simples para obtener un índice con el que se trata de reflejar la evolución de una magnitud no fácil de definir concretamente, por ejemplo: coste de vida, nivel de salarios, comercio exterior, etc.

Estas son medidas estadísticas que se emplean frecuentemente para

mostrar los cambios que se ocasionan en una variable o grupo de variables que se relacionan entre si.

Es una medida estadística diseñada para poner de relieve cambios en una variable o en un grupo de variables relacionadas con respecto al tiempo, situación geográfica, ingreso o cualquier otra característica. Una colección de Números Índice para diferentes años, lugares, etc.; se llama a veces serie de índices.

Los Números Índices miden el tamaño o la magnitud de algún objeto en un punto determinado en el tiempo, como el porcentaje de una base o referencia en el pasado

En realidad, los números índices relacionan una o varias variables de un período dado con la misma variable o variables en otro período, llamado período base.

Los Números Índices son estadística descriptiva y como tal deben ser fáciles de comprender y evaluar. En consecuencia, las fórmulas de los Números Índices utilizan generalmente expresiones matemáticas relativamente sencillas y basadas en la aplicación de porcentajes.

### 2.3.1. El Uso de los Números Índice

En los estudios sobre las series de tiempo con frecuencia se comparan datos en un período con los datos de otro período. Tales comparaciones deben hacerse cuidadosamente, ya que las condiciones económicas van cambiando con el paso del tiempo. Tales cambios dificultan el análisis de los datos comerciales o la interpretación de las variables económicas. Las

comparaciones directas de un período de tiempo con el siguiente, con frecuencia se tornan engañosas.

El uso de números índice puede proporcionar, a quienes toman las decisiones, un panorama más preciso del comportamiento de las variables económicas a través del tiempo y hacer comparaciones a través de periodos más significativos. Un número índice relaciona un valor en un período de tiempo, denominado período de base, con un valor en otro período, denominado periodo de referencia (o actual).

### 2.3.2. Tipos de Números Índice

Los números índices son tal vez los resultados mejor conocidos del pensamiento estadístico.

#### **Índice Simple**

Un índice de precios simple caracteriza la relación entre el precio de un producto o servicio en un período base, con el precio del mismo producto o servicio en el período de referencia.

La función principal de un índice simple es transformar los valores absolutos de una variable en valores relativos, para facilitar la comparación de los cambios en la variable con respecto al tiempo.

Para efectuar este cálculo, se divide el precio del producto en el período de referencia (más actual) por su precio en el período base y se multiplica por cien.

$$\text{I.P.S.} = \frac{P_R}{P_B} * 100$$

I.P.S. = índice de precios simple

$P_R$  = Precio del período de referencia

$P_B$  = Precio en el período base

### Índice de Precios Agregativos

Con frecuencia se desea calcular un índice de precio para varios productos. Esto se denomina índice de precios agregativos. Las empresas que manufacturan dos o más artículos se interesan en los índices agregativos.

El índice agregativo se calcula sumando el precio de cada uno de los bienes en el año de referencia y dividiendo por la suma de tales precios en el año base. El resultado entonces se multiplica por 100.

$$\text{I.P.A.} = \frac{\sum P_R}{\sum P_B} * 100$$

I.P.A. = Índice de precios agregativos

$P_R$  = Precio del período de referencia

$P_B$  = Precio en el período base

### Índice de Precios Agregativos Ponderado

El índice de precios agregativos ponderado se calcula asignando ponderaciones diferentes a los precios individuales. Estas ponderaciones se establecen para tener en cuenta las cantidades vendidas de cada producto. Esto proporciona un reflejo más exacto del costo verdadero de la canasta típica del consumidor. Este tipo de índice puede calcularse a través de las 3 fórmulas siguientes:

$$\text{Índice de Laspeyres} = \frac{\sum(P_R * Q_B)}{\sum(P_B * Q_B)} * 100$$

El índice de Laspeyres utiliza las ponderaciones del período base (cantidades) en su cálculo. El fundamento es que tales cantidades NO cambiarán de un cálculo al siguiente, permitiendo por tanto, comparaciones más significativas con el tiempo. Este índice utiliza las cantidades vendidas en el período base como factor de ponderación.

Cabe destacar que para cada año calculado el denominador será siempre el mismo bajo este método.

$$\text{Índice de Paasche} = \frac{\sum(P_R * Q_R)}{\sum(P_B * Q_R)} * 100$$

El índice de Paasche utiliza como ponderaciones las cantidades vendidas en cada uno de los años de referencia. Esto tiene la ventaja de que el índice se basa en los patrones de comportamiento del consumidor corriente. A medida que ocurren variaciones en la tendencia de consumo,

estas variaciones se reflejan en este índice; el mismo, pierde relevancia cuando no son considerados bienes de poco consumo, ya que presenta una leve desviación.

El índice de Paasche utiliza como factor de ponderación las cantidades vendidas en el período de referencia.

$$\text{Índice Ideal de Fisher} = \sqrt{\text{Índice de Laspeyres} * \text{Índice de Paasche}}$$

El índice de Laspeyres tiende a sobreestimar el valor del índice y el de Paasche a subestimarlo; por tal razón el índice de Fisher constituye una mejor aproximación al verdadero índice, aunque su construcción es más difícil.

El índice de Laspeyres tiende a sobreponderar los bienes cuyos precios aumentan, debido que este incremento en el precio va acompañado de una reducción en la cantidad, que no se ve reflejado en el índice de Paasche el cual utiliza cantidades con base fija como ponderación. Por otra parte, el índice de Paasche tiende a sobreponderar los productos cuyos precios bajan. por lo tanto, el índice ideal de Fisher se sugiere para algunos casos.

### **Índices Específicos**

El uso de un índice específico depende de quién está calculándolo y qué factores tienen en cuenta en su formulación. Quizá la serie de índice más conocida es el Índice de Precios al Consumidor. Existen otros tales como: Índices de Precios al Productor, Índice de Paridad, Relación de

Intercambio, Deflación de las Ventas, el de Dow Jones y Standard & Poor`s.

### **Índices Nacional de Precios al Consumidor (INPC)**

Este índice es altamente útil para medir la inflación, medir los cambios en los valores monetarios eliminando el impacto de los cambios en el precio, así en un grado limitado sirviendo como índice del costo de vida.

El INPC mide los cambios conjuntos en los precios de una canasta fija de bienes y servicios en el mercado nacional.

El INPC en Venezuela es el resultado del convenio firmado por el BCV y el INE para obtener un Índice Nacional de Precios al Consumidor representativo del país.

### **2.4. Técnicas De Suavizamiento**

Una propuesta alternativa al análisis de series temporales es el suavizamiento, que pretende liberarse de la componente aleatoria o irregular de la serie, pero no se interesa en los detalles de la tendencia, los patrones estacionales o los efectos cíclicos.

La mayoría de los métodos de suavizamiento proporcionan pronósticos que son, en un sentido o en otro, promedios de valores pasados. Si los datos muestran una tendencia pronunciada, estos pronósticos tienden por lo mismo a estar retardados en relación con la tendencia. Además, estos métodos ignoran por lo general los factores estacionales.

Se analizarán dos métodos comunes de suavizamiento de datos de



series de tiempo, los cuales son Promedios Móviles ó Medias Móviles y Suavizamiento Exponencial:

#### 2.4.2. Promedios Móviles ó Medias Móviles

Un promedio móvil (PM) tendrá el efecto de “suavizamiento” de los datos, produciendo un movimiento con menos picos y valles. Se calcula promediando los valores en las series de tiempo sobre un número fijo de períodos. El mismo número de períodos se mantiene para cada promedio, eliminando la observación mas antigua y recogiendo la mas reciente. Este método incluye el cálculo de la mediana, en vez de la media, de los  $M$  valores más recientes de la serie temporal. En vista de que la mediana, a diferencia de la media, no se ve afectada por valores extremos. A continuación, mostramos los datos (Tabla N° 4) a los cuales se aplicará esta técnica.

**TABLA N° 4. Compras (Unidades)**

<b>Mes</b>	1	2	3	4	5	6	7	8	9	10	11	12
<b>Compras</b>	86	80	85	93	96	102	97	89	96	87	82	81
<b>Mes</b>	13	14	15	16	17	18	19	20	21	22	23	24
<b>Compras</b>	87	89	97	88	95	87	81	79	82	85	96	93
<b>Mes</b>	25	26	27	28	29	30	31	32	33	34	35	36
<b>Compras</b>	99	103	96	85	78	83	90	96	85	82	89	96

**Fuente:** Los autores.

Los siguientes cálculos corresponden a las predicciones de las medias móviles de 3 meses y de 5 meses, y en la Figura N° 4 de la parte inferior, se muestran estos valores contrastados con los valores reales.

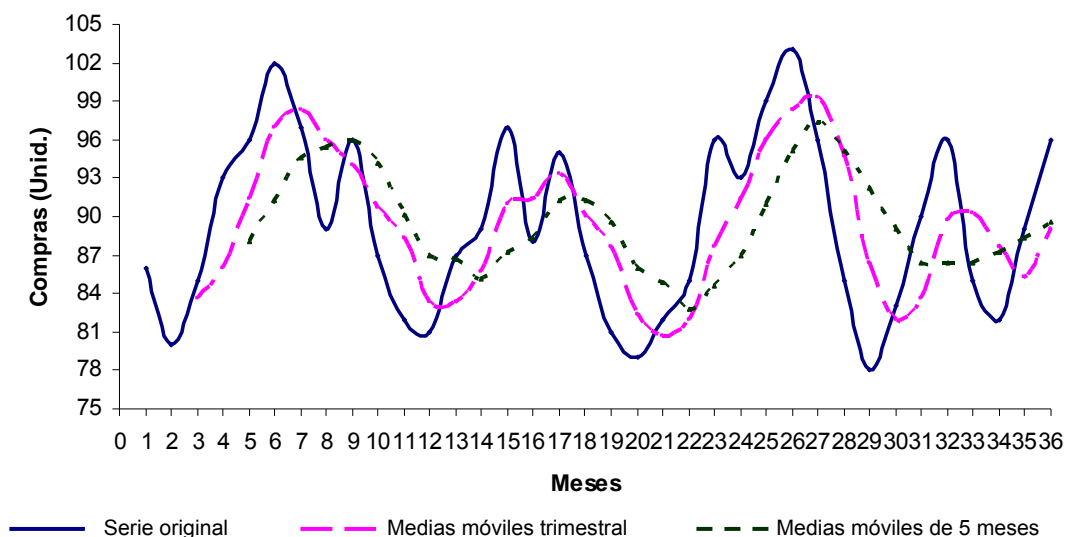
Media móvil trimestral. El valor de 83,7 correspondiente al 1er. trimestre se obtiene de dividir las compras de los tres primeros meses entre tres  $(86+80+85)/3$ .

		83,7	86,0	91,3	97,0	98,3	96,0	94,0	90,7	88,3	83,3
83,3	85,7	91,0	91,3	93,3	90,0	87,7	82,3	80,7	82,0	87,7	91,3
96,0	98,3	99,3	94,7	86,3	82,0	83,7	89,7	90,3	87,7	85,3	89,0

Media móvil de 5 meses. El valor de 88,0 correspondiente a los primeros 5 meses y se obtiene de dividir las compras de estos cinco primeros meses entre cinco  $(86+80+85+93+96)/5$ .

				88,0	91,2	94,6	95,4	96,0	94,2	90,2	87,0
86,6	85,2	87,2	88,4	91,2	91,2	89,6	86,0	84,8	82,8	84,6	87,0
91,0	95,2	97,4	95,2	92,2	89,0	86,4	86,4	86,4	87,2	88,4	89,6

**FIGURA Nº 4. Medias o Promedios Móviles (Suavizamiento)**



**Fuente:** Los autores.

### 2.4.3. Suavizamiento Exponencial

Como su nombre lo indica, el suavizamiento exponencial tiene el efecto de suavizar la serie. También proporciona un medio efectivo de predicción. Suavizamiento exponencial de primer orden se utiliza cuando los datos no presentan ningún patrón de tendencia. El modelo contiene un mecanismo de autocorrección que ajusta los pronósticos en dirección opuesta a los errores pasados.

Es una herramienta de proyección en la cual el pronóstico se basa en un promedio ponderado de los valores actuales y anteriores.

Suavizamiento Exponencial	$F_{t+1} = \alpha A_t + (1 - \alpha) F_t$
---------------------------	---

En donde  $F_{t+1}$  es el pronóstico para el siguiente período

$A_t$  es el valor real observado para el período corriente

$F_t$  es la proyección hecha previamente para el período corriente

$\alpha$  es una “constante de suavizamiento” a la cual se le asigna un valor entre 0 y 1.

Como ejemplo, se supone que es último día hábil del mes de octubre. Las ventas totales de autos del mes fueron de Bs. 110.000,00 se ha decidido pronosticar las ventas para el mes de noviembre.

Debido que antes no se han hecho predicciones, entonces  $F_t$  es desconocido. Lo usual es utilizar simplemente el valor real del período

anterior, septiembre en este caso, para la primera proyección. Los registros de la empresa muestran que las ventas de autos para el mes de septiembre fueron de Bs. 105.000,00 se asume un valor de 0,3 para  $\alpha$ , entonces el pronóstico para noviembre será:

$$F_{t+1} = \alpha A_t + (1 - \alpha) F_t$$

$$F_{t+1} = \alpha A_{\text{Oct}} + (1 - \alpha) F_{\text{Oct}}$$

$$F_{t+1} = (0,3)(110) + (0,7)(105)$$

$$F_{t+1} = \text{Bs. } 106.500,00 \text{ como ventas proyectadas para noviembre.}$$

**TABLA Nº 5. Ventas de Autos**

Mes	Proyección	Real	Error ( $F_t - A_t$ )
Septiembre	0	105.000,00	
Octubre	105.000,00	110.000,00	- 5.000,00
Noviembre	106.500,00	107.000,00	- 500,00
Diciembre	106.650,00	112.000,00	- 5.350,00

**Fuente:** Los autores.

### Cuadro Comparativo de los Métodos de Suavizamiento de las Series Temporales

Medias o Promedios Móviles	Exponencial
Otorgan el mismo peso a valores relativamente viejos y a los valores recientes.	Proporciona un mayor peso a los valores mas recientes.
Los pronósticos tienden a ser lentos en sus respuestas a las tendencias o ciclos de corta duración.	A menudo puede ser mas efectivo en su reacción frente a desplazamientos o ciclos en la serie.

## MÉTODOS DE ANÁLISIS DE SERIES TEMPORALES

Dentro de éstos, se conocen dos tipos, los cuales se explican detalladamente a continuación:

### 2.5.1. Análisis de Tendencia

El análisis de series temporales comprende métodos que ayudan a interpretar este tipo de datos, extrayendo información representativa, tanto referente a los orígenes o relaciones subyacentes como a la posibilidad de extrapolar y predecir su comportamiento futuro.

De hecho uno de los usos más habituales de las series de datos temporales es su análisis para predicción y pronóstico. Por ejemplo de los datos climáticos, o de las acciones de bolsa, o las series pluviométricas.

Si una serie de tiempo tiene una tendencia ascendente o descendente a largo plazo, el análisis de tendencia puede ser útil para desarrollar pronósticos. Es decir, si una tendencia está presente debido a que los datos no varían alrededor de algún promedio a largo plazo, los métodos de suavizamiento –como el promedio móvil y el suavizamiento exponencial– no son apropiados. En su lugar, se puede estimar una recta de tendencia, utilizando las técnicas para la regresión simple.

Veamos a continuación las fórmulas para llevar a cabo el análisis de tendencia.

Recta de tendencia

Utilizando Regresión Simple

$$Y_i = a + bx$$

Suma de cuadrados de X	$SC_x = \sum X^2 - \frac{(\sum X)^2}{n}$
------------------------	--

Suma de cuadrados de productos cruzados	$SC_{xy} = \sum XY - \frac{(\sum X)(\sum Y)}{n}$
--	--

Pendiente de la recta de tendencia	$(bx) = \frac{SC_{xy}}{SC_x}$
------------------------------------	-------------------------------

Intercepción de la recta de tendencia	$(a) = \bar{Y} - bx\bar{X}$
---------------------------------------	-----------------------------

A través del siguiente ejemplo, pondremos en práctica el análisis de tendencia.

“Vallas Publicidad”, hace publicidad de un nuevo producto. Para determinar la tendencia en el número de clientes, ésta consulta los registros de la compañía y encuentra los datos que aparecen a continuación. “Vallas” desea pronosticar el número de clientes para los períodos futuros.

Tomando como origen el mes de diciembre de 2006, para realizar el pronóstico del año 2008, específicamente el del mes de agosto (período 21) presentamos a continuación los datos de la Tabla N° 6.

**TABLA N° 6. Pronostico de Clientes**

Período (Mes)	t(X)	Clientes (Y)	X.Y	X <sup>2</sup>
---------------	------	--------------	-----	----------------

Enero 2007	1	41	41	1
Febrero	2	43	86	4
Marzo	3	39	117	9
Abril	4	37	148	16
Mayo	5	42	210	25
Junio	6	35	210	36
Julio	7	30	210	49
Agosto	8	31	248	64
Septiembre	9	32	288	81
Octubre	10	30	300	100
Noviembre	11	28	308	121
Diciembre	12	28	336	144
Enero 2008	13	29	377	169
Febrero	14	26	364	196
	<hr/>	<hr/>	<hr/>	<hr/>
	105	471	3.243	1.015

$$SC_x = 1.015 - \frac{(105)^2}{14} = 227,50$$

$$SC_{xy} = 3.243 - \frac{(105)(471)}{14} = -289,50$$

$$b_x = \frac{SC_{xy}}{SC_x} = \frac{-289,50}{227,50} = -1,27 \text{ pendiente de la recta}$$

$$a = \bar{Y} - b_x \bar{X} = 33,64 - (-1,27)(7,5) = 43,20$$

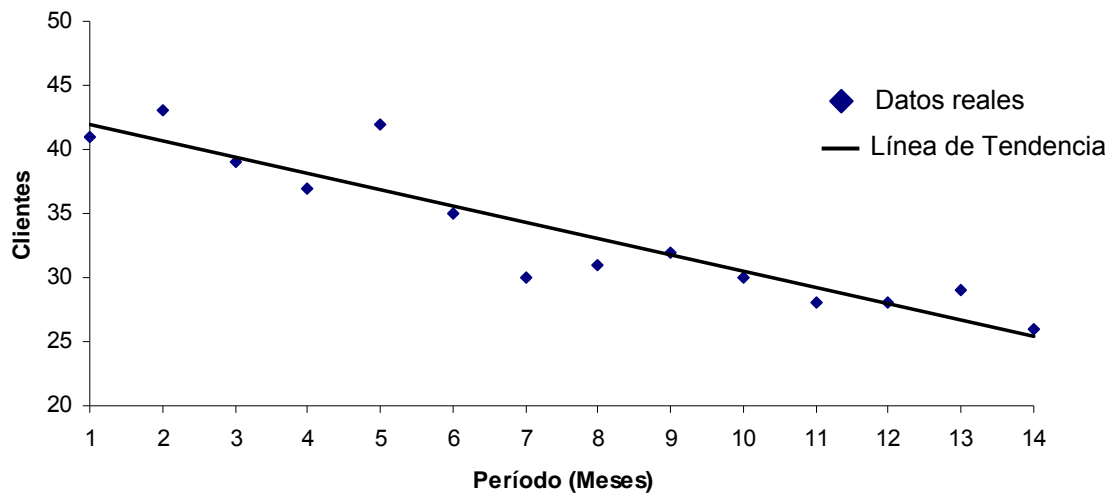
$$\hat{Y}_i = a + b_x X = 43,20 - 1,27 X_i \quad \text{Ecuación para la recta de tendencia}$$

Si Vallas Publicidad, desea pronosticar el número de clientes que puede conseguir su empresa en el mes de agosto (período 21) de 2008:

$$\hat{Y}_{Ago} = 43,20 - 1,27_{(21)} = 16,53 \approx 17 \text{ clientes}$$

El coeficiente negativo para  $b$  de  $-1,27$  indica a Vallas Publicidad que el negocio está descendiendo a una tasa de 1,27 clientes por cada período (mes).

**FIGURA Nº 5. Análisis de Tendencia**



**Fuente:** Los autores.



### 2.5.2. Descomposición de las Series de Tiempo

Con frecuencia es útil descomponer una serie de tiempo o una serie temporal “desglosando” cada uno de sus cuatro componentes. Así, cada componente se puede examinar individualmente. La tendencia histórica puede reflejar patrones anteriores de comportamiento, permitiendo ganar discernimiento en cuanto a los movimientos a largo plazo de las variables que se desean examinar. Esto permite el desarrollo de modelos de tendencia, útiles en la proyección y pronóstico. Al analizar el factor estacional se puede determinar si la actividad comercial presenta alguna variación de este tipo, que pueda considerarse para formular planes futuros. Por ejemplo, si se comercializan disfraces o algún otro producto estacional, se puede encontrar que las ventas se incrementan en el primer trimestre del año para nuestro país. Además, el desempeño cíclico puede también influir en la orientación de la planeación del negocio.

## CAPÍTULO III

### ANÁLISIS DE REGRESIONES

#### 3.1. Regresión

El primero en desarrollar el análisis de regresión fue el científico inglés Sir Francis Galton (1822 – 1911). Sus primeros experimentos con regresión comenzaron con un intento de analizar los patrones de crecimiento hereditarios de los guisantes. Animado por los resultados, Sir Francis extendió su estudio para incluir los patrones hereditarios en la estatura de las personas adultas. Descubrió que los niños que tienen padres altos o bajos tendían a “regresar” a la estatura promedio de la población adulta. Con este modesto inicio el uso del análisis de regresión se dió a conocer convirtiéndose en una de las herramientas estadísticas más poderosas que se encuentran disponibles actualmente.

En el análisis de regresión, los datos históricos de las variables relevantes se utilizan para desarrollar y evaluar una ecuación de predicción. La variable que se está prediciendo con la ecuación es la variable dependiente. Cualquier variable que se utilice para hacer la predicción es una variable independiente.

La idea básica del análisis de regresión es utilizar los datos de una variable cuantitativa independiente para predecir o explicar la variación de una variable cuantitativa dependiente. Podemos distinguir entre predicción (referencia a valores futuros) y explicación (referencia a valores actuales o pasados). Las virtudes de la retrospectiva indican que la explicación es más fácil que la predicción. No obstante, a menudo es más claro utilizar el término

“predicción” para incluir ambos casos.

Para que la predicción (o explicación) tenga sentido, debe haber alguna conexión entre la variable que estamos prediciendo (la variable dependiente) y la variable que estamos utilizando para hacer la predicción (la variable independiente). La predicción requiere de una unidad de asociación. Debe haber una entidad que relacione las dos variables. Para los datos relativos a las series temporales, la unidad de asociación puede ser simplemente el tiempo.

### **3.2. Objetivos Del Análisis De Regresión**

- 1) Los datos se pueden utilizar para obtener una ecuación de predicción.
- 2) Los datos se pueden utilizar para estimar la magnitud de la variabilidad o incertidumbre en torno a la ecuación.
- 3) Como los datos son solo una muestra, se pueden llevar a cabo inferencia acerca de los verdaderos valores poblacionales de los parámetros considerados en la regresión.
- 4) La ecuación de predicción se puede utilizar para predecir un rango razonable de valores futuros de la variable dependiente.
- 5) Los datos se pueden utilizar para estimar el grado de correlación entre las variables dependiente e independiente, el cual indica el grado de la relación entre ellas.

### **3.3. Regresión Simple**

En la regresión simple, se establece que  $Y$  es una función de solo una variable independiente. Con frecuencia se le denomina regresión bivariada

porque solo hay dos variables, y esta se representa con la fórmula:

Y es una función de X	$Y = f(x)$
-----------------------	------------

**"Y está regresando por X"**

En el modelo de regresión es muy importante identificar cuál es la variable dependiente y cuál es la variable independiente.

La variable dependiente es la variable que se desea explicar, predecir. También se le llama regresando ó variable de respuesta.

La variable independiente **X** se le denomina variable explicativa ó regresor y se le utiliza para explicar **Y**.

El estudio de la relación funcional entre dos variables poblacionales, presenta la siguiente notación:

$Y_i = a + bx + e$
--------------------

“**a**” es el valor de la ordenada donde la línea de regresión se intercepta con el eje **Y**. “**b**” es el coeficiente de regresión poblacional (pendiente de la línea recta). “**e**” es el error probabilístico.

**Y<sub>i</sub>** = Valores de la variable dependiente.

**x** = Valores de la variable independiente.

**e** = Errores aleatorios no observables.

### 3.4. Modelos De Regresión Simple

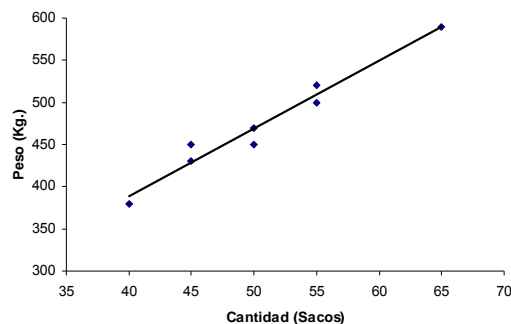
Un modelo de regresión es un medio formal de expresar dos de los ingredientes esenciales de una relación estadística:

- 1) La tendencia de la variable dependiente ( $Y_i$ ) a variar con la(s) variable(s) independiente(s) en forma sistemática, como las descritas por una línea o curva que muestre una relación estadística.
- 2) La dispersión de las observaciones en el entorno de la curva o de la línea que muestra la relación estadística.

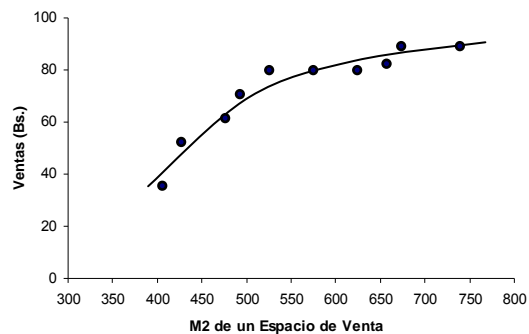
Estudiaremos dos tipos de modelos de regresión bivariantes:

- a) **Lineal**. Si hay sólo una variable independiente y la función de regresión es lineal (Ver Figura N° 6).
- b) **No Lineal**. Si hay sólo una variable independiente y la función de regresión no es lineal (Ver Figura N° 7).

**FIGURA N° 6. Relación Estadística Lineal**



**Fuente:** Los autores.

**FIGURA N° 7. Relación Estadística No Lineal**

**Fuente:** Los autores.

El modelo (a) se dice que es simple, y es lineal en la variable independiente y en los parámetros. Simple cuando hay una variable independiente, lineal en la variable independiente porque esta variable tiene una potencia de orden uno, y lineal en los parámetros, porque ningún parámetro aparece como potencia o es multiplicado o dividido por otro parámetro. Para que exista la linealidad en el modelo se requiere que sea lineal en la(s) variable(s) independiente(s) y lineal en los parámetros. Los modelos no lineales en su forma externa, lo son en su forma intrínseca. La linealidad intrínseca se obtiene a través de una transformación.

### 3.5. Determinación Del Modelo De Regresión

Si el investigador quiere conocer cual es el modelo que mejor se ajusta a una colección de datos, debe proceder de la siguiente manera:

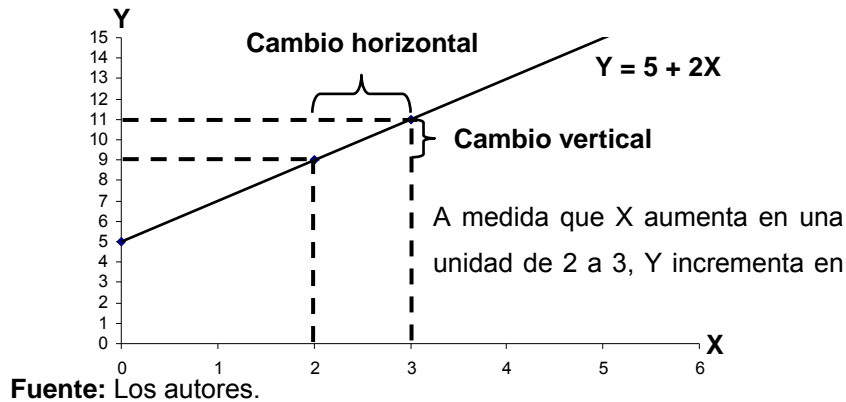
1.– Representar los datos en un diagrama de dispersión para conocer que tipo de función se ajusta mejor a los datos.

2.– Este primer diagrama de dispersión se hace en una escala aritmética. Si resulta una curva y no una recta, se grafica luego en una escala semilogarítmica o logarítmica, para de ese modo obtener una recta, en cualquiera de esta dos escalas. Obtenida la recta por el método gráfico, en cualquiera de las escalas usadas, estamos ante el modelo general de regresión que mejor se ajusta a los datos. Estos modelos son lineales o no lineales, como anteriormente se señaló. Pero algunos modelos no lineales son intrínsecamente lineales.

Partiendo de la ecuación de la recta,  $Y_i = a + bx$ , en donde “ $a$ ” es el intercepto y “ $bx$ ” es la pendiente de la recta. Por ejemplo se tiene que  $Y = 5 + 2X$  entonces, como se observa en la Figura N° 8, la recta intercepta el eje vertical en 5. Además, la pendiente de la recta se halla:

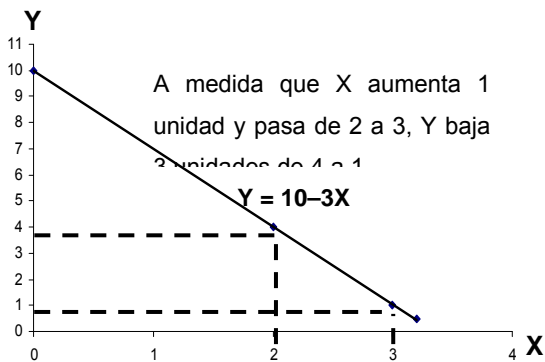
$$b_1 = \frac{\Delta vertical}{\Delta horizontal} = \frac{2}{1} = 2$$

Por cada cambio de una unidad en  $X$ ,  $Y$  cambia en dos unidades.

**FIGURA N° 8. Línea Recta con Pendiente Positiva**

Veamos ahora la pendiente menor que cero,  $b_1 < 0$ , por ejemplo: (ver Figura N° 9A)  $Y = 10 - 3X$ .

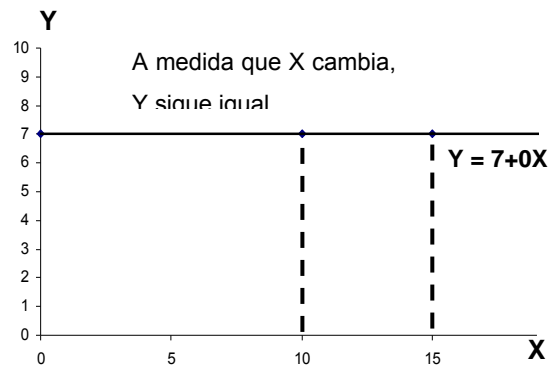
Revela que existe una relación negativa entre “X” e “Y” que por cada incremento (reducción) de una unidad en “X”, “Y” reducirá (aumentará) en 3 unidades si la pendiente de la recta es  $b_1 = 0$  (ver Figura N° 9B), entonces un cambio de “X” no tiene relación con un cambio en “Y”. Por tanto, “X” no puede utilizarse como variable explicativa de “Y”.

**FIGURA N° 9A. Línea Recta con Pendiente Negativa**

**Fuente:** Los autores.



**FIGURA N° 9B. No Existe Ninguna Relación entre X e Y**



**Fuente:** Los autores.

Las relaciones entre variables son o **determinísticas** o **estocásticas** (aleatorias). Una relación determinística puede expresarse mediante la fórmula que convierte la velocidad expresada en m/h a km/h. ya que una milla es aproximadamente 1,6 Km, este modelo es  $1 \text{ m/h} = 1,6 \text{ Km/h}$ . Por tanto, una velocidad de  $5 \text{ m/h} = 5(1,6) \text{ km/h} = 8,0 \text{ km/h}$ . Este es un modelo determinístico porque la relación es exacta y no hay error (salvo la aproximación).

Infortunadamente, muy pocas relaciones en el mundo de los negocios son así de exactas. Con frecuencia se encuentra que al utilizar una variable para explicar otra, existe alguna variación en la relación. Por ejemplo, se supone que la gerencia de Cumaná, Inc., distribuidores de productos para la salud, desea desarrollar un modelo de regresión en el cual se utiliza la publicidad para explicar los ingresos por concepto de ventas. Probablemente encontrarán que cuando hacen publicidad y ésta se fija en cierta cantidad  $X_i$ , las ventas tendrán algún valor  $Y_i$ . Sin embargo, la próxima vez que se fije la publicidad en la misma cantidad, las ventas pueden producir otro valor. La

variable dependiente (ventas, en este caso) presenta algún grado de aleatoriedad. Por tanto, habrá algún *error* en el intento por explicar o predecir las ventas. Se dice que un modelo de esta naturaleza es estocástico, por la presencia de la variación aleatoria y puede expresarse como:

**Un modelo lineal**

$$Y_i = A + Bx + \mathcal{E}$$

La fórmula anterior, es la relación poblacional (o verdadera) según la cual se hace regresión de **Y** sobre **X**.

Además, **A+Bx** es la porción determinística de la relación, mientras que  $\mathcal{E}$  representa el carácter aleatorio que muestra la variable dependiente y por tanto denota el término del error en la expresión.

Los parámetros “**A**” y “**B**”, lo mismo que la mayoría de los parámetros, permanecerán desconocidos y se pueden estimar sólo con los datos muestrales. Esto se expresa así:

**Un modelo lineal**

**Con base en datos muestrales**

$$Y = a + bx + e$$

en donde los valores “**a**” y “**b**” son estimaciones de “**A**” y “**B**”, respectivamente, y “**e**” es el término aleatorio. Habitualmente, se le denomina *residual* cuando se utilizan datos muestrales, **e** reconoce que no todas las observaciones caen exactamente en una línea recta. Si se supiera el valor exacto de “**e**”, se podría calcular de manera precisa “**Y**”. Sin embargo, debido a que “**e**” es aleatoria, “**Y**” sólo puede estimarse. El modelo de regresión por ende toma la forma de:

Un Modelo de	$\hat{Y} = a + bx$
Regresión Estimada	

en donde  $\hat{Y}$  es el valor *estimado* de “ $Y$ ”, y “ $a$ ” y “ $b$ ” son el intercepto y la pendiente de la recta de regresión estimada. Es decir,  $\hat{Y}$  simplemente es el valor estimado para las ventas con base en el modelo de regresión.

### 3.6. Método De Mínimos Cuadrados

Se basa en la propiedad que dice: “La suma de las desviaciones cuadráticas de las observaciones, con relación a la media aritmética es un mínimo”  $\sum (X_i - \bar{X})^2 = \text{mínimo}$ . Obsérvese en las Figuras N° 6 y N° 7 que el modelo de estimación no da una relación perfecta entre las observaciones y la recta ajustada, sino que esta recta viene a ser una media de la relación entre las variables, pues tiene observaciones por encima y por debajo de la misma. La diferencia entre los valores observados y los valores estimados se denomina desviaciones. La suma cuadrática de estas desviaciones con relación al modelo de estimación será mínimo por la propiedad antes citada; y por ello el modelo a obtener será el modelo que mejor se ajusta a las observaciones.

El problema del análisis de regresión es encontrar la mejor recta de predicción. El criterio para decidir cual es “la mejor” se basa en el cuadrado del error de predicción. Lo que hacemos es encontrar la ecuación de la recta de predicción, es decir la pendiente “ $b$ ” y la ordenada en el origen “ $a$ ”, que

minimice el total de los cuadrados de los errores de predicción. El método que logra esta meta se conoce como el método de los Mínimos Cuadrados, porque selecciona los valores “**a**” y “**b**” que minimizan la cantidad. La recta resultante presenta dos características importantes:

1. Es nula la suma de las desviaciones verticales de los puntos a partir de la recta de ajuste  $\sum (Y_r - Y) = 0$ .
2. Es mínima la suma de los cuadrados de dichas desviaciones. Ninguna otra recta daría una suma menor de las desviaciones elevadas al cuadrado  $\sum (Y_r - Y)^2 \rightarrow 0$  (mínima).

El procedimiento consiste entonces en minimizar los residuos al cuadrado  $C_i^2$

$$\sum C_i^2 = \sum (Y_i - \hat{Y})^2 \text{ reemplazando } \hat{Y} \text{ nos queda}$$

$$\sum C_i^2 = \sum [Y_i - (a+bx)]^2$$

La obtención de los valores de “**a**” y “**b**” que minimizan esta función es un problema que se puede resolver recurriendo a la derivación parcial de la función en términos de “**a**” y “**b**”. Ilamemos “**G**” a la función que se va a minimizar:

$$G = \sum (y - a - bx)^2$$

Tomemos las derivadas parciales de “**G**” respecto de “**a**” y “**b**” que son

las incógnitas y las igualamos a cero; de esta forma se obtienen dos ecuaciones llamadas ecuaciones normales del modelo que pueden ser resueltas por cualquier método ya sea igualación o matrices para obtener los valores de “**a**” y “**b**”.

$$G = \sum (y - a - bx)^2$$

Derivamos parcialmente la ecuación respecto de “**a**” y obtenemos nuestra 1ra. ecuación normal:

$$\sum y = n \cdot a + b \sum x$$

Derivamos parcialmente la ecuación respecto de “**b**” obtenemos nuestra 2da. ecuación normal:

$$\sum xy = a \sum x + b \sum x^2$$

Los valores de “**a**” y “**b**” se obtienen resolviendo el sistema de ecuaciones resultante. Veamos el siguiente ejemplo:

En un estudio económico se desea saber la relación entre el nivel de instrucción de las personas y el ingreso.

Se toma una muestra aleatoria de 8 ciudades de una región geográfica de 13 departamentos y se determina por los datos del censo el porcentaje de graduados en educación superior y la mediana del ingreso de cada ciudad, los resultados son los siguientes:

Ciudad	1	2	3	4	5	6	7	8
% de (X) Graduados:	7,2	6,7	17,0	12,5	6,3	23,9	6,0	10,2
Ingreso (Y) Mediana:	4,2	4,9	7,0	6,2	3,8	7,6	4,4	5,4

Tenemos las ecuaciones normales:

$$\sum y = n.a + b \sum x$$

$$\sum xy = a \sum x + b \sum x^2$$

Debemos encontrar los términos de las ecuaciones  $\sum y$ ,  $\sum x$ ,  $\sum x.y$ ,  $\sum x^2$ , por tanto, procedemos de la siguiente forma empleado el método de igualación:

Y	X	X.Y	X <sup>2</sup>
4,2	7,2	30,24	51,84
4,9	6,7	32,83	44,89
7,0	17,0	119	289
6,2	12,5	77,5	156,25
3,8	6,3	23,94	39,69
7,6	23,9	181,64	571,21
4,4	6,0	26,4	36
5,4	10,2	55,08	104,04
<b>43,50</b>	<b>89,80</b>	<b>546,63</b>	<b>1.292,92</b>

Sustituyendo en las ecuaciones los resultados obtenidos tenemos:

$$43,50 = 8a + 89,80b$$

$$546,63 = 89,80a + 1.292,92b$$

Multiplicamos la primera ecuación por  $(-89,80)$  y la segunda por  $(8)$  así:

$$43,50 = 8a + 89,80b \quad (-89,8)$$

$$546,63 = 89,80a + 1.292,92b \quad (8)$$

$$-3.906,30 = -718,40a - 8.064,04b$$

$$4.373,04 = 718,40a + 10.343,36b$$

$$466,74 = -0 - 2.279,32b$$

$$\mathbf{b = \frac{466,74}{2.279,32} = 0,20477}$$

Este valor de "**b**" lo reemplazamos en cualquiera de las ecuaciones para obtener "**a**" así:

Reemplazando "**b**" = 0,20477 en la primera ecuación normal

$$43,50 = 8a + 89,80(0,20477)$$

$$43,50 = 8a + 18,3880$$

$$43,50 - 18,3880 = 8a$$

$$25,1120 = 8a$$

$$\mathbf{a = \frac{25,120}{8} = 3,139}$$

Tenemos entonces que los coeficientes de regresión son: **a** = 3,139 y **b** = 0,20477. Por tanto, la ecuación de regresión nos queda:

$$\hat{Y} = 3,1390 + 0,20477x$$

Significa entonces que por cada incremento en una unidad en “**x**” el valor de  $\hat{Y}$  se aumenta en 0,20477.

Esta ecuación permite estimar el valor de  $\hat{Y}$  para cualquier valor de “**x**”, por ejemplo: Una ciudad que tiene un porcentaje de graduados a nivel superior del 28% la mediana de ingreso para la ciudad será:

$$\hat{Y} = 3,1390 + 0,20477(28)$$

$$\hat{Y} = 8,87 \text{ (Bs. 88.700,00)}$$

Los valores “**a**” y “**b**” también se pueden obtener de la siguiente forma: partiendo de las ecuaciones normales:

Si dividimos todos los términos de la ecuación (1) entre “**n**” nos queda:

$$\frac{\sum y}{n} = \frac{n.a}{n} + \frac{b \sum x}{n}$$

Tenemos entonces que el primer término es  $\bar{Y}$  el segundo término es la incógnita “**a**” y el tercer término es la incógnita “**b**” multiplicada por  $\bar{X}$  por tanto nos queda:

$$\bar{Y} = a + b. \bar{X} , \text{ entonces, } a = \bar{Y} - b. \bar{X}$$



“**b**” se puede obtener a través de la fórmula siguiente:

$$\mathbf{b} = \frac{\sum xy - n\bar{Y}\bar{X}}{\sum x^2 - n\bar{X}^2}$$

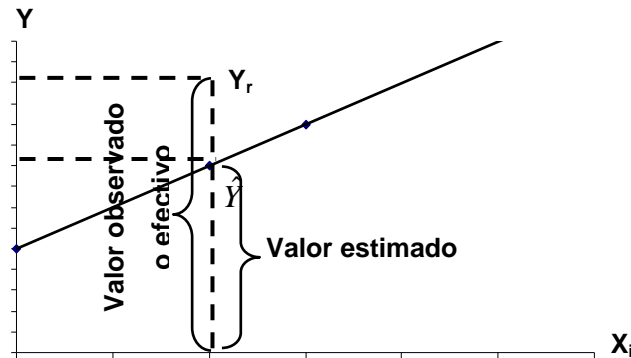
$$\mathbf{b} = \frac{546,63 - 8(5,4375)(11,2250)}{1.292,92 - 8(11,2250)^2} = \frac{58,3425}{284,9150} = 0,20477$$

$$\mathbf{a} = 5,4375 - 0,20477(11,2250) = 5,4375 - 2,2985 = 3,139$$

Se debe tener presente la diferencia entre el valor de  $\hat{Y}$  obtenido con la ecuación de regresión y el valor de “**Y**” observado. Mientras  $\hat{Y}$  es una estimación y su bondad en la estimación depende de lo estrecha que sea la relación entre las dos variables que se estudian;  $Y_r$  es el valor efectivo, verdadero obtenido mediante la observación del investigador. En el ejemplo  $Y_r$  es el valor mediano del ingreso que obtuvo el investigador utilizando todos los ingresos observados en cada ciudad y  $\hat{Y}$  es el valor estimado con base en el modelo lineal utilizado para obtener la ecuación de regresión.

Los valores estimados y observados pueden no ser iguales por ejemplo la primera ciudad tiene un ingreso mediano observado de  $Y_r = 4,2$  al reemplazar en la ecuación el porcentaje de graduados obtenemos un  $\hat{Y}$  estimado de  $\hat{Y} = 3,1390 + 0,20477(7,2) = 4,61$ , gráficamente, lo anterior se puede mostrar así:

**FIGURA N° 10. Valores Observados y Estimados**



**Fuente:** Los autores.

Claramente se observa en la gráfica que hay una diferencia entre el valor efectivo de  $Y_r$  y el valor estimado; esta diferencia se conoce como error en la estimación, este error se puede medir. A continuación se verá el procedimiento.

### 3.7. Error Estándar En La Estimación

El error estándar de la estimación designado por  $S_{YX}$  mide la diferencia "promedio" entre los valores observados y los valores estimados o esperados de "Y" de acuerdo al modelo, puede considerarse como un indicador del grado de precisión con que la ecuación de regresión, describe la relación entre las dos variables. Este error estándar se ve afectado por las unidades y sus cambios ya que es una medida absoluta, pues, se da en la misma unidad de medida que esta dada la variable "Y". Para su determinación se utiliza la siguiente fórmula:

$$S_{yx} = \sqrt{\frac{\sum (Y_i - \hat{Y})^2}{n - 2}}$$

Es importante recalcar, que la fórmula anterior es aplicable a las muestras, ya que en un elevado porcentaje de los casos, el total de los elementos de la población (o sus parámetros) son de difícil determinación. Además, podemos decir que en el denominador el número 2, indica los estadísticos utilizados para llevar a cabo la estimación de los parámetros.

Debemos entonces calcular los valores de  $\hat{Y}$  para cada ciudad sustituyendo en la ecuación los valores de los porcentajes de graduados de cada ciudad estudiada.

$$Y = 3,1390 + 0,20477(x)$$

Y	X	$\hat{Y}$	$Y_i - \hat{Y}$	$(Y_i - \hat{Y})^2$
4,2	7,2	4,6	-0,4	0,17
4,9	6,7	4,5	0,4	0,15
7,0	17,0	6,6	0,4	0,14
6,2	12,5	5,7	0,5	0,25
3,8	6,3	4,4	-0,6	0,40
7,6	23,9	8,0	-0,4	0,19
4,4	6,0	4,4	0,0	0,00
5,4	10,2	5,2	0,2	0,03
				<b>1,33</b>

$$S_{yx} = \sqrt{\frac{\sum (Y_i - \hat{Y})^2}{n - 2}} = \sqrt{\frac{1,33}{8 - 2}} = \sqrt{\frac{1,33}{6}} = \sqrt{0,222}$$

$$S_{yx} = 0,47 \text{ (Bs. 4.700,00)}$$

Como esta medida trata de resumir la disparidad entre lo observado y lo estimado, es decir, trata de medir la diferencia promedio entre lo observado y lo estimado ó esperado de acuerdo al modelo, puede considerarse como un indicador del grado de precisión con que la ecuación de regresión, describe la relación entre las dos variables. Este error estándar se ve afectado por las unidades y sus cambios ya que es una medida absoluta, pues, se da en la misma unidad de medida que esta dada la variable "Y"; en el ejemplo 0,46 serán decenas de miles de bolívares, razón por la cual no es posible comparar con las relaciones de variables dadas en distinta unidad de medida. Es necesario entonces calcular una medida que interprete o mida mejor el grado de relación entre las variables.

### **3.8. Análisis De Correlación**

Cuando se analizan datos, el interés del analista suele centrarse en dos grandes objetivos: comparar grupos y estudiar relaciones.

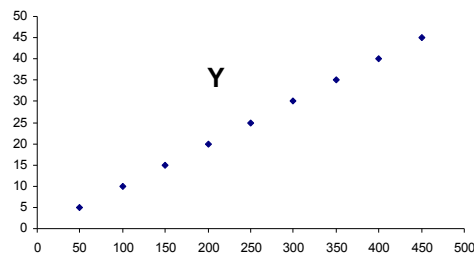
Suele decirse que los sujetos más frustrados son también más agresivos; que cuanto mayor es el nivel educativo, mayor es el nivel de renta; que los niveles altos de colesterol en sangre suelen ir acompañados de dietas alimenticias ricas en grasas; que los sujetos muestran más interés por una tarea cuanto mayor es el tamaño de la recompensa que reciben; etc. En todos los ejemplos mencionados se habla de la relación entre dos variables. En este capítulo se estudian algunos índices estadísticos que permiten cuantificar el grado de relación existente entre dos variables.

### 3.8.1. Correlación Lineal Simple

El concepto de relación o correlación se refiere al grado de variación conjunta existente entre dos o más variables. En este apartado nos vamos a centrar en el estudio de un tipo particular de relación llamada lineal y nos vamos a limitar a considerar únicamente dos variables (simple). Una relación lineal positiva entre dos variables  $X_i$  e  $Y_i$  indica que los valores de las dos variables varían de forma parecida: los sujetos que puntúan alto en  $X_i$ , tienden a puntuar alto en  $Y_i$ , y los que puntúan bajo en  $X_i$ , tienden a puntuar bajo en  $Y_i$ . Una relación lineal negativa significa que los valores de las dos variables varían justamente al revés: los sujetos que puntúan alto en  $X_i$ , tienden a puntuar bajo en  $Y_i$ , y los que puntúan bajo en  $X_i$ , tienden a puntuar alto en  $Y_i$ .

La forma más directa e intuitiva de formarnos una primera impresión sobre el tipo de relación existente entre dos variables es a través de un diagrama de dispersión. Un diagrama de dispersión es un gráfico en el que una de las variables ( $X_i$ ) se coloca en el eje de abscisas, la otra ( $Y_i$ ) en el de las ordenadas y los pares  $(x_i, y_i)$  se representan como una nube de puntos. La forma de la nube de puntos nos informa sobre el tipo de relación existente entre las variables.

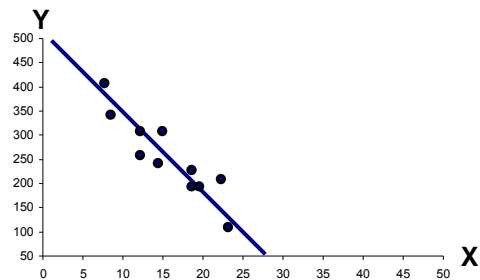
**FIGURA Nº 11A. Relación Lineal Positiva Perfecta**



Fuente: Los autores.

La figura superior (11A) muestra una situación en la que cuanto mayores son los valores en una de las variables, mayores son también los valores en la otra; cuando ocurre ésto, los puntos se sitúan en una línea recta ascendente y hablamos de relación lineal positiva.

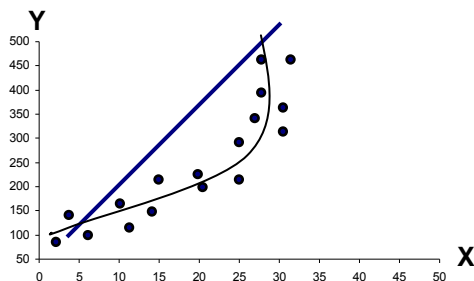
**FIGURA Nº 11B. Relación Lineal Negativa**



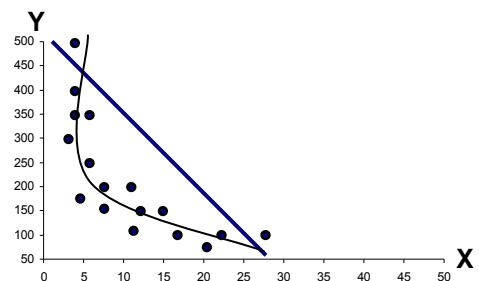
**Fuente:** Los autores.

La figura 11B (arriba), representa una situación en la que cuanto mayores son los valores en una de las variables, menores son los valores en la otra; en este caso, los puntos se sitúan en una línea recta descendente y hablamos de relación lineal negativa.

**FIGURA Nº 11C. Relación Exponencial o Curvilínea**



**Fuente:** Los autores.

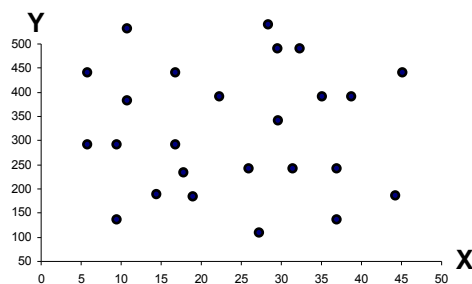


**Fuente:** Los autores.

En la situación representada en la Figura N° 11C de la parte superior, también existe una pauta de variación clara, pero no es lineal: los puntos no dibujan una línea recta.

En la Figura N° 11D de la parte inferior, no parece existir ninguna pauta de variación clara, lo cual queda reflejado en una nube de puntos dispersa, muy lejos de lo que podría ser una línea recta.

**FIGURA N° 11D. No Existe Relación**



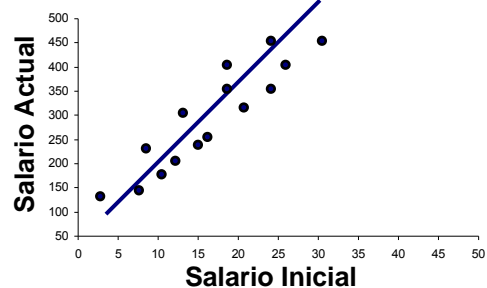
**Fuente:** Los autores.

Vemos, pues, que un diagrama de dispersión nos permite formarnos una idea bastante aproximada sobre el tipo de relación existente entre dos variables. Pero, además, observando los diagramas de las figuras 11A–11D, podemos ver que un diagrama de dispersión también puede utilizarse como una forma de cuantificar el grado de relación lineal existente entre dos variables: basta con observar el grado en el que la nube de puntos se ajusta a una línea recta.

Sin embargo, utilizar un diagrama de dispersión como una forma de cuantificar la relación entre dos variables no es, en la práctica, tan útil como puede parecer a primera vista. Esto es debido a que la relación entre dos

variables no siempre es perfecta o nula: habitualmente no es ni lo uno ni lo otro. Consideremos los diagramas de dispersión de la parte de abajo, la **Figura N° 11E**, los puntos, aun no estando situados todos ellos una línea recta, se aproximan bastante a ella. Podríamos encontrar una línea recta ascendente que representara de forma bastante aproximada el conjunto total de los puntos del diagrama, lo cual indica que la relación entre las variables salario inicial y salario actual es lineal y positiva: a mayor salario inicial, mayor salario actual.

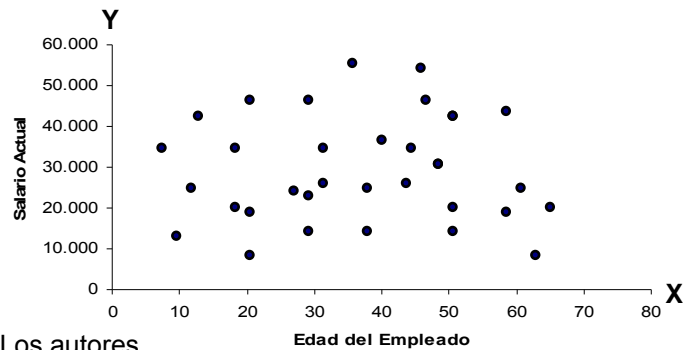
**FIGURA N° 11E. Relación Lineal Positiva**



**Fuente:** Los autores.

En el diagrama siguiente (**Figura N° 11F**), por el contrario, da la impresión de que no hay forma de encontrar una recta a la que poder aproximar los puntos. Al margen de que entre las variables edad y salario actual pueda existir algún tipo de relación, parece claro que la relación no es de tipo lineal.



**FIGURA Nº 11F. Independencia Lineal**

**Fuente:** Los autores.

Estas consideraciones sugieren que hay nubes de puntos a las que es posible ajustar una línea recta mejor de lo que es posible hacerlo a otras. Por lo que el ajuste de una recta a una nube de puntos no parece una cuestión de todo o nada, sino más bien de grado (más o menos ajuste).

Lo cual nos advierte sobre la necesidad de utilizar algún índice numérico capaz de cuantificar ese grado de ajuste con mayor precisión de lo que nos permite hacerlo una simple inspección del diagrama de dispersión.

Estos índices numéricos suelen denominarse coeficientes de correlación y poseen la importante propiedad de permitir cuantificar el grado de relación lineal existente entre dos variables cuantitativas. Por supuesto, al mismo tiempo que permiten cuantificar el grado de relación lineal existente entre dos variables, también sirven para valorar el grado de ajuste de la nube de puntos a una línea recta.

### 3.8.2. Coeficientes de Correlación

Este coeficiente como ya se dijo mide la fuerza de la relación entre las variables. El coeficiente tiene el signo que tiene “**b**” y su valor estará  $-1 \leq r \leq 1$ . El signo menos en el índice significa una relación negativa y un signo más una correlación positiva. El coeficiente se obtiene sacando la raíz cuadrada al coeficiente de determinación y se simboliza con “**r**”. Veamos la fórmula del método directo a continuación:

$$r = \sqrt{\frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y_r - \bar{Y})^2}}$$

$$r = \sqrt{\frac{11,78}{13,2710}} = \sqrt{0,8876} = 0,9421$$

En este caso el coeficiente “**r**” tiene signo positivo ya que toma el valor de “**b**” obtenido con las ecuaciones normales.

A continuación se da, a modo de orientación, como podrían interpretarse los valores de “**r**” (positivo o negativo).

De 0,0 a 0,2	Correlación muy débil, despreciable.
De 0,2 a 0,4	Correlación débil. Bajo.
De 0,4 a 0,7	Correlación moderada.
De 0,7 a 0,9	Correlación fuerte, alto, importante.
De 0,9 a 1,0	Correlación muy fuerte, muy alto.

La correlación entre los valores de dos variables es un hecho. El que lo consideremos satisfactorio o no, depende de la interpretación. Otro problema que representa la correlación es cuando se pregunta si una variable, de algún modo causa o determina a la otra. La correlación no implica causalidad. Si las variables “*x*” e “*y*” están correlacionadas, ésto puede ser por que “*x*” causa a “*y*”, o porque “*y*” causa a “*x*” o porque alguna otra variable afecta tanto a “*x*” como a “*y*”, o por una combinación de todas estas razones; o puede ser que la relación sea una coincidencia.

Pueden seleccionarse uno o más de los siguientes tres coeficientes de correlación:

### 3.8.3. Pearson

El coeficiente de correlación de Pearson (1896) es, quizá, el mejor coeficiente y el más utilizado para estudiar el grado de relación lineal existente entre dos variables cuantitativas. Se suele representar por “*r*” y se obtiene tipificando el promedio de los productos de las puntuaciones diferenciales de cada caso (desviaciones de la media) en las dos variables correlacionadas, veamos como funciona mediante el uso de la fórmula del método abreviado.

$$r_{xy} = \frac{n \sum x_i \cdot y_i - \sum x_i \cdot \sum y_i}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2] [n \sum y_i^2 - (\sum y_i)^2]}}$$

(*x<sub>i</sub>* e *y<sub>i</sub>* se refieren a las puntuaciones diferenciales de cada par; “*n*” al número de casos; y *S<sub>x</sub>* ó  $[n \sum x_i^2 - (\sum x_i)^2]$  y *S<sub>y</sub>* ó  $[n \sum y_i^2 - (\sum y_i)^2]$  a las

desviaciones típicas de cada variable).

El coeficiente de correlación de Pearson toma valores entre  $-1$  y  $1$ : un valor de  $1$  indica relación lineal perfecta positiva; un valor de  $-1$  indica relación lineal perfecta negativa (en ambos casos los puntos se encuentran dispuestos en una línea recta); un valor de  $0$  indica relación lineal nula (lo que ocurre, por ejemplo, en las figuras 11C y 11D). El coeficiente " $r$ " es una medida simétrica: la correlación entre  $X_i$  e  $Y_i$ , es la misma que entre  $Y_i$  y  $X_i$ .

Es importante señalar que un coeficiente de correlación alto no implica causalidad. Dos variables pueden estar linealmente relacionadas (incluso muy relacionadas) sin que una sea causa de la otra.

#### 3.8.4. Características del Coeficiente de Correlación

- a. El valor del coeficiente de correlación es independiente de cualquier unidad usada para medir las variables.
- b. El valor del coeficiente de correlación se altera de forma importante ante la presencia de un valor extremo, como sucede con la desviación típica. Ante estas situaciones conviene realizar una transformación de datos que cambia la escala de medición y modera el efecto de valores extremos (como la transformación logarítmica).
- c. El coeficiente de correlación mide solo la relación con una línea recta. Dos variables pueden tener una relación curvilínea fuerte, a pesar de que su correlación sea pequeña. Por tanto, cuando analicemos las relaciones entre dos variables debemos representarlas gráficamente y posteriormente calcular el coeficiente de correlación.
- d. El coeficiente de correlación no se debe extrapolar más allá del rango de valores observado de las variables a estudio ya que la relación existente

entre “ $x$ ” e “ $y$ ” puede cambiar fuera de dicho rango.

- e. La correlación no implica causalidad. La causalidad es un juicio de valor que requiere más información que un simple valor cuantitativo de un coeficiente de correlación.

El coeficiente de correlación de Pearson ( $r$ ) puede calcularse en cualquier grupo de datos, sin embargo la validez del test de hipótesis sobre la correlación entre las variables requiere en sentido estricto 4: a) que las dos variables procedan de una muestra aleatoria de individuos. b) que al menos una de las variables tenga una distribución normal en la población de la cual la muestra procede. Para el cálculo válido de un intervalo de confianza del coeficiente de correlación de  $r$  ambas variables deben tener una distribución normal. Si los datos no tienen una distribución normal, una o ambas variables se pueden transformar (transformación logarítmica) o si no se calcularía un coeficiente de correlación no paramétrico (coeficiente de correlación de Spearman) que tiene el mismo significado que el coeficiente de correlación de Pearson y se calcula utilizando el rango de las observaciones.

El cálculo del coeficiente de correlación “ $r$ ” entre peso y talla de 20 niños varones se muestra en la tabla 1. La covarianza, que en este ejemplo es el producto de peso (kg) por talla (cm), para que no tenga dimensión y sea un coeficiente, se divide por la desviación típica de “ $x$ ” (talla) y por la desviación típica de “ $y$ ” (peso) con lo que obtenemos el coeficiente de correlación de Pearson que en este caso es de 0,885 e indica una importante correlación entre las dos variables. Es evidente que el hecho de que la correlación sea fuerte no implica causalidad. Si elevamos al cuadrado el coeficiente de correlación obtendremos el coeficiente de determinación ( $r^2 = 0,783$ ) que nos indica que el 78,3% de la variabilidad en el peso se explica por la talla del niño. Por lo tanto, existen otras variables que modifican y

explican la variabilidad del peso de estos niños. La introducción de más variable con técnicas de análisis multivariado nos permitirá identificar la importancia de que otras variables pueden tener sobre el peso.

Ejemplo: Cálculo del Coeficiente de correlación de Pearson entre las variables talla y peso de 20 niños varones

Y Peso (Kg)	X Talla (cm)	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X}) * (Y - \bar{Y})$
9	72	5,65	1,40	7,91
10	76	9,65	2,40	23,16
6	59	-7,35	-1,60	11,76
8	68	1,65	0,40	0,66
10	60	-6,35	2,40	-15,24
5	58	-8,35	-2,60	21,71
8	70	3,65	0,40	1,46
7	65	-1,35	-0,60	0,81
4	54	-12,35	-3,60	44,46
11	83	16,65	3,40	56,61
7	64	-2,35	-0,60	1,41
7	66	-0,35	-0,60	0,21
6	61	-5,35	-1,60	8,56
8	66	-0,35	0,40	-0,14
5	57	-9,35	-2,60	24,31
11	81	14,65	3,40	49,81
5	59	-7,35	-2,60	19,11
9	71	4,65	1,40	6,51
6	62	-4,35	-1,60	6,96
10	75	8,65	2,40	20,76
				$\Sigma$ 290,80

X(Media de  $\bar{X}$  = 66,35)

Y(Media de  $\bar{Y}$  = 7,60)

$$\text{Covarianza} = \frac{\sum(\bar{X} - X) * (\bar{Y} - Y)}{n - 1} = \frac{290,80}{19} = 15,31$$

$$r = \frac{\text{covarianza}}{S_x * S_y} = \frac{15,30}{8,087 * 2,137} = 0,886$$

Sx = Desviación típica x = 8,087

Sy = Desviación típica y = 2,137

Tras realizar el cálculo del coeficiente de correlación de Pearson "**r**" debemos determinar si dicho coeficiente es estadísticamente diferente de cero. Para dicho cálculo se aplica un test basado en la distribución de la **t de Student**.

$$\text{Error estándar de } r = \sqrt{\frac{1 - r^2}{n - 2}}$$

Si el valor del "**r**" calculado (en el ejemplo previo  $r = 0,886$ ) supera al valor del error estándar multiplicado por la **t de Student** con  $n-2$  grados de libertad, diremos que el coeficiente de correlación es significativo.

El nivel de significación viene dado por la decisión que adoptemos al buscar el valor en la tabla de la **t de Student**.

En el ejemplo previo con 20 niños, los grados de libertad son 18 y el valor de la tabla de la **t de Student** para una seguridad del 95% es de 2,10 y para un 99% de seguridad el valor es 2,88.

$$\text{Error estándar de } r = \sqrt{\frac{1 - (0,886)^2}{20 - 2}} = 0,109$$

Como quiera que  $r = 0,886 > a 2,10 * 0,109 = 0,229$  podemos asegurar que el coeficiente de correlación es significativo ( $p < 0,05$ ). Si aplicamos el valor obtenido en la tabla de la *t* de Student para una seguridad del 99% ( $t = 2,88$ ) observamos que como  $r = 0,886$  sigue siendo  $> 2,88 * 0,109 = 0,314$  podemos a su vez asegurar que el coeficiente es significativo ( $p < 0,001$ ). Este proceso de razonamiento es válido tanto para muestras pequeñas como para muestras grandes. En esta última situación podemos comprobar en la tabla de la *t de Student* que para una seguridad del 95% el valor es 1,96 y para una seguridad del 99% el valor es 2,58.

### 3.8.5. Tau-B de Kendall

Es un coeficiente de correlación por rangos, inversiones entre dos ordenaciones de una distribución normal bivalente. Este coeficiente de correlación es apropiado para estudiar la relación entre variables ordinales. Se basa en el número de inversiones y no inversiones entre casos. Toma valores entre  $-1$  y  $1$ , y se interpreta exactamente igual que el coeficiente de correlación de Pearson.

La utilización de este coeficiente tiene sentido si las variables no alcanzan el nivel de medida de intervalo y/o no podemos suponer que la distribución poblacional conjunta de las variables sea normal.



### 3.8.6. Spearman

El coeficiente de correlación *rho* de Spearman (1904) es el coeficiente de correlación de Pearson, pero aplicado después de transformar las puntuaciones originales en rangos. Toma valores entre  $-1$  y  $1$ , y se interpreta exactamente igual que el coeficiente de correlación de Pearson.

Al igual que ocurre con el coeficiente Tau-B de Kendall, el de Spearman puede utilizarse como una alternativa al de Pearson cuando las variables estudiadas son ordinales y/o se incumple el supuesto de normalidad.

Para las variables cuantitativas y distribuidas según una distribución normal, seleccionaremos el coeficiente de correlación de Pearson, mientras que si los datos no están normalmente distribuidos o tienen categorías ordenadas, seleccionaremos la Tau-B de Kendall o el coeficiente de correlación por rangos de Spearman, que miden la asociación entre órdenes de rangos.

## 3.9. Limitaciones Del Análisis De Regresión

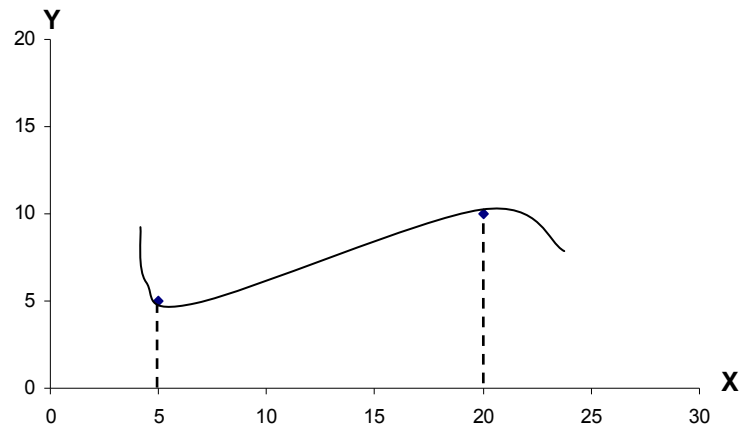
Aunque los análisis de regresión y correlación con frecuencia han demostrado ser de utilidad en la toma de decisiones para una gran variedad de negocios y de asuntos económicos, existen ciertas limitaciones en su aplicación e interpretación. Estos no pueden determinar relaciones causa-efecto. La correlación no implica causalidad. Este punto fue elaborado dramáticamente por un estadístico británico quien "probó" que las cigüeñas traían a los bebés. Él recolectó datos sobre las tasas de natalidad y el número de cigüeñas en Londres y descubrió una correlación muy alta –algo

así como  $r = 0,92$ . Por tanto, él concluyó que la historia sobre las cigüeñas y los bebés era cierta.

Sin embargo, como ya se habrá sospechado, ésta no es la forma como funciona la correlación. Parece que a esta especie de cigüeñas les gusta anidar en la parte superior de la chimenea de los londinenses. Por tanto, cuando la población era densa y la tasa de natalidad era alta, habían muchas chimeneas para atraer a estas aves, de allí la alta correlación entre la tasa de natalidad y las cigüeñas. En realidad, tanto las cigüeñas como los nacimientos eran *causados* por un tercer factor, la densidad poblacional, que el investigador ignoró a su conveniencia. Vale la pena recordar que la correlación no significa causalidad.

Adicionalmente, se debe tener cuidado de no utilizar el modelo de regresión para predecir “ $y$ ” para valores de “ $x$ ” que estén fuera del rango del conjunto original de datos. Los valores para “ $x$ ” en el conjunto de datos de Cumaná, Inc. oscilan desde tan bajo como 5 hasta tan alto como 20. Se ha aislado la relación entre “ $x$ ” y “ $y$ ” sólo para ese rango de valores “ $x$ ”. No se sabe cuál es la relación fuera de ese rango. Todo lo que se sabe es que puede aparecer como en la figura 12. Como se puede observar, para los valores fuera del rango de 5 a 20, la relación  $X-Y$  es totalmente diferente de lo que se puede esperar dado el ejemplo.

**FIGURA 12. Una Posible Relación X –Y**



**Fuente:** Los autores.

Otra falla del análisis de correlación y de regresión se hace evidente cuando dos variables no relacionadas parecen presentar alguna relación. Se asume que se desea analizar la correlación entre el número de monos nacidos en el zoológico de Barquisimeto y las toneladas de sardina que recogen los pescadores del Golfo de Cariaco. Se halla  $r = 0,91$ . ¿Se concluirá que existe una relación?, tal conclusión es obviamente disparatada. A pesar del valor “ $r$ ”, la simple lógica indica que no hay relación entre estas dos variables. No se ha abarcado la correlación espúrea, que es la correlación que ocurre simplemente por suerte. No hay sustituto para el sentido común en el análisis de correlación ni en el de regresión.

## CONCLUSIONES

Las herramientas estadísticas son, hoy en día, vitales para los empresarios y el mundo de los negocios, incluso hasta para otras disciplinas distintas a la Economía, ya que permiten una visión más amplia para la toma de decisiones con una base, aunque no exacta, pero si razonable.

En la actualidad, existe una amplia gama de aplicaciones de la Estadística a la Economía. Ya se encuentran en el mercado programas informáticos que ayudan a llevar a cabo el estudio con mayor facilidad y rapidez. No obstante, sugerimos consultar la bibliografía existen, para complementar su formación.

A través del estudio del comportamiento pasado de una variable, podemos planificar y prevenir lo que ocurrirá con esa misma variable en el futuro. La aplicación de este estudio requiere el registro histórico de los cambios, pero para eventos de reciente ocurrencia, resulta inoperante este análisis.

Los Económetras se han visto en la obligación de aplicar, cada vez más, la inferencia estadística para llevar a cabo sus estudios, ya que en la mayoría de los casos, las poblaciones son infinitas o difíciles de precisar, por lo que generalmente resulta más práctico hacer las estimaciones de los parámetros poblacionales.

## RECOMENDACIONES

Se recomienda que al estudiar dos variables para regresión y correlación, es prudente asegurarse de que éstas guarden una relación lógica.

Aplicar, en cuanto sea posible, los programas estadísticos existentes en el mercado, para facilitar el manejo de las grandes poblaciones (o muestras) y la rapidez en la obtención de los resultados.

Efectuar o llevar registros de los eventos que pudieran ser relevantes para una empresa u organismo para posteriormente poder aplicar las herramientas estadísticas proporcionadas en el contenido de este trabajo.

Al iniciar estudios de inferencia estadística, se aconseja poner especial cuidado al momento de determinar el tamaño de la muestra, conociendo que esta sea representativa de la población de la cual se extrajo, ya que de ello dependerá la mayor precisión del estudio o las desviaciones resultantes.

Graficar los datos, para que así tanto el investigador como el lector, puedan observar de una manera más rápida el comportamiento de las variables, pero se sugiere efectuar el estudio más a fondo para llegar a las conclusiones con mayor exactitud y poder determinar el modelo a aplicar.

## BIBLIOGRAFÍA

ALLARD, R. J. Introducción a la Econometría. 1ra. Edición. Editorial Limusa, S.A. 1980. México D.F., México.

HILDEBRAND, David K. y OTT, R. Lyman. Estadística Aplicada a la Administración y la Economía. 3ra. Edición. Editorial Addison–Wesley Iberoamericana. 1997. Delaware, EUA.

SALVATORE, Dominick y REAGLE, Derrick. (2004) Estadística y Econometría. 2da. Edición, Editorial McGraw–Hill.

SHAO, Stephen P. Estadística para Economistas y Administradores de Empresas. 1ra. Edición 1960. Editorial Herrero Hermanos, Sucs., S.A.

Universidad de Oriente (UDO)–Núcleo de Sucre. Escuela de Administración, Comisión de Trabajo de Grado. Instructivo para la Elaboración de Proyectos de Trabajo de Grado. Octubre 2006.

Universidad Pedagógica Experimental Libertador (UPEL). Manual de Trabajos de Grado de Especialización y Maestría y Tesis Doctorales. 4ta. Edición 2006. Reimpresión 2007. Aragua, Venezuela.

FERNÁNDEZ GARCÍA, José y DÍAZ DE URDANIVIA, Claramartha Adalid.

**Para una breve historia de la econometría.** *Política y Cultura* [en línea] 2000, (013):[fecha de consulta: 24 de junio de 2008] Disponible en:

<<http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=26701302>> ISSN 0188-7742.

MATA MOLLEJAS, Luis, Roger Da Costa S. **Notas metodológicas: De la econometría clásica al análisis no lineal.** *Revista Venezolana de Análisis de Coyuntura* [en línea] 2004, X (001):[fecha de consulta: 24 de junio de 2008] Disponible en:

<<http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=36410102>>  
ISSN 1315-3617.

VENTOSA SANTAULÀRIA, Daniel. ¿Qué es la Econometría?. *Acta Universitaria* [en línea] 2006, 16 (003):[fecha de consulta: 24 de junio de 2008] Disponible en:

<<http://redalyc.uaemex.mx/redalyc/src/inicio/ArtPdfRed.jsp?iCve=41600305>> ISSN 0188-6266.

<http://www.unifr.ch/ddp1/derechopenal/articulos/pdf/Denisanurio.pdf>

# **Hoja de Metadatos**



## Hoja de Metadatos para Tesis y Trabajos de Ascenso – 1/5

<b>Título</b>	Aplicaciones Estadísticas a la Economía (Econometría)
<b>Subtítulo</b>	

### Autor(es)

<b>Apellidos y Nombres</b>	<b>Código CVLAC / e-mail</b>	
<b>Pérez Loaiza, Daniel Enrique</b>	<b>CVLAC</b>	<b>V-11.424.175</b>
	<b>e-mail</b>	<a href="mailto:deperezloaiza@hotmail.com">deperezloaiza@hotmail.com</a>
	<b>e-mail</b>	<a href="mailto:deperezloaiza@cantv.net">deperezloaiza@cantv.net</a>
<b>Sánchez Rodríguez, Jesús Emilio</b>	<b>CVLAC</b>	<b>V-14.596.818</b>
	<b>e-mail</b>	<b>Jesusemilio08@gmail.com</b>
	<b>e-mail</b>	<b>Je_san01@yahoo.es</b>
	<b>CVLAC</b>	
	<b>e-mail</b>	
	<b>e-mail</b>	
	<b>CVLAC</b>	
	<b>e-mail</b>	
	<b>e-mail</b>	

### Palabras o frases claves:

<b>Econometría, Economía Aplicada, Regresiones, Series de tiempo,</b>
<b>Correlación, Diagrama de Dispersión.</b>

# Hoja de Metadatos para Tesis y Trabajos de Ascenso – 2/5

## Líneas y sublíneas de investigación:

Área	Subárea
Contaduría – Estadística	Econometría
	Inferencia Estadística

## Resumen (abstract):

La Econometría es una técnica de estudio de fenómenos económicos, aplicados por economistas con el fin de poder predecir y comprender futuros acontecimientos, a partir del análisis estadístico y matemático, de esta manera poder sugerir medidas de políticas económicas conforme a objetivos deseados. Para la puesta en práctica de esta técnica se utilizan varias herramientas estadísticas, como las series de tiempo, que es una sucesión de observaciones cuantitativa de un fenómeno ordenadas en el tiempo, esta metodología se basa en descomponer la series en varias partes: tendencia, variaciones estacionales, cíclicas e irregulares. Otra herramienta utilizada es el análisis de regresión y correlación, la idea básica del análisis de regresión es utilizar los datos de una variable cuantitativa independiente para predecir o explicar la variación de una variable cuantitativa dependiente, utilizando estas variables para desarrollar una ecuación de predicción, y el análisis de correlación, se refiere al grado de variación existente entre dos o más variables. En el desarrollo de este trabajo se aplicó una investigación de tipo descriptivo y el diseño fue de tipo documental, utilizando fuentes de información secundarias.

# Hoja de Metadatos para Tesis y Trabajos de Ascenso – 3/5

## Contribuidores:

Apellidos y Nombres	ROL / Código CVLAC / e-mail	
Romero, Miguel	ROL	CA <input type="checkbox"/> AS <input type="checkbox"/> TU <input checked="" type="checkbox"/> JU <input type="checkbox"/>
	CVLAC	V-8.879.006
	e-mail	mtreves@hotmail.com
	e-mail	
	ROL	CA <input type="checkbox"/> AS <input type="checkbox"/> TU <input type="checkbox"/> JU <input type="checkbox"/>
	CVLAC	
	e-mail	
	e-mail	

Fecha de discusión y aprobación:

Año      Mes      Día

2008	11	07
------	----	----

Lenguaje: Esp.

# Hoja de Metadatos para Tesis y Trabajos de Ascenso – 4/5

**Archivo(s):**

Nombre de archivo	Tipo MIME
TESIS-PEREZYSANCHEZ.DOC	Aplication / Word

**Alcance:**

**Espacial:** \_\_\_\_\_ (Opcional)

**Temporal:** \_\_\_\_\_ (Opcional)

**Título o Grado asociado con el trabajo:**

**Licenciado en Contaduría Pública** \_\_\_\_\_.

**Nivel Asociado con el Trabajo: Licenciatura** \_\_\_\_\_.

**Área de Estudio: Contaduría** \_\_\_\_\_.

**Institución(es) que garantiza(n) el Título o grado:**

**Universidad de Oriente (UDO) – Núcleo de Sucre** \_\_\_\_\_.

\_\_\_\_\_

\_\_\_\_\_

# Hoja de Metadatos para Tesis y Trabajos de Ascenso

– 5/5

## Derechos:

Nosotros, los autores, damos nuestro consentimiento para que este material sea divulgado única y exclusivamente con fines didácticos.

---

---

---

---



**Pérez Loaiza, Daniel Enrique**

C.I.: V-11.424.175

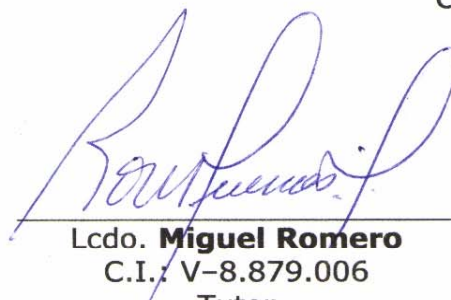
Autor



**Sánchez Rodríguez, Jesús Emilio**

C.I.: V-14.596.818

Autor



**Lcdo. Miguel Romero**

C.I.: V-8.879.006

Tutor

Por la Subcomisión de Trabajo de Grado:



**Lcda. Yenny J. Alzolar H.**

C.I.: V-9.978.152