

Universidad de Valparaíso
Facultad de Ingeniería
Escuela de Ingeniería Civil Industrial



**“Modelo Econométrico para la proyección de los ingresos
en los casinos de juegos en Chile”**

Por:

Hernán Felipe Vargas Droguett

Kevin Osman Parraguirre Cid

Trabajo de Título para optar al Grado de
Licenciado en Ciencias de la Ingeniería y título de
Ingeniero Civil Industrial

Prof. Guía José Irrazabal

Mayo, 2017

Dedicatoria**De Hernán:**

A mi hijo Martín, que en la vida que se te avecina encuentres siempre la felicidad. A mi mujer, Ivonne Sánchez, por todo el amor y cariño que me brinda. A mi madre, Ana por dar siempre lo mejor de sí. A mi padre Hernán, por incentivarme constantemente en la disciplina del estudio. A mi hermano, Esteban y su mujer Fernanda por ser una pareja maravillosa conmigo e Ivonne. A John, mi tío, por su incondicional apoyo en cada etapa de mi vida. A mi compañero Kevin, por su leal amistad.

A la memoria de Rocko y Thomas, mis fieles mascotas.

De Kevin:

A mis padres Daniel y Julia por su cariño, apoyo y paciencia (sobre todo paciencia) que me han tenido desde siempre. A mi hermano Fabrizio y a su esposa Alejandra por ser los mejores padres. A mi tío Fernando por su asesoría en la elección de la carrera. A Emerson por tener la calma que yo nunca tuve en su momento. A Hernán por su amistad en las mejores y peores circunstancias. A Gabriel, por enseñarme que no todo aprendizaje proviene de los libros. A mis amigos de universidad porque sin ellos mi vida como estudiante habría sido muy desagradable.

Índice

ÍNDICE	3
GLOSARIO.....	8
LISTA DE ABREVIATURAS	10
LISTA DE ILUSTRACIONES.....	11
LISTA DE TABLAS	13
1.1 INTRODUCCIÓN	14
1.1.1.1 Ámbito de Desarrollo	14
1.1.1.2 Definición de la situación	16
1.2 OBJETIVOS.....	18
1.2.1 Objetivo General	18
1.2.2 Objetivos Específicos	18
1.2.3 Alcance	18
1.2.4 Resultado Esperado.....	18
1.3 METODOLOGÍA	19
2.1 REGRESIÓN LINEAL MÚLTIPLE.....	20
2.2 MODELOS LOGARÍTMICOS.....	22
2.2.1 Modelo Doble log.....	23
2.2.2 Modelo Semilog	23

2.3	ESTIMACIÓN POR MÍNIMOS CUADRADOS.....	25
2.4	ESTIMACIÓN POR MÁXIMA VEROSIMILITUD.....	28
2.5	SUPUESTOS PARA LA GENERALIZACIÓN DE REGRESIÓN LINEAL.....	29
2.5.1	Supuesto De Normalidad.....	29
2.5.2	Supuesto De La Matriz De Los Errores Multiplicado Por Su Traspuesta.....	30
2.6	VALIDACIÓN DE LOS COEFICIENTES DE DETERMINACION.....	31
2.6.1	Razonamiento Informal.....	31
2.6.2	Razonamiento Formal.....	32
2.6.3	Coefficiente De Determinación o R^2.....	33
2.6.4	Coefficiente De Determinación Ajustado o R^2.....	35
2.7	VALIDACIÓN DEL MODELO.....	36
2.7.1	Multicolinealidad.....	36
2.7.1.1	Test De Multicolinealidad.....	37
2.7.2	Homocedasticidad.....	39
2.7.2.1	Test De Homocedasticidad.....	39
2.7.3	Autocorrelación De Los Errores.....	40
2.7.3.1	Test De Autocorrelación.....	40
2.7.3.2	Corrección De Autocorrelación Mediante Cochrane-Orcutt.....	42
2.8	OTRAS VARIABLES.....	44
2.8.1	Variables Binarias.....	44
2.8.2	Variables Retardadas.....	44

2.9	ELECCIÓN DE MÉTODO	45
2.9.1	Justificación Del Modelo Lineal	45
2.9.1.1	Justificación Del Método De Mínimos Cuadrados Ordinarios	45
3.1	ÍNDICE DE PRECIOS AL CONSUMIDOR.....	46
3.1.1	Cálculo del IPC.....	48
3.2	ÍNDICE DE REMUNERACIONES (IR)	51
3.2.1	Calculo del IR	53
3.3	ÍNDICE DEL COSTO DE LA MANO DE OBRA (ICMO).....	54
3.3.1	Calculo del ICMO	54
3.3.2	Recolección de información	55
3.4	TASA DE DESEMPLEO	56
3.5	ÍNDICE DE PERCEPCIÓN DEL CONSUMIDOR (IPECO).....	56
3.6	ÍNDICE MENSUAL DE ACTIVIDAD ECONÓMICA (IMACEC)	57
3.7	MÉTODOS DE PRONÓSTICO DE LAS VARIABLES EXPLICATIVAS	57
3.7.1	Promedio Móvil Simple	57
3.7.2	Suavización exponencial de Holt-Winters.....	58
5.1	VARIABLE REGRESADA.....	67
5.1.1	Ingresos brutos de los casinos de juegos	67
5.2	VARIABLES REGRESORAS	68

5.2.1	Índice Mensual de Actividad Económica.....	69
5.2.2	Tasa de Desempleo	69
5.2.3	Reservas en Dólares del Banco Central.....	69
5.2.4	Dólar Observado	70
5.2.5	Visitas totales.....	70
5.2.6	Índice de remuneraciones	70
5.2.7	Índice de costo de mano de Obra.....	71
5.2.8	Índice Precios del Consumidor	71
5.2.9	Índice de percepción del consumidor	71
5.2.10	Existencia de la ley de tabaco	72
6.1	MARCO DE ACCIÓN	80
6.1.1	Test de Multicolinealidad	85
6.1.2	Contraste De White.....	86
6.1.3	Test de Durbin Watson	87
6.1.3.1	Método de Cochare-Orcutt.....	88
6.1.3.2	Definición del Modelo.....	89
6.2	ANÁLISIS DE RESULTADOS.....	90
6.3	PROYECCIÓN DE INGRESOS.....	95
6.3.1	Proyección de la variable Ley de tabaco	95
6.3.2	Proyección de total de visitas	96
6.3.3	IR General	97

6.3.4	Subdivisión del IPC, IPC Recreacional	98
6.4	PROYECCIÓN DE INGRESOS.....	99
7.1	CONCLUSIONES.....	100
9.1	EXTRACTO DEL COMPENDIO ESTADÍSTICO INE.....	103
9.2	INGRESOS MEDIOS DE LOS ASALARIADOS.....	104
9.3	RESULTADOS PROYECCIÓN HOLT Y WINTERS.....	105

Glosario

Ceteris paribus: Locución latina que significa “dadas las mismas circunstancias”

Distribución normal: Se dice que una distribución de probabilidad de una variable aleatoria es una función en la cual se le asigna a cada valor una probabilidad de ocurrencia. Dentro de los tipos de distribución se encuentra la distribución normal la cual es la de mayor uso en la estadística debido a que se usa para estudiar fenómenos físicos (temperatura, precipitación fluvial, etc.) como también por el hecho de que mientras más grande sea una muestra está siempre tenderá a comportarse como indica la distribución normal. Se define la distribución de probabilidad como:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$$

Donde:

x : variable aleatoria que sigue la distribución normal.

σ : desviación estándar de la distribución normal.

μ : media de la distribución normal.

Distribución t Student: Cuando se realiza un ejercicio de distribución, existen casos en el que se desconoce el valor de la desviación estándar, para estos casos se necesita usar la función de distribución t Student de probabilidad. Se define la variable para la distribución como:

$$T = \frac{Z}{\sqrt{X/v}}$$

Donde:

Z : Variable aleatoria con distribución normal con media cero y varianza constante.

X : Variable aleatoria con distribución χ^2 con v grados de libertad.

Enterar en arcas fiscales: Ingresar dineros al fisco.

Función de densidad de probabilidad conjunta: En ciertos casos es necesario que para realizar un análisis estadístico se tomen en cuenta no solo un tipo de variable, sino que dos o más sean estudiadas mediante la misma distribución de probabilidad, es a estas funciones que se les conoce como de probabilidad conjunta.

Hipótesis alternativa: En una prueba de hipótesis, se define como un estado alternativo el cual se utiliza para rechazar o no rechazar la hipótesis nula.

Hipótesis nula: En una prueba de hipótesis, se define como el estado actual.

Matriz: Arreglo rectangular de números.

Variable explicada: También llamada variable dependiente o regresada, para un analista es el objeto de estudio, es decir, aquello que se intenta deducir mediante otras variables.

Variable explicativa: También llamada variable independiente o regresora, son para un analista los elementos con los que se va a estudiar una variable explicada.

Lista de abreviaturas

R^2 : Coeficiente de determinación.

\bar{R}^2 : Coeficiente de regresión ajustado.

K: Cantidad de variables que posee una función

N: Cantidad de observaciones que se utilizaran para confeccionar un modelo

FIV: Factor de inflación de la varianza.

χ^2 : Chi cuadrado.

IPC: Índice de precios al consumidor.

IR: Índice de remuneraciones.

ICMO: Índice del costo de mano de obra.

IPECO: Índice de percepción del consumidor.

IMACEC: Índice mensual de actividad económica

SII: Servicio de impuestos internos.

BC: Banco central

Lista de Ilustraciones

Ilustración 1: Ingresos Brutos de los Casinos en Chile	16
Ilustración 2 Docima de Hipótesis a dos colas	32
Ilustración 3 Subdivisión IPC	48
Ilustración 4 Tabla de subdivisión IPC	49
Ilustración 5 Estructura conceptual IR.....	52
Ilustración 6 Segmentaciones de los asistentes.....	60
Ilustración 7 Segmentaciones de los asistentes.....	61
Ilustración 8 Segmentaciones de los asistentes.....	62
Ilustración 9 Percepción de la situación económica personal actual	63
Ilustración 10 Percepción de la situación económica personal futura.....	63
Ilustración 12 Intereses personales en cada tema	64
Ilustración 11 En que gasta si existiese dinero sobrante.....	64
Ilustración 13 Como describirían su propia personalidad	65
Ilustración 14 Variable regresada	67
Ilustración 15 Variables regresoras.....	73
Ilustración 16 Variables regresoras.....	74
Ilustración 17 Variables regresoras.....	75
Ilustración 18 Variables regresoras.....	76
Ilustración 19 Variables regresoras.....	77
Ilustración 20 Variables regresoras.....	78
Ilustración 21 Variables regresoras.....	79
Ilustración 23 Resultados Gretl	82
Ilustración 24 Resultados Gretl	85
Ilustración 25 Resultados Gretl	86
Ilustración 26 Resultados Gretl	87
Ilustración 27 Resultados Gretl	88
Ilustración 28 Resultados Gretl	90
Ilustración 29 Resultados Gretl	91
Ilustración 30 Resultados Gretl	91
Ilustración 31 Resultados Gretl	92
Ilustración 32 Resultados Gretl	92

Ilustración 33 Resultados Gretl	93
Ilustración 34 Resultados Gretl	93
Ilustración 35 Resultados Gretl	94
Ilustración 36 Resultados Gretl	94
Ilustración 37 Proyección Visitas	96
Ilustración 38 IR General	97
Ilustración 39 IPC Recreacional	98

Lista de Tablas

Tabla 1 Principales estadísticos variable regresada	68
Tabla 2 Principales estadísticos variables regresoras.....	73
Tabla 3: Principales estadísticos variables regresoras.....	74
Tabla 4 Principales estadísticos variables regresoras.....	75
Tabla 5 Principales estadísticos variables regresoras.....	76
Tabla 6 Principales estadísticos variables regresoras.....	77
Tabla 7 Principales estadísticos variables regresoras.....	78
Tabla 8 Principales estadísticos variables regresoras.....	79
Tabla 9 Análisis de significancia	83
Tabla 10 Proyección de Ingresos.....	99

1.1 Introducción

1.1.1.1 Ámbito de Desarrollo

Una de las tareas frecuentes en la planificación de un proyecto, una empresa o ejercicio económico cualesquiera, es la definición de la proyección de ventas, ingresos o comportamiento futuro que tendrá lo que se desea planificar, dicho de otra forma, es la cuantificación de la respuesta del mercado ante la incorporación de un nuevo actor, ante el cambio en el escenario o el comportamiento de la evolución natural del ciclo de negocio. Para tales efectos existen diversas herramientas con distintos niveles de complejidad, que nos permiten preparar con un argumento técnico la proyección. Se pueden distinguir dos tipos de análisis, uno de ellos sustenta la proyección los valores de una variable a través del estudio de su conducta anterior, es decir, según el comportamiento que ha presentado hasta el momento de la planificación, se proyecta con algún modelo suponiendo que su conducta en los periodos futuros esta explicada en periodos pasados, este modelo tiene intrínsecamente un estado “ceteris paribus”, es decir, supone que las demás variables del ecosistema que afectan al desarrollo de lo que se desea estudiar se comportarán tal como se han comportado hasta el momento de la proyección, ósea, supone un estado estático¹ de las variables que afectan de una u otra forma el desarrollo del mismo, o dicho de otra forma, supone los factores que han llevado a la variable estudiada a comportarse de la manera que lo ha hecho, seguirán desenvolviéndose de la misma manera y afectando de la misma forma para efectos de la proyección.

Existen además métodos que explican el comportamiento de la variable estudiada, también llamada en este caso, variable a explicar o variable dependiente, relacionándolas con otras variables que incidan de algún modo, variables dependientes o explicativas. Esta forma de explicar el comportamiento o dependencia de una variable se conoce como Econometría. *“La econometría puede definirse como el análisis cuantitativo de fenómenos económicos reales basados en el desarrollo simultaneo de la teoría y la observación, relacionándolos mediante métodos apropiados de inferencia”* [Tintner68]. Esta herramienta supone que el comportamiento de la variable estudiada, desde ahora “variable explicada”, está relacionado por

¹ Este estado estático se refiere a que el comportamiento que ha evidenciado la variable en el pasado será el mismo que presentara en el futuro, es decir, supone que no habrá inflexiones de comportamiento en las variables macroeconómicas que inciden en lo que se proyecta.

el comportamiento de otras variables, desde ahora “variables explicativas”, dicho de otra manera, supone que existen factores de alguna manera cuantificables que tienen incidencia en el desarrollo de lo que queremos planificar, además que estos factores se puedan relacionar en una función matemática, y el comportamiento de esta función debe ser razonablemente similar a la variable a proyectar con el fin de asentar la proyección de los periodos venideros sobre esta función.

Dicho lo anterior, en el ámbito de la planificación elaborar proyecciones fiables, dentro de un contexto macroeconómico que posee aspectos de incertidumbre y especulación que complejizan y muchas veces invalidan los pronósticos, es un factor de éxito clave, al proyectar, por ejemplo, en cuanto variarán los ingresos futuros se puede elaborar los requerimientos necesarios para cubrir tal demanda, cuanto serán los costos esperados y cuanta dotación de personal requiere la compañía, además de dar la posibilidad de calcular la rentabilidad del ejercicio o proyecto. Elaborar métodos efectivos de proyección de ingresos o demandas en cualquier industria entrega una herramienta que agrega valor a la planificación y operación de la misma.

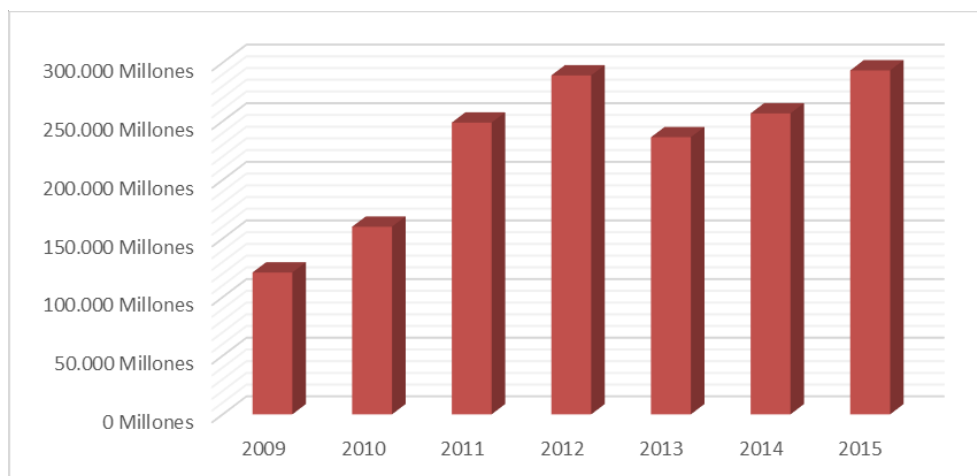
Dicho esto, se decidió efectuar un estudio en la industria de casinos de juegos de Chile, debido principalmente a la disponibilidad de información y el conocimiento sobre la industria que debido a circunstancias puntuales se disponen con antelación dando una ventaja clave para el desarrollo de este trabajo.

Adentrándonos en la industria, actualmente existen 16 casinos de juegos en Chile los que se ubican en las comunas de Calama, Antofagasta, Copiapó, Rinconada, San Antonio, Mostazal, Santa Cruz, Talca, Talcahuano, Los Ángeles, Temuco, Valdivia, Osorno, Castro, Coyhaique y Punta Arenas., además de otro que está en etapa de construcción en la comuna de Ovalle, estos son operados a través de concesiones municipales y se rigen bajo el amparo de la Ley N°19.990, son fiscalizados por la superintendencia de casinos de juegos, quien además publica la información oficial de estos, información que será parte fundamental de la base de datos en la cual se estructura.

Durante el último año la industria generó ingresos brutos por \$293.450 millones pesos lo que se tradujo en una recaudación de \$48.641 millones por impuesto específico al juego y \$46.853 millones por concepto de IVA. El impuesto específico se distribuye en dos partes

iguales una de ellas entra al patrimonio de la municipalidad correspondiente a la comuna donde se ubica este se destina a obras de desarrollo, el otro 50% se incorpora al patrimonio del gobierno regional correspondiente a la región donde se ubica el casino y también se destina a obras de desarrollo. El IVA recaudado por el juego se entera en arcas fiscales.

Ilustración 1: Ingresos Brutos de los Casinos en Chile



Fuente: Elaboración propia

1.1.1.2 Definición de la situación a desarrollar

Para la fecha en la cual se elaboró el presente proyecto de tesis, la industria de casinos se encontraba en proceso de licitación con las municipalidades, dentro de ese contexto y para dar cumplimiento a las bases técnicas desarrollamos una herramienta para la elaboración de las proyecciones del nuevo periodo de la licitación, esta herramienta aporta validez técnica en el aspecto económico y financiero de la postulación, específicamente en el presupuesto, flujos financieros y cálculos del valor presente neto. En este ámbito presentamos nuestro proyecto

a altos mandos de diferentes casinos de los cuales el tesorero del grupo de casinos Mundodreams, Don Jaime Barrueto², y el Director de Marketing de Sun Monticello Don Carlos Solar³ nos solicitaron los resultados de este trabajo para parangonar con los que elaboro el departamento de finanzas, que posteriormente serán presentados como sustento técnico de la postulación. Dado lo anterior se decidió elaborar un modelo genérico que capture la dinámica de la industria y que pueda ser adaptable a la realidad de ambos casinos.

² Jaime Barrueto, Tesorero General, Mundo Dreams, jbarrueto@mundodreams.com

³ Carlos Solar C., Direct Markeing Gran Casino & Entertainment World, carlos.solar@suninternational.com

1.2 Objetivos

1.2.1 Objetivo General.

Formular un modelo econométrico que describa el comportamiento de los flujos de ingresos de la industria de los casinos de juegos.

1.2.2 Objetivos Específicos

- Seleccionar y validar las variables claves para el desarrollo del modelo.
- Seleccionar y validar la relación funcional entre las variables.
- Efectuar análisis económicos que expliquen las variables seleccionadas.

1.2.3 Alcance

El alcance del presente trabajo de tesis se define en la creación de un modelo econométrico que describa los ingresos del casino con un margen de error razonable, con variables macroeconómicas que, en un sentido lógico, tengan relación en los flujos de la industria, validando económicamente de la incidencia de la variable en el comportamiento de los ingresos.

1.2.4 Resultado Esperado

Obtener un modelo econométrico cuyo comportamiento sea razonablemente similar a los ingresos totales de los casinos en Chile.

1.3 Metodología

Para abordar el trabajo de memoria, se definió la siguiente metodología:

- 1) Comprensión de la situación
- 2) Recolección de datos e información
- 3) Limpieza de datos
- 4) Análisis de posibles variables involucradas
- 5) Selección de variables involucradas.
- 6) Elaboración de modelos econométricos que proyecten los ingresos.
 - a. Planteamiento de la teoría o de la hipótesis
 - b. Especificación del modelo matemático de la teoría
 - c. Especificación del modelo econométrico o estadístico de la teoría
 - d. Estimación de los parámetros del modelo econométrico
 - e. Pruebas de Hipótesis
- 7) Testeo del modelo elaborado.
- 8) Análisis y conclusiones

Lo anterior permitió darle una estructura formal y un orden lógico a los esfuerzos que se invirtieron para desarrollar el trabajo de tesis. Esta metodología está estructurada a partir de la metodología planteada en Pronósticos de Negocios, 8va ed – John E. Hanke y Econometría 5ta Ed – Damodar Gujarati.

Capítulo 2. Marco Teórico

2.1 Regresión Lineal Múltiple

En la econometría aplicada, la forma funcional que más se utiliza en la práctica para representar la relación causal entre variables dependientes e independientes es la función lineal, que en su forma más general puede expresarse de la siguiente manera [Cole96]:

Sean:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

- y : la variable regresada
- x_i : las variables regresoras, con $i = 1, 2, \dots, k$
- β_i : los coeficientes parciales de pendientes para la variable i , es decir, si el resto de las variables permanecen constantes β_i representaría la tasa de cambio entre x_i e y . Con $i = 1, 2, \dots, k$

β_0 : coeficiente de intersección, es decir, si los valores de las variables son cero, entonces obligatoriamente el valor de y sería β_0 , dicho de otra forma, es el valor que corta al eje de las regresadas.

Esta expresión es la forma de simbolizar la evolución de la serie de tiempo ya que esta ecuación representa a cada uno de los valores que puede tomar la variable regresada, dicho de otra forma, la ecuación representa de manera simultánea los n valores que puede tomar la variable y :

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + u_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + u_2$$



$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + u_n$$

Se puede expresar de una forma matricial si, estas n ecuaciones con k variables se escriben de la siguiente manera [Cole96], [Gujarati&Porter10]:

:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Que en una notación matricial simplificada se define de la siguiente forma:

$$Y = X\beta + u$$

Donde:

- Y : matriz de dimensión $n \times 1$ que contiene a todo el conjunto de variables regresadas, es decir, y_t con $t = 1, 2, \dots, n$.
- X : matriz de dimensión $n \times (k+1)$ que contiene a todo el conjunto de las variables regresoras, es decir, x_{ti} con $t = 1, 2, \dots, n$ e $i = 1, 2, \dots, k$ más una columna que contiene solamente números 1, para dar cabida al término de intercepto.
- β : matriz de dimensión $(k+1) \times 1$ que contiene la totalidad de los valores de los coeficientes parciales de las pendientes más el coeficiente de intercepto (situado en la posición β_0).
- u : matriz de los errores de dimensión $n \times 1$ que contiene los errores u_j (o desviaciones) de cada observación con $j = 1, 2, \dots, n$.

2.2 Modelos Logarítmicos

Si bien el modelo funcional más reconocido de las funciones de regresión son las regresiones lineales también existen “variaciones” que, dependiendo de la situación, pueden ser más útiles que el modelo de regresión lineal. Pero como el método de encontrar estimadores para los coeficientes de regresión parcial se realiza mediante mínimos cuadrados ordinarios, y este a su vez se realiza cuando la función es lineal, se deben realizar ciertos cambios a las estructuras de las regresiones que serán presentadas a continuación. En este punto necesario tener claro que se considera intrínsecamente lineal.

Se considera a la siguiente función de regresión como lineal:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

Cuando su expresión es un polinomio de primer grado tanto en las variables como en los parámetros, es decir, no existen potencias de variables en la expresión ni tampoco hay productos entre ellas, cosa que también se puede decir de los parámetros. Sin embargo, se pueden construir funciones que no cumplan con tales características, por ejemplo:

$$y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \dots x_k^{\beta_k} u$$

Esta expresión no se puede considerar lineal en su forma más pura, porque las variables se encuentran expresadas como potencia y además forman producto entre ellas, pero con una debida transformación se puede reescribir para que se considere lineal solamente en los parámetros. Por lo tanto, se dice que esta función es intrínsecamente lineal [Gujarati&Porter10].

2.2.1 Modelo Doble log

Se tiene la siguiente función de regresión:

$$y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} \dots x_k^{\beta_k} u$$

Que al aplicar una transformación logarítmica queda expresada (después de usarse las propiedades de los logaritmos) de la siguiente forma:

$$\begin{aligned} \ln(y) &= \ln(\beta_0 x_1^{\beta_1} x_2^{\beta_2} \dots x_k^{\beta_k} u) \\ \ln(y) &= \ln(\beta_0) + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \dots + \beta_k \ln(x_k) + \ln(u) \end{aligned}$$

De esta forma se observa que esta forma funcional es intrínsecamente lineal y por tanto es posible encontrar los estimadores de los coeficientes de regresión parcial mediante mínimos cuadrados ordinarios. Una de las ventajas de este modelo es que los ponderadores hacen las veces de coeficientes de elasticidad para cada variable, y como este no varía, se dice que es un modelo de elasticidad constante[Cole96], [Gujarati&Porter10].

2.2.2 Modelo Semilog

Se tienen las siguientes funciones de regresión:

$$\begin{aligned} y &= \beta_0 \beta_1^{x_1} \beta_2^{x_2} \dots \beta_k^{x_k} u \\ e^y &= \beta_0 x_1^{\beta_1} x_2^{\beta_2} \dots x_k^{\beta_k} u \end{aligned}$$

Que al aplicar una transformación logarítmica se pueden dejar expresadas de la siguiente manera:

$$\begin{aligned} \ln(y) &= \ln(\beta_0) + \ln(\beta_1) x_1 + \ln(\beta_2) x_2 + \dots + \ln(u) \\ y &= \ln(\beta_0) + \beta_1 \ln(x_1) + \beta_2 \ln(x_2) + \dots + \beta_k \ln(x_k) + \ln(u) \end{aligned}$$

De esta forma se observa que ambas formas funcionales son intrínsecamente lineales para los logaritmos de los coeficientes en el primer caso y para los coeficientes en el segundo. Por lo tanto, es factible encontrar sus respectivas estimaciones mediante el método de los mínimos cuadrados ordinarios. Para el primer modelo (log-lin) se dice un cambio en una unidad de x_i supone un cambio porcentual en y de $\ln(\beta_i)$, mientras que para el segundo modelo (lin-log) se dice que un cambio porcentual de x_i supone un cambio absoluto en y de β_i unidades.

Aparte de las variantes del modelo de regresión lineal acá presentados existen otros modelos como los hiperbólicos y los log-hiperbólicos, pero que son descartados de plano en esta investigación por la naturaleza asintótica de sus variables (situación que no se presentó en todas las potenciales variables a utilizar en el modelo), [Gujarati&Porter10].

2.3 Estimación Por Mínimos Cuadrados

El objetivo de este método es que, dado una muestra de la población, se puedan encontrar valores estimados de β supongamos la función:

$$Y_j = \hat{\beta}_0 + \hat{\beta}_1 x_{j1} + \hat{\beta}_2 x_{j2} + \dots + \hat{\beta}_k x_{jk} + \hat{u}_j$$

Esta generalización, que en este caso sería el total de la muestra observada, matricialmente es representada de la siguiente forma [Cole96]:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} + \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}$$

Se puede reescribir de una forma más compacta de la siguiente manera:

$$Y = X\hat{\beta} + \hat{u}$$

Donde:

- $\hat{\beta}$: matriz de dimensión $(k+1) \times 1$ que contiene cada una de las estimaciones de los términos de intercepto, es decir, $\hat{\beta}_i$ con $i = 1, 2, \dots, k$ más el estimado del término de intercepto $\hat{\beta}_0$.
- \hat{u} : matriz de dimensión $n \times 1$ que contiene los errores de estimación de la función de regresión muestral, es decir, \hat{u}_j con $j = 1, 2, \dots, n$.

Dadas las ecuaciones anteriores se introduce y define el error cuadrático de la siguiente manera:

$$\hat{u}'\hat{u} = [\hat{u}_1 \quad \hat{u}_2 \quad \dots \quad \hat{u}_n] \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix} = \hat{u}_1^2 + \hat{u}_2^2 + \dots + \hat{u}_n^2 = \sum \hat{u}_j^2$$

Que representa la sumatoria de los cuadrados de los errores de la función de regresión muestral. Esto también se pueden escribir en la forma escalar de la siguiente manera [cole96]:

$$\hat{u}'\hat{u} = \sum \hat{u}_j^2 = (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{j1} - \hat{\beta}_2 x_{j2} - \dots - \hat{\beta}_k x_{jk})^2$$

Ahora el siguiente paso para encontrar los estimadores de los coeficientes β , es minimizar el error cuadrático y para esto se debe derivar parcialmente con respecto a cada uno de los coeficientes (dado que x e y son valores conocidos).

$$\frac{\partial \hat{u}'\hat{u}}{\partial \beta_0} = -2 \sum (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k) = 0$$

$$\frac{\partial \hat{u}'\hat{u}}{\partial \beta_1} = -2 \sum (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k) x_1 = 0$$

$$\frac{\partial \hat{u}'\hat{u}}{\partial \beta_2} = -2 \sum (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k) x_2 = 0$$



$$\frac{\partial \hat{u}'\hat{u}}{\partial \beta_k} = -2 \sum (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k) x_k = 0$$

Que en su forma simplificada se puede escribir de la siguiente manera:

$$\sum \hat{u} = 0$$

$$\sum x_1 \hat{u} = 0$$

$$\sum x_2 \hat{u} = 0$$



$$\sum x_k \hat{u} = 0$$

A este conjunto de ecuaciones se les conoce como las ecuaciones normales, que escrito de manera matricial queda representado como la matriz traspuesta de los regresores multiplicada (por la derecha) por la matriz de los errores, siendo igualado a una matriz de dimensión $(k+1) \times 1$ donde todos sus elementos poseen el valor cero, es decir:

$$X' \hat{u} = 0$$

Dado que la matriz de los errores se describe como la diferencia entre el valor observado y la función de regresión (sin el término de error), entonces la ecuación se expresa de la siguiente manera:

$$X'(Y - X\hat{\beta}) = 0$$

$$X'Y - X'X\hat{\beta} = 0$$

$$X'X\hat{\beta} = X'Y$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

2.4 Estimación Por Máxima Verosimilitud

Si bien el método de mínimos cuadrados ordinarios sirve para encontrar estimadores de los ponderadores, no es el único método con el que se cuenta para encontrarlos, otra forma muy aceptada por la literatura es el método de máxima verosimilitud y su nombre se debe a que, descrito de manera básica, se intenta que los valores estimados sean tales que maximicen la probabilidad de encontrar (u observar) los valores de la variable regresada.

De manera formal, se tiene una variable regresada y un conjunto de variables regresoras arregladas de la siguiente manera:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

Donde el conjunto de observaciones de la variable y son independientes y normalmente distribuidas con media igual a la función lineal mostrada y varianza σ^2 . Dados estos supuestos se tiene que la función de densidad de probabilidad conjunta se expresa como se muestra a continuación [Gujarati&Porter10]:

$$f(y_1, y_2, \dots, y_n | \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k, \sigma^2) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} \text{EXP}_e \left\{ -\frac{1}{2} \sum \frac{y - \beta_0 - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_k x_k}{\sigma^2} \right\}$$

ahora para encontrar los ponderadores, como se dijo anteriormente, hay que buscar estimadores que maximicen la probabilidad de encontrar los valores de la variable regresada, para lo cual hay que derivar parcialmente con respecto a todas las incógnitas de la función de densidad conjunta, es decir, los β_i y la varianza σ^2 . Después de derivar y reordenar las ecuaciones estas quedan:

$$\begin{aligned} \sum (Y_t - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k) &= 0 \\ \sum (Y_t - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k) x_1 &= 0 \\ \sum (Y_t - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k) x_2 &= 0 \end{aligned}$$



$$\sum (Y_t - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k) x_k = 0$$

$$\frac{1}{n} \sum (Y_t - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k) = \hat{\sigma}^2$$

Donde las primeras $k + 1$ ecuaciones son exactamente iguales al conjunto de ecuaciones normales que se obtienen mediante el método de los mínimos cuadrados ordinarios mientras que la última igualdad representa un estimado de la varianza, se debe tener en cuenta que esta forma de estimar la varianza arroja un resultado asintóticamente insesgada por lo que requiere un gran volumen de observaciones.

2.5 Supuestos Para La Generalización De Regresión Lineal

El objetivo de esta sección es desarrollar procedimientos para testear hipótesis sobre los coeficientes del modelo lineal. Para esto, debemos hacer ciertas suposiciones sobre el comportamiento estadístico de los errores. Los dos supuestos más importantes en el modelo clásico de regresión lineal son los siguientes [Cole96]:

2.5.1 Supuesto De Normalidad

Cada termino de error o perturbación sigue una distribución normal con media cero y varianza constante, es decir, $u \sim N(0, \sigma^2)$. Lo que implica de este primer supuesto es que, ya que las medias poseen el valor cero entonces la esperanza del vector de los errores es igual a una matriz de dimensión $n \times 1$ donde cada termino es cero, dicho de otra forma:

$$E(\hat{u}) = 0$$

2.5.2 Supuesto De La Matriz De Los Errores Multiplicado Por Su Traspuesta

Se define como supuesto que el valor esperado de la matriz de los errores multiplicado por su traspuesta es igual a una matriz de dimensión $n \times n$ en donde los elementos de su diagonal principal tienen la misma varianza σ^2 , es decir:

$$E(uu') = \sigma^2 I$$

Cabe decir que esta matriz tiene como elemento característico a $u_j u_i$. por tanto, suponer que el valor esperado de uu' es una matriz escalar equivale a suponer lo siguiente [Cole 96]:

- 1) $E(u_j u_i) = 0$ para $j \neq i$. esta declaración quiere decir que todos los elementos que no pertenecen a la diagonal principal son iguales a cero, por lo tanto, los errores son independientes entre sí.
- 2) $E(u_j^2) = \sigma^2$ para $j = i$, es decir, que todos los elementos de la diagonal principal son σ^2 . Dicho de otra forma, que la varianza es igual para todos los elementos pertenecientes a la matriz de los errores.

2.6 Validación De Los Coeficientes de Determinación

2.6.1 Razonamiento Informal

Supóngase que se desea validar si uno de los coeficientes de regresión parcial es significativo, y se entiende de manera intuitiva que un coeficiente es significativo cuando la variable regresora (asociada a dicha pendiente), al aumentar o disminuir, hace que la variable regresada aumente o disminuya (o viceversa), esto es posible, solo cuando el coeficiente de regresión parcial es distinto de cero.

También se entiende de manera intuitiva de que un coeficiente de regresión parcial es significativo cuando su varianza es lo suficientemente pequeña, ya que en el caso contrario su valor, al tomar valor con un grado de dispersión mayor, se convertiría en “poco fiable”.

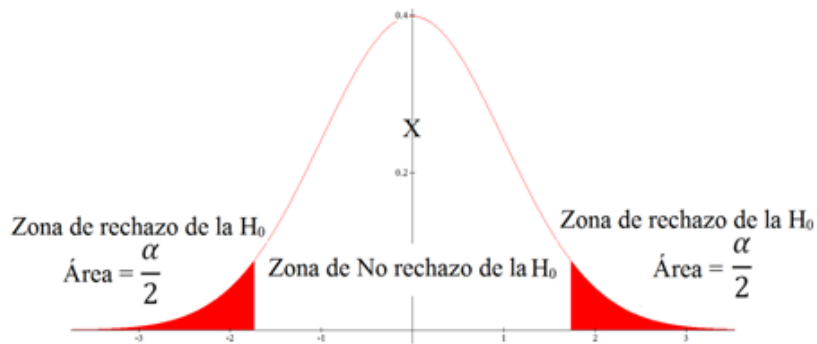
Por lo tanto, se puede deducir que dado un coeficiente de regresión lo suficientemente alejado del cero dividido por una desviación estándar lo suficientemente pequeña sería un indicador que demuestra que el coeficiente es un estimador significativo, o de manera inversa, que un coeficiente de regresión lo suficientemente cerca de cero, o cero en un caso extremo, dividido por una desviación estándar lo suficientemente grande indica que el indicador no es confiable.

2.6.2 Razonamiento Formal

Se tienen las siguientes hipótesis nula e hipótesis alternativa:

- $H_0: \beta = 0$
- $H_1: \beta \neq 0$

Ilustración 2 Docima de Hipótesis a dos colas



Fuente: Pruebas de Hipótesis Mario Suárez

La primera hipótesis hace referencia a que la variable X_i de la función de regresión muestral no posee incidencia en la función de regresión muestral mientras que la hipótesis alternativa expresa todo lo contrario, que la variable X_i , si posee una incidencia sobre la función de regresión muestral. Ahora, el estadístico de prueba que sirve para contrastar dichas hipótesis es la t de Student que se calcula de la siguiente manera:

$$t = \frac{\hat{\beta}_i}{S(\hat{\beta}_i)}$$

Donde $S(\hat{\beta}_i)$: representa la desviación típica del coeficiente parcial de la regresión de cada una de las variables con $i = 1, 2, \dots, k$. Ahora para encontrar dichas desviaciones se calculan las raíces cuadradas de cada uno de los valores de la diagonal principal de la matriz de varianza-covarianza:

$$S^2(X'X)^{-1}$$

Donde S^2 representa el estimado de la varianza poblacional, que se calcula de la siguiente manera:

$$S^2 = \frac{\hat{u}'\hat{u}}{n-k-1} = \frac{\sum \hat{u}_j^2}{n-k-1}$$

2.6.3 Coeficiente De Determinación o R^2

Como ya se ha calculado la recta de regresión y se ha revisado si la cada uno de los coeficientes de regresión parcial son estadísticamente significativos mediante las pruebas de hipótesis, en orden para que se vea el nivel de “concordancia con el desarrollo del modelo” se realizara la siguiente pregunta ¿Cómo es posible medir el nivel de exactitud el modelo? O dicho de otra forma ¿Cuán precisas con las variables regresoras al momento de estimar la variable regresada? Para resolver dichas interrogantes es que se desarrolló en concepto de coeficiente de determinación o R^2 .

En principio se tiene que cada observación de la variable regresada puede ser descrita como la suma de su componente estimado más su termino de error.

$$Y_j = \hat{Y}_j + \hat{u}_j$$

Ahora restamos a cada lado de la ecuación el promedio de las observaciones de la variable regresada para luego elevar al cuadrado ambos lados de la ecuación.

$$(Y_j - \bar{y})^2 = [(\hat{Y}_j - \bar{y}) + \hat{u}_j]^2 = (\hat{Y}_j - \bar{y})^2 + 2(\hat{Y}_j - \bar{y})\hat{u}_j + \hat{u}_j^2$$

Y de la misma manera se puede representar la sumatoria de las observaciones de la variable regresada.

$$\sum (Y_j - \bar{y})^2 = \sum (\hat{Y}_j - \bar{y})^2 + 2 \sum (\hat{Y}_j - \bar{y})\hat{u}_j + \sum \hat{u}_j^2$$

Ahora, al centrarse en el siguiente termino, podemos descomponerlo de la siguiente manera:

$$\sum (\hat{Y}_j - \bar{y}) \hat{u}_j = \sum \hat{Y}_j \hat{u}_j - \bar{y} \sum \hat{u}_j$$

Y, haciendo uso de las ecuaciones normales descritas al momento de encontrar los valores de la matriz $\hat{\beta}$ se procede a eliminar el término de la derecha con lo cual la igualdad queda de la siguiente manera:

$$\sum (\hat{Y}_j - \bar{y}) \hat{u}_j = \sum \hat{Y}_j \hat{u}_j = \sum (\hat{\beta}_0 + \hat{\beta}_1 x_{j1} + \hat{\beta}_2 x_{j2} + \dots + \hat{\beta}_k x_{jk}) \hat{u}_j$$

$$\sum (\hat{\beta}_0 + \hat{\beta}_1 x_{j1} + \hat{\beta}_2 x_{j2} + \dots + \hat{\beta}_k x_{jk}) \hat{u}_j = \hat{\beta}_0 \sum \hat{u}_j + \hat{\beta}_1 \sum x_{j1} \hat{u}_j + \hat{\beta}_2 \sum x_{j2} \hat{u}_j + \dots + \hat{\beta}_k \sum x_{jk} \hat{u}_j$$

Y gracias al conjunto de ecuaciones normales, el resultado de todos estos términos se puede igualar a cero reduciendo la ecuación original como se muestra a continuación:

$$\sum (Y_j - \bar{y})^2 = \sum (\hat{Y}_j - \bar{y})^2 + \sum \hat{u}_j^2$$

Las consecuencias que acarrear la interpretación de esta igualdad son de vital importancia ya que muestra que la sumatoria del cuadrado de la variación de la observación con respecto a su promedio se encuentra solo directamente relacionado con la sumatoria del cuadrado de la variación entre el valor estimado de la regresada y el promedio más la sumatoria del cuadrado del error, que expresado de otra forma, la variación total del termino dependiente se encuentra compuesta por la variación “explicada” por la función de regresión muestral más la variación “no explicada”. Para simplificar aún más la expresión anterior esta se expresa a manera de porcentaje.

$$R^2 = \frac{\sum (\hat{Y}_j - \bar{y})^2}{\sum (Y_j - \bar{y})^2} = 1 - \frac{\sum \hat{u}_j^2}{\sum (Y_j - \bar{y})^2}$$

2.6.4 Coeficiente De Determinación Ajustado o \bar{R}^2

Ahora, si bien se habló que este método sirve para ver en qué grado la variable regresada estimada se “acerca” a la variable regresada observada, hay que tener en cuenta que el coeficiente de determinación no es una buena guía para verificar la exactitud de un modelo de regresión, ya que por el simple hecho de agregar una regresora más a la función de regresión (poblacional o muestral), el coeficiente está necesariamente obligado a aumentar su valor, o como mínimo no disminuir. Es bajo esta línea de pensamiento que se desarrolló⁴ el criterio del coeficiente de determinación ajustado o \bar{R}^2 desarrollado por Henri Theil [Cole96], [Gujarati&Porter10]:

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{S_Y^2}$$

Donde:

- $\hat{\sigma}^2$: Varianza residual, nótese de que es el valor estimado de la verdadera σ^2
- S_Y^2 : Varianza muestral de la regresada observada.

Para que el cálculo de este indicador sea más sencillo, es que se plantea de la siguiente manera:

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

Esta ecuación pone de manifiesto el hecho de que el coeficiente de determinación es necesariamente mayor o igual (aunque difícil) al coeficiente de determinación ajustado, pero que nunca el coeficiente \bar{R}^2 será superior a su original. Además, \bar{R}^2 , puede tomar valores negativos, pero esta condición no posee interpretación alguna, para tal caso se asume que \bar{R}^2 es igual a cero.

⁴ Existen otros métodos para resolver la problemática, pero en orden de simplificar el análisis, se optó por usar el coeficiente de determinación ajustado. Para aprender los otros métodos consulte la literatura especializada.

2.7 Validación del Modelo

2.7.1 Multicolinealidad

Dicho de manera informal, un conjunto de variables regresoras posee multicolinealidad cuando estas se encuentran “fuertemente relacionadas entre sí”, lo que conlleva a la siguiente irregularidad que en un principio resulta paradójica: dada una función de regresión muestral se obtiene que el coeficiente de determinación arroja un resultado alto (imagínese sobre 0,7), pero que ninguno de los coeficientes parciales de regresión, después de sus respectivas pruebas de hipótesis, muestran que son estadísticamente significativos [Cole96]. Como se mencionó anteriormente, esta situación resulta paradójica ya que, si el coeficiente de determinación es alto, entonces como mínimo uno de los coeficientes parciales de regresión debe ser estadísticamente significativo. Entonces, cuando se manifiesta esta situación aparentemente imposible es que se presencia un caso de multicolinealidad.

Dicho de manera formal se define la multicolinealidad perfecta cuando existen $k+1$ valores λ , donde por lo menos uno de ellos es distinto de cero donde:

$$\lambda_0 x_0 + \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k = 0$$

Pero resulta que en la mayoría de los casos (si no en todos), es difícil que esta situación se presente, se da a conocer la multicolinealidad “no perfecta” o “imperfecta” que se encuentra definida cuando existen $k+1$ valores λ donde por lo menos uno es distinto de cero y además un término v , de error estocástico (aleatorio) en el que se cumple la siguiente condición:

$$\lambda_0 x_0 + \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k + v = 0$$

Nótese que estas dos igualdades solo analizan la multicolinealidad o grado de multicolinealidad, en el caso de haber una relación imperfecta, cuando las variables regresoras presentan una relación lineal entre ellas.

Como conclusión, este es un problema de información, ya que no podemos saber a priori el grado en que las variables se encuentran relacionadas. Lo que se puede hacer en los casos de que exista una fuerte (o perfecta relación en el peor de los casos), es cambiar una o más variables regresoras, pero se debe tener en cuenta el hecho de que deben ser estadísticamente significativas y que también el coeficiente de correlación debe ser lo suficientemente alto. Como consideración final, el hecho de cambiar una o más variables regresoras, no hace que necesariamente el problema planteado desaparezca, es exige plantar pruebas para medir si el grado de multicolinealidad es lo suficientemente bajo [Cole 96].

2.7.1.1 Test De Multicolinealidad

Como se vio en el apartado anterior, el problema no se trata de que exista o no multicolinealidad perfecta o imperfecta, porque esto es inevitable dada su propia naturaleza, el verdadero problema radica en que si el grado de multicolinealidad sea demasiado alta para que la función de regresión tenga que ser replanteada. Para revisar esto se toma como base la siguiente observación según Klein, citado por [Cole96]:

“La multicolinealidad o inter-correlación entre las variables no es siempre un problema, a menos, que sea alta en relación con el grado general de correlación múltiple entre todas las variables (de la regresión).”

Bajo esta línea de pensamiento, es que se desarrolla la regla de Klein⁵ que dice que la multicolinealidad es un problema cuando:

$$R_i^2 > R^2$$

Donde R_i^2 es el coeficiente de correlación de la variable x_i con respecto a las demás variables regresoras.

Otro método para probar si el grado de multicolinealidad es lo suficientemente grande como para invalidar el modelo de regresión es el factor de inflación de la varianza, que viene dado por la siguiente ecuación:

⁵ G. S. Maddala, *Econometría* (Madrid: McGraw-Hill, 1985), p. 195

$$Var(\hat{\beta}_j) = \frac{S^2}{(n-1)Var(x_i)} \left(\frac{1}{1-R_i^2} \right)$$

Donde:

- $Var(\hat{\beta}_j)$: Varianza muestral de $\hat{\beta}_j$.
- $Var(x_i)$: Varianza muestral de x_i

El término que se encuentra a la derecha se llama “factor de inflación de la varianza” (FIV), y la interpretación que se le da es que mientras más grande el FIV entonces la varianza de los coeficientes de regresión también será más grande. Partiendo de esa base, para que el FIV de cada variable crezca, entonces el coeficiente de correlación de la respectiva variable necesariamente debe crecer, lo que indica una mayor correlación entre las regresoras. Para finalizar la literatura indica que un FIV superior a 10 ($FIV > 10$) indica que la multicolinealidad entre las variables regresoras es lo suficientemente grande para que el modelo deba ser replanteado.

Se aprecia entonces que la multicolinealidad es básicamente un problema de información. Lo que sucede es que estamos pidiendo a los datos más de lo que nos pueden decir. La muestra no contiene suficientes observaciones como para estimar el efecto separado de cada variable explicativa. Puesto que se trata de un problema muestral, es muy poco lo que puede hacerse para resolver el problema si no es posible obtener información adicional. Por ejemplo, se podría pensar en descartar algunas variables explicativas, para romper así la multicolinealidad. Sin embargo, si la variable descartada es una variable relevante, entonces esto podría agravar los problemas, ya que se producirá un sesgo en los coeficientes de las otras x 's. En este caso los coeficientes de las variables no descartadas recogerán su propio efecto, más parte del efecto la variable descartada.

Puesto que el problema multicolinealidad es en última instancia un problema de información insuficiente, se desprende que la única solución real consistirá en obtener más información.

[Cole96]

2.7.2 Homocedasticidad

El hecho de que no exista homocedasticidad (heterocedasticidad) hace relación con el segundo supuesto de la función generalizada de la regresión lineal, específicamente viola el supuesto de que $E(uu') = \sigma^2 I$ haciendo que la varianza σ^2 no sea considerado como una constante y que esta aumente o disminuya a lo largo de las observaciones.

Si bien la violación de este supuesto no afecta de forma directa a los coeficientes parciales de regresión, trae consecuencias desastrosas ya que, si la varianza obtenida no es “confiable”, entonces la prueba de hipótesis con el estadístico t de Student también pierde validez y no se podría decir a ciencia cierta si los coeficientes de regresión parcial son realmente significativos o no.

2.7.2.1 Test De Homocedasticidad

Para asegurar que exista heterocedasticidad (varianza constante), es que se desarrolló el test de White [Cole 96], [Gujarati&Porter10]. Para desarrollarlo se siguen los siguientes pasos:

- Se construye la función de regresión a analizar
- Se calculan los residuos de dicha regresión
- Se construye una regresión, cuya variable regresada son los cuadrados de los residuos mientras que las variables regresoras son las variables regresoras de la función original, los cuadrados de las mismas y los productos cruzados de ellas.
- Se calcula el coeficiente de correlación R^2 de dicha regresión y se multiplica por la cantidad de observaciones n.
- Finalmente, si el producto nR^2 sigue una distribución χ^2 se puede decir entonces que la función de regresión muestral posee heterocedasticidad.

Comúnmente para solucionar este problema la opción más utilizada es la agregación de más variables, de no ser posible, existe un método llamado ajuste de White que lo poseen la mayoría de los paquetes de software estadísticos[Cole96].

2.7.3 Autocorrelación De Los Errores

Existen dos tipos de autocorrelación, pero dada la naturaleza del proyecto solo se hablará de autocorrelación temporal mientras que la espacial será dejada de lado.

Según lo que se habló en la generalización de supuestos de regresión lineal, se desprende del segundo supuesto que la correlación entre dos términos de error debe ser igual a cero, más concretamente:

$$E(u_j u_i) = 0$$

Pero esta situación no siempre se produce, es más, por lo general es bastante extraño o casi imposible que ocurra este caso de forma espontánea debido a la propia naturaleza de los números, como también puede ser causado por una variable o un conjunto de variables no identificadas. Este fenómeno aparece con aun más fuerza en los casos en que los datos poseen una distribución temporal, es decir, se obligan a ordenarse en el tiempo en que la observación fue tomada, de modo de que se produce el efecto de que la observación de una variable regresora tomada en un periodo anterior puede afectar el comportamiento de la variable regresada en periodos actuales. A este fenómeno se le conoce como Autocorrelación temporal. Una forma simple de apreciar esto es analizar los errores con respecto al tiempo y ver si crecen o disminuyen cada cierta periodicidad (autocorrelación positiva) o de lo contrario, si un error produce un efecto en magnitud inversa en periodos posteriores.

2.7.3.1 Test De Autocorrelación

Para detectar la autocorrelación se utilizó uno de los métodos más empleados llamado "test de Durbin-Watson"

Matemáticamente se dice que un modelo posee autocorrelación con el periodo anterior cuando:

$$u_j = \rho u_{j-1} + v_j$$

Donde:

- ρ : coeficiente de autocorrelación, se dice que si $\rho > 0$ entonces se produce autocorrelación positiva, y de manera inversa, si $\rho < 0$ entonces se produce autocorrelación negativa. El intervalo en que se mueve el coeficiente se encuentra entre $|\rho| < 1$
- v : variable aleatoria de media cero y varianza constante

Ahora, estadísticamente, se dice que existe autocorrelación entre los errores cuando al tener la siguiente prueba de hipótesis se rechaza la H_0 , es decir:

- $H_0: \rho = 0$
- $H_1: \rho \neq 0$

Ahora, para rechazar o no la hipótesis, se usa el estadístico de Durbin-Watson, cuyo coeficiente se calcula de la siguiente manera:

$$d = \frac{\sum_{j=2}^n (\hat{u}_j - \hat{u}_{j-1})^2}{\sum_{j=1}^n \hat{u}_j^2} = \frac{\sum_{j=2}^n \hat{u}_j^2 - 2 \sum_{j=2}^n \hat{u}_j \hat{u}_{j-1} + \sum_{j=1}^{n-1} \hat{u}_j^2}{\sum_{j=1}^n \hat{u}_j^2}$$

Este estadístico se compara con las tablas de puntos críticos de Durbin-Watson y si resulta que este estadístico se encuentra entre 0 y 2 se desprende que:

- Si $d > d_U$ entonces no se rechaza H_0 , dando a entender que no existe autocorrelación positiva entre los errores.
- Si $d > d_L$, entonces se rechaza H_0 , dando a entender que si existe autocorrelación positiva entre los errores.
- Si $d_L < d < d_U$, entonces la prueba no es concluyente.

Mientras que si $d > 2$, entonces:

- Si $4 - d > d_U$, entonces no se rechaza H_0 , dando a entender que no existe autocorrelación negativa entre los errores.
- Si $4 - d < d_L$, entonces se rechaza H_0 , dando a entender que existe autocorrelación negativa entre los errores.
- Si $d_L < 4 - d < d_U$, entonces la prueba no es concluyente.

Para finalizar este apartado, cabe indicar que, al rechazar la hipótesis nula, no necesariamente significa que exista autocorrelación, esto puede indicar un problema en la identificación de variables al omitir alguna regresora o que la forma funcional de la regresión se encuentra mal planteada [Cole96].

2.7.3.2 Corrección De Autocorrelación Mediante Cochrane-Orcutt.

Como se mencionó, este proceso se recomienda hacerlo cuando ya se han cambiado variables o cuando ya se ha cambiado la forma funcional de la función. El método de Cochrane-Orcutt sigue una serie de pasos iterativos, pero antes de entrar a describirlos hay que entrar a definir algunos aspectos.

Se supone que se quiere encontrar la siguiente función de regresión lineal poblacional:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + u_t$$

Y, que con existencia de autocorrelación, los errores de cada periodo t quedan definidos de la siguiente forma

$$u_t = \rho u_{t-1} + v_t$$

Si se retarda en un periodo la función de regresión lineal, para luego ser multiplicada a ambos lados de la ecuación por el coeficiente de autocorrelación la igualdad queda:

$$\rho y_{t-1} = \rho \beta_0 + \rho \beta_1 x_{(t-1)1} + \rho \beta_2 x_{(t-1)2} + \dots + \rho \beta_k x_{(t-1)k} + \rho u_{t-1}$$

Si se toma la ecuación original y se resta por su versión retardada entonces tenemos

$$y_t - \rho y_{t-1} = \beta_0 (1 - \rho) + \beta_1 (x_t - \rho x_{(t-1)1}) + \beta_2 (x_{t2} - \rho x_{(t-1)2}) + \dots + \beta_k (x_{tk} - \rho x_{(t-1)k}) + v_t (*)$$

Como se puede ver en esta ecuación, el término de error v_t es aleatorio. Ahora los pasos del método de Cochrane-Orcutt son los siguientes:

- Estimar los coeficientes de la función de regresión muestral mediante el método de los mínimos cuadrados.
- Estimar el coeficiente de autocorrelación mediante:

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

- Sustituir el valor $\hat{\rho}$ en la ecuación (*) para estimar los nuevos coeficientes de regresión y de intercepto ($\beta_i(x_t - \rho x_{(t-1)i})$ y $\beta_0(1 - \rho)$ respectivamente, con $i = 1, 2, \dots, k$).
- Repetir el proceso hasta que el valor del coeficiente de autocorrelación sea inferior a 0,001, cuando se alcanza ese valor se puede decir que es lo suficientemente bajo para que la autocorrelación no sea un problema grave.

El problema al utilizar este método es que hay veces que los coeficientes de regresión parcial contradicen todo razonamiento intuitivo o teórico al quedar con los signos cambiados, pero aun así estos nuevos coeficientes siguen siendo insesgados.

2.8 Otras Variables

2.8.1 Variables Binarias

Hasta ahora se han tratado las variables como si todas ellas tuviesen un carácter cuantitativo, pero hay casos, en el que un modelo de regresión, para que sea suficientemente adecuado a las circunstancias, este debe contemplar un componente cualitativo, es por esto que se desarrollaron las variables binarias, también llamadas dicotómicas, debido al hecho de que estas muestran un aspecto de la realidad que se pueden responder “con un sí o un no”.

Matemáticamente, las variables binarias se tienden a representar de la siguiente manera:

$$x = \begin{cases} 1 & \text{si se cumple alguna condicion} \\ 0 & \text{si no se cumple} \end{cases}$$

Como se ve, la expresión no admite que ambas situaciones se den de manera simultánea en un mismo momento, por eso se dice que las variables binarias son de carácter dicotómico.

2.8.2 Variables Retardadas

Por lo general, para ilustrar el comportamiento de una función lineal (serie temporal en este caso), se simplifica el origen o naturaleza de la variable, por lo tanto, se considera que la variable regresada solo se ve influenciada por las regresoras del mismo periodo. Pero existen casos en que no posee sentido usar para un análisis las variables de un mismo periodo por diversas razones, por ejemplo, la incidencia del IPC, se conoce que este indicador sube o baja dependiendo de las circunstancias, pero una “persona normal” no tiene conciencia de cuanto fue la variación en el periodo actual hasta que este acabe, por lo tanto, esta persona se prepara con la información disponible hasta el momento, es decir, con los valores rezagados o retardados.

2.9 Elección De Método

2.9.1 Justificación Del Modelo Lineal

La razón por la cual se eligió el modelo de regresión lineal por sobre los demás planteados radica en dos fundamentos:

- **Perdida de información:** Para estudios con grandes volúmenes de datos se tiene una regla no formal de que existen datos que fueron tomados de forma errónea o que estos sean lo suficientemente atípicos como para no deban ser incluidos en la función de regresión. A esta situación se suma el hecho que se tiene una fuente de información que no es completamente fiable (en cualquier caso ninguna lo es), en la cual se aplicarán transformaciones logarítmicas donde se aproximarán o truncaran cifras alejando aún más el valor de la estimación del coeficiente muestral con su homónimo poblacional.
- **Aspecto de propósito:** El objetivo último de esta investigación es obtener valores que sirvan como pronóstico para los ingresos de los casinos, por lo tanto no existe propósito en analizar las propiedades de las variantes de la regresión lineal ya que estos solo aportan de manera académica (en este caso).

2.9.1.1 Justificación Del Método De Mínimos Cuadrados Ordinarios

Como quedó demostrado, el método de máxima verosimilitud consigue llegar exactamente a las ecuaciones normales obteniendo las mismas estimaciones para los ponderadores de regresión, además el hecho de que su varianza es asintóticamente insesgada coloca a ambos métodos en una posición de indiferencia si la razón de su elección se debiese únicamente a que tan precisos pudiesen llegar a ser, pero al final se decantó por los mínimos cuadrados ordinarios por razones de familiaridad con el método además de la mayor simplicidad.

Capítulo 3. Variables Regresoras

3.1 Índice de Precios al Consumidor

Antes de entrar en definiciones más precisas, primero se hablará informalmente que significa el índice de precios al consumidor. Como se tiene entendido, cada mes los precios de la mayoría de los productos (sino todos) poseen un cierto movimiento, que puede ser hacia la baja, abaratando el coste de la vida, que no necesariamente es bueno, o hacia el alza, causando que el coste de la vida aumente. A estos fenómenos se les conoce como deflación e inflación respectivamente.

Formalmente, a la inflación se le conoce como el aumento sostenido del nivel de precios, tómese como ejemplo que el kilo de pan al principio del mes de agosto cueste \$1000, por concepto de inflación al principio del mes de septiembre el mismo kilo de pan ya no cuesta el precio original, sino que \$1050, mientras que al inicio del mes de octubre el mismo kilo de pan cuesta \$1103. A lo que se quiere llegar es que al finalizar el primer mes y al comenzar el siguiente, se percibió un aumento en el precio del kilo de pan en un porcentaje de 5%, de la misma forma, se percibe de nuevo un aumento del 5% para el último mes. Si bien este es un ejemplo sencillo ya que no contempla los porqués del aumento del precio del pan (precio de la harina, electricidad, agua, bencina, etc.), es lo suficientemente ilustrativo para describir la mecánica percibida.

Una de las consecuencias más evidentes de la inflación es que las personas, para un mismo nivel de ingreso y un mismo nivel de “cosas” a consumir en un mes, tienen que destinar más recursos, o bien buscar sustitutos más económicos, o simplemente dejar de consumir ciertos bienes que no sean apreciados con el carácter de urgentes para destinar el dinero a otros más necesarios.

Bajo este marco es que nace el índice de precios al consumidor (IPC), cuyos dos grandes propósitos son:

- “Monitorear la evolución de los precios de consumo a lo largo del tiempo (inflación) puesto que su comportamiento impacta en: la política monetaria definida por la autoridad pertinente, la evolución de la Unidad de Fomento (UF), Unidad Tributaria Mensual

(UTM), las fluctuaciones en las tarifas de servicios y la deflactación⁶ del consumo de los hogares en las cuentas nacionales, entre otros usos”.

- “Servir como proxy del costo de vida, lo que permite la indexación de contratos”, dicho de otra forma, que el valor del IPC que toma cada mes no tiene mucho valor en sí, pero sus derivados e implicaciones son necesarias para el diario vivir, que en el caso acá propuesto, habla de “actualizar los contratos” mediante la tasa de inflación de un periodo.

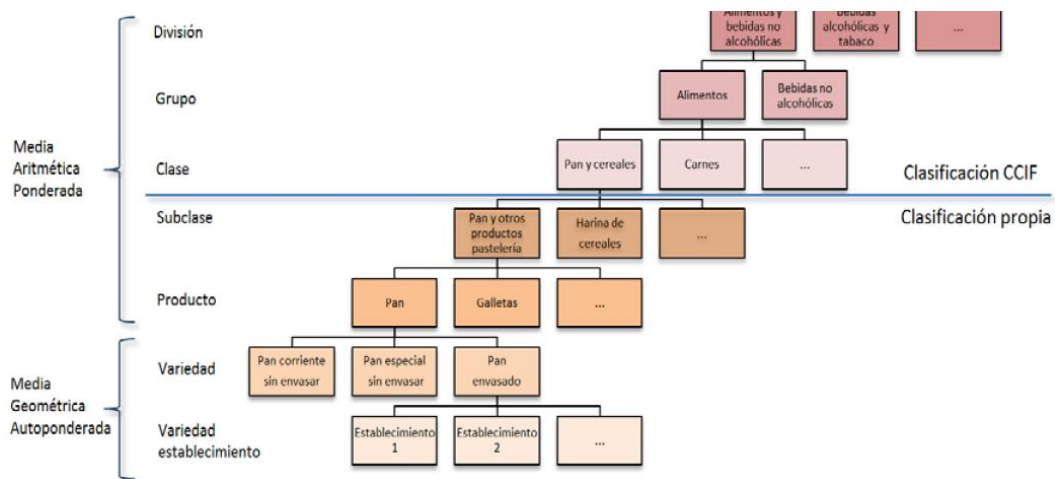
⁶ La deflactación consiste en transformar un valor en pesos corrientes a pesos constantes, es decir, tomar un valor y dividirlo por la tasa de inflación más un entero. En el caso de Chile la inflación viene dada por el IPC.

3.1.1 Cálculo del IPC

Facilitando el cálculo, los conceptos asociados al cálculo del IPC no suenan complejos, ya que dichos de manera simplificada se trata de calcular un porcentaje mediante una división entre la sumatoria de los precios de un periodo y la sumatoria de los precios un periodo base. Actualmente este periodo base es el año 2013.

Si se quisiese ser más riguroso, primero se elige un conjunto de productos, llámese carne de vacuno, snacks, harina, cereal, arroz, etc. La decisión de que producto elegir se hace en base a las preferencias de las personas específicamente, estos productos se reúnen en una sola sub-clase, posteriormente un conjunto de sub-classes se reúnen en una clase, que a su vez definen un grupo, para que finalmente un conjunto de grupos afines se identifique en una división. Al total de divisiones, con todas sus partes (grupos, clases, sub-classes y productos) es lo que se llama la canasta de productos. Se debe tener en cuenta que, si bien existe un producto, como por ejemplo el pan, este también se puede diferenciar por variedades de pan, ya sea hallulla, marraqueta, baguete, frica, sin sal, amasado, de fibra, kosher, etc. Y a su vez estas variedades pueden ser diferenciadas por el establecimiento de venta.

Ilustración 3 Subdivisión IPC



Fuente: Manual metodológico IPC, Instituto nacional de estadísticas

Como se puede apreciar en la imagen existen dos tipos de clasificaciones:

- Clasificación del consumo individual de finalidades (CCIF): sistema de categorización desarrollada por el sistema de cuentas nacionales (SCN) del año 1993, a su vez desarrollado por la comisión de estadística de las naciones unidas. Esta clasificación propone los tres primeros conjuntos para la confección del IPC, es decir, división, grupo y clase.
- Clasificación propia: son los grupos que fueron creados por el instituto nacional de estadísticas. Como se puede apreciar, este modo de clasificar abarca los siguientes cuatro conjuntos, a saber: sub-clase, producto, variedad y variedad establecimiento.

A día de hoy, existen, con base a la última información relativa entregada, existen 12 divisiones, 41 grupos, 88 clases, 137 sub-clases y 321 productos. Para efectos de este estudio, solo se considerarán los valores del IPC separado por divisiones, las cuales son:

Ilustración 4 Tabla de subdivisión IPC

DIVISIÓN	Ponderación canasta IPC (%)
1. Alimentos y bebidas no alcohólicas	19,05855
2. Bebidas alcohólicas y tabaco	3,31194
3. Vestuario y calzado	4,48204
4. Vivienda y servicios básicos	13,82810
5. Equipamiento y mantención del hogar	7,02041
6. Salud	6,44131
7. Transporte	14,47381
8. Comunicaciones	5,00064
9. Recreación y cultura	6,76121
10. Educación	8,08996
11. Restaurantes y hoteles	4,37454
12. Bienes y servicios diversos	7,15749

Fuente: Manual metodológico IPC, Instituto nacional de estadísticas

Como se puede apreciar en la figura, las divisiones se ponderan para el cálculo final del IPC.

El proceso de construcción del IPC se divide en dos etapas:

Primera etapa: antes de poder calcular cualquier tipo de indicador ya sea el índice global (IPC) o los que separan por grupos, se hace necesario tener la información mensual sobre las compras de un conjunto de familias o entidades, luego estas se asocian según el orden que ya ha sido establecido. Después de este paso, se construyen los primeros indicadores llamados “agregados elementales” que no son más que “un micro índice que corresponde a un ratio (relativo) de precios”⁷. Estos agregados elementales se calculan mediante una media geométrica que matemáticamente se encuentra escrito de la siguiente manera:

$$p_i^t = \prod_{j=1}^n \left(\frac{v_j^t}{v_j^{t-1}} \right)^{1/n}$$

Donde:

- p_i^t : Precios en el periodo t de los bienes y servicios i.
- v_j^t : Precio de la variedad j en el periodo t.
- v_j^{t-1} : Precio de la variedad j en el periodo t-1.
- n : Número de variedades.

Segunda etapa: el siguiente paso consiste simplemente en agrupar esos índices elementales mediante el índice de Laspeyres, que no es más que un promedio ponderado que se escribe de la siguiente manera:

$$p^t = \frac{\sum_{i=1}^N p_i^t q_i^0}{\sum_{i=1}^N p_i^0 q_i^0}$$

Donde:

- p_i^0 : Precios en el periodo base de los bienes y servicios i.

⁷ Véase http://www.ine.cl/canales/chile_estadistico/estadisticas_precios/ipc/base_2013/antecedentes_metodologicos/manual_metodologico_IPC_base_2013.pdf

- q_i^0 : Cantidad de bienes y servicios en el periodo base de los bienes y servicios i.
- N : Número total de bienes de la canasta.

Dados los indicadores construidos en estos procesos, se puede decir que estos son los valores del IPC separado por grupos. Para construir el indicador global se ocupa la siguiente formula

$$IPC(t) = I^t = \left[\sum_{i=1}^N \frac{p_i^t}{p_i^0} w_i \right] * 100 = \left[\sum_{i=1}^N I_i^t w_i \right] * 100$$

Donde:

- w_i : Ponderación del gasto del producto i. los valores para cada grupo se pueden observar en la tabla anterior.
- I_i^t : Relativo de precios del producto i en el periodo t

3.2 Índice de remuneraciones (IR)

Se define como:

- “este índice mide la evolución mensual de las remuneraciones ordinarias por horas ordinarias pagadas a los trabajadores con contratos definidos o indefinidos en las empresas”.

Entendiéndose como remuneraciones ordinarias⁸ solo aquellas que involucran los pagos que se realizan de forma regular, descontando todos aquellos pagos que no se consideren regulares.

⁸ En conformidad con http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/remuneraciones/050310/Manual_Metodologico_IR_base_anual_2009_V2_marzo2016.pdf

Ilustración 5 Estructura conceptual IR



Fuente: Manual metodológico IPC, Instituto nacional de estadísticas

Mientras que la hora ordinaria⁹ se define como las horas pagadas en el mes al trabajador por prestar efectivamente sus servicios en conformidad al contrato. La jornada ordinaria de trabajo, no puede exceder las cuarenta y cinco horas semanales.

⁹ En conformidad con http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/remuneraciones/050310/Manual_Metodologico_IR_base_anual_2009_V2_marzo2016.pdf

3.2.1 Cálculo del IR

Para el cálculo del índice de remuneraciones se ocupa la siguiente fórmula

$$IR^t = \sum_{i=1}^n \left[\left(\frac{(RO/HO)_i^t}{(RO/HO)_i^0} \right) * \left(\frac{(RO/HO)_i^0 * HO_i^0}{\sum_{i=1}^n (RO/HO)_i^0 * HO_i^0} \right) \right] = \sum_{i=1}^n \left(\frac{(RO/HO)_i^t}{(RO/HO)_i^0} \right) * W_i$$

Donde¹⁰:

- $(RO/HO)_i^t$: remuneraciones ordinarias por horas extraordinarias pagadas en el periodo t, de los grupos de trabajadores i
- $(RO/HO)_i^0$: Remuneraciones ordinarias por horas ordinarias pagadas en el periodo base (año 2009 = 100), de los grupos de trabajadores i
- HO_i^0 : Horas ordinarias pagadas en el año base, para los grupos de trabajadores i
- W_i : ponderación de las remuneraciones ordinarias pagadas en el año 2009, de los grupos de trabajadores i.

Si se realiza un análisis simplificado de la ecuación, mientras más grande sea remuneración para el periodo t, manteniendo la cantidad de horas para el mismo periodo fijas, con respecto al año base el aumento del IR será más grande y a su vez, si la remuneración disminuye entonces el IR también disminuiría. Si se hace un análisis similar pero esta vez al aumentar o disminuir las horas ordinarias para el periodo t mientras se dejan fijas las remuneraciones del mismo periodo, con respecto a un año base el IR disminuirá o aumentará correspondientemente. Concretamente, la remuneración se encuentra en relación directa con el índice mientras que las horas se encuentran en relación inversa. Para finalizar este apartado se debe indicar que el IR posee un uso potencial dentro del modelo ya que, si este indicador sube o baja, significa que los potenciales clientes de los casinos de juego tendrán más o menos capacidad para para gastar dinero dentro de las instalaciones.

¹⁰ En conformidad con http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/remuneraciones/050310/Manual_Metodologico_IR_base_anual_2009_V2_marzo2016.pdf

3.3 Índice del costo de la mano de obra (ICMO)

Se define¹¹ este índice como

- “Evolución mensual del costo de la mano de obra por horas totales pagadas. Gastos en que incurren las empresas en mantener a sus trabajadores. El cual incluye las remuneraciones ordinarias más costos no salariales.”

3.3.1 Cálculo del ICMO

Para calcular el índice se utiliza la siguiente fórmula:

$$ICMO^t = \sum_{i=0}^n \left[\left(\frac{(C/HT)_i^t}{(C/HT)_i^0} \right) * \left(\frac{(C/HT)_i^0 * HT_i^0}{\sum_{i=1}^n (C/HT)_i^0 * HT_i^0} \right) \right] = \sum_{i=0}^n \left(\frac{(C/HT)_i^t}{(C/HT)_i^0} \right) * W_i$$

Donde

- $(C/HT)_i^t$: Costo de la mano de obra por horas totales en el periodo t, del grupo de trabajadores i.
- $(C/HT)_i^0$: Costo de la mano de obra en el periodo base para el grupo de trabajadores i.
- HT_i^0 : Horas totales pagadas en el periodo base para el grupo de trabajadores i.
- W_i : Ponderación de las horas ordinarias pagadas en el año 2007 para los grupos de trabajadores i

De la ecuación se desprende que si al aumentar el costo de la mano de obra para un periodo dado mientras se mantiene fijo las horas totales del mismo periodo, con respecto al año base el aumento del índice será mayor, y a su vez, si este mismo costo de la mano de obra disminuye, con respecto al año base el índice disminuye. De la misma forma se puede realizar un análisis similar, pero con la cantidad de horas totales, pero con la diferencia de que

¹¹ En conformidad con http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/remuneraciones/050310/Manual_Metodologico_IR_base_anual_2009_V2_marzo2016.pdf

se presentaría una relación inversa, es decir, al aumentar las horas el índice disminuye y viceversa. Esta variable posee un potencial uso dentro del modelo ya que, dicho de manera informal, si aumenta o disminuye el ICMO, los posibles usuarios de los casinos pueden gastar más o menos dentro de las instalaciones (no debe olvidarse que el ICMO contempla las remuneraciones).

3.3.2 Recolección de información

Para que el INE pueda dar cuenta del índice, el método de recolección de información se realiza mediante encuestas que se realizan a las empresas que forman parte de la muestra seleccionada, en estas se consulta por las remuneraciones brutas de los trabajadores que se encuentren bajo contrato. Además, se consulta por los costos asociados a los trabajadores (por mano de obra), la cantidad de hombres y mujeres trabajando en la empresa y por horas ordinarias y extraordinarias. Luego se agrupan los trabajadores de cada empresa por las funciones que desempeñan. Nótese que estos datos no solo sirven para generar un índice global, sino que también para generar índices segmentados, es más el índice de remuneración general se compone mediante estos índices segmentados. El proceso completo, es decir, desde el levantamiento de la información hasta la publicación de los resultados posee un desfase aproximado de un mes, por ejemplo: si se quiere publicar el índice de remuneración del mes de enero, la recolección se realiza en febrero para finalmente publicarse los resultados los primeros días de marzo.

3.4 Tasa de desempleo

Para medir el nivel de desempleo o desocupación en el país de una manera que sea suficientemente simple sin perder significancia se desarrolló el concepto de la tasa de desempleo cuya formulación matemática se escribe como:

$$TD = \frac{PD}{FT}$$

Donde:

- TD: Tasa de desempleo o desocupación
- PD: Población desocupada. Población que se encuentra en edad de trabajar y que al momento de realizar la encuesta no se encontraban ocupadas. Nótese que estas personas han estado buscando trabajo durante las últimas cuatro semanas incluyéndose la semana cuando se realiza la encuesta y que a su vez, se encuentran disponibles para trabajar dentro de las dos semanas posteriores a la encuesta.
- FD: Fuerza de trabajo. Personas que cumplen con la edad para poder trabajar y que cumplen con la categoría de desocupada u ocupada.

Cabe señalar que la tasa de desocupación puede segmentarse por región, sexo, actividad económica, etc. Pero para efectos de esta investigación solo se considerará el índice global.

3.5 Índice de percepción del consumidor (IPeCo)

Este índice, lo realiza cada mes el centro de estudios de economía y negocios, también conocido como CEEN (dependiente de la universidad del desarrollo) en conjunto con Mall Plaza desde diciembre de 2001. Lo que el IPeCo muestra cada periodo, como su nombre lo indica, es la confianza que poseen los consumidores en el consumo de bienes y servicios tomando como periodo base el mes de diciembre del año 2001. Si bien el aspecto más reconocido del índice es la parte global también se puede segmentar por estrato socioeconómico, por región, etc. Para efectos de esta investigación se considerará solo el índice global.

3.6 Índice mensual de actividad económica (IMACEC)

Este índice, como su nombre lo indica, mide la evolución mensual de la economía del país, y para lo cual utiliza alrededor del 90% de los bienes y servicios que componen el producto interno bruto (también conocido como PIB). Para el cálculo del indicador se toma como periodo base el año 2008. Como sucede con los demás indicadores antes mencionado, el IMACEC puede ser segregado, esta vez por actividad económica, pero solo se considerará el índice global para efectos de esta investigación

3.7 Métodos de pronóstico de las variables explicativas

Uno de los problemas al analizar los valores futuros de una función de regresión, es que los valores de las regresoras no se conocen para periodos futuros, por lo tanto, estas variables también deben ser estimadas mediante técnicas de pronóstico. Para el caso de esta investigación se considera que son de carácter autoexplicado o bien se consideran como series de tiempo cuyo término independiente, valga la redundancia, es el tiempo.

Se considerarán las siguientes formas de pronóstico:

3.7.1 Promedio Móvil Simple

Se conoce el promedio o media móvil como la suma de n observaciones donde n es la cantidad de periodos hacia atrás en el tiempo. Matemáticamente se describe como:

$$F_{t+1} = \frac{A_t + A_{t-1} + \dots + A_{t-n}}{n}$$

Donde:

- F_{t+1} : Valor pronosticado para el periodo $t+1$.
- A_{t-i} : Observaciones para cada periodo anterior al pronóstico t con $i = 0, 1, 2, \dots, n$.
- n : Cantidad de observaciones.

nótese que mientras más grande la cantidad de observaciones, la variabilidad del pronóstico F_t va a ser menor, tendiendo hacia la media, cosa que no necesariamente sea mejor, por tanto, se deben probar con cantidades de observaciones distintas para de encontrar una que se acerque lo suficiente, es decir que el error de pronóstico no sea considerado un impedimento para el investigador.

3.7.2 Suavización exponencial de Holt-Winters

La ecuación presentada a continuación, se conoce como la suavización exponencial simple:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t)$$

Donde:

- \hat{Y}_{t+1} : Valor pronosticado para el periodo t+1 (periodo siguiente).
- α : Constante de suavizamiento, este valor se encuentra entre 0 y 1.
- Y_t : Observación para el periodo t (periodo actual).

Definido lo anterior, se da a conocer la suavización exponencial de Holt-Winters:

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)S_{t-1}$$

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p}$$

Donde a las tres primeras ecuaciones se les conoce como los estimados de nivel, tendencia y estacionalidad respectivamente y a la última como la ecuación de pronóstico [Hanke&Wichern06]. Además:

- L_t : Estimación de nivel para el periodo t (actual)
- α : Constante de suavizamiento para el estimado de nivel, este valor se encuentra entre 0 y 1.

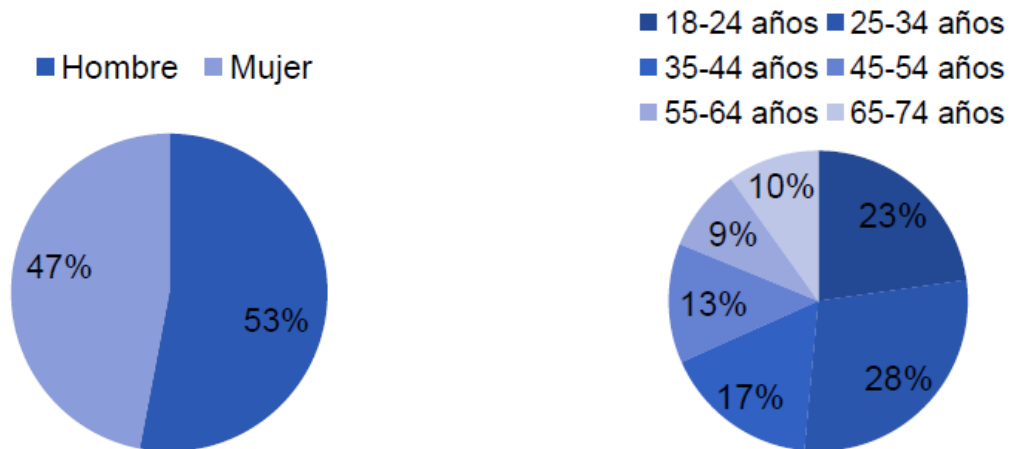
- Y_t : Observación para el periodo t
- T_t : Estimado de tendencia para el periodo t
- β : Constante de suavizamiento para el estimado de tendencia, este valor se encuentra entre 0 y 1.
- S_t : Estimado de estacionalidad para el periodo t
- γ : Constante de suavizamiento para el estimado de estacionalidad, este valor se encuentra entre 0 y 1.
- p : Número que indica la cantidad de periodos en el futuro a pronosticarse.
- \hat{Y}_{t+p} : Pronostico para p periodos en el futuro.

Capítulo 4. Sobre los asistentes a los Casinos

Los siguientes análisis fueron realizados en base a un estudio realizado por la consultora Visión Humana llamado Radiografía de los asistentes a casinos de juegos 2009 y el mismo estudio para el año 2011. Ese informe fue basado otro estudio que realiza la misma consultora llamado Chilescoopio.

Los principales elementos estadísticos que describen el perfil de los asistentes a casinos de juegos son los siguientes:

Ilustración 6 Segmentaciones de los asistentes



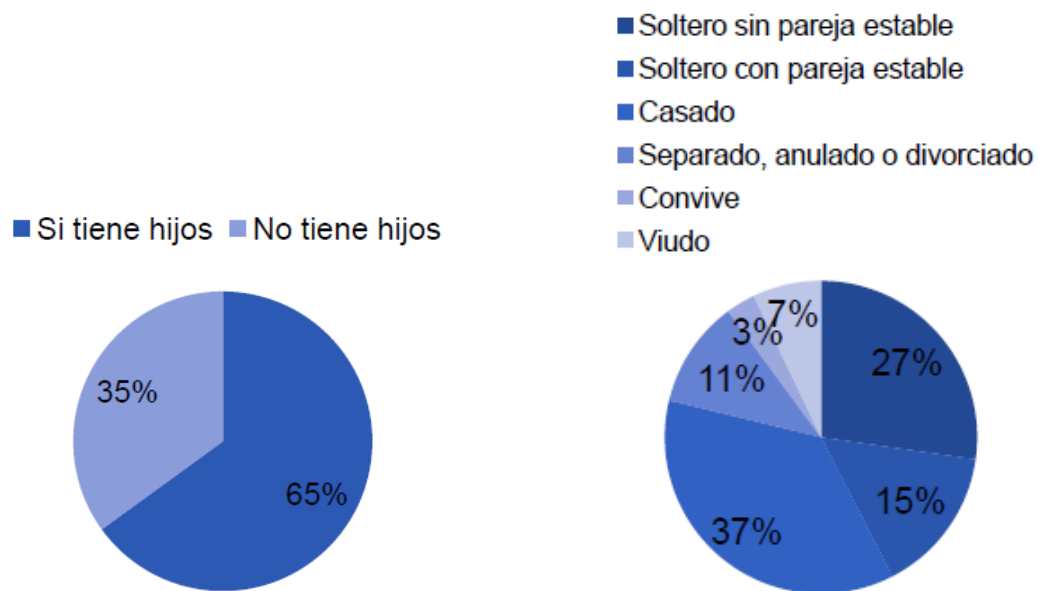
Fuente: Radiografía de los asistentes a casinos de juegos

Para el siguiente análisis primero es necesario tener en cuenta lo siguiente, según el compendio estadístico¹² de la población considerada como posibles asistentes, esto es, personas mayores a 18 el porcentaje de Hombres y de mujeres es de 49% y 51% respectivamente

¹² Compendio estadístico 2012 elaborado por el INE, población total de 18 años o más estimada al 30 de junio de 2012

en el país, en otros términos, existe un mayor porcentaje de mujeres en el público objetivo, lo anterior realza el hecho de que los asistentes sean mayoritariamente varones. De la siguiente segmentación se extrae que en el rango etario de 18 a 54 se encuentra el 81% de los asistentes, considerando además un informe del instituto de estadísticas el cual define, siguiendo además la lógica de lo que se podría suponer previamente, que el rango etario entre 25 y 54 poseen un mayor poder adquisitivo¹³ lo que los convierte en el público clave del rubro.

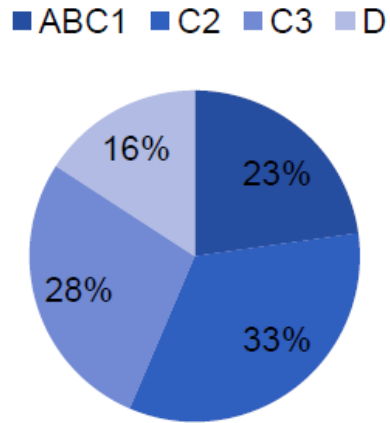
Ilustración 7 Segmentaciones de los asistentes



Fuente: Radiografía de los asistentes a casinos de juegos

Los datos obtenidos de estos gráficos se condicen con lo esperado con respecto a la tenencia de hijos debido al rango etario, 25 y 54 años, anteriormente definida, con respecto al estado civil los resultados extraídos del mencionado estudio indican que en la mayoría de los asistentes se encuentra casado.

¹³ Ingresos de hogares y personas 2009, instituto nacional de estadísticas de Chile

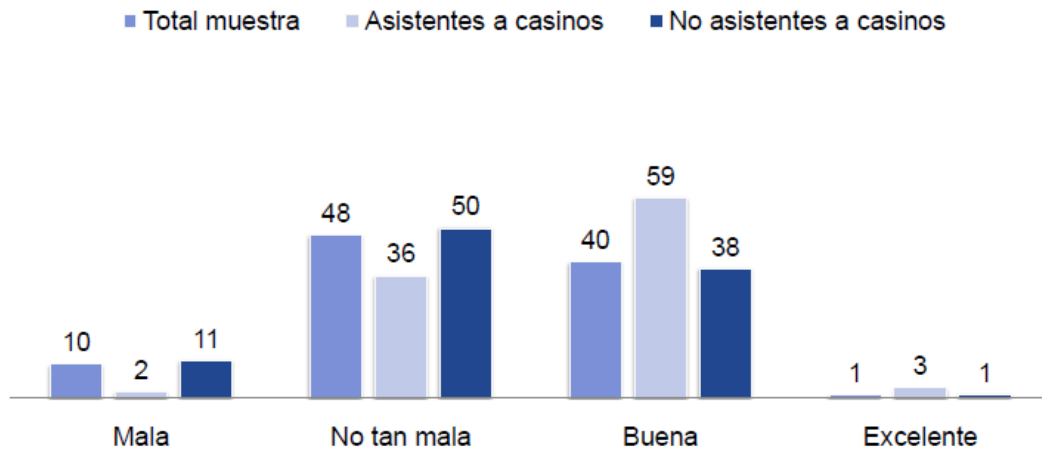
Ilustración 8 Segmentaciones de los asistentes

Fuente: Radiografía de los asistentes a casinos de juegos

Este gráfico da un resultado racionalmente esperado, pues bajo un análisis de lógica intuitiva podemos en primera instancia entender o, mejor dicho, asociar a los gastos que asumen los asistentes a los casinos como gastos de diversión o esparcimiento, lo que por consiguiente son gastos de segunda necesidad, se podría debatir acerca de esta afirmación pues los juegos de azar pueden provocar en algunos casos adicción¹⁴ y para un asistente que padezca esto, el gasto en el casino para él, puede considerarse como un gasto de primera necesidad y priorizar este por sobre otras necesidades más apremiantes, pero excluyendo este caso del análisis, las personas que tienen la oportunidad de cubrir sus necesidades básicas y también sus necesidades secundarias entre estas, las actividades de esparcimiento, son aquellas que poseen un mejor nivel económico más acomodado.

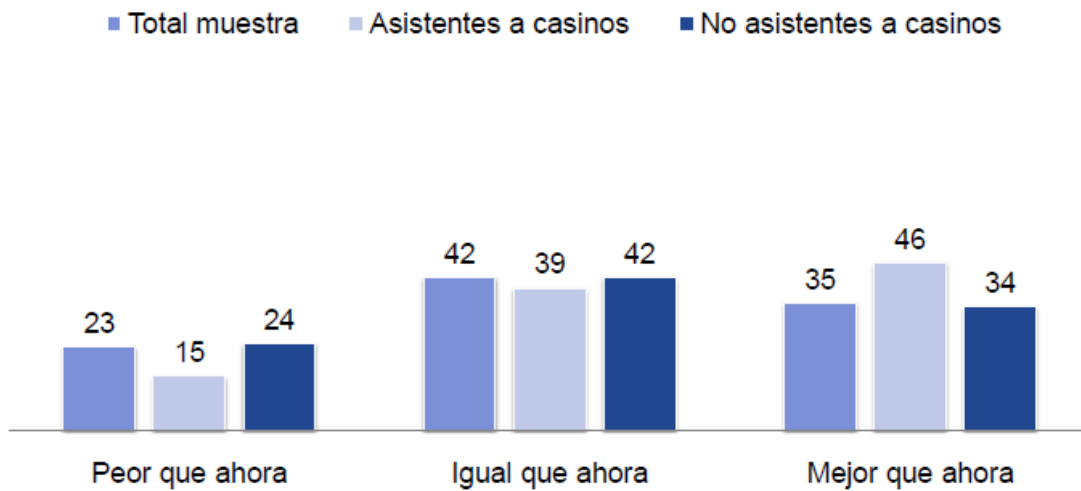
¹⁴ Ludopatía es la adicción a los juegos de azar

Ilustración 9 Percepción de la situación económica personal actual



Fuente: Radiografía de los asistentes a casinos de juegos

Ilustración 10 Percepción de la situación económica personal futura

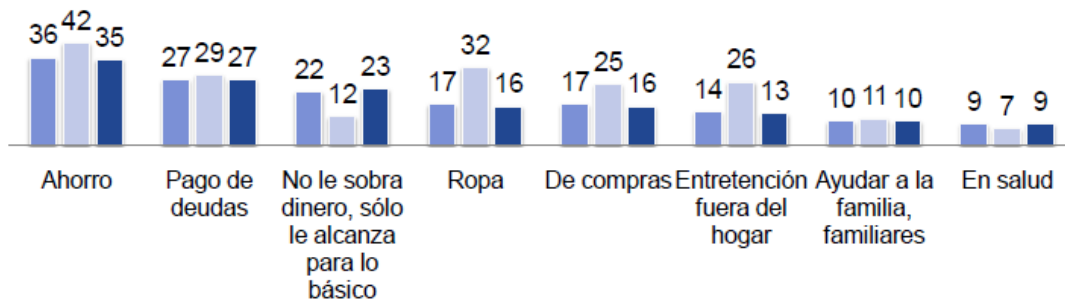


Fuente: Radiografía de los asistentes a casinos de juegos

Los análisis anteriores acerca de la percepción actual y futura sobre la situación económica personal de los asistentes a los casinos de juegos son complementarios con la ilustración 8 y reafirman que los asistentes a los casinos son y se perciben como personas con un situación económica capaz de gastar e invertir dinero en actividades de recreación y esparcimiento y además se agrega una nueva información, que se manifiestan en general de manera optimista con respecto a la situación económica futura, esta información nos llevó a analizar índices de percepción económica de los consumidores.

Ilustración 12 En que gasta si existiese dinero sobrante

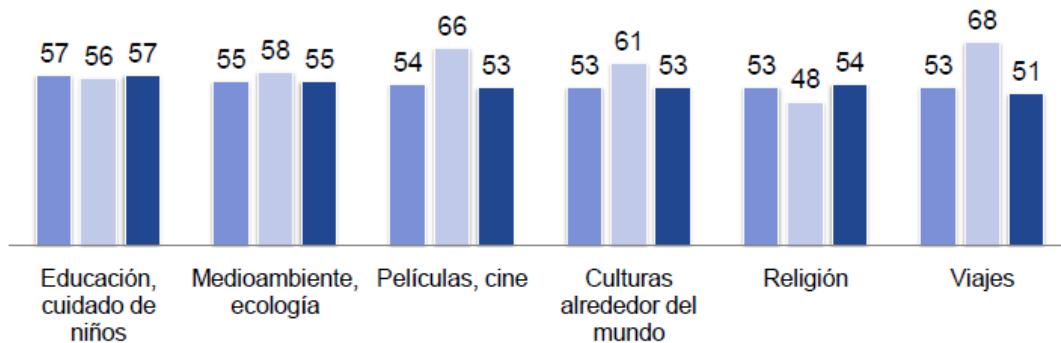
■ Total ■ Si ha asistido al casino ■ No ha asistido al casino



Fuente: Radiografía de los asistentes a casinos de juego

Ilustración 11 Intereses personales en cada tema

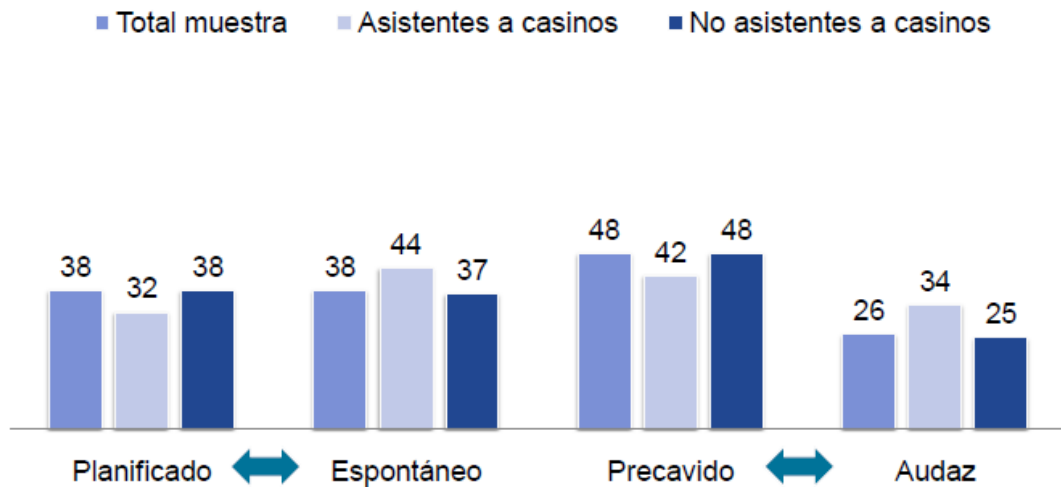
■ Total muestra ■ Asistentes a casinos ■ No asistentes a casinos



Fuente: Radiografía de los asistentes a casinos de juego

Los anteriores gráficos nos dan una información complementaria acerca de los intereses particulares de los asistentes a los casinos de juegos, intereses sobre los cuales se podría argumentar de manera cualitativa la selección o descarte de posibles variables regresoras, como, por ejemplo, se podrían formar argumentos cualitativos a favor de asociar al modelo la variable de IPC de Recreación y cultura.

Ilustración 13 Como describirían su propia personalidad



Fuente: Radiografía de los asistentes a casinos de juego

La información contenida en el anterior gráfico nos da un bosquejo de la personalidad del asistente al casino, que extrapolado a términos financieros nos habla de un consumidor compulsivo, aquí hay que aclarar que se pueden clasificar las compras del consumidor en dos tipos [Monzo00], la primera planificada, en la cual podríamos definir de la siguiente manera: *“El consumidor actúa de forma racional, busca argumentos sólidos y convincentes para justificar la compra de un producto. La relación calidad-precio es fundamental. Fidelidad al producto/marca que ha ganado su confianza”*¹⁵. Y la compra compulsiva de la siguiente forma: *“El consumidor actúa de forma impulsiva, compra por su atractivo estético o cualquier otra circunstancia sensible que lleve asociada. Predomina una conducta emotiva, tiende a desplazar*

¹⁵ <http://www.mailxmail.com/cursos-manual-gestion-comercio-segunda-parte/compra-razonada-compra-impulsiva>

la decisión de compra hacia aspectos simbólicos del producto (receptores idóneos de campañas de publicidad de imagen del producto y de anuncios con líderes de opinión en lo que identificarse)”¹⁶. Se pueden entonces afirmar con cierto grado de confianza que los asistentes a los casinos de juegos son mayormente propensos a realizar compras compulsivas y sobre esta conducta también se pueden asociar variables que sirvan para formular el modelo.¹⁷

¹⁶ <http://www.mailxmail.com/curso-manual-gestion-comercio-segunda-parte/compra-razonada-compra-impulsiva>

¹⁷ En términos prácticos no se utilizó este raciocinio para definir o buscar variables puesto que por ejemplo, se podría haber planteado la teoría que los ingresos de los casinos podrían asociarse a los ingresos de la heladerías o bien al comportamiento de la industria textil juvenil en el país, pero dichas variables representarían un nuevo desafío al momento de proyectarla con un cierto grado de confianza para que el modelo finalmente prediga los ingresos de los casinos, razón por la cual se privilegió el uso de variables que se pudieran proyectar sin que esta proyección demande un esfuerzo desmedido y que la incertidumbre de la misma reste confianza a la proyección objetivo.

Capítulo 5. Análisis de las variables

Las siguientes variables, regresoras y regresadas serán con las que se estructurará el modelo econométrico

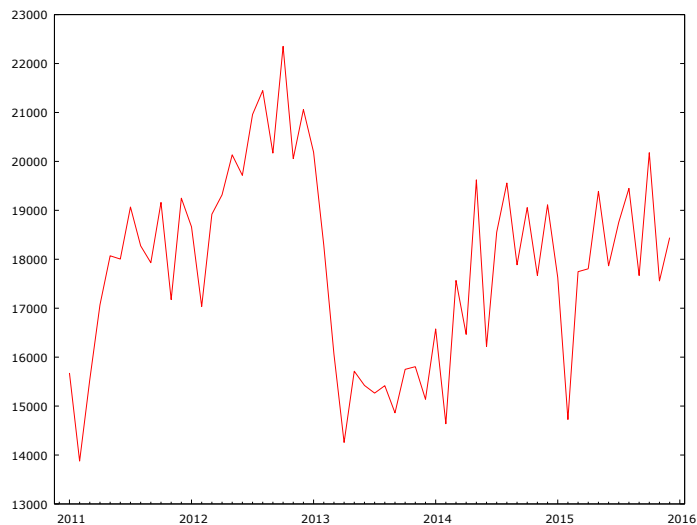
5.1 Variable regresada

La variable regresada o a explicar, queda definida al momento de la definición de la situación, es el objetivo del trabajo y los esfuerzos están dirigidos a la mejor explicación de esta.

5.1.1 Ingresos brutos de los casinos de juegos

Para el presente modelo representa como variable regresada a los ingresos brutos de los 8 casinos con mayores ventas de los últimos 5 años, y fue denominada “*Ingresos Top MM*”, La unidad en la cual esta representa es de Millones de Pesos (MM).

Ilustración 14 Variable regresada



Fuente: Elaboración Propia

Tabla 1 Principales estadísticos variable regresada

Estadístico	Valor
<i>Media</i>	17.854
<i>Mediana</i>	17.908
<i>Mínimo</i>	13.876
<i>Máximo</i>	22.351
<i>Desviación típica</i>	1.959,2

Fuente: Elaboración Propia

5.2 Variables regresoras

Dentro del universo de variables posibles que se pueden testar para la elaboración de un modelo se seleccionan todas aquellas que podrían estar relacionadas con la variable dependiente, regresora o a explicar, es decir, bajo un análisis cualitativo se plantea que dentro de un ecosistema económico compuesto por diversos factores, ¿Cuáles podrían o no estar relacionadas con el nivel de ingresos?, por ejemplo, se podría de alguna plantear que el nivel del agua caída, lluvia, incide en los ingresos que pueda percibir la industria, y quizás en cierta medida en algún caso extremo, en un mes extremadamente lluvioso podrían verse mermados los ingresos, pero en general no pareciera haber una mayor relación entre el nivel de lluvia con los ingresos que pueda tener la industria, en cambio, un indicador que refleje los ingresos de sus asistentes bajo cualquier análisis económico pareciera si tener una importante incidencia en los posibles ingresos que podría tener la industria. Bajo este criterio se buscaron y estructuraron distintas variables macroeconómicas para formular un modelo que permitiese proyectar los ingresos del casino, es necesario tener en consideración que las variables seleccionadas deben ser susceptibles a ser proyectadas con un cierto de nivel de confianza, en la medida de lo posible, ya que esto permitirá al modelo proyectar los ingresos disminuyendo posibles sesgos arrastrados desde las variables regresoras.

5.2.1 Índice Mensual de Actividad Económica

Indicador Mensual de Actividad Económica, IMACEC, incluye el 90 por ciento de los bienes y servicios que integran el Producto Interior Bruto (PIB).¹⁸ Fue considerada una posible variable pues es un índice que pretende reflejar la actividad económica del país en el cual opera la industria operada, y una contracción o una expansión de la misma podría reflejar una peor o mejor perspectiva para el desarrollo de la industria.

5.2.2 Tasa de Desempleo

Cálculo: Población de 15 años y más que no está trabajando y busca trabajo dividido por la población económicamente activa de 15 años y más (ocupados más desocupados).¹⁹ Se consideró como posible variable debido a que, bajo una perspectiva económica el nivel de personas que perciban o no dinero podrían afectar la variable estudiada.

5.2.3 Reservas en Dólares del Banco Central

Las reservas internacionales son los activos líquidos en moneda extranjera que mantiene el Banco para apoyar su política monetaria y cambiaria. Así, se constituye en uno de los instrumentos de que dispone para cumplir con el objetivo permanente de velar por la estabilidad de la moneda y el normal funcionamiento de los pagos internos y externos.²⁰ Se consideró bajo las posibles variables explicativas porque mide en cierta medida la estabilidad del país.

¹⁸ <http://www.hacienda.cl/glosario/imacec.html>

¹⁹ celade.cepal.org/redatam/PRYESP/SISPPI/Webhelp/tasa_de_desempleo.htm

²⁰ http://www.bcentral.cl/es/faces/ofinanciaras/greservaint/resena?_adf.ctrl-state=lrzs01zxn_4&_afLoop=85642793461821&_afWindowMode=0&_afWindowId=lrzs01zxn_1#!

5.2.4 Dólar Observado

El dólar observado es el promedio de todas las transacciones ejecutadas el día anterior por los bancos y empresas en Chile. Esta variable al igual que las anteriores miden desde una perspectiva la realidad económica del país, siendo considerada en las posibles variables macroeconómicas involucradas.

5.2.5 Visitas totales

Si entendemos a los ingresos del casino como el producto del número de visitas por el gasto promedio por visitas se explica la existencia de esta variable entre las regresoras debido a que su ausencia obligaría a las demás variables proyectar tanto el aumento del gasto por visitante como el flujo de visitantes

5.2.6 Índice de remuneraciones

El índice de remuneraciones muestra la tasa de cambio mensual que poseen las remuneraciones ordinarias por hora ordinaria de trabajo. La razón por la cual este índice se consideró para el análisis, es porque si este aumenta, se podría esperar que las personas al verse con una mayor capacidad adquisitiva, opten por reservar una mayor parte de sus ingresos en actividades recreativas, como por ejemplo el tiempo destinado a juegos de azar que ofrecen los casinos de juegos.

5.2.7 Índice de costo de mano de Obra

Siguiendo con la misma línea de pensamiento, el índice de costo de mano de obra presenta una interpretación similar al índice de remuneraciones, lo que lo hace un candidato para ser una variable dentro del modelo. Este indicador se puede interpretar como una tasa de cambio que indica la evolución de los costos en los que incurren las empresas para mantener sus empleadores, incluyendo remuneraciones ordinarias en adición a costos no salariales. De manera informal, se espera que el dinero destinado a la recreación sea directamente proporcional a este índice, justificando su potencial inclusión al modelo.

5.2.8 Índice Precios del Consumidor

Mencionado anteriormente, el índice de precios del consumidor (o IPC) refleja un aumento sostenido en el coste de vida de las personas, de forma intuitiva, se espera que cada vez que aumente, los asistentes a los establecimientos guarden cierto recelo al invertir dinero en esta forma de entretenimiento causando una disminución en los ingresos de la industria, justificando su posible inclusión al modelo de regresión.

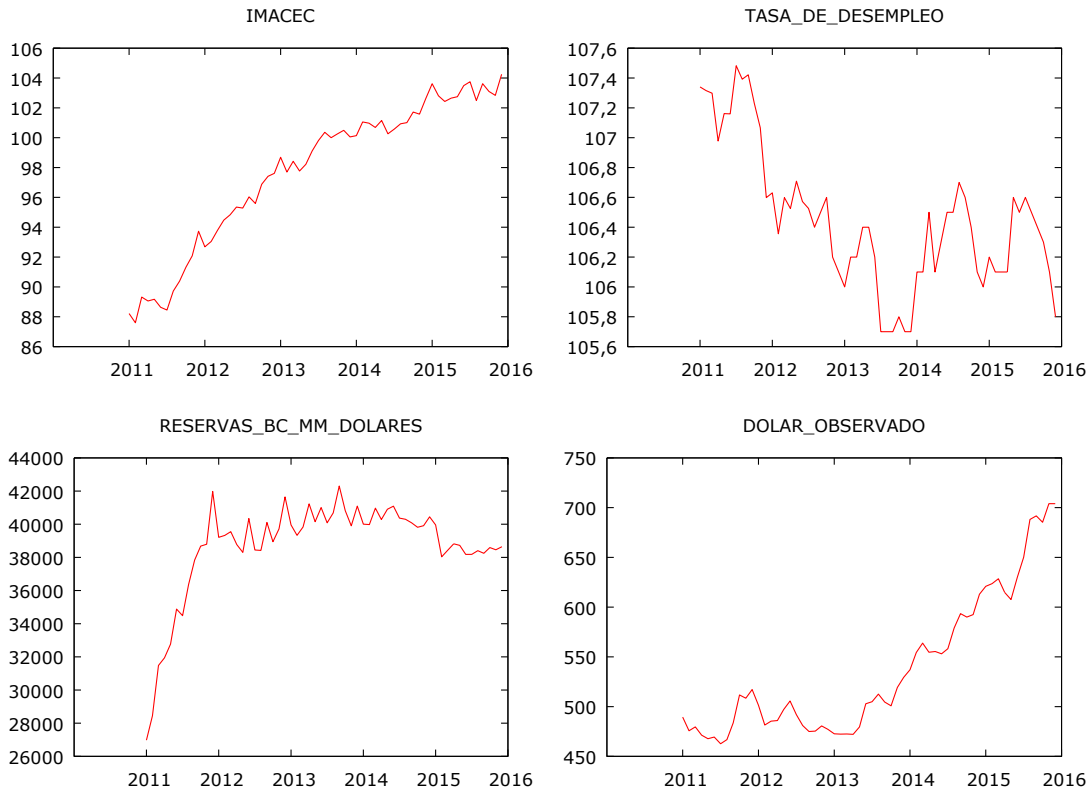
5.2.9 Índice de percepción del consumidor

el índice de percepción del consumidor es realizado con la finalidad de evaluar la idea que posee la gente en relación a la situación económica, el desempleo y los ingresos futuros. Se considera como una posible variable dentro del modelo debido a su posible implicación con la industria, de estar en relación directa con los ingresos de los casinos de juegos.

5.2.10 Existencia de la ley de tabaco

Como se puede apreciar, existe un considerable desplome en los ingresos de los casinos de juego en el año 2013, año en que se promulgó la ley N° 20.600 que modifica la ley N° 19.419 (Regula actividades que indica relacionadas con el tabaco) en donde obliga de forma expresa que no se debe fumar en casinos de juego, específicamente el literal h del artículo 11 señala que no se puede fumar en pubs, restaurantes, discotecas y casinos de juego con excepción de sus patios o espacios al aire libre. Se considera que esta variable binaria incide de forma inversa (convirtiéndola en candidata), es decir, que la modificación de la ley produjo la baja en los ingresos.

Ilustración 15 Variables regresoras



Fuente: Elaboración Propia

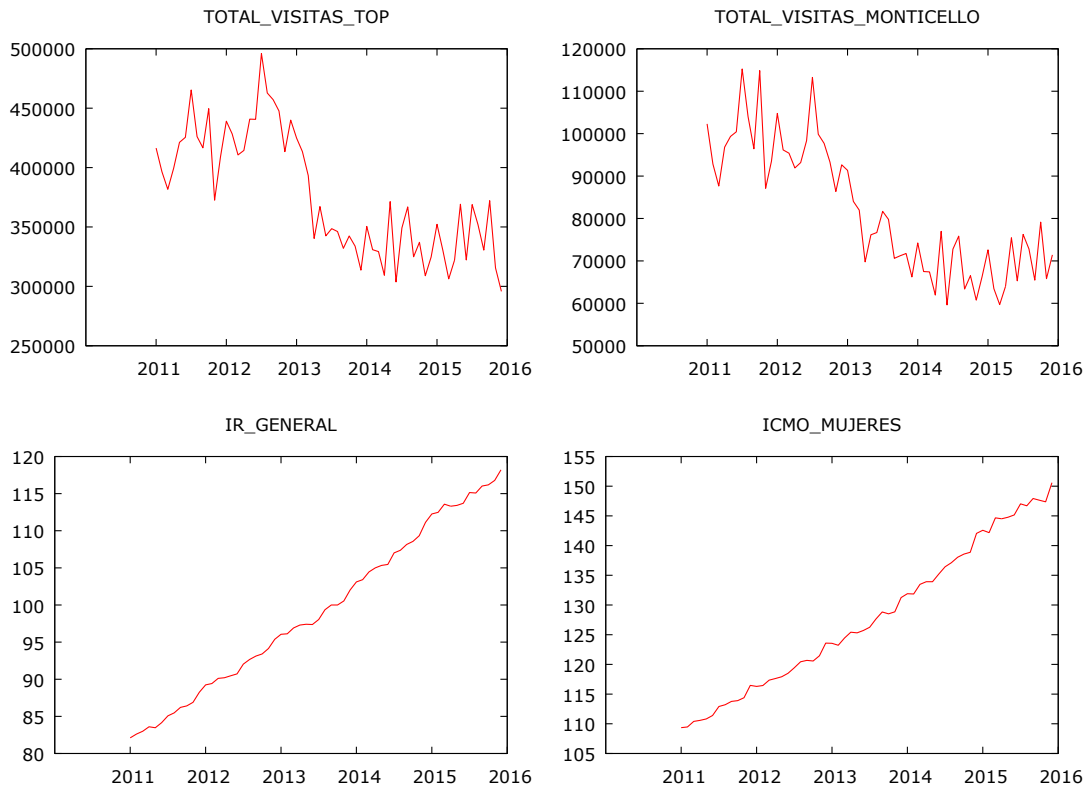
Tabla 2 Principales estadísticos variables regresoras

Variable	Media	Mediana	Mínimo	Máximo	Des Típica
<i>Imacec</i>	97,701	99,462	87,596	104,25	4,9120
<i>Tasa de Desempleo</i>	106,45	106,40	105,70	107,48	0,47510
<i>Reservas BC</i>	38.677	39440	26977	42303	2985,8
<i>Dólar Observado</i>	537,97	507,03	462,64	704,00	70,626

Fuente: Elaboración Propia

Fuente: Elaboración Propia

Ilustración 16 Variables regresoras



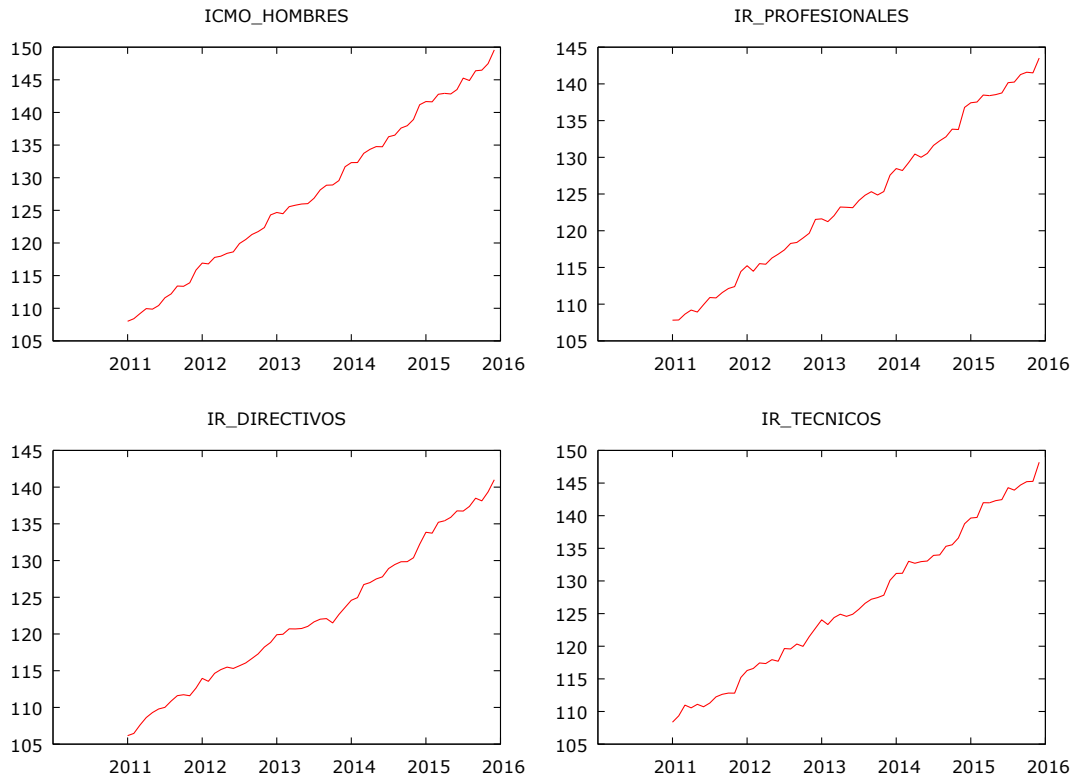
Fuente: Elaboración Propia

Tabla 3: Principales estadísticos variables regresoras

Variable	Media	Mediana	Mínimo	Máximo	Des Típica
<i>Total Top Visitas</i>	376.880	368.990	295.750	495.960	50.812
<i>Total Visitas Mont.</i>	81.989	78.057	59.626	115.250	15.251
<i>IR General</i>	99,229	97,975	82,110	118,20	10,831
<i>ICMO Mujeres</i>	127,97	125,98	109,35	150,59	12,283

Fuente: Elaboración Propia

Ilustración 17 Variables regresoras



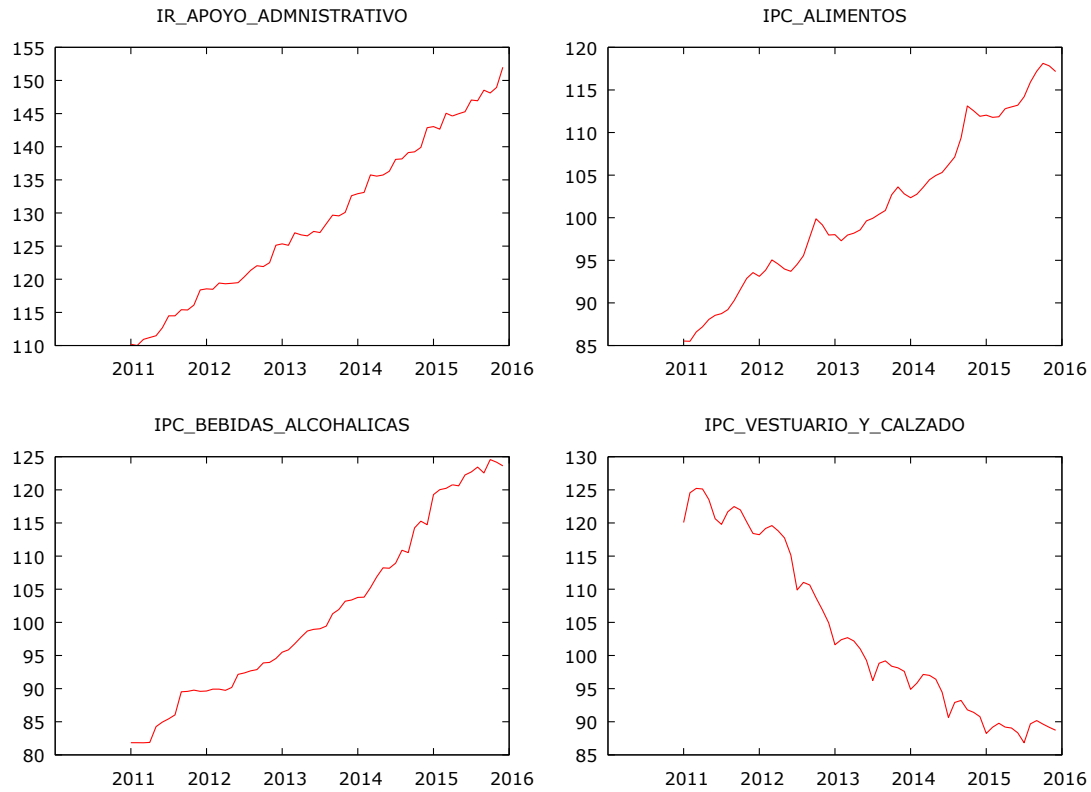
Fuente: Elaboración Propia

Tabla 4 Principales estadísticos variables regresoras

Variable	Media	Mediana	Mínimo	Máximo	Des Típica
<i>ICMO Hombres</i>	127,6	126,44	108,01	149,58	12,034
<i>IR Profesionales.</i>	124,58	123,68	107,82	143,52	10,611
<i>IR Directivos</i>	122,42	121,28	106,16	141,00	9,7138
<i>IR Técnicos</i>	126,73	125,29	108,36	148,19	11,421

Fuente: Elaboración Propia

Ilustración 18 Variables regresoras



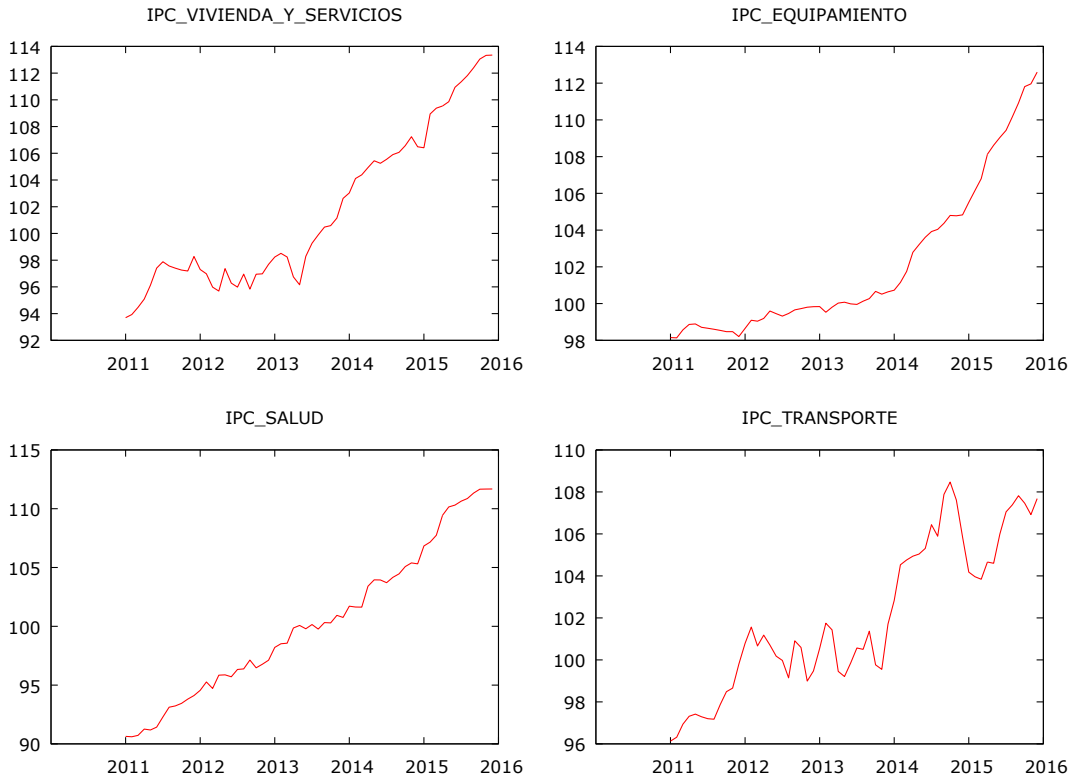
Fuente: Elaboración Propia

Tabla 5 Principales estadísticos variables regresoras

Variable	Media	Mediana	Mínimo	Máximo	Des Típica
<i>IR Administrativo</i>	129,13	127,13	110,00	152,00	12,057
<i>IPC Alimentos.</i>	101,26	99,901	85,493	118,12	9,3323
<i>IPC Beb. Alcohólicas</i>	101,58	98,987	81,833	124,56	13,310
<i>IPC Vestuario</i>	103,61	99,227	86,830	125,22	12,726

Fuente: Elaboración Propia

Ilustración 19 Variables regresoras



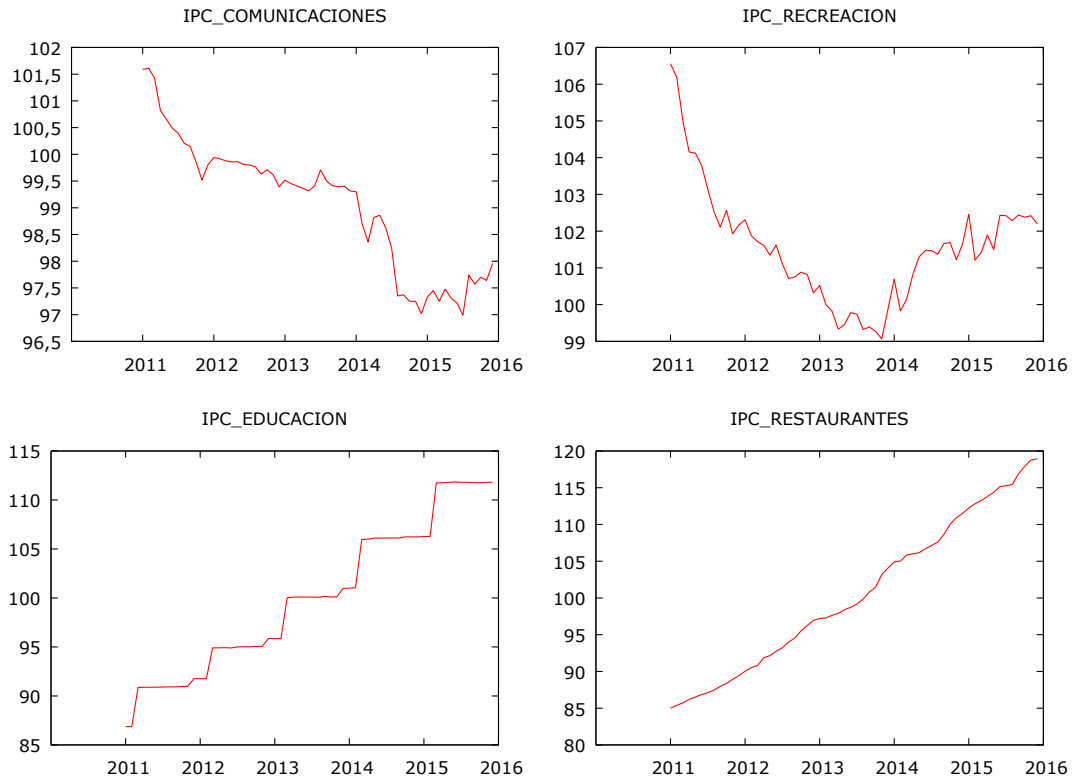
Fuente: Elaboración Propia

Tabla 6 Principales estadísticos variables regresoras

Variable	Media	Mediana	Mínimo	Máximo	Des Típica
<i>IPC Vivienda</i>	101,70	98,887	93,693	113,35	5,8700
<i>IPC Equipamiento</i>	102,13	100,05	98,135	112,60	4,1332
<i>IPC Salud</i>	100,32	99,965	90,606	111,68	6,3379
<i>IPC Transporte</i>	102,03	101,04	96,133	108,47	3,5506

Fuente: Elaboración Propia

Ilustración 20 Variables regresoras



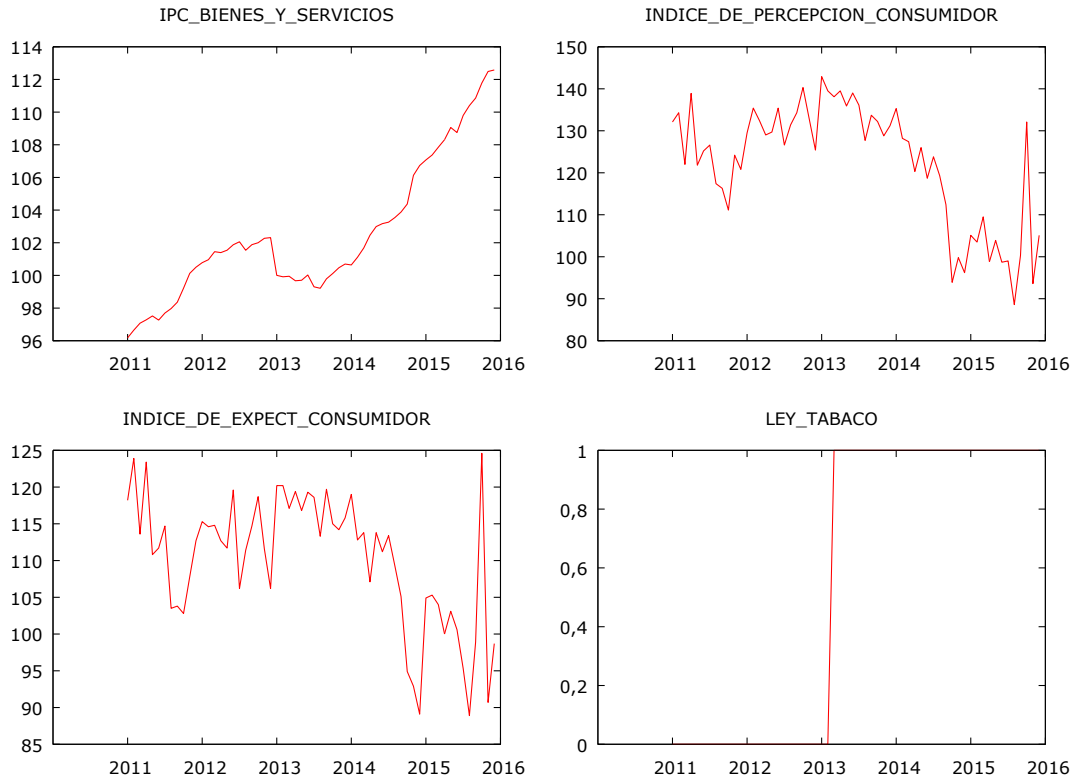
Fuente: Elaboración Propia

Tabla 7 Principales estadísticos variables regresoras

Variable	Media	Mediana	Mínimo	Máximo	Des Típica
<i>IPC Comunicación</i>	99,061	99,41	96,99	101,61	1,23
<i>IPC Recreación</i>	101,56	101,49	99,07	106,56	1,57
<i>IPC Educación</i>	100,08	100,09	86,87	111,84	7,61
<i>IPC Restaurantes</i>	100,58	98,95	84,99	118,93	10,32

Fuente: Elaboración Propia

Ilustración 21 Variables regresoras



Fuente: Elaboración Propia

Tabla 8 Principales estadísticos variables regresoras

Variable	Media	Mediana	Mínimo	Máximo	Des Típica
<i>IPC Bienes y Serv</i>	102,52	101,50	96,177	112,58	4,2487
<i>Perc. Consumidor</i>	122,28	126,60	88,60	142,90	14,451
<i>Expect. Consumidor</i>	110,19	112,70	88,90	124,60	8,8303
<i>Ley Tabaco (Bin)</i>	0,567	1,00	0,00	1,00	0,49972

Fuente: Elaboración Propia

Capítulo 6. Elección de Variables

6.1 Marco de Acción

Después de trabajar en el primer entendimiento de la industria, siguiendo la metodología que se utilizó para enfrentar el siguiente trabajo de tesis, de “Comprensión de la situación” que consistió en realizar esfuerzos para comprender la estructura y el método global de funcionamiento de la industria en el país, considerando siempre que la percepción global puede variar y presentar diferencias al analizar cada caso en particular. Dentro de este trabajo de comprensión se concluyó lo siguiente:

- a) Los gastos en que incurren los clientes del casino está considerado como “gastos de esparcimiento” este tipo de gastos presentan una demanda elástica, es decir, son susceptibles a cambios debido factores externos, entre ellos la priorización del consumidor, y las expectativas de estos con respecto a sus ingresos.
- b) En el año 2010 debido a un fuerte terremoto que afectó la zona centro-sur del país, varios casinos producto de los daños sufridos a raíz de este, no pudieron operar de manera normal, esto sumado a que produjo una contracción económica y un posterior rebote asociado a la recuperación de la infraestructura ²¹ hace valido plantear que los efectos reales de este evento sobre el comportamiento de la variable estudiada son difíciles de medir y cuantificar y una mala interpretación podría traducirse en sesgos e imprecisiones del modelo, motivo por el cual se decidió presentir del mencionado año.
- c) Se consideró de relevancia que en la búsqueda de variables regresoras estas tuvieran la virtud de contar con proyecciones públicas o en su defecto que las proyecciones de estas no requirieran un esfuerzo desmedido o bien, que los resultados obtenidos de la proyección de la variable regresora tuvieran un nivel aceptable de confianza.

²¹ Alberto Ramos, economista senior en Goldman Sachs, Nueva York.

- d) Con el fin de evitar sesgos en los análisis, y homogenizar la muestra se consideraron los 8 casinos más importantes del país, en términos de flujo de ingresos brutos.

una vez conformadas las bases de datos con todas las variables previamente elegidas se procede a conformar el modelo predictivo de la siguiente manera:

1. Se cargaron todas las bases al software estadístico Gretl.
2. Se escogió la forma funcional de la ecuación a formar.
3. Se escogió el método para encontrar los parámetros de la función.
4. Se ejecutó el método
5. Se analizarán las pruebas estadísticas del valor P y se comparan con la significancia esperada
6. Se descartan las variables que obtengan menos significancia que lo deseado
7. Se conforma el modelo con las variables que aprueban
8. Se evalúa la bondad del modelo
9. De haber inconsistencia en el modelo realizan los ajustes pertinentes
10. Se define el Modelo

Como primera medida se definió un valor de significancia de 0.01, posteriormente se ejecutó por primera vez el software arrojando los siguientes resultados

Ilustración 22 Resultados Gretl

Modelo 26: MCO, usando las observaciones 2011:01-2015:12 (T = 60)
Variable dependiente: INGRESOSTOPMM

	Coefficiente	Desv. Típica	Estadístico t	Valor p
const	-124807	89522,6	-1,394	0,1729
IMACEC	138,556	188,020	0,7369	0,4665
TASADEESEMPL	185,606	537,698	0,3452	0,7322
ReservasBCMillon~	0,131180	0,111246	1,179	0,2470
DAlarpromedioobs~	1,71309	8,14979	0,2102	0,8348
TOTALVISITASTOP	0,0365242	0,00432156	8,452	1,17e-09 ***
IRGENERAL	-1205,35	584,883	-2,061	0,0475 **
ICMOMUJERES	645,323	348,176	1,853	0,0731 *
ICMOHOMBRES	821,489	664,680	1,236	0,2255
IRPROFESIONALES	27,2878	355,017	0,07686	0,9392
IRDIRECTIVOS	208,624	255,665	0,8160	0,4205
IRTECNICOS	-574,446	239,173	-2,402	0,0223 **
IRAPOYOADMNIS	-113,894	313,437	-0,3634	0,7187
IPCALIMENTOS	385,336	109,960	3,504	0,0014 ***
IPCBEBIDASALCOHA~	-49,2041	130,131	-0,3781	0,7078
IPCVESTUARIYOYCAL~	63,7490	64,7423	0,9847	0,3322
IPCVIVIENDAYSERV~	127,206	126,364	1,007	0,3217
IPCEQUIPAMIENTOY~	-296,109	281,598	-1,052	0,3009
IPCSALUD	-416,236	215,124	-1,935	0,0619 *
IPCTTRANSPORTE	-124,904	117,901	-1,059	0,2973
IPCCOMUNICACIONES	80,9687	446,050	0,1815	0,8571
IPCRECREACIANYCU~	607,201	228,677	2,655	0,0122 **
IPCEDUCACIAN	569,161	97,7375	5,823	1,81e-06 ***
IPCRESTaurantesy~	188,221	236,172	0,7970	0,4313
IPCBIENESYSERVIC~	-306,397	188,996	-1,621	0,1148
INDICEDEPERCEPCI~	97,1085	39,7287	2,444	0,0202 **
INDICEDEXPECTAT~	-112,628	49,8912	-2,257	0,0309 **
LEYTABACOBIN	-3896,15	776,315	-5,019	1,88e-05 ***
Media de la vble. dep.	17853,75	D.T. de la vble. dep.	1959,219	
Suma de cuad. residuos	10729482	D.T. de la regresión	579,0478	
R-cuadrado	0,952624	R-cuadrado corregido	0,912650	
F(27, 32)	23,83126	Valor p (de F)	1,66e-14	
Log-verosimilitud	-447,9612	Criterio de Akaike	951,9223	
Criterio de Schwarz	1010,564	Crit. de Hannan-Quinn	974,8603	
rho	-0,292665	Durbin-Watson	2,425297	

Fuente: Elaboración Propia

Como se puede apreciar en la columna 5 se observa el valor p de contraste, el cual, si es muy elevado, ósea mayor a 0.01 (1%), indica que existe una probabilidad mayor de que la hipótesis nula sea cierta, recordemos que la hipótesis nula es la suposición de que el parámetro asociado a la variable sea cero, es decir, la probabilidad de que el valor para el parámetro β_i sea igual a cero, es mayor que el tolerado.

Con el fin de reducir al mínimo las opciones de descartar variables significantes que su aporte este siendo opacado por variables que no presentan trascendencia se optó por la estrategia de descartar una a una cada variable, descartando a la que presente el valor p más elevado.

Con el fin de reducir la presentación de la elaboración del modelo se agrupo los valores P de cada variable en una tabla de resumen:

Tabla 9 Análisis de significancia

Variable	Valor P
IMACEC	0,6376
Tasa de desempleo	0,1986
Reservas BC (Millones de dólares)	0,2442
Dólar promedio observado	0,5978
Total Visitas Top Casinos	9,98e-08**
IR General	1,84e-013**
ICMO Mujeres	0,0643
ICMO Hombres	0,0002*
IR Profesionales	0,9392
IR Directivos	0,7543
IR Técnicos	0,0526
IR Apoyo administración	0,8824
IPC Alimentos	0,0001*
IPC Bebidas Alcohólicas y tabaco	0,6607
IPC Vestuario y Calzado	0,3885
IPC Vivienda y servicios básicos	0,6365
IPC Equipamiento y mantención del hogar	0,3631
IPC Salud	2,44e-06*
IPC Transporte	0,3300
IPC Comunicaciones	0,0010*
IPC Recreación y cultura	1,06e-05**
IPC Educación	2,20e-09*
IPC Restaurantes y hoteles	0,1885
IPC Bienes y servicios	0,0169
Índice de percepción del consumidor	0,3533
Índice de Expectativas	0,2564
Ley tabacos (Binaria)	2,52e-05**

Fuente: Elaboración Propia

Acá se nos presentó una decisión, puesto que 9 variables resultaron ser lo suficientemente nobles como para proyectar el modelo, pero el objetivo de abrir en divisiones la variable IPC fue de encontrar una más específica al modelo propuesto, ya que entre ellas presentan una considerable colinealidad, mismo caso sucede con las variables ICMO hombres e IR general, motivo por el cual se escogieron bajo un criterio racional aquellas que tienen relación con la variable a proyectar, Se seleccionó primeramente el IR General, por sobre el ICMO puesto que el IR es un índice que refleja el aumento de remuneración país, lo que bajo un análisis económico simple, se concluye que ante un aumento en este índice, se debiera esperar un aumento en los ingresos, así mismo se descartó ICMO puesto que es una variable que además de medir el costo de la mano de obra, también indica el aumento del costo productivo, es decir, si bien ambas tienden a reflejar un aumento en las remuneraciones, el ICMO además agrega otros factores en relación del costo de mano de obra. Se seleccionó también a IPC de recreación y cultura puesto que refleja el aumento del precio en los costos de recreación que están relacionados con los gastos de esparcimientos de los casinos de juegos. Ahora bien, cualesquiera de estas nueve variables presentaban un nivel de significancia aceptable para incluirse en el modelo, la elección de uno sobre otro radica en la percepción del equipo para argumentar de mejor manera una defensa económica sobre la elección de las mismas.

6.1.1 Test de Multicolinealidad

Como se definió en el marco teórico la multicolinealidad, ósea la correlación entre variables se puede medir usando el factor de inflación de la varianza, arrojando los siguiente resultados

Ilustración 23 Resultados Gretl

```
Factores de inflación de varianza (VIF)
Mínimo valor posible = 1.0
Valores mayores que 10.0 pueden indicar un problema de colinealidad

      LEY_TABACO      8,169
TOTAL_VISITAS_TOP    4,753
      IR_GENERAL_3    3,755
      IPC_RECREACION_1 1,409

VIF(j) = 1/(1 - R(j)^2), donde R(j) es el coeficiente de correlación múltiple
entre la variable j y las demás variables independientes
```

Fuente: Elaboración Propia

Tal como se aprecia, no existe una correlación significativa

6.1.2 Contraste De White

Con el fin de medir la Heterocedasticidad del modelo, se utilizó el contraste de White, el cual arroja un valor de contraste de 21,6639, y para el número de observación el valor crítico de una variable chi-cuadrado con 13 grados de libertad y 5% en cola derecha es 22.362 por lo que se concluye que no existe heterocedasticidad en esta regresión.

Ilustración 24 Resultados Gretl

Contraste de heterocedasticidad de White
MCO, usando las observaciones 2011:04-2015:12 (T = 57)
Variable dependiente: uhat^2

	Coefficiente	Dev. Típica	Estadístico t	Valor p
const	3,65364e+09	2,52103e+09	1,449	0,1545
LEY_TABACO	-2,37910e+08	2,50008e+08	-0,9516	0,3466
TOTAL_VISITAS_TOP	285,726	942,239	0,3032	0,7632
IR_GENERAL_3	1,67209e+06	4,82988e+06	0,3462	0,7309
IPC_RECREACION_1	-7,42259e+07	4,75580e+07	-1,561	0,1259
X2_X3	16,1251	64,0787	0,2516	0,8025
X2_X4	549866	486988	1,129	0,2651
X2_X5	1,78341e+06	2,32951e+06	0,7656	0,4481
sq_TOTAL_VISITAS~	-0,000211371	0,000243760	-0,8671	0,3907
X3_X4	-2,58645	1,72942	-1,496	0,1421
X3_X5	1,17182	8,66645	0,1352	0,8931
sq_IR_GENERAL_3	-30375,0	12677,5	-2,396	0,0210 **
X4_X5	48001,3	60612,6	0,7919	0,4327
sq_IPC_RECREAC~_1	340082	220884	1,540	0,1310

ATENCIÓN: ;matriz de datos casi singular!

R-cuadrado = 0,380070

Estadístico de contraste: $TR^2 = 21,663981$,
con valor p = $P(\text{Chi-cuadrado}(13) > 21,663981) = 0,060792$

Fuente: Elaboración Propia

6.1.3 Test de Durbin Watson

En la prueba de Durbin Watson, con el fin de detectar si existiese autocorrelación en los residuos del modelo se encontró la prueba no era concluyente, esto es, que el test no era capaz de asegurar que no habían errores con retardo, si bien la presencia de autocorrelación, más aún consideran la escala de tiempo que se consideró para la proyección, no es un problema infranqueable, se procedió de todas maneras a sanear esta inconsistencia.

Ilustración 25 Resultados Gretl

Modelo 9: MCO, usando las observaciones 2011:04-2015:12 (T = 57)
Variable dependiente: INGRESOS_TOP_MM

	Coefficiente	Desv. Típica	Estadístico t	Valor p	
LEY_TABACO	-3363,94	639,119	-5,263	2,61e-06	***
TOTAL_VISITAS_TOP	0,0267857	0,00501980	5,336	2,01e-06	***
IR_GENERAL_3	211,376	23,5541	8,974	3,22e-012	***
IPC_RECREACION_1	-106,847	30,7398	-3,476	0,0010	***
Media de la vble. dep.	18002,32	D.T. de la vble. dep.	1886,522		
Suma de cuad. residuos	47384145	D.T. de la regresión	945,5371		
R-cuadrado	0,997462	R-cuadrado corregido	0,997319		
F(4, 53)	5208,009	Valor p (de F)	4,53e-68		
Log-verosimilitud	-469,3558	Criterio de Akaike	946,7116		
Criterio de Schwarz	954,8838	Crit. de Hannan-Quinn	949,8876		
rho	0,234541	Durbin-Watson	1,528749		

Contraste de heterocedasticidad de White -
Hipótesis nula: No hay heterocedasticidad
Estadístico de contraste: LM = 21,664
con valor p = P(Chi-cuadrado(13) > 21,664) = 0,0607919

Fuente: Elaboración Propia

6.1.3.1 Método de Cochare-Orcutt

Se procedió entonces a eliminar el problema detectado anteriormente utilizando este método, los resultados de este fueron satisfactorios, permitieron eliminar la autocorrelación y no presento diferencias en los test de durbin-watson, multicolinealidad y en el contraste de White

Ilustración 26 Resultados Gretl

Realizando el cálculo iterativo de rho...

ITERACIÓN	RHO	SCR
1	0,23454	4,40324e+007
2	0,29001	4,38539e+007
3	0,30128	4,38466e+007
4	0,30354	4,38463e+007
5	0,30400	4,38462e+007
6	0,30409	4,38462e+007
7	0,30411	4,38462e+007

Modelo 6: Cochrane-Orcutt, usando las observaciones 2011:05-2015:12 (T = 56)

Variable dependiente: INGRESOS_TOP_MM

rho = 0,304106

	Coefficiente	Desv. Típica	Estadístico t	Valor p
IR_GENERAL_3	220,936	32,3216	6,836	9,05e-09 ***
IPC_RECREACION_1	-139,609	38,7714	-3,601	0,0007 ***
LEY_TABACO	-2948,01	724,508	-4,069	0,0002 ***
TOTAL_VISITAS_TOP	0,0324272	0,00503151	6,445	3,79e-08 ***

Estadísticos basados en los datos rho-diferenciados:

Media de la vble. dep.	18018,99	D.T. de la vble. dep.	1899,354
Suma de cuad. residuos	43846240	D.T. de la regresión	918,2575
R-cuadrado	0,779690	R-cuadrado corregido	0,766980
F(4, 52)	2642,335	Valor p (de F)	2,32e-59
rho	-0,058292	Durbin-Watson	2,097282

Fuente: Elaboración Propia

6.1.3.2 Definición del Modelo

El modelo resultante fue el siguiente:

$$Y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t-3} + \beta_4 x_{4t-1}$$

En donde:

x_{1t} : Presencia de la ley de tabaco en el periodo t

x_{2t} : Total de visitas en el periodo t

x_{3t} : Índice general de remuneraciones en el periodo t

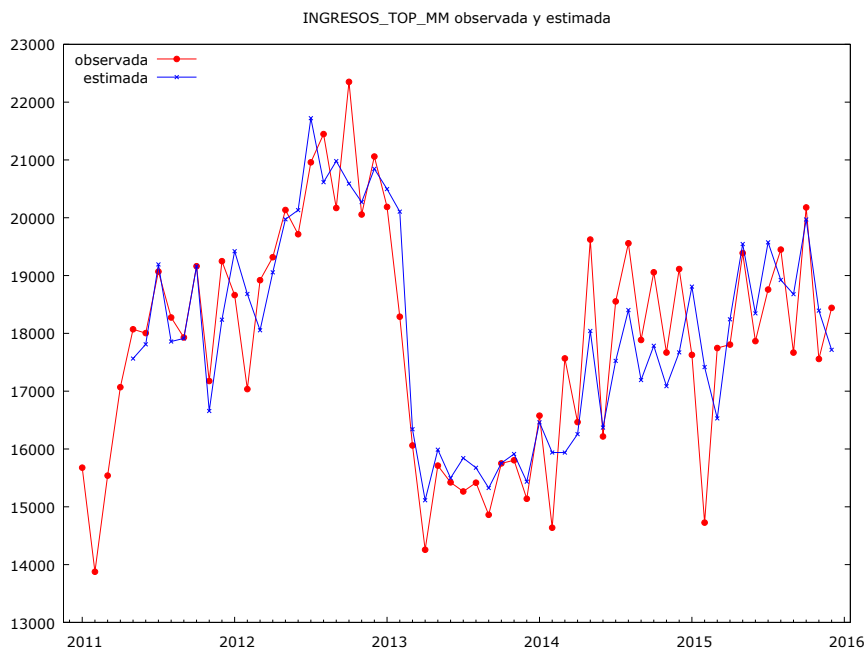
x_{4t} : Subdivisión del IPC de recreación y cultura en el periodo t

$\beta_1, \beta_2, \beta_3$ y β_4 : Son los parámetros que indican la proporción de cambio de cada una de las variables con respecto a la variable explicada manteniendo a las demás constantes.

6.2 Análisis de Resultados

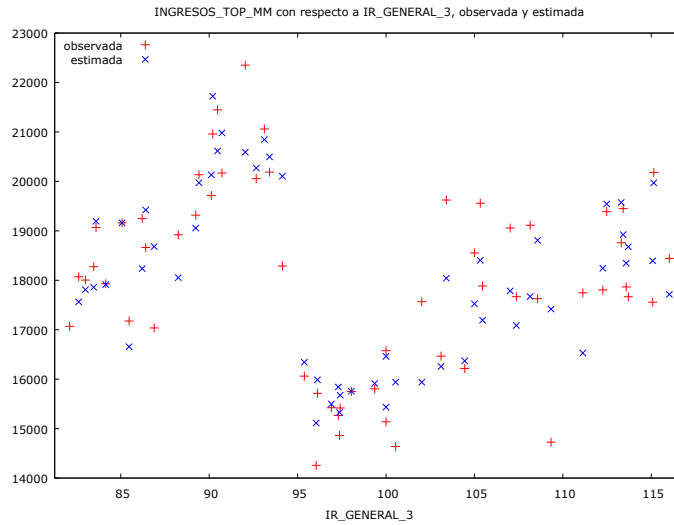
El modelo resultante al entregar un coeficiente de determinación ajustado de 74,83 %, esto quiere decir, que con el conjunto de las variables seleccionadas como regresoras, explican el 74,83 % de los cambios que se visualizan en el flujo de ingresos, el otro 25,17% está explicado en variables y factores que no fueron involucrados en la ecuación.

Ilustración 27 Resultados Gretl



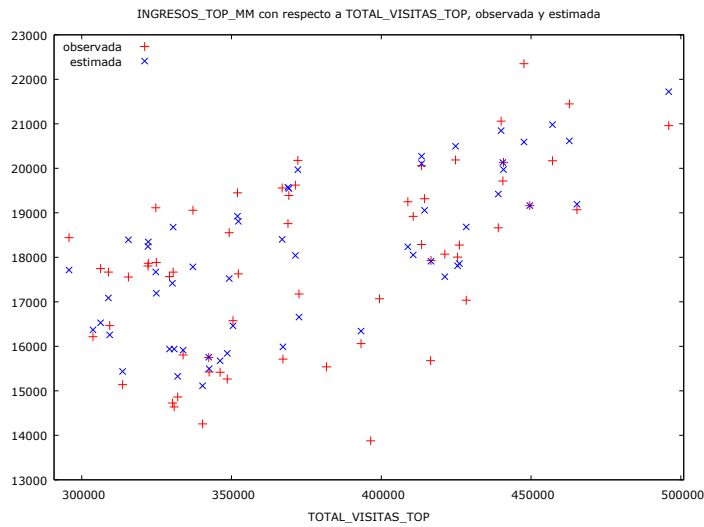
Posteriormente se presentan graficos de dispersión que analizan la variable regresada vs cada una de las variables regresoras, aquí se pueden observar si el modelo funciona mejor con alguna magnitud en particular, ademas de, observar tendencias y relaciones positivas y negativas

Ilustración 28 Resultados Gretl



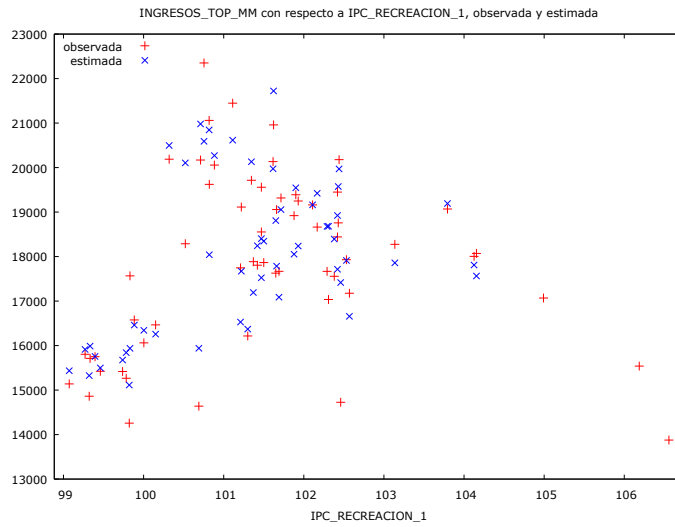
Fuente: Elaboración Propia

Ilustración 29 Resultados Gretl



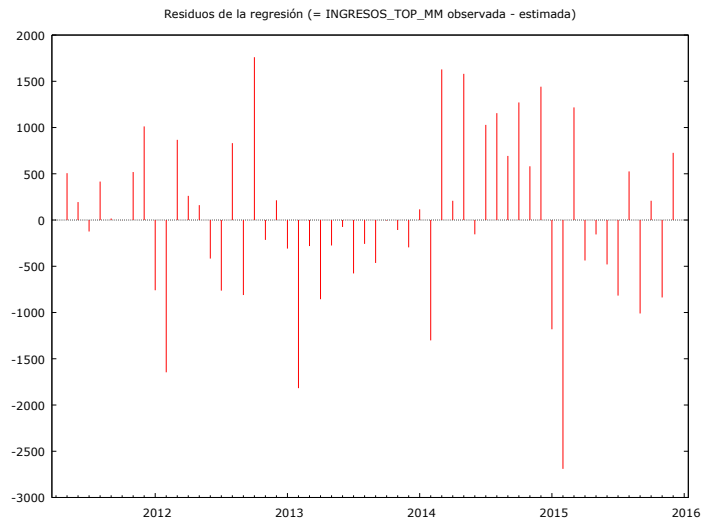
Fuente: Elaboración Propia

Ilustración 30 Resultados Gretl



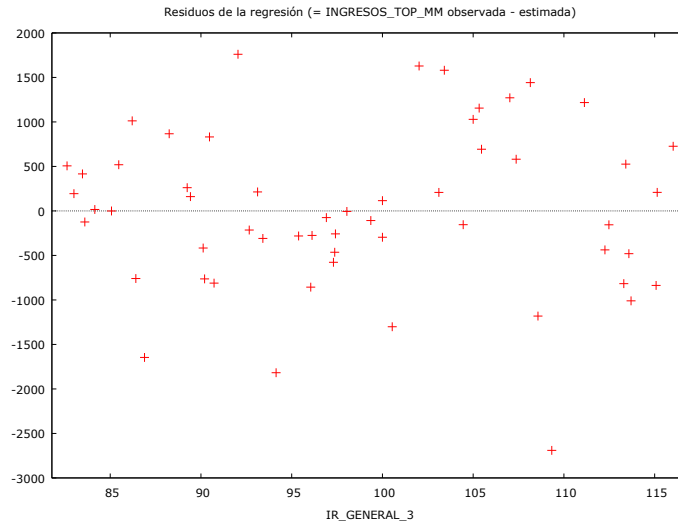
Fuente: Elaboración Propia

Ilustración 31 Resultados Gretl



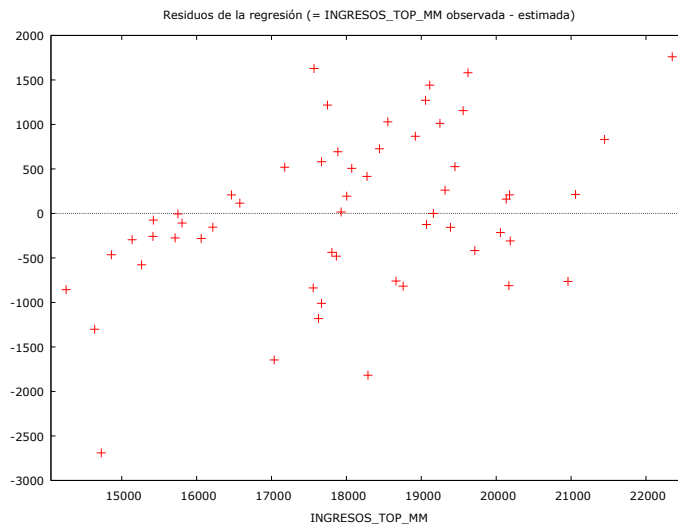
Fuente: Elaboración Propia

Ilustración 32 Resultados Gretl



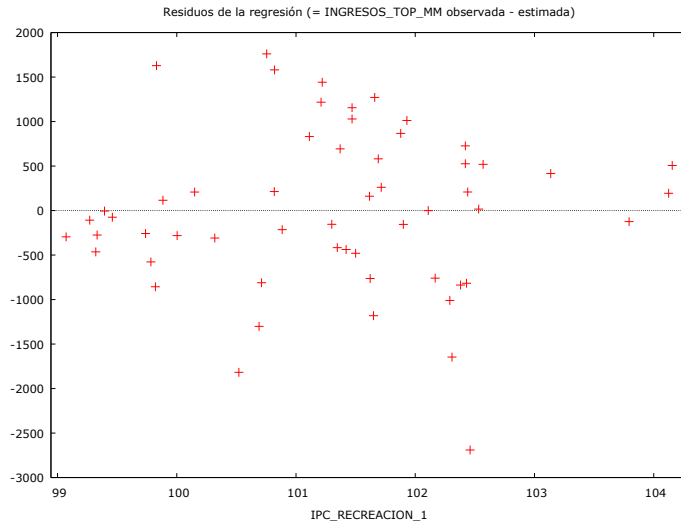
Fuente: Elaboración Propia

Ilustración 33 Resultados Gretl



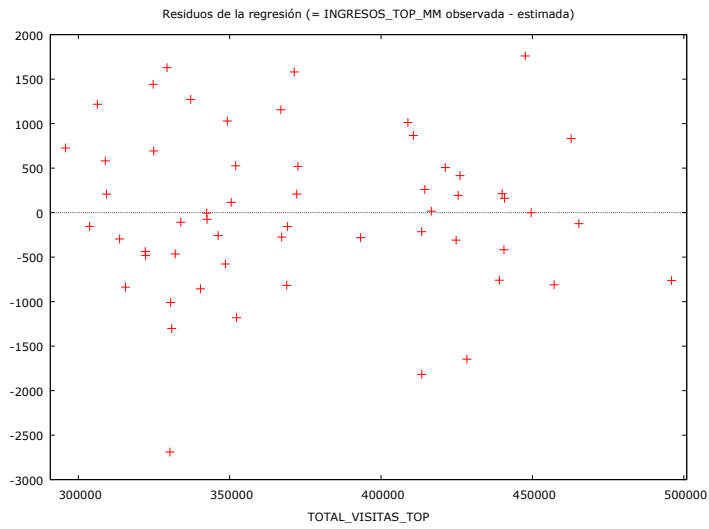
Fuente: Elaboración Propia

Ilustración 34 Resultados Gretl



Fuente: Elaboración Propia

Ilustración 35 Resultados Gretl



Fuente: Elaboración Propia

6.3 Proyección de ingresos

Una vez verificado el grado de bondad del modelo, se procede a analizar la factibilidad de usar el modelo econométrico para predecir los ingresos futuros, dicha factibilidad está dada por la posibilidad de proyectar las variables sobre las cuales se sustenta el modelo, hay que tener en consideración que los errores en la predicción de las variables regresoras se sumarán al error del modelo, es precisamente este nuevo margen de error, provocado por la suma de los errores recién mencionados, el que requiere ser analizado , si este nuevo margen de error es tolerable se considera que el modelo tiene capacidad predictiva.

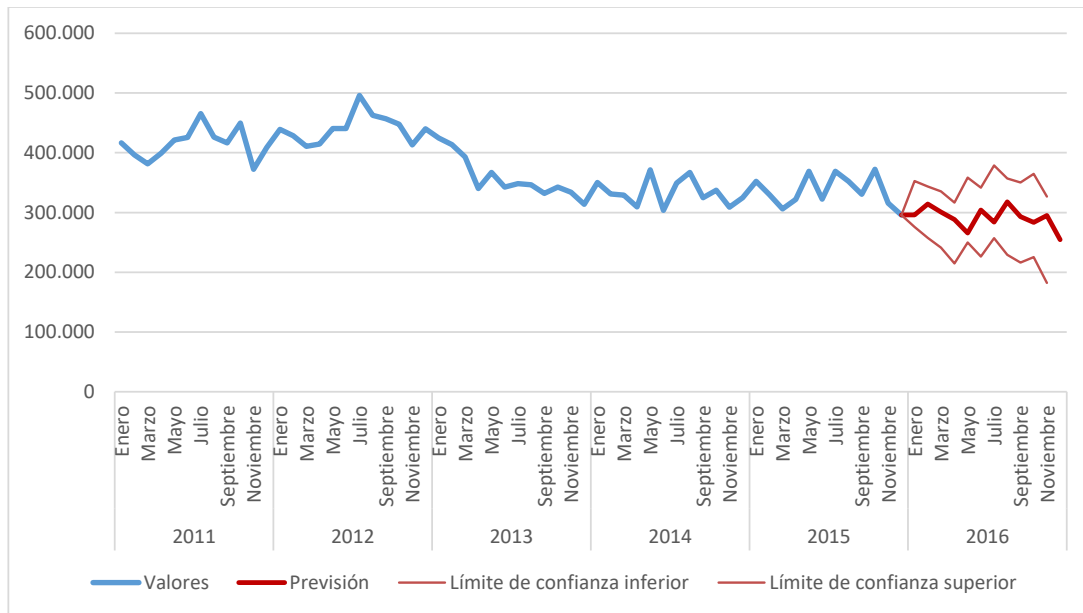
6.3.1 Proyección de la variable Ley de tabaco

La presencia o no de la existencia de la ley del tabaco fue definida como una variable binaria, es decir, su ausencia se representó con un 0 y su presencia con un 1, con lo anterior presente, la proyección de esta variable está en base a que no existe algún precedente que indique que esta pueda sufrir modificaciones, entonces se asume hasta que exista algún elemento que indique un posible cambio sobre esta, que seguirá estando vigente, razón por la cual se proyecta con valor 1.

6.3.2 Proyección de total de visitas

Para la proyección de visitas se utilizó el método de suavización exponencial de Holt-Winters debido a que presentaba una cierta estacionalidad además de un factor de crecimiento (o disminución) lineal, obteniéndose los siguientes resultados:

Ilustración 36 Proyección Visitas



Fuente: Elaboración Propia

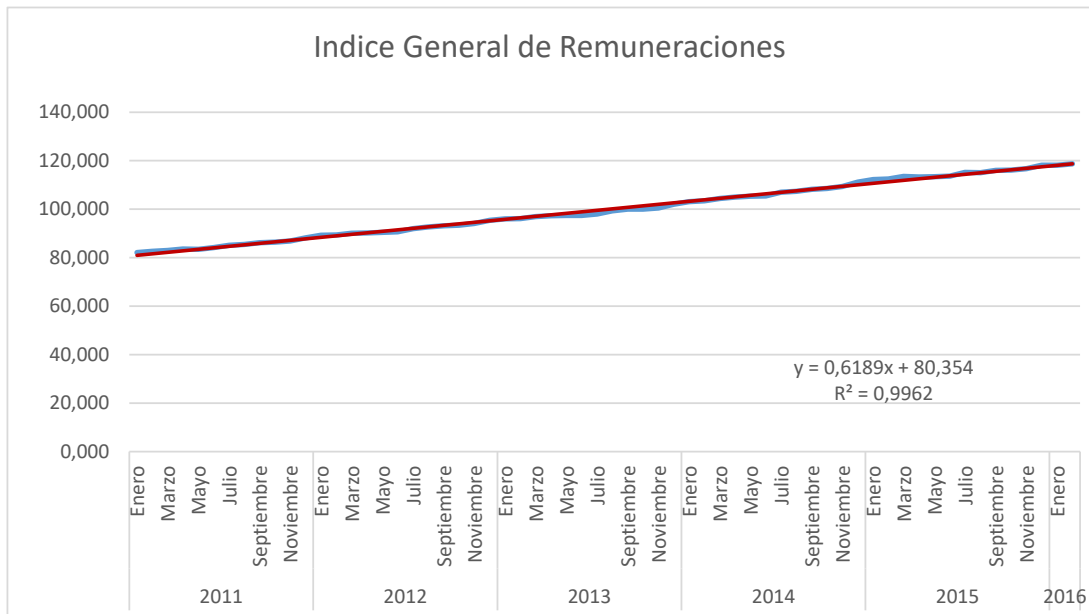
<i>Estadística</i>	<i>Valor</i>
<i>Alpha</i>	0,5000
<i>Beta</i>	0,0010
<i>Gamma</i>	0,0010
<i>MASE</i>	0,72
<i>SMAPE</i>	0,05
<i>MAE</i>	16.031,65
<i>RMSE</i>	20.766,90

Esta proyección permitió trabajar bajo 3 escenarios, uno base, uno pesimista y otro optimista, siendo este último el que arrojo mejores resultados al compararlos con enero y febrero del 2016, razón por la cual para la proyección se utilizó dicho escenario.

6.3.3 IR General

Debido al comportamiento, del Índice general de remuneraciones, que presenta un aumento sostenido y prácticamente lineal en el tiempo, es que se consideró una regresión lineal simple para su proyección.

Ilustración 378 IR General



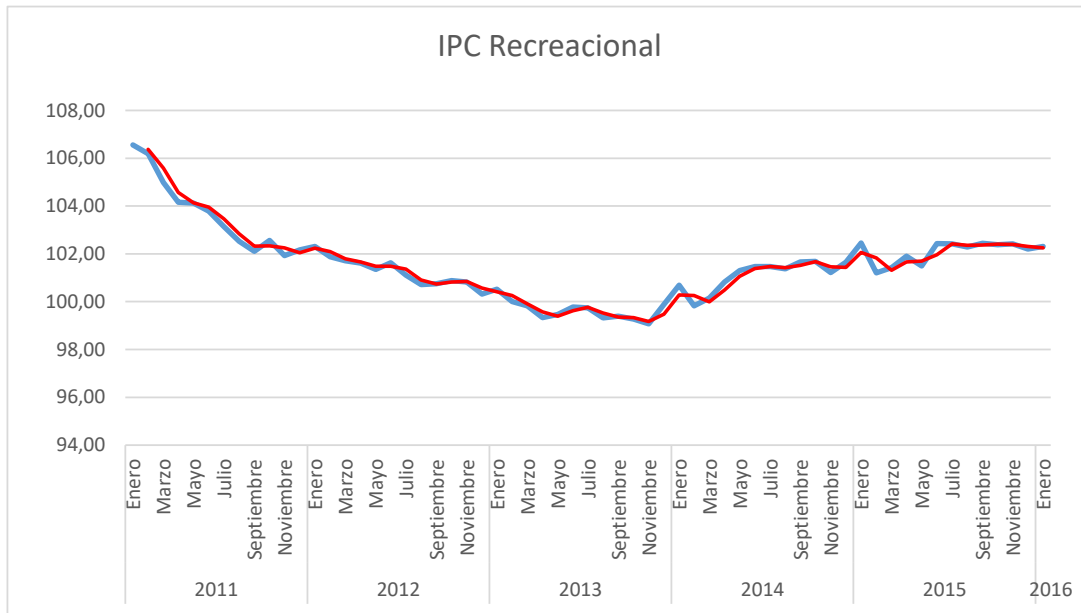
Fuente: Elaboración Propia

La regresión lineal presentada tiene como pendiente el valor 0,6189 y como coeficiente de intercepto 80,354. Cabe destacar que en el modelo, esta variable presenta un retardo de tres periodos.

6.3.4 Subdivisión del IPC, IPC Recreacional

Debido al comportamiento que presenta de la subdivisión del índice IPC, “IPC Recreacional” es que se utilizó un promedio móvil simple de dos periodos, arrojando los siguientes resultados.

Ilustración 389 IPC Recreacional



Fuente: Elaboración Propia

6.4 Proyección de Ingresos

Resultados de la proyección de ingresos, utilizando pronósticos en las variables regresoras:

Tabla 10 Proyección de Ingresos

Año	Mes	Modelo	Dif	Dif%	Real	X1	B1	X2	B2	X3	B3	X4	B4
2015	Enero	18271,0	643,3	3,6%	17627,7	1	-2948,01	352274	0,0324272	112,260	220,9360	102,46	-139,6090
2015	Febrero	17612,4	2886,1	19,6%	14726,3	1	-2948,01	330272	0,0324272	112,480	220,9360	101,21	-139,6090
2015	Marzo	17406,6	-340,2	-1,9%	17746,8	1	-2948,01	306282	0,0324272	113,580	220,9360	101,42	-139,6090
2015	Abril	18139,9	334,2	1,9%	17805,7	1	-2948,01	322098	0,0324272	113,300	220,9360	101,90	-139,6090
2015	Mayo	19646,9	257,4	1,3%	19389,5	1	-2948,01	369139	0,0324272	113,410	220,9360	101,50	-139,6090
2015	Junio	18424,2	558,4	3,1%	17865,8	1	-2948,01	322218	0,0324272	113,700	220,9360	102,43	-139,6090
2015	Julio	19744,4	986,7	5,3%	18757,7	1	-2948,01	368842	0,0324272	115,140	220,9360	102,42	-139,6090
2015	Agosto	19223,5	-225,8	-1,2%	19449,3	1	-2948,01	351986	0,0324272	115,080	220,9360	102,29	-139,6090
2015	Septiembre	18609,3	941,7	5,3%	17667,6	1	-2948,01	330509	0,0324272	116,020	220,9360	102,44	-139,6090
2015	Octubre	20256,2	77,9	0,4%	20178,3	1	-2948,01	372132	0,0324272	116,180	220,9360	102,38	-139,6090
2015	Noviembre	18417,3	860,5	4,9%	17556,8	1	-2948,01	315575	0,0324272	116,780	220,9360	102,42	-139,6090
2015	Diciembre	17976,6	-464,9	-2,5%	18441,5	1	-2948,01	295750	0,0324272	118,200	220,9360	102,20	-139,6090
2016	Enero	19888,5	-1209,5	-5,7%	21098,04	1	-2948,01	352674	0,0324272	118,107	220,9360	102,31	-139,6090
2016	Febrero	19708,8	397,3	2,1%	19311,54	1	-2948,01	343518	0,0324272	118,726	220,9360	102,26	-139,6090

Fuente: Elaboración Propia

Los resultados analizados para enero y febrero del 2016 resultaron ser tolerables, por lo que con cierto nivel de confianza podemos afirmar que el modelo tiene capacidad predictiva.

Conclusiones

Como conclusión del trabajo, primero hay que resaltar que se logró el objetivo general, ya que se construyó un modelo que según los análisis estadísticos tienen un grado de bondad suficiente como para proyectar los ingresos con él.

Se incorporaron variables macroeconómicas que bajo un sentido lógico tienen incidencia en los resultados del ejercicio, vale decir, se probó desde una perspectiva estadística y económica que la relación de ingresos del casino está dada por variables inflacionarias, niveles de ingresos, número de visitas y la ley del tabaco.

Además, se lograron todos los objetivos específicos, con respecto a este punto es bueno señalar que las variables seleccionadas son satisfactorias desde un punto de vista estadístico, y económico, en este mismo ámbito destacar que el efectuar análisis adicionales sobre los retardos de las variables da al modelo una cota superior de acercamiento a la realidad, pues económicamente, los efectos que un cambio de una variable puede tener sobre otra no son inmediatos y comúnmente se encuentran distribuidos en varios periodos.

Adicionalmente el modelo respondió de buena manera a los tests estadísticos de rigor efectuados, lo que nos indica que no presenta sesgos, autocorrelaciones, y los parámetros estimados, son estimadores insesgados. Con el fin de testear la capacidad predictiva del modelo se proyectaron los ingresos de dos periodos los cuales arrojaron resultados admisibles con márgenes considerablemente bajos -5,7% y 2,1% respectivamente, lo que nos lleva a afirmar que el modelo tiene capacidad predictiva.

El presente modelo fue entregado a las empresas Mundodreams y Sun Monticello con el fin de dar robustez a la estructuración financiera a la postulación para la licitación de concesiones, licitaciones que fueron efectuadas el segundo semestre del 2016, por el carácter confidencial de dicha postulación desconocemos si efectivamente fue utilizado y los resultados de las mismas.

Bibliografía

Escrivá Monzó, Joan; Clar Bononad, Federico (2000) «Unidad 2 Conocimiento del cliente» Marketing en el punto de venta (Primera edición) Aravaca (Madrid, España): McGraw-Hill Interamericana de España.

[Gujarati&Porter10] Gujarati, Damodar N.; Porter, Dawn C. Econometría (Quinta edición). Mexico: McGraw-Hill de Mexico, 2010.

[Cole96] Cole, Julio. Elementos de la econometría aplicada (Tercera edición). Guatemala: J&G Ediciones 2014.

Esteban, M. Victoria; Moral, M. Paz; Orbe, Susan; Regúlez, Marta; Zarraga, Ainhoa; Zubia, Marian. Análisis de regresión con Gretl. País Vasco: Departamento de Economía Aplicada III, Econometría y Estadística, Facultad de ciencias Económicas y Empresariales, Universidad del País Vasco

Canavos, George C. Probabilidad y Estadística, Aplicaciones y Métodos (Primera edición). Mexico: McGraw-Hill de Mexico, 1987.

[Hanke&Wichern06] Hanke, John E; Wichern, Dean W. Pronósticos en los Negocios (Octava edición). Mexico: Pearson Prentice Hall, 2006 Cap.4, p. 105-133.

Índice de Precios al Consumidor (IPC) Base Anual 2013. Santiago. Departamentos de estudios de precios, Instituto nacional de estadísticas. 2013.

Manual Metodológico del Índice de Remuneraciones (IR) Índice de Costo de Mano de Obra (ICMO) Base anual 2009 = 100. Santiago. Subdirección de operaciones, proyecto de actualización IR – ICMO. Subdirección técnica, departamento de investigación y desarrollo, Instituto nacional de estadísticas. 2016.

Centro De Estudios En Economía y Negocios Ceen, Universidad del Desarrollo Índice de Percepción del Consumidor. (disponible vía web en <http://negocios.udd.cl/ceen/estudios-y-publicaciones/ipeco/> visitada en Agosto del 2016).

Banco Central de Chile. Indicador Mensual de Actividad Económica (disponible vía web en <http://si3.bcentral.cl/estadisticas/Principal1/Metodologias/CCNN/imacec/Metodologias.html> visitada en Agosto del 2016)

Banco Central de Chile. Estadísticas en Excel Indicador Mensual de Actividad Económica (disponible vía web en <http://si3.bcentral.cl/estadisticas/Principal1/Excel/CCNN/imacec/excel.html> visitada en Agosto del 2016).

Capítulo 7. Anexos

7.1 Extracto del compendio estadístico INE

1.2.1-03 POBLACIÓN TOTAL DE 18 AÑOS O MÁS, ESTIMADA AL 30 DE JUNIO, POR SEXO. PAÍS 1992-2012

AÑO	Población de 18 años o más ¹		
	Ambos sexos	Hombres	Mujeres
1992	8.868.411	4.317.383	4.551.028
1993	9.052.632	4.408.779	4.643.853
1994	9.236.857	4.500.175	4.736.682
1995	9.421.079	4.591.570	4.829.509
1996	9.597.331	4.678.871	4.918.460
1997	9.773.590	4.766.173	5.007.417
1998	9.949.846	4.853.474	5.096.372
1999	10.126.098	4.940.774	5.185.324
2000	10.302.358	5.028.078	5.274.280
2001	10.506.435	5.129.761	5.376.674
2002	10.710.523	5.231.451	5.479.072
2003	10.914.603	5.333.135	5.581.468
2004	11.118.687	5.434.822	5.683.865
2005	11.322.769	5.536.507	5.786.262
2006	11.537.176	5.643.072	5.894.104
2007	11.751.584	5.749.639	6.001.945
2008	11.965.990	5.856.202	6.109.788
2009	12.180.403	5.962.769	6.217.634
2010	12.394.813	6.069.337	6.325.476
2011	12.584.252	6.162.768	6.421.484
2012	12.773.697	6.256.202	6.517.495

¹ Estimación realizada con antecedentes hasta censo 2002.

FUENTE: INE, Chile, Proyecciones y Estimaciones de población por sexo y edad, 1990-2020.

7.2 Ingresos Medios de los Asalariados

4. INGRESOS MEDIOS DE LOS ASALARIADOS¹

4.1 SEGÚN TRAMOS DE EDAD

4.1-01 INGRESO MEDIO MENSUAL DE LOS ASALARIADOS SEGÚN EDAD. 2009 (*)(**)(***)

	2009			
	Total	Mujer	Hombre	Brecha (***)
TOTAL	350.800	321.501	368.165	-12,7%
15 - 24	191.451	185.868	194.977	-4,7%
25 - 34	320.339	310.331	327.682	-5,3%
35 - 44	389.663	345.033	416.636	-17,2%
45 - 54	406.509	353.787	436.255	-18,9%
55 - 64	413.249	420.024	410.556	2,3%
65 y más	344.103	408.966	330.271	23,8%

(*): Pesos de Octubre de cada año.

(**): Incluye solo sueldos y salarios.

(***) : Excluye Personal de Servicio Doméstico.

(****): Brecha=(Ingreso promedio mensual mujeres / Ingreso promedio mensual hombres-1)*100

FUENTE: Encuesta Suplementaria de Ingresos, INE.

7.3 Resultados proyección Holt y Winters

Valores para la ecuación:

<i>Estadística</i>	<i>Valor</i>
<i>Alpha</i>	0,5000
<i>Beta</i>	0,0010
<i>Gamma</i>	0,0010
<i>MASE</i>	0,72
<i>SMAPE</i>	0,05
<i>MAE</i>	16.031,65
<i>RMSE</i>	20.766,90

Resultados del ejercicio:

<i>Escala de Tiempo</i>	<i>Previsión</i>	<i>Límite inferior</i>	<i>Limite Superior</i>
61	314303	275933	352674
62	300601	257685	343518
63	288320	241279	335360
64	265845	214999	316691
65	304272	249873	358671
66	284068	226321	341816
67	317841	256917	378765
68	293064	229110	357019
69	283251	216393	350110
70	294957	225304	364610
71	254540	182190	326890
72	269917	194957	344876
73	288470	210970	365971
74	274768	194807	354730
75	262487	180129	344844