



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

REGRESIÓN LOGÍSTICA APLICADA EN LA
CONFIGURACIÓN DE ESTADOS EN ALGUNOS PAÍSES DE
AMÉRICA LATINA

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIA

P R E S E N T A:

DIANA CITLALLI MÁRQUEZ PALACIOS



DIRECTOR DE TESIS:
M. EN A. P. MARÍA DEL PILAR ALONSO REYES
2011



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**REGRESIÓN LOGÍSTICA APLICADA EN LA CONFIGURACIÓN DE ESTADOS EN
ALGUNOS PAÍSES DE AMÉRICA LATINA**

Diana Citlalli Márquez Palacios

1. Datos del alumno
Márquez
Palacios
Diana Citlalli
53 53 47 32
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
302109878
2. Datos del tutor
M. en A. P.
María del Pilar
Alonso
Reyes
3. Datos del sinodal 1
M. en C.
José Antonio
Flores
Díaz
4. Datos del sinodal 2
M. en A. P.
Elizabeth Esperanza
Gámez
López
5. Datos del sinodal 3
Act
Francisco
Sánchez
Villarreal
6. Datos del sinodal 4
Act
Edna Gabriela
López
Estrada
7. Datos del trabajo escrito
Regresión logística aplicada en la configuración de Estados en algunos países de
América Latina
101 p
2011

REGRESIÓN LOGÍSTICA APLICADA EN LA CONFIGURACIÓN DE ESTADOS EN
ALGUNOS PAÍSES DE AMÉRICA LATINA
Diana Citlalli Márquez Palacios

Es complicado enumerar en un espacio tan pequeño, los reconocimientos hacia todas las personas que con su comprensión y consejos me apoyaron durante la realización de este trabajo; de manera particular, es un verdadero placer:

Agradecer de manera especial y sincera a la Mtra. María del Pilar Alonso Reyes por haber aceptado la dirección de esta tesis. Su apoyo, paciencia y orientación, tanto en la presente investigación como en mi etapa estudiantil, han sido un apoyo invaluable en mi formación profesional.

Quiero expresar también mi más profunda gratitud a mis padres, por creer en mí y siempre impulsarme en los momentos más difíciles durante todos estos años, gracias a ustedes esta meta ha sido alcanzada.

CONTENIDO

1.	RÉGIMEN DE BIENESTAR Y LA GLOBALIZACIÓN	1
1.1.	RÉGIMEN DE BIENESTAR O WELFARE STATE	1
1.1.1.	¿RÉGIMEN O ESTADO DE BIENESTAR?	1
1.1.2.	CARACTERÍSTICAS DE LOS DISTINTOS REGÍMENES	3
1.1.3.	UTILIZACIÓN EN EUROPA Y AMÉRICA	6
1.2.	GLOBALIZACIÓN	17
1.2.1.	DEFINIENDO EL CONCEPTO	17
2.	REGRESIÓN LOGÍSTICA	21
2.1.	CONCEPTOS ASOCIADOS A LOS MODELOS DE REGRESIÓN LOGÍSTICA	22
2.1.1.	VARIABLES	22
2.1.2.	ESCALAS DE MEDICIÓN PARA LAS VARIABLES	24
2.1.3.	MOMIOS (ODDS)	27
2.1.4.	FUNCIÓN EXPONENCIAL	29
2.1.5.	FUNCIÓN LOGARÍTMICA NATURAL O NEPERIANO	30
2.1.6.	FAMILIA EXPONENCIAL DE DISTRIBUCIONES	30
2.1.7.	MODELOS LINEALES GENERALIZADOS	31
2.2.	MODELOS DE REGRESIÓN LOGÍSTICA	36

**REGRESIÓN LOGÍSTICA APLICADA EN LA CONFIGURACIÓN DE ESTADOS EN
ALGUNOS PAÍSES DE AMÉRICA LATINA**

Diana Citlalli Márquez Palacios

2.2.1. MODELO DE REGRESIÓN LOGÍSTICA CON VARIABLE DE RESPUESTA BINARIA.....	37
2.2.2. EJEMPLO DEL MODELO SIMPLE DE REGRESION LOGÍSTICA CON VARIABLE DE RESPUESTA BINARIA	51
2.2.3. MODELO DE REGRESIÓN LOGÍSTICA MÚLTIPLE CON VARIABLE DE RESPUESTA BINARIA.....	53
2.2.5. MODELO POLITÓMICO DE REGRESIÓN LOGÍSTICA.....	69
3. APLICACIÓN DE UN MODELO DE REGRESIÓN LOGÍSTICA EN AMERICA LATINA A LOS PROCESOS DE GLOBALIZACIÓN	72
CONCLUSIONES.....	92
ANEXO A.....	96
REFERENCIAS	99

INTRODUCCIÓN

El presente trabajo busca realizar el análisis de los estados de bienestar existentes en América Latina. Para ello es considerada la clasificación europea, así como los estudios concernientes al caso particular del continente americano.

Con base en lo anterior, el resultado buscado en esta investigación, es la corroboración de los estados considerados como existentes en esta región.

Para lograr dicha meta, es utilizado el método estadístico de regresión logística, debido a la posibilidad de evaluar la influencia que cada una de las variables independientes sobre la respuesta, así como, la opción de controlar el efecto de aquellas que no se observan en ese momento.

Derivado de los puntos señalados, el objetivo de este trabajo de investigación, tiene como fin, mostrar una aplicación del método de regresión logística y determinar si la estructura de los estados de bienestar considerada en América Latina, se ha mantenido desde las más recientes investigaciones.

El análisis es presentado en los capítulos de este trabajo, manifestando en el primero de ellos, las características fundamentales de los estados del bienestar o welfare states y la globalización.

Posteriormente, en el segundo capítulo se hace una breve y concisa explicación acerca de la Regresión Logística, sus modelos y aplicaciones.

Por último, el tercer capítulo engloba la aplicación del modelo de Regresión Logística Múltiple en la configuración de los estados del bienestar en América Latina.

1. RÉGIMEN DE BIENESTAR Y LA GLOBALIZACIÓN

1.1. RÉGIMEN DE BIENESTAR O WELFARE STATE

1.1.1. ¿RÉGIMEN O ESTADO DE BIENESTAR?

A finales de los años veinte y primeros de los treinta, en el periodo del capitalismo de la posguerra, con la desolación, la frustración y la reconstrucción del orden mundial, en diversas partes de Europa y el norte de América, se buscó la reforma del Estado, introduciendo una nueva constitución de ideas, que durante la guerra fueran solamente ideas, gestando principios de igualdad, empleo y bienestar social; y, asociando conceptos como: “New Deal” de Roosevelt, “Patria de los Pueblos” de la socialdemocracia sueca y “Welfare State”, para denominar las políticas sociales enfocadas a la distribución de protección al trabajador, frente a las incertidumbres del mercado, como: el despido, enfermedad, pobreza e ignorancia, entre otras.

De esta manera, al realizar la traducción de las locuciones inglesas anteriores, se encuentra la distinción de las características que abordaban, al relacionarlas con dos conceptos: régimen y estado.

Primeramente la palabra “régimen” de acuerdo a la Real Academia Española es asociado al conjunto de normas que gobiernan o rigen una actividad; siendo en el caso de un sistema político, aquellas medidas por las que se manda una nación.

Por otro lado el término “estado”, según la misma fuente anterior, proveniente el latín status, es el indicado para nombrar la situación en que se encuentra alguien o algo, y en especial cada uno de sus sucesivos modos de ser o estar. Cabe mencionar que, vinculado a la política, el concepto de estado, denota el conjunto de los órganos de gobierno de un país; y de esta manera, en un régimen federal, será la porción de

territorio cuyos habitantes se rigen por leyes propias, aún cuando estén sometidos en ciertos asuntos a las decisiones de un gobierno común.

Ante este panorama, diversos autores han determinado el concepto de acuerdo a sus estudios, en el presente trabajo, se considera “régimen de bienestar” como la locución indicada para determinar las reformas económicas y sociales encaminadas a la repartición del bienestar en una nación.

La anterior decisión, se fundamenta en los estudios realizados por la Organización para la Cooperación y el Desarrollo Económico (OCDE); en los cuales, dicha expresión fue elegida para “describir los cambios entre la política social, modelos económicos y estrategias domésticas”, como lo escribe el autor Carlos Barba, en su texto dedicado al régimen de bienestar mexicano.

Adicionalmente; es conveniente aclarar que, el introducir la palabra régimen, se basa en el papel tan importante, que globalización funge en el bienestar; como más adelante se verá, es complicado definir un estado en una sociedad global.

Cabe mencionar que, el término “régimen de bienestar”, fue acuñado por primera vez por el sociólogo Esping Andersen; y lo atribuyó a todos aquellos procesos relacionados con el bienestar, encontrados a nivel global y local; así como, en una combinación de ambos, es decir, en una sociedad interconectada.

Desde esa perspectiva, un régimen de bienestar será el modo en que se reparte y produce el bienestar por parte del estado, la familia, el mercado y la sociedad civil, mismo que dependerá de las clases sociales, políticas y el legado histórico de cada país.

Aunado a esto, el estado del bienestar está definido como el conjunto de las instituciones estatales que garantizan la distribución de derechos sociales, a través de

políticas apoyadas en la idea del trabajador y sus necesidades asociadas; con el objeto de nivelar la pobreza, la desigualdad y la seguridad económica y social.

Como se dijo anteriormente, según investigaciones recientes, realizadas por Esping-Andersen, se considera que la construcción histórica concreta del concepto, tiene sus bases en las décadas referentes a los años 1930 y 1940, desarrollándose principalmente para la reintegración de la relación entre el gobierno y la ciudadanía.

Así mismo, la historia del régimen de bienestar de la posguerra tuvo sus cimientos en el comportamiento de la familia y del mercado laboral de esas épocas; derivando en el cambio de la percepción de problemas, que eran considerados como insubstanciales en aquellos años, a muy importantes en la actualidad.

1.1.2. CARACTERÍSTICAS DE LOS DISTINTOS REGÍMENES

Entre las características de un régimen de bienestar, se puede mencionar, la manera en que se instaura y distribuye el bienestar en el Estado, mercado laboral y la familia; por ello, los regímenes de bienestar, son comúnmente reconocidos como la interrelación de los tres núcleos antes señalados y que cada uno de éstos son igualmente importantes en el desarrollo del Bienestar en un Estado.

Una explicación ampliamente abordada es la creada por Esping-Andersen; donde presenta tres prototipos de régimen de bienestar: liberal, conservador y socialdemócrata; derivados todos ellos de la política europea clásica y representando la evolución histórica de dichas políticas y su culminación en las décadas setenta y ochenta.

El *régimen de bienestar liberal*, concentra a los países anglosajones: Estados Unidos de América, Reino Unido, Nueva Zelanda y Australia. Con las características de tener un estado en el cual se considera la individualización de los riesgos y de la promoción

de soluciones de mercado. En el que se define la posibilidad de que un individuo acceda a los beneficios brindados por la sociedad, desarrollando los medios por los cuales fuese posible determinar sus necesidades, de igual forma que analiza los problemas que deberán ser apreciados como prioritarios en la sociedad, siendo, por ejemplo, los relacionados con las madres solteras los que se lleven el mayor grado de importancia.

Para llevar a cabo las ideas anteriores, como el apoyo a la población de escasos recursos, el mercado utiliza una serie de estudios socioeconómicos, los cuales, entre otras de sus funciones es la de limitar los derechos sociales.

El *régimen de bienestar socialdemócrata*, agrupa a los países escandinavos: Suecia, Noruega, Dinamarca y Finlandia. El objetivo de este régimen, es sustituir cualquier sistema, en el cual, las políticas conduzcan a condiciones de desigualdad e inadecuación, por aquellos que busquen la protección pública integral; cubriendo así, de manera universal a los ciudadanos e intentando un igualitarismo entre ellos.

Por lo cual, Esping-Andersen le atribuye características como: solidario, universalista, desmercantilizador; esto último derivado de que el individuo se logra independizar del mercado.

Por último, el *régimen de bienestar conservador*, reúne a los países de Europa continental: Alemania, Holanda, Austria, Francia, Italia, Grecia, Portugal y España. La postura de este régimen, se basa primordialmente en proteger de manera individual a los trabajadores mediante políticas de contratación y despido, es decir, la base de la familia queda determinada por el varón que es el sustento del recurso material y la seguridad del hogar; de manera que, el Estado tiene una mínima responsabilidad hacia aquellos ciudadanos que no obtengan el bienestar por medio del mercado y provoca una mayor integración familiar.

Resultado de lo anterior, este Régimen se encuentra íntimamente relacionado con las clases o el status, así como, dejando a las familias como las encargadas de asumir los riesgos sociales.

Todos los regímenes antes considerados, expresan las condiciones prevalecientes en una economía basada en la producción industrial, en la cual el obrero masculino y la familia estable con sólo una fuente de ingresos, el prototipo de vida de esa época; por tal razón, Esping-Andersen, hace referencia a un cuarto mundo adicional, conformado por: las Antípodas, el Mediterráneo y Japón.

El Cuarto Mundo de las Antípodas o Australiano, denominado también, régimen de bienestar de los asalariados, puesto que, en Australia el seguro de salud (Medicare) es un programa de derechos, en los que, los subsidios asistenciales son otorgados primariamente a aquellas familias que cuenten con hijos y posteriormente a las personas solteras. Otra característica que se manifiesta en este mundo, es que el tope máximo de ingresos para la asignación de los apoyos es el nivel medio, no considerando los límites de pobreza, buscando cubrir a un mayor número de la población.

El Cuarto Mundo del Mediterráneo, definido principalmente por la búsqueda de delimitar un régimen especial para los países mediterráneos y no considerarlos en Europa Continental, como lo es Portugal; autores como Ferrara fundamentan la argumentación de este mundo, en la distribución de los subsidios sociales y en las aplicaciones políticas, como en el caso Italiano. Por otro lado, Leibfried hace referencia que la importancia de este cuarto mundo radica en el programa de asistencia social, y de manera más específica en el valor asignado a la familia por parte de los países del Sur de Europa, ya que apoyaron su régimen en dicha unidad, por ser la entidad más sólida en la vida de las personas y en la cual se otorgaba el mayor apoyo social.

El Cuarto Mundo de Asia Oriental, correspondiente a Japón, Corea y Taiwán, es una versión única del capitalismo, la cual combina aspectos de dos regímenes, el liberal y el conservador; del primero, conserva la asignación de recursos a una población específica. Debido a sus características tan singulares; Japón ha constituido una “Sociedad del Bienestar” considerablemente desarrollada, que en consecuencia, no resulta ser denominada Estado del Bienestar; uno de los factores causantes de esto último, es que considera que el mercado y la familia son suficientes para lograr el bienestar en la ciudadanía.

1.1.3. UTILIZACIÓN EN EUROPA Y AMÉRICA

Derivado de las múltiples investigaciones realizadas, la autora Deborah Mitchell¹ denota la existencia de cinco puntos viables para el análisis de la aplicación de los regímenes de bienestar, los cuales consisten principalmente en:

1. Situación política del país estudiado, esto es, analizar las condiciones en las que se está desarrollando el régimen por estudiar.
2. Situación de la población, en especial elementos que determinan modificaciones en el sistema social.
3. Principales sistemas de producción.
4. Descripción de los procesos para la adquisición de beneficios y/o servicios.
5. Resultados que se han derivado de la aplicación del régimen estudiado.

AMÉRICA

De acuerdo a los regímenes establecidos en el continente europeo, se crea la incógnita de lo que sucede en el continente Americano o específicamente en América Latina; siendo este problema abordado por diversos autores, analizando tres premisas

¹ <http://www2.rgu.ac.uk/publicpolicy/introduction/wstate.htm>

inadecuadas para la conceptualización del régimen de bienestar en América, las cuales radican en:

- ★ la primera en que pueden aplicarse en Latinoamérica de manera directa los regímenes establecidos con base a economías desarrolladas,
- ★ la segunda es que se considera que el continente Americano es una región homogénea, en la que se manifiesta la existencia de un Régimen de bienestar, y
- ★ la tercera es considerar que al establecer un régimen de bienestar, los resultados serán los mismos que en los países europeos.

Por otra parte, algunos autores consideran que no es posible suponer la existencia de regímenes de bienestar en este continente, ya que en un preciso sentido, éstos nunca han existido, ya que las políticas sociales no se han basado primordialmente en los derechos sociales de carácter universal y solidario; lo que es más, ninguna de las ideas del capitalismo avanzado, como son:

- ★ ciudadanía social universal y solidaria,
- ★ democracia plena,
- ★ reconocimiento y consolidación del sindicalismo y de los sistemas industriales modernos,
- ★ el derecho a la educación, y
- ★ la expansión de los modernos sistemas de educación.

Con base al análisis de los regímenes del bienestar estructurados por la Dra. Juliana Martínez Franzoni, existen cuatro regímenes identificados en los países de América Latina, y son:

Régimen de bienestar estatal-proteccionista

Caracterizado principalmente por el papel desempeñado por el Estado y su control en la prestación de servicios a la sociedad.

1. Brasil
2. Costa Rica
3. Panamá
4. México
5. Uruguay

Régimen de bienestar productivista-informal

El sector privado es el protagonista, con la peculiaridad de que en comparación con un sistema privado, los recursos utilizados son públicos.

1. Argentina
2. Chile
3. Panamá

Régimen de bienestar familiarista

Derivado de su nombre, tiene la peculiaridad de concretar la dependencia entre la población y los acuerdos económicos familiares que se establezcan para el bienestar.

1. Ecuador
2. El Salvador
3. Guatemala
4. Colombia
5. Venezuela
6. Perú
7. República dominicana

Régimen de bienestar altamente familiarista

Se diferencia del anterior por su grado más alto de dependencia de la distribución del bienestar en el núcleo familiar.

1. Honduras
2. Nicaragua
3. Bolivia
4. Paraguay

Para la aparición de dichos regímenes, han sido necesarios diversos hechos históricos-sociales en los cuales el impacto de la repartición de recursos es la principal repercusión; ante este panorama, para comprender los estados en la actualidad; es necesario entender, los periodos de cambios significativos y los legados históricos de cada región.

Para concretar los antecedentes, a continuación se definirán algunos puntos importantes dentro de la historia de algunos países de América Latina.

MEXICO

No obstante, se considera la larga trayectoria que los sistemas de prestaciones y políticas sociales han tenido en América, tal es la cuestión mexicana, en donde concurren ciertas polémicas conforme al momento en que se desarrolló un régimen de bienestar, contemplando el periodo transcurrido entre la promulgación de la Constitución de 1917 y la crisis económica de 1982, para dar solución a esas polémicas, el caso mexicano se analiza por medio de los periodos siguientes:

- ☆ 1917 - 1940, definido por algunos autores como la fase de constitución de la política social pos-revolucionaria.
- ☆ 1940 - 1982, correspondiente a la articulación de la política social y el modelo ISI.
- ☆ 1982 - 1988, fase de crisis de la articulación entre la política social y el modelo ISI.
- ☆ 1988 - , cambio de paradigma de bienestar.

De los anteriores lapsos de tiempo, se observa que los pilares de la aparición de un régimen de bienestar en México se encuentran en el periodo 1917 – 1940, reiterando las características de los países europeos, es decir, que como consecuencia de movimientos populares fue que se inició la instauración o la propuesta de políticas sociales encaminadas hacia un régimen de bienestar.

Como consecuencia de los cambios realizados en los años señalados, se tiene que el régimen de bienestar mexicano cuenta con un auge mayor en la década de 1990, siendo razones fundamentales:

- ★ la búsqueda de la estabilidad macroeconómica,

- ★ el inicio de los procesos de privatización de los activos estatales y
- ★ la apertura comercial.

Lo anterior, en razón de que se buscaba la renegociación de la deuda externa, la estabilización monetaria y la reducción del déficit público; utilizando estrategias que reducían el gasto público y de acuerdo con el Fondo Monetario Internacional (FMI), medidas que estabilizaban la economía mexicana.

De igual forma que el análisis de los años 1917 - 1982, la década de 1990 es estudiada en 2 etapas diferenciadas por la crisis económica de 1994 - 1995, en que se promovió la apertura comercial, la desincorporación de paraestatales, el avance de la desregularización financiera, la modificación del régimen de la tenencia de la tierra y la transformación del sistema de pensiones del Instituto Mexicano del Seguro Social (IMSS).

CHILE

Integrante del régimen estatal proteccionista, en este país, después de haber existido una dictadura por 16 años y las condiciones de bajos ingresos y desempleo resultantes de ese sistema, se provocó que el bienestar fuera visto como un producto con el que se debía comercializar, conformando un régimen de carácter liberal y de características similares a las del existente en Estados Unidos.

Posteriormente, la forma de gobernar dicho país se vio influenciada por dos corrientes muy fuertes: la dictadura militar chilena (1973-1990) y las correcciones propuestas por el gobierno de centro-izquierda de la concentración de Partidos Demócratas (1990-2010); dichas corrientes, en la búsqueda de desmercantilizar la política pública del país.

Producto de estos cambios, en el país se logró una reducción significativa de la pobreza; sin embargo, en contra de ese adelanto, el equilibrio en la distribución de los

ingresos no era un punto en el que se tuviera avance; de tal manera que, las características más representativas de este tipo de gobierno, fueron:

- ★ Reflejar una distribución desigual de los ingresos.
- ★ Mostrar el alto grado de desempleo, a pesar de la disminución del número de desempleados.
- ★ Indicar la gran inestabilidad del trabajo asalariado.
- ★ Manifiestar la estratificación de la población entre quienes tenían el derecho a recibir los beneficios del sistema de seguridad social y los que no.

Estas particularidades, atentando contra los principios de la seguridad social, sobre todo contra su carácter de universal en salud, vivienda, educación y seguros de vejez.

A lo largo de los años subsecuentes, la política social se fue desarrollando en busca de mejorar la educación y salud existentes en Chile; hasta que en los años 70, el Estado retiró su apoyo y promoción de la protección social, provocando que la unidad familiar fuera la encargada de acoger a todas las personas que no logran integrarse al mercado o aquellas que sólo permanecen en él de manera temporal.

Así bien, los gobiernos demócratas, se enfocaron en intentar reducir algunos de todos los efectos negativos, tal fue el caso del gobierno de Alwyn (1990-1994) en el cual, los pactos sociales buscaron la mejoría de los derechos de los trabajadores y sus sindicatos; adicionalmente, se aumentó la inversión destinada a la población con más bajos recursos.

Unos años más tarde, durante la administración de Ricardo Lagos (2000-2006), fue aplicado el Plan de Acceso Universal y Garantías Explícitas de Salud, destinado a garantizar un sistema público de salud.

COSTA RICA

Siendo un país en el cual la prestación pública de servicios es una prioridad, sobre todo enfocada hacia el sector medio, se espera que el manejo colectivo de riesgos sea también de principal interés, como lo puede reflejar la resistencia a las reformas en apoyo del mercado durante la última década.

El modelo que impera en la actualidad, es aquel que se basa en las reformas que imponen la individualización y en donde los beneficios son significativamente pequeños y de manera complementaria.

Para los años ochenta, el estilo de administración se enfocó en la atracción por la inversión extranjera, la liberación comercial y la promoción de las exportaciones.

Una década más tarde, el impulso por la exportación fue en aumento y con ello las instituciones a favor de las importaciones fueron disminuyendo de manera gradual.

Adicionalmente con las anteriores características relativas a uno de los regímenes del bienestar más desarrollado en Latinoamérica, le propician que sea uno de los que menos cuenta con recursos.

Durante su pasado, Costa Rica ha sido uno de los países con un crecimiento económico alto, así como la gran inclusión de nuevos productos derivados de la diversificación de las exportaciones y el gran número de inmigrantes.

Aunado a esto, el nivel de desempleo fue el mayor dentro de las dos décadas anteriores al 2003 y ello, debido a que las exportaciones no tradicionales y el turismo no lograron suplir la ausencia de empleos de casi todos los sectores, con excepción del sector salud, educación y pensiones.

De esta manera, el empleo ha sido estratificado en formal e informal, teniendo un mayor porcentaje el segundo, es decir, en el país, la cantidad de empleos formales de alta calidad son menos que los informales y de escasa calidad.

De acuerdo a las especificaciones anteriores, la concentración de los ingresos fue mermando la equidad existente en la distribución de recursos.

Para compensar las fallas encontradas en el manejo de los recursos, el gobierno experimentó diversas reformas, sobre todo en los sectores de pensiones, educación y salud; conservando ante todo el carácter de universal y solidario; con dichas modificaciones las instituciones a su vez, tuvieron que adaptarse, otorgando el papel de supervisoras a las instituciones nuevas y a las previas, la actividad de financiar y prestar servicios.

En repercusión de estos cambios, las inversiones públicas fueron en aumento; los sectores de educación y de pensiones se vieron mejorados con la excepción del sector salud que tuvo una caída considerable.

El impulso en la educación continuó durante el periodo 1999-2003, en el cual, el aumento en el personal y en el número de escuelas fue notable. Así mismo, en el aspecto de la vivienda, una manera de repartir el subsidio con el que se buscaba solucionar los problemas relacionados, fue el dar bonos, pero no obstante, esta idea fue desviada hacia las políticas clientelistas.

Este último aspecto, fue determinado principalmente por los pagos sobretodo ilegales por conseguir una rápida atención y una mejor calidad, sobre todo en el ámbito de la salud.

En general, por todas las características de ese modelo, las condiciones de vida en este país son aceptables pero no de igual manera la inversión social actual.

Por último, una de los grandes impedimentos se basa en la consideración de la familia como la unidad tradicional, con el hombre como proveedor y la mujer como cuidadora; no considerando la salida de la mujer del hogar.

ECUADOR

Mostrando un régimen de bienestar, resultado de la combinación de la mercantilización e individualización del bienestar, el modelo ecuatoriano deriva de un proceso de desarrollo y modernización trunco, que le ha dejado a la población escasas opciones para sobrevivir con excepción de las salidas familiaristas.

Por otro lado, la inestabilidad política no hizo más que agudizar las dificultades hacia la óptima participación ciudadana y la solución a las demandas laborales.

A principios del siglo, el aumento del Producto Interno Bruto (PIB) y la demanda de bienes y servicios produjeron que en Ecuador se presentara una leve recuperación; dicho bajo nivel, derivado de que el mercado laboral se encontraba en precarización, informalidad y segmentación.

Este último aspecto relacionado con el empleo, es definido por la no creación de empleos formales, sino por el contrario el auge de empleos informales que simplemente derivan en los incrementos del desempleo y subempleo.

Dentro de este país, un aspecto de gran impacto para la constitución de su régimen de bienestar, es la presencia asalariada en general y en especial en el sector público de las mujeres.

De igual manera, los ingresos actuales son principalmente derivados del mercado transnacional, así como, las remesas aportan divisas equivalentes a las exportaciones conjuntas realizadas en dicho país.

Este intercambio entre países no sólo fue llevado a través de productos, ya que el aumento de la población emigrante cambió de manera significativa, dado que la población emigrante se volvió tanto rural como urbana y de ambos sexos; convirtiendo a la emigración en una estrategia de sobrevivencia colectiva.

Con todo lo anterior, el país experimentó un proceso de transición en el cual las tasas de mortalidad y fecundidad fueron descendiendo; con ello, la asignación autorizada de recursos se enfrentó a diversos problemas para cubrir la oferta y la demanda.

Así mismo, a pesar de que el gasto es progresivo y enfocado a la población de escasos recursos, conservando su carácter universal, es un plan insuficiente para detener el empobrecimiento y atender a las demandas sociales históricas como las producidas en la última crisis en este país.

Entre los principales programas sociales focalizados se encuentran:

- ★ El Bono de Desarrollo Humano y
- ★ El Programa de Alimentación y Nutrición.

Por otro lado entre los programas universales destacan:

- ★ El Programa de Maternidad Gratuita y Atención infantil y
- ★ El Aseguramiento, derivado de la Ley de Seguridad Social.

NICARAGUA

Siendo un elemento de un régimen que no tiene muchas opciones de manejo de los riesgos, con excepción de los comerciales, es que la población tiene que buscar la forma de obtener recursos para sobrevivir, encontrando una opción en la migración y el apoyo familiar.

Todo este escenario resultado de dos hechos de gran trascendencia: la revolución sandinista (1979), en la cual se derrocó a la dinastía de los Somoza que gobernó desde los años treinta; y de la transición democrática entre el Partido Liberal Constitucionalista y el Frente Sandinista de Liberación Nacional (FSLN), de los cuales, el primero gobernaba el país.

Dichos aspectos formaron un abismo entre las necesidades sociales y la oferta pública, ya que conforme la población aumentaba, el grado en inversión social no lograba alcanzar dicha rapidez. Con lo que Nicaragua continúa siendo uno de los países del continente Americano con el menor grado de inversión pública.

Conforme fue siendo más visible la falta de una inversión más fuerte, las expectativas no eran nada favorables, ya que los objetivos de reducción del gasto públicos contraídos con el FMI implicaban una reducción en el gasto público y un aumento en el gasto para reducir la pobreza.

A partir de ese momento, los recursos disponibles, fueron canalizados principalmente a la educación, salud y potabilidad del agua, distribuidos principalmente por la participación comunitaria y la coordinación entre las instituciones y el gobierno.

Así, las medidas tomadas para la transformación de la asignación de los recursos, fueron orientadas a la población de extrema pobreza, sin embargo, de acuerdo a que la inversión provenía de recursos externos, el endeudamiento de la población fue a su vez mayor, con un estado incapacitado para ayudarles.

1.2. GLOBALIZACIÓN

1.2.1. DEFINIENDO EL CONCEPTO

Dentro de la historia, se ha podido observar la utilización de diversos términos, los cuales, en un principio no habían sido definidos como se conocen en la actualidad, tal es el caso del concepto de “Globalización”; cuyos inicios se remontan a los años sesenta y setenta, con la investigación realizada por educador, filósofo y estudioso canadiense Marshall McLuhan, quien acuñó el término de “aldea global”, con el objeto de expresar el cambio que en el mundo estaba ocurriendo; dicha variación se vio manifestada en medios de transporte, nuevas tecnologías de comunicación e información, sistemas de mercado y flujos financieros, los que han provocado que el planeta se encuentre interrelacionado.

Otro concepto que no fue totalmente aceptado fue el de “sistema mundial”, introducido por Wallerstein, en 1979; el cual, se creó considerando que la expansión del mercado provocaba la creación de la economía-mundo capitalista, en el cual la división del trabajo no es sólo buscar la funcionalidad, sino también la situación geográfica.

De esta manera, así como los términos anteriores se encuentran relacionados, la acepción de globalización, no puede ser desligada a los quehaceres políticos, ya que, brinda un poder negociador entre las empresarios y sus trabajadores.

Asimismo, se integran en el contexto, el “globalismo”, entendido como el encargado de poner en una sola dimensión económica la globalización, buscando la minimización de los costos y la maximización de las ganancias; y, la “globalidad”, que dirige la mezcla de las formas económicas, culturales y políticas que interactúan al darse la globalización.

Ante este panorama, la globalización, es comprendida como la sociedad mundial; estando ausentes, un Estado y gobierno mundiales; permitiendo la pérdida de las fronteras, y la incursión en formas de vida transnacionales. Derivando de dichos resultados, se encuentra otra definición de la globalización, explicada como politización.

Cabe señalar, que la expresión “transnacional”, está íntimamente ligada con la capacidad de los hombres para llevar una vida social, donde las relaciones de intercambio, pasan por alto las distancias físicas existentes; resultantes de las actividades internacionales.

Resultado de lo anterior, los Estados transnacionales, también pueden ser denominados como “glocales”, que se analizan y comprenden al ver a la sociedad mundial como una sola región.

Entre las actividades relacionadas, es posible visualizar que, existen algunos puntos en peligro derivados de esta forma de vida, entre los que se encuentra, el sistema de seguridad; debido a que, los conflictos de la pobreza, étnicos y nacionalistas, adquieren una importancia al re-localizarlos en un mundo global y no regional.

Aunado al punto anterior, cabe señalar que, la globalización forma nuevas comunidades con la característica de ser transnacionales y transcontinentales; en las cuales, los derechos y obligaciones adquieren una nueva distribución; en la que, se ve acentuada a nivel mundial, la polarización y estratificación económica de la población.

Tal es el caso, que los Estados del Bienestar, son limitados para ejercer sus políticas, cambiándolas por reglas que optimizan la repartición de los trabajos en relación a la maximización de los beneficios. De acuerdo con esa relación, los conocimientos, habilidades, tecnología y capital, determinan el bienestar de un país dentro de la

sociedad global; dejando pendiente la afirmación de la existencia de un sistema de previsión social de carácter transnacional.

Es posible puntualizar que, el concepto de “globalización” comprende un aspecto muy importante, la interdependencia a nivel mundial, considerando que este aspecto ya existía en otros tiempos y que lo realmente valioso y novedoso se establece en que la interconexión entre las sociedades tiene rápidos medios de comunicación y de transporte.

De lo anterior, se evocan nociones como: la unidad, las conexiones y la proximidad; no obstante, la globalización es un concepto diverso y a veces complicado pues rompe con el concepto de unidad, con expresiones tales como:

- ★ La creación de desigualdades
- ★ La manifestación de múltiples formas.
- ★ La incrementación de la reivindicación de la diferencia cultural.

Concluyéndose que así como genera cercanía en la distancia, también crea distancias en la cercanía; considerando una gran variedad de ideas, así como, respuestas ante éstas, entre las que se pueden resumir en:

- ★ Interconexiones eficientes gracias a los medios de comunicación y transporte existentes.
- ★ La expansión de las economías locales, en las que la globalización no es constante, es un proceso que avanza y retrocede.
- ★ Las conexiones son muy variadas.
- ★ Las conexiones son de distintas naturalezas, refiriéndose a éstas como conjuntos sin límites, que se encuentran en diversos campos como: económicos, tecnológicos, culturales, mediáticos, étnicos, etc.

- ★ La diversidad que existe, derivada de todos los pueblos del mundo, en el que no se debe imponer una cierta tendencia, sino considerar a todos en la expansión de la economía de mercado. Por consiguiente, la globalización no es unitaria y homogeneizadora, sino que es variada y diversa, creadora de iniciativas y estrategias.
- ★ El propiciar el reforzamiento de los ámbitos locales, adquiriendo un nuevo protagonismo.

Con base en todos los aspectos antes citados, se puede considerar que la economía ha formado parte fundamental en la globalización, no obstante, la política nunca ha perdido sus fuerzas, ya que se ha visto la incursión del Estado en el mercado, como por ejemplo, la protección brindada hacia los sectores más poderosos de dicho mercado.

El término de globalización comprende un proceso de internacionalización de muchos aspectos de tipo económico, tecnológico, financiero, cultural, etc., expansión del mercado relacionado con la dominación del capitalismo como sistema económico y social, el no aseguramiento de la igualdad entre personas, países o zonas del planeta, una modificación en la utilización de los recursos naturales, entre otros.

Por lo anterior, globalización se entiende como la nueva regionalización, hablando en niveles sub y supranacionales.

2. REGRESIÓN LOGÍSTICA

Dentro de los estudios realizados en el campo de la estadística, se encuentra una amplia variedad de conceptos, sin embargo cabe destacar que entre los de mayor importancia por sus aportes a esta ciencia se encuentra el de “regresión”.

Este concepto tuvo sus inicios gracias a Legendre, Gauss, Laplace y Francis Galton; en los albores de 1800. Con el paso del tiempo y los avances en las investigaciones el significado de análisis de regresión ha evolucionado el conjunto de funciones matemáticas que permiten conocer la relación que es posible encontrar entre el valor de una variable dependiente o de respuesta (Y) y un conjunto de variables (X_1, X_2, \dots, X_n) denominadas predictoras o regresoras, covariantes o factores; para finalmente tener la posibilidad de predicción y explicación.

Buscando cumplir con el objetivo anterior, se encuentran diversos problemas, como: la elección de las variables a incluir para la construcción del modelo más sencillo y efectivo posible; así como, las estimaciones lo más precisas posibles o predicciones más acertadas.

Por las características de los datos se genera otro tipo de regresión, basado en el método lineal generalizado (GLM, generalized linear model), la “regresión logística”; es ampliamente utilizada en estudios médicos y campos de la salud; misma que fue retomada para investigaciones por: Bartlett (1937, análisis de proporciones), Fisher y Yates (1938, datos binarios), Joseph Berkson (1944, término logit), Jerome Cornfield (aplicación de la RL), David R. Cox (1970, The Analysis of Binary Data), John Nelder y R.W.M. Wedderburn (1972, modelos lineales generalizados GLM).

El objetivo del método de regresión logística, es encontrar y ajustar el mejor modelo de manera parsimoniosa. Se distingue del modelo de regresión lineal en la variable respuesta, ya que ésta, puede ser dicotómica o politómica.

Derivado de dicha clasificación del suceso (variable) a estudiar, es que se obtienen 2 modelos dentro de la regresión logística: el “modelo simple de regresión logística” y el “modelo politómico de regresión logística” que en las páginas siguientes se presentarán.

2.1. CONCEPTOS ASOCIADOS A LOS MODELOS DE REGRESIÓN LOGÍSTICA

2.1.1. VARIABLES

Proveniente del latín *variabilis*, que significa aquello que varía o puede variar, se define en las matemáticas, el concepto de **variable** como una función que asocia a cierto elemento de un conjunto, una cierta característica; es decir, la relación entre un componente y los valores que le serán asignados dentro de un cierto universo (rango).²

Dentro de esta conceptualización, se pueden encontrar las siguientes clases:³

- ★ **Variables dependientes o de respuesta.**- Aquellas que como su nombre describe, son determinadas por los valores que posean otras variables, también conocidas como variables respuesta o explicada. Definida por Hayman, en 1974, como: “Propiedad o característica que se trata de cambiar mediante la manipulación de otra”.
- ★ **Variables independientes o explicativas.**- Denominación para las variables manipuladas por el investigador y que con ello determinan el valor y el posible comportamiento de otras, de ahí que en otros contextos sean conocidas como explicativas.

² <http://www.scribd.com/doc/4924597/DEFINICION-DE-VARIABLES-ESTADISTICAS>

<http://definicion.de/variable/>

³ http://es.wikipedia.org/wiki/Variable_estad%C3%ADstica

La anterior clasificación refleja el grado de influencia que puede llegar a existir de unas variables sobre otras.

Por otro lado, la clasificación determinada por el tipo de valores asignados es:

- ★ **Variables categóricas o cualitativas.**- Son las variables cuyos “valores” asignados, son elementos característicos denominados “cualidades”, mismos que son representados por nombres o expresiones alfanuméricas.

Derivado de las características anteriores, se constituye una nueva organización dentro de las variables categóricas:⁴

- ☆ **Variables categóricas nominales.**- En este grupo de variables se pueden encontrar aquéllas que por las características de sus valores no es posible realizar una ordenación, por ejemplo en el caso de las elecciones, en las que la variable “voto” tiene como posibles valores los nombres de los partidos políticos en cuestión.
- ☆ **Variables categóricas ordinales.**- Este otro grupo de variables tienen la cualidad de una posible organización u ordenación, ya que aún cuando sus “valores” son categorías, entre éstas, puede existir una relación, por ejemplo de magnitud no determinada exactamente pero insinuada: nunca, raramente, la mitad de las veces, generalmente, siempre, etc.
- ★ **Variables numéricas o cuantitativas.**- Por el contrario de las variables cualitativas, los “valores” asignados a este tipo de variables son

⁴ <http://www.uaq.mx/matematicas/estadisticas/xu2.html>

números, con los cuales es posible identificar qué variable cuenta con mayor “valor” que otra.

Dentro de este tipo de variables se distingue una subclasificación, derivada del posible número de “valores” que asignados; de ahí que, si éstos pertenecen a un conjunto (un intervalo) una nueva clasificación se conforma:

- ☆ **Variables discretas.**- Cuando el número de valores por tomar dentro de cierto intervalo es finito, es decir, sus valores son números enteros o numéricamente fijos.
- ☆ **Variables continuas.**- Son denominadas continuas cuando existe un número infinito de posibles valores por asignarles, considerando entre éstos, números enteros, fraccionarios o irracionales.

2.1.2. ESCALAS DE MEDICIÓN PARA LAS VARIABLES

En muchos estudios científicos, sobre todo en las Ciencias Naturales, por el investigador Stevens, en el año 1951, la medición compone un concepto muy completo en el cual se consideran: la asignación de números a objetos y eventos según determinadas reglas o la vinculación de conceptos abstractos con indicadores empíricos en el caso de las Ciencias Sociales, estudiadas por Carmines y Zeller en 1979, procesos de suma importancia para el desarrollo de cualquier estudio.⁵

Ante este panorama, Therese L. Baker, en 1997 constituyó cuatro escalas de medición para la clasificación de las variables, dos para las variables categóricas (escalas nominal y ordinal) y dos para las variables numéricas (de intervalo y de razón).

⁵ <http://www.eumed.net/libros/2006c/203/1v.htm>

Dichas escalas de medición asociadas a las variables tienen entre sus objetivos: resumir y presentar los datos, así como determinar las pruebas que podrían llevarse a cabo.⁶

- ★ **Escala nominal.**- Dentro de esta escala se consideran aquellos valores que sólo pueden ser clasificados o contados, en donde no existe un orden natural o específico. Por otro lado, las categorías en las que se clasifican son *mutuamente excluyentes* y *colectivamente exhaustivas*. Asimismo, si se asignan números a las clasificaciones, éstos solamente servirán para la identificación de los elementos sin que el orden de los mismos afecte la medición, sin referente cuantitativo. Derivadas de esta escala se conocen las variables categóricas nominales.
- ★ **Escala ordinal.**- La cualidad cuantitativa es la principal característica que cumple esta escala, permitiendo el orden de elementos, en función de la posesión de un atributo o característica o atributo. Al igual que la escala nominal, esta escala contempla categorías mutuamente excluyentes y colectivamente exhaustivas; sin embargo, los valores concentrados tienen un orden natural, considerando como característica peculiar, que la magnitud existente entre cada posición de dicho orden, no se puede determinar con exactitud. Las variables categóricas ordinales provienen de esta escala.
- ★ **Escala intervalor.**- Con todas las características de la escala nominal y ordinal, esta escala contempla categorías mutuamente excluyentes y colectivamente exhaustivas; sin embargo, los valores concentrados tienen un orden natural, estableciendo la distancia entre una medida y otra. Las variables continuas generalmente aplican esta escala, en la cual, se carece de un valor absoluto del cero.

⁶ Universidad Panamericana, Estadística I. Prof. Andrés Sandoval H., Documento Word.

- ★ **Escala de razón.**- Absorbiendo todas las características de las escalas ordinal, nominal y de intervalo, esta medición se diferencia con la determinación de la distancia entre los intervalos de una categoría y como característica especial adicional: “el cero como ausencia de valor”.

Un ejemplo de las escalas anteriores, es posible identificarlo en la contratación de personal de una empresa; donde los resultados de evaluación de los candidatos son los siguientes:

Candidato	Sexo (Mujer=1, Hombre=2)	Edad	Orden de término de pruebas de conocimientos	Calificación del desempeño en las pruebas
Persona 1	1	24	5°	9.0
Persona 5	2	28	2°	9.6
Persona 2	2	23	4°	8.5
Persona 4	1	30	3°	9.7
Persona 3	2	26	1°	8.1

En el caso del género de las personas, se encuentra presente la escala nominal, en la que se asigna un número a cada una, dependiendo de si es hombre o mujer, valores que son claramente arbitrarios; con respecto a la edad de los candidatos, se aprecia la escala de razón, donde se puede ver la diferencia entre los valores de la escala y el cero como ausencia de valor.

Por otro lado, la escala ordinal se ve en el orden de finalización de las pruebas, ya que en ella únicamente se determina un valor para la rapidez con que contestan 2 candidatos; sin embargo, no es posible establecer exactamente si la diferencia de tiempo es grande o pequeña.

Finalmente, con la calificación de las pruebas de conocimientos, se manifiesta la escala intervalar, que representa a una variable continua y en la que si una persona obtiene cero no significa exactamente que carezca de conocimientos.

2.1.3. MOMIOS (ODDS)

Sean:

E : el evento a estudiar.

$P(E)$: la probabilidad de ocurrencia del evento.

$M(E)$: MOMIO relacionado al evento.

Se denomina “*momio*” a la razón existente entre la probabilidad de que ocurra un cierto evento y la probabilidad de que no ocurra, explicando cuanto más probable es que se produzca el evento contra que no ocurra.

$$M(E) = \frac{P(E)}{1 - P(E)}$$

La escala de medición de esta razón es:

$M(E) < 1$	Tiene mayor probabilidad de no ocurrencia.
$M(E) = 1$	Tiene igual probabilidad de ocurrencia que de no ocurrencia.
$M(E) > 1$	Tiene mayor probabilidad de ocurrencia.

Derivado de la ecuación anterior, si se considera conocido el valor de $M(E)$, se tiene el

resultado siguiente:

$$P(E) = \frac{M(E)}{1 + M(E)}$$

Igualdad equivalente a la ecuación del *momio* que de igual forma representa que tan probable es que un evento se produzca.

Cabe señalar que dicha ecuación tiene una relación directa a la ecuación del momio, donde se cumple que:

- ★ Si $M(E)$ aumenta también lo hará $P(E)$,
- ★ si $P(E) = 0$ entonces $M(E) = 0$; y,
- ★ si $P(E) \rightarrow 1$ entonces $M(E) \rightarrow \infty$.

COCIENTE DE MOMIOS (ODDS RATIO)

Relacionado con la $P(E)$ se encuentra el concepto de *riesgo relativo* (RR), el cual representa la probabilidad de que se presente un evento E , por causa del factor A o B .

$$RR(E) = \frac{P_A(E)}{P_B(E)}$$

Por otro lado, dado que $M(E)$ y $P(E)$ son equivalentes y la existencia del riesgo relativo, se tiene el “*cociente de momios*” (CE) constituido por lo siguiente:

$$CM(E) = \frac{M_A(E)}{M_B(E)} = \frac{\frac{P_A(E)}{1 - P_A(E)}}{\frac{P_B(E)}{1 - P_B(E)}}$$

Donde:

A – presencia / exposición al factor A .

B – no presencia / exposición al factor A .

Los resultados de esta relación de momios, explican que tan probable es que el evento E ocurra por uno de los factores, considerando lo siguiente:

$CM(E) < 1$	Tiene mayor probabilidad de ocurrencia el evento E , por consecuencia del factor B .
$CM(E) = 1$	Tiene igual probabilidad de ocurrencia el evento tanto por el factor A como por el B .
$CM(E) > 1$	Tiene mayor probabilidad de ocurrencia el evento E , por consecuencia del factor A .

Entre la mayoría de los estudios involucrados en este concepto la utilización de las tablas de contingencia es común, denominando al CM como “razón de productos cruzados”.

2.1.4. FUNCIÓN EXPONENCIAL

Es aquella función que a cada número real x le asigna el número e (constante de Euler) elevado a la potencia x .

$$\exp(x) = e^x$$

Cuenta con las siguientes características:

- ★ La función es positiva, sea cual sea el valor de x .
- ★ El exponencial de 0 es 1 y los valores para los siguientes intervalos:
 - ☆ $\exp(x) = e^x$ con $x < 0$, sus valores son menores que 1.
 - ☆ $\exp(x) = e^x$ con $x > 0$, sus valores son mayores que 1.
- ★ la función que lo representa es creciente y se incrementa más rápidamente conforme $x \rightarrow \infty$.

★ $e^x e^y = e^{x+y}$

★ $e^{-x} = 1/e^x$

2.1.5. FUNCIÓN LOGARÍTMICA NATURAL O NEPERIANO

Expresada con las letras “ln” (logaritmo neperiano), la función logarítmica es la función inversa a la exponencial, es decir, si $\exp(x) = y$ entonces $\ln(y) = x$; $\ln(\exp(x)) = x$; $y, \exp(\ln(x)) = x$.

Cuenta con las siguientes características:

- ★ Es positiva para $x > 1$, negativa para $x < 1$ y para $x = 1$ su valor es 0.
- ★ Es creciente.
- ★ Sólo está definida para $x > 0$
- ★ $\ln(xy) = \ln(x) + \ln(y)$
- ★ $\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)$

2.1.6. FAMILIA EXPONENCIAL DE DISTRIBUCIONES

Sea Y una variable aleatoria discreta o continua, con función de probabilidad o densidad respectivamente, que depende del parámetro θ . Se dice que la distribución de Y pertenece a la familia exponencial si puede expresarse de la forma:

$$\begin{aligned} f(Y; \theta, \phi) &= \exp\left[\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + h(Y, y_i)\right] = \exp\left[\frac{y_i \theta_i}{a(\phi)} - \frac{b(\theta_i)}{a(\phi)} + h(Y, \phi)\right] \\ &= \exp\left[\frac{y_i \theta_i}{a(\phi)}\right] \exp\left[-\frac{b(\theta)}{a(\phi)}\right] \exp[h(Y, \phi)] \end{aligned}$$

Donde ϕ es un parámetro de escala y θ_i es el parámetro natural de localización.

Dentro de esta familia, se define lo siguiente:

- ★ $\mu = E(Y) = \frac{db(\theta_i)}{d(\theta_i)}$
- ★ $Var(Y) = \frac{d^2b(\theta_i)}{d\theta_i^2} a(\phi) = \frac{d\mu}{d\theta_i} a(\phi)$
- ★ $Var(\mu) = \frac{Var(Y)}{a(\phi)} = \frac{d\mu}{d\theta_i}$; representación de la dependencia de la varianza de la variable respuesta con su media. Cuando $a(\phi) = 1$ queda determinada la igualdad: $\frac{d\theta_i}{d\mu} = \frac{1}{Var(Y)}$. Esta característica se presenta en todos los integrantes de la familia exponencial, con excepción de la distribución normal.

Entre las distribuciones referentes a esta familia se encuentran:

- ★ Normal
- ★ Poisson
- ★ Binomial
- ★ Exponencial
- ★ Gamma
- ★ Pareto

2.1.7. MODELOS LINEALES GENERALIZADOS

El modelo lineal generalizado (MLG), es un modelo unificador de las características de modelos de regresión lineal y no lineal, con variables respuesta numéricas y categóricas, dando apertura a las distribuciones de respuesta no normales.

En este modelo es requisito que la variable respuesta sea un miembro de la familia exponencial, cuya ecuación está dada por: $f(y_i; \theta_i) = \exp[a_i(y_i)b_i(\theta_i) + c_i(\theta_i)d_i(y_i)]$ y donde $b_i(\theta_i)$ es el parámetro natural.

El MLG está compuesto por los tres elementos siguientes:

- ★ **Componente aleatoria o variables de respuesta:** vector aleatorio $Y = (Y_1, \dots, Y_n)$, correspondiente a la variable de respuesta, con componentes Y_i independientes e idénticamente distribuidos, cuya distribución es un elemento de la familia exponencial, se encuentra en forma canónica y depende sólo de un parámetro θ , es decir, es de la forma:

$$f(y_i; \theta_i) = \exp[a_i(y_i)b_i(\theta_i) + c_i(\theta_i)d_i(y_i)]$$

- ★ **Componente sistemática o predictor lineal:** matriz cuyas entradas (η_i) relacionan los parámetros con las variables explicativas, cuya representación está dada por el producto $(g[E(Y_i)] = \eta_{(n \times 1)} = X_{n \times (k+1)}\beta_{(k+1) \times 1})$, donde:

- ★ Un vector de parámetros: $\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix}$

- ★ y una matriz de variables explicativas: $X = \begin{pmatrix} 1 & X_{11} & \dots & X_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{pmatrix}$.

- ★ Resultando: $\eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 X_{11} + \dots + \beta_k X_{1k} \\ \vdots \\ \beta_0 + \beta_1 X_{n1} + \dots + \beta_k X_{nk} \end{pmatrix}$.

- ★ **Función liga, de vínculo o de enlace:** función (g) monótona y diferenciable, que relaciona los componentes aleatorio y sistemático, es decir, el predictor lineal (η_i) y el valor esperado de la variable de respuesta ($\eta_i = E(Y_i); i = 1, \dots, n$).

$$g(\mu_i) = \sum_{j=0}^k \beta_j x_{ij} \quad ; \quad i = 1, \dots, n$$

La relación entre $E(Y)$ y el predictor lineal está dada por:

$$\mu_i = g^{-1}(\eta_i) = g^{-1}(X_i\beta\eta_i).$$

Por otro lado, las ligas canónicas son aquellas funciones que transforman la esperanza de Y_i al parámetro natural ($\eta_i = \theta_i$), entre las existentes para el MLG se encuentran⁷:

Liga:	
Identidad	$\eta_i = \mu_i$ (para la distribución normal)
Logarítmica	$\eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$ (para la distribución binomial)
Logarítmica	$\eta_i = \ln(\lambda)$ (para la distribución Poisson)
Recíproca	$\eta_i = \frac{1}{\lambda_i}$ (para las distribuciones exponencial o gamma)
Probit	$\eta_i = \Phi^{-1}[E(Y_i)]$
Log-log complementaria	$\eta_i = \ln\{\ln[1 - E(Y_i)]\}$
De la familia de potencias	$\eta_i = \ln\begin{cases} E(Y_i)^\lambda; \lambda \neq 0 \\ \ln[E(Y_i)]; \lambda = 0 \end{cases}$

Correspondientes a este grupo de MLG, se encuentran los modelos:

- ★ Log-lineal
- ★ Probit
- ★ Logístico
- ★ De regresión lineal

MODELO DE REGRESIÓN POISSON

En el grupo de los modelos concernientes a los MLG, se encuentra, el que tiene escenarios en los cuales la distribución de la variable respuesta no es una distribución normal; tal es el caso, del modelo de regresión de Poisson, que se manifiesta cuando la

⁷ Montgomery, Peck, Vining; Introducción al Análisis de Regresión Lineal. 3ª edición, Editorial CECSA.

variable de respuesta representa la incidencia de un evento, el cual es relativamente raro.

Sea Y_i la variable respuesta, la cual busca ser un contador de un evento, por lo cual cuenta con la distribución Poisson: $f(Y) = \frac{\lambda^y e^{-\lambda}}{y!}$; $Y = 0, 1, \dots$; $\lambda > 0$. Con esperanza y varianza: $E(Y) = \lambda$ y $Var(Y) = \lambda$ respectivamente. De donde se supondrá $E(Y_i) = \lambda_i$; y como función liga, la función identidad: $g(\lambda_i) = \eta_i = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k = X_i' \beta$, con relación entre la esperanza y el predictor lineal: $\lambda_i = g^{-1}(\eta_i) = g^{-1}(X_i' \beta)$.

Cabe destacar que existen 2 funciones de liga que pueden ser usadas con la distribución Poisson, la arriba citada y:

- ★ La liga logarítmica $g(\lambda_i) = \ln(\lambda_i) = X_i' \beta$, con relación entre la esperanza y el predictor lineal: $\lambda_i = g^{-1}(X_i' \beta) = e^{X_i' \beta}$.

Esta función de liga cuenta con la propiedad especial de asegurar que los valores predichos de la variable respuesta serán mayor que cero.

El modelo de regresión de Poisson queda determinado por:

$$Y_i = E(Y_i) + \varepsilon_i ; i = 1, \dots, n$$

ESTIMACIÓN DE PARÁMETROS

Considerando el método de máxima verosimilitud, una muestra aleatoria de n observaciones de la respuesta Y y X los predictores; la función de máxima verosimilitud se conforma por:

$$L(Y, \beta) = \prod_{i=1}^n f_i(Y_i) = \prod_{i=1}^n \frac{e^{-\lambda_i} \lambda_i^{Y_i}}{Y_i!} = \prod_{i=1}^n \frac{\lambda_i^{Y_i} \exp(-\sum_{i=1}^n \lambda_i)}{\prod_{i=1}^n Y_i!}$$

Llevando a cabo el desarrollo correspondiente, el cual resultará ser similar al del modelo de regresión logística; se tiene, que una vez encontrados los estimados β^* , el modelo ajustado queda determinado por:

$$\lambda_i = g^{-1}(X_i' \hat{\beta}) = X_i' \hat{\beta} \text{ (utilizando la función liga identidad)}$$

$$\lambda_i = g^{-1}(X_i' \hat{\beta}) = e^{X_i' \hat{\beta}} \text{ (utilizando la función liga logarítmica)}$$

2.2. MODELOS DE REGRESIÓN LOGÍSTICA

Partiendo del modelo de regresión lineal simple y buscando evaluar el efecto que un cierto factor representado por la variable X tenga sobre la variable (dicotómica) dependiente Y , la cual puede asumir los valores 0 y 1, con los que refleja la ocurrencia ($Y = 1$) o la no ocurrencia del evento ($Y = 0$).

$$Y = \beta_0 + \beta_1 x$$

Se presenta en algunos casos el problema de la falta de representación en la realidad, debido a su caracterización lineal y a la no consideración de los valores de respuesta al intervalo $[0,1]$.

Por tal motivo, se busca el cambio del modelo de regresión simple, dicho cambio es logrado mediante la transformación logística de la probabilidad P de que el evento E ocurra.

Derivado de que el momio y la probabilidad de ocurrencia de un evento son equivalentes, la transformación sería aplicar el logaritmo al momio de esta probabilidad.

Conformando el modelo de regresión logística, relacionando una variable binaria de respuesta y un grupo de variables independientes, se tienen los siguientes modelos:

1. Para la regresión logística simple (una sola variable independiente):

$$Y = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x$$

2. Para la regresión logística múltiple (k variables independientes):

$$Y = \ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

2.2.1. MODELO DE REGRESIÓN LOGÍSTICA CON VARIABLE DE RESPUESTA BINARIA

Este modelo evalúa el efecto que un cierto evento tenga en la ocurrencia de otro, para lo cual se necesitan:

Y : variable dependiente o variable en la que se quiere observar la ocurrencia, cuyos valores posibles serán asumidos del conjunto $\{0,1\}$, asociando el valor de uno con la ocurrencia del evento y cero en el caso contrario.

De esta manera, la variable Y tiene un comportamiento Bernoulli con parámetro π , lo que se interpreta como $P(Y = 1) = \pi$; $P(Y = 0) = 1 - \pi$ y con las propiedades:

- ★ Valor esperado: $E(Y) = 1(\pi) + 0(1 - \pi) = \pi$
- ★ $f(Y; \pi) = \pi^Y (1 - \pi)^{1-Y} = (1 - \pi) \exp \left[Y \log \left(\frac{\pi}{1-\pi} \right) \right]$;

función generalizada de probabilidad, que incluye la función de liga logit, $\log \left(\frac{\pi}{1-\pi} \right)$.

x : única variable explicativa que determinará la ocurrencia del evento medido por Y .

β_i : parámetros asociados al modelo.

Elementos con los que el modelo de regresión logística simple (una sola variable) queda constituido como:

$$Y = \ln \left(\frac{\pi}{1 - \pi} \right) = \beta_0 + \beta_1 x$$

De donde despejando la probabilidad y aplicando un poco de álgebra, se obtiene que:

$$\pi = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x)}$$

AJUSTE DEL MODELO

Al intentar resolver algunos casos a partir del modelo de regresión lineal simple ($Y = \beta_0 + \beta_1 x + \varepsilon$), con variable de respuesta binaria, es posible encontrar diversas conjeturas, como:

- ★ Dado que y_i es una variable binaria, el error sólo puede ser expresado por:

$$\varepsilon = \begin{cases} 1 - x\beta_1; & \text{con } Y = 1 \\ -x\beta_1; & \text{con } Y = 0 \end{cases}$$

Por lo que esta representación no permite la normalidad.

- ★ Otro punto a resaltar, es que el valor de la varianza de las observaciones depende de la media.

Afirmación derivada de que definidos la esperanza y el error como $E(Y) = \pi$ y $\varepsilon = Y - \pi$ respectivamente, se tiene que:

$$\sigma_Y^2 = E[Y - E(Y)]^2 = \pi(1 - \pi) = E(Y)[1 - E(Y)]$$

Con lo que se igualan las varianzas del error y de Y .

- ★ Por último, la restricción $0 \leq E(y_i) = \pi_i \leq 1$, complica la selección de los datos, puesto que éstos siempre deben mantener los valores de la variable respuesta (0 y 1).

Con base a los supuestos anteriores, se propone una función sigmoide, cuya gráfica tiene una loma y un valle, con la característica de ser monótona creciente (decreciente).

Sea $E(Y) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}} = \pi$ la función propuesta, donde el predictor lineal (elemento del MLG) es $\eta = \beta_0 + \beta_1 X$. De esta forma, se tiene que:

$$\pi(1 + e^{\beta_0 + \beta_1 x}) = e^{\beta_0 + \beta_1 x} = \pi + \pi(e^{\beta_0 + \beta_1 x})$$

Despejando:

$$\pi = e^{\beta_0 + \beta_1 x} - \pi(e^{\beta_0 + \beta_1 x}) = (e^{\beta_0 + \beta_1 x})(1 - \pi)$$

Para concluir con la expresión:

$$\frac{\pi}{1 - \pi} = e^{\beta_0 + \beta_1 x}$$

Y como consecuencia $\ln\left(\frac{\pi}{1 - \pi}\right) = \beta_0 + \beta_1 x$; que resulta ser la transformación logit o función liga logarítmica $\eta = \ln\left(\frac{\pi}{1 - \pi}\right)$ correspondiente al momio que representa la proporción de que un evento suceda.

Logrando así, la definición del modelo de regresión logística como:

$$Y = \ln\left(\frac{\pi}{1 - \pi}\right) = \beta_0 + \beta_1 x + \varepsilon_i$$

Cabe destacar que:

- ★ Si $x \rightarrow \infty$; $\pi(x) \rightarrow 0$ cuando $\beta_1 < 0$ y $\pi(x) \rightarrow 1$ cuando $\beta_1 > 0$.
- ★ La curva tiende a ser una horizontal, cuando $\beta_1 \rightarrow 0$.
- ★ $E(Y) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$ alcanza su valor máximo cuando $\pi(X) = \frac{1}{2}$ y $x = \frac{-\beta_0}{\beta_1}$; este resultado puede ser comprobado al efectuar la primera y segunda derivada y posteriormente resolviendo dicho sistema.

ESTIMACIÓN DE PARÁMETROS POR MÁXIMA VEROSIMILITUD

Dentro de la estimación de parámetros, existen diversos métodos que pueden ser empleados en las regresiones, como el de mínimos cuadrados y máxima verosimilitud, entre otros.

Sin embargo, el método de mínimos cuadrados al basarse en encontrar aquellos valores de los coeficientes que redujeran notablemente la suma de los residuales (diferencias entre lo observado y lo esperado), era mucho más práctico al referirse a una regresión lineal puesto que el problema recaía en un sistema de ecuaciones lineales.

Dadas las características de la regresión logística, el método generalmente usado para la estimación de parámetros es el de *máxima verosimilitud*.

Sea el modelo de regresión logística:

$$Y = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x + \varepsilon_i = \underline{X}\underline{\beta} + \varepsilon$$

Donde $\underline{X} = (1, x)$ y $\underline{\beta} = (\beta_0, \beta_1)$ el vector de parámetros que se estimará por medio del método de máxima verosimilitud.

Considerando que $y_i \sim \text{Bernoulli}(\pi)$, cada variable y_i tiene como función de distribución de probabilidades:

$$f_i(y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}; i = 1, \dots, n$$

Por lo cual, para las variables aleatorias independientes, la función de verosimilitud queda determinada por:

$$L(\underline{Y}; \underline{\beta}) = \prod_{i=1}^n f_i(y_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

Seguindo el método de máxima verosimilitud, se requiere maximizar el logaritmo de la verosimilitud, para lo cual se tiene:

$$\begin{aligned} \ln L(\underline{Y}; \underline{\beta}) &= \sum_{i=1}^n \{y_i \ln(\pi_i) + (1 - y_i) \ln(1 - \pi_i)\} \\ &= \sum_{i=1}^n \{y_i (\ln(\pi_i) - \ln(1 - \pi_i)) + \ln(1 - \pi_i)\} \\ &= \sum_{i=1}^n \left\{ y_i \left(\ln \left(\frac{\pi_i}{1 - \pi_i} \right) \right) \right\} + \sum_{i=1}^n \{ \ln(1 - \pi_i) \} \\ &= \sum_{i=1}^n \left\{ y_i \left(\underline{X}_i \underline{\beta} \right) + \ln(1 + e^{\underline{X}_i \underline{\beta}})^{-1} \right\} = \sum_{i=1}^n \left\{ y_i \left(\underline{X}_i \underline{\beta} \right) \right\} - \sum_{i=1}^n \left\{ \ln(1 + e^{\underline{X}_i \underline{\beta}}) \right\} \end{aligned}$$

Posteriormente derivando:

$$\frac{\partial \ln L(\underline{\beta})}{\partial \beta_0} = \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{e^{\underline{X}_i \underline{\beta}}}{1 + e^{\underline{X}_i \underline{\beta}}}$$

$$\frac{\partial \ln L(\underline{\beta})}{\partial \beta_1} = \sum_{i=1}^n y_i x_i - \sum_{i=1}^n \ln \left(\frac{e^{\underline{X}_i \underline{\beta}}}{1 + e^{\underline{X}_i \underline{\beta}}} \right) x_i$$

Que puede ser visto como:

$$\frac{\partial \ln L(\underline{\beta})}{\partial \underline{\beta}} = \sum_{i=1}^n y_i \underline{X}_i - \sum_{i=1}^n \ln \left(\frac{e^{\underline{X}_i \underline{\beta}}}{1 + e^{\underline{X}_i \underline{\beta}}} \right) \underline{X}_i = \sum_{i=1}^n y_i \underline{X}_i - \sum_{i=1}^n \ln \left(\frac{1}{1 + e^{-(\underline{X}_i \underline{\beta})}} \right) \underline{X}_i$$

De donde se puede observar que $\frac{\partial \ln L(\underline{\beta})}{\partial \underline{\beta}} = 0 \Rightarrow \underline{X}_i \pi_i = y_i \underline{X}_i$; con lo que y_i es igual a la probabilidad estimada de $P(Y = 1)$; sea \hat{y}_i dicha estimación, el sistema a resolver será $\underline{X}(y_i - \hat{y}_i) = 0$.

Por medio del método de Newton-Raphson, primero obteniendo la derivada respecto a los parámetros, es decir, $\frac{\partial \underline{X}(y_i - \hat{y}_i)}{\partial \underline{\beta}}$ que de otra manera sería como buscar:

$$\frac{\partial \underline{X}(\pi_i)}{\partial \underline{\beta}} = - \left(\frac{\partial \pi_i}{\partial \underline{\beta}} \right).$$

De donde se constituye una matriz diagonal V, tal que,

$$\begin{aligned} V(\hat{\underline{\beta}}) &= \left[\sum_{i=1}^n \frac{e^{(-\underline{X}\hat{\underline{\beta}})}}{\{1 + e^{(-\underline{X}\hat{\underline{\beta}})}\}^2} \underline{X}\underline{X}' \right]^{-1} = \left[\sum_{i=1}^n \left(\frac{1}{1 + e^{(-\underline{X}\hat{\underline{\beta}})}} \right) \left(\frac{e^{(-\underline{X}\hat{\underline{\beta}})}}{1 + e^{(-\underline{X}\hat{\underline{\beta}})}} \right) \underline{X}\underline{X}' \right]^{-1} \\ &= \left[\sum_{i=1}^n \hat{\pi}_i(1 - \hat{\pi}_i) \underline{X}\underline{X}' \right]^{-1} = \sum_{i=1}^n y_i \underline{X} - \sum_{i=1}^n \ln \left(\frac{1}{1 + e^{(-\underline{X}\hat{\underline{\beta}})}} \right) \underline{X} \end{aligned}$$

Cuyos valores en la diagonal son de la forma $\hat{\pi}_i(1 - \hat{\pi}_i)$.

Ahora bien, tomando el estimador inicial $\underline{\beta}^{(0)}$, el cual puede ser encontrado por diversos métodos y entre ellos usando la expresión:

$$\underline{\beta}^{(0)} = \begin{bmatrix} \underline{\beta}_0^{(0)} \\ \underline{\beta}_1^{(0)} \end{bmatrix} = \begin{bmatrix} \ln \left(\frac{\hat{\omega}_1}{\hat{\omega}_0} \right) - 0.5 \left(\frac{\hat{\mu}_1^2 - \hat{\mu}_0^2}{\hat{\sigma}^2} \right) \\ \frac{\hat{\mu}_1^2 - \hat{\mu}_0^2}{\hat{\sigma}^2} \end{bmatrix}$$

Donde:

- ★ $\hat{\omega}_i$ es la estimación de $P(Y = i)$, definido por el promedio de los valores de Y , cuando $Y = i$.

- ★ $\hat{\mu}_i$ es el promedio de los valores de X , cuando $Y = i$.
- ★ $\hat{\sigma}^2 = \frac{[(n_0-1)s_0^2 + (n_1-1)s_1^2]}{n_0+n_1-2}$, s_i^2 definido como la varianza muestral y n_i el tamaño de la muestra correspondientes a cuando $Y = i$.

Por lo que, continuando el método de Newton-Raphson para cada iteración, se tendrá:

$$\underline{\beta}^{(l+1)} = \underline{\beta}^{(l)} + (\underline{X}'V^{(l)}\underline{X})^{-1}\underline{X}'(y_i - \hat{\pi}_i^{(l)})$$

$$\hat{\pi}_i^{(l)} = \left[\frac{1}{1 + e^{(-\underline{X}\underline{\beta}^{(l)})}} \right]$$

$$V^{(l)} = \text{diag}[\hat{\pi}_i^{(l)}(1 - \hat{\pi}_i^{(l)})]$$

Donde $\hat{\pi}_i^{(l)}$ denota la estimación del parámetro π en la iteración l . El proceso finaliza en el momento en que $\underline{\beta}^{(l+1)} \cong \underline{\beta}^{(l)}$.

INTERPRETACIÓN DE PARÁMETROS

Sea el caso base en el cual el predictor lineal sólo consta de un regresor, es decir $\hat{\eta}(x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$, que con valor ajustado en x_{i+1} es: $\hat{\eta}(x_{i+1}) = \hat{\beta}_0 + \hat{\beta}_1 x_{i+1}$ y cuya diferencia entre ambos valores es :

$$\hat{\eta}(x_{i+1}) - \hat{\eta}(x_i) = \hat{\beta}_1$$

Por otro lado, considerando que $\hat{\eta}(x_j) = \ln\left(\frac{\pi_j}{1+\pi_j}\right)$, la diferencia anterior es equivalente a:

$$\ln\left(\frac{\frac{\pi_{X_{i+1}}}{1 + \pi_{X_{i+1}}}}{\frac{\pi_{X_i}}{1 + \pi_{X_i}}}\right) = \hat{\beta}_1$$

De donde al observar conforme a la expresión del cociente de momios:

$$CM(E) = \frac{M_A(E)}{M_B(E)} = \frac{\frac{\pi_A(E)}{1-\pi_A(E)}}{\frac{\pi_B(E)}{1-\pi_B(E)}}$$

se constituye lo siguiente:

$$\ln(CM) = \beta(x_{i+1} - x_i)$$

$$CM = e^{\beta(x_{i+1}-x_i)}$$

Particularmente, cuando $x_{i+1} = x_i + 1$:

$$\ln(CM) = \hat{\beta}_1(x_{i+1} - x_i) = \hat{\beta}_1[x_i - (x_i - 1)] = \hat{\beta}_1$$

Se obtiene una expresión para $\hat{\beta}_1$, de la cual resulta otra para CM :

$$\hat{M}_R = \frac{\frac{\pi_{X_{i+1}}}{1 + \pi_{X_{i+1}}}}{\frac{\pi_{X_i}}{1 + \pi_{X_i}}} = e^{\hat{\beta}_1}$$

Resultado cuya interpretación muestra el aumento estimado en la probabilidad de éxito asociado con un cambio en el valor de la variable, que en general, con una variación de w unidades en X_i , $\hat{M}_R = e^{w\hat{\beta}_1}$.

Para el caso del modelo logístico múltiple, $e^{\hat{\beta}_j}$ resulta ser el cociente para el regresor X_j , considerando que todos los X_k con $k \neq j$ son constantes.

NOTAS:

- ★ Siempre se considera que la diferencia entre x_j y x_i debe ser de uno.

- ★ $\beta_1 = 0$ significaría que para cualquier valor x_i se tendrá la misma probabilidad de que $(Y = 1)$ o $(Y = 0)$.
- ★ Por otro lado, $\beta_i = 0$; con $i \neq 0$:

$$\Rightarrow \pi = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x)} = \frac{1}{1 + e^{-\beta_0}}$$

Con lo que se observa que π sigue siendo una constante que no depende de x_i , con lo que se afirma la independencia entre la variable Y y x_i (cualquiera que sea su valor).

PRUEBAS DE SIGNIFICANCIA

Al igual que en la regresión lineal, la regresión logística requiere de una evaluación que brinde confianza en el momento de la aplicación de cada modelo, para tal efecto, son considerados: el estadístico log-verosimilitud y el estadístico de Wald.

Procedimientos que resultan ser muy similares entre sí cuando las muestras tienen un tamaño considerable, sin embargo, en el caso contrario, los textos recomiendan la aplicación del método de razones de verosimilitud.

ESTADISTICO LOG-VEROSIMILITUD Y RAZÓN DE VEROSIMILITUD

Recordando que la función de verosimilitud tiene la forma:

$$L(y_1, \dots, y_n; \beta) = \prod_{i=1}^n f_i(y_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

Defínase la devianza como:

$$\delta = -2 \ln L$$

Donde:

- ★ Si: $L < 1 \Rightarrow \ln L < 0 \Rightarrow \delta > 0$
- ★ El modelo será exitoso si: $L = 1 \Rightarrow \delta = 0$; en contraste será fallido si existe una lejanía grande.
- ★ Considerando las características del modelo nulo, la lejanía asociada a este último, será siempre mayor que la correspondiente a aquel que involucre una o más variables independientes.

Por otro lado, cuando se construye un modelo de regresión entran en relación dos lejanías, las correspondientes al ajustado y al nulo, este último descrito por $\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0$, que para el caso de la regresión logística simple implica que la única variable independiente no tiene relación alguna con la dependiente.

Sea el cociente o la razón de verosimilitud⁸ (RV) definida como la diferencia entre la lejanía del modelo ajustado (δ) y la del nulo (δ_0); misma que, mide la aportación que la variable independiente tiene o de otra manera, la afectación que su valor puede tener sobre el valor de la variable respuesta.

$$RV = \delta_0 - \delta = -2\ln(L_0) - [-2\ln(L)] = -2[\ln(L_0) - \ln(L)] = -2\ln\left(\frac{L_0}{L}\right)$$

Esta diferencia entre lejanías, se distribuye ji-cuadrada con n grados de libertad, los cuales, son determinados por el número de variables independientes en el modelo ajustado; que para el caso de la regresión logística simple es un grado.

Así mismo, la evaluación del modelo será determinada por la prueba con hipótesis:

H_0 : El modelo ajustado no es adecuado vs H_1 : El modelo ajustado es adecuado

⁸ En diversos textos, la razón de verosimilitud es denominada con la palabra inglesa "deviance", término acuñado en 1972 por Nelder y Wedderburn.

Que implicarán que:

- ★ Si $RV \leq \chi_{\alpha, n-p}^2$, se rechaza H_0 .
- ★ En otro caso, si $RV > \chi_{\alpha, n-p}^2$, la hipótesis H_0 no se rechaza.

PRUEBA DE HIPÓTESIS DE MÁXIMA VEROSIMILITUD PARA LA VARIABLE INDEPENDIENTE

Como se pudo apreciar en lo anterior, la razón de verosimilitudes es de gran utilidad en el momento de determinar si una variable debe ser incluida en el modelo como explicativa del comportamiento de la dependiente o no.

Desde otro punto de vista, el análisis necesario para la elección de la variable se basa en la decisión de si el coeficiente que acompaña a ésta, es cero o distinto a cero.

Considerando la hipótesis: $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

Sean A y B conjuntos, A conformado por la única variable explicativa ($A = \{x\}$) y B asociado al modelo nulo (ausencia de variable independiente), siendo este último el resultado de suponer cierta la hipótesis $H_0: \beta_1 = 0$.

Cuyas respectivas funciones de verosimilitud serán consideradas como: L_A y L_B ; y, donde la razón de verosimilitudes será dada por: $RV = -2\ln\left(\frac{L_B}{L_A}\right)$ con una distribución Ji-cuadrada con 1 grado de libertad (*grados de libertad = número de variables en el conjunto A - número de variables en el conjunto B*).

De donde, los criterios de decisión quedan determinados como:

- ★ Si $\lambda(\beta_2|\beta_1) \geq \chi_{\alpha, r}^2$, se rechaza H_0
- ★ Si $\lambda(\beta_2|\beta_1) < \chi_{\alpha, r}^2$, no se rechaza H_0

Lo que simboliza que la presencia de la variable explicativa aumenta o disminuye el nivel de la explicación brindada por el modelo.

ESTADÍSTICO DE WALD

Dentro de la regresión logística, otro método ampliamente conocido para la evaluación de mismo, es la *prueba de Wald*.

El fin de este procedimiento de igual forma que aquel que utiliza el estadístico de máxima verosimilitud, consiste en determinar si el coeficiente que acompaña a la variable independiente es significativamente distinto de cero.

Para llevar a cabo el objetivo anterior, es utilizado el estadístico:

$$Z_{Wald} = \frac{\hat{\beta}_i - \beta_i}{\hat{E}E(\hat{\beta}_i)}$$

el cual sigue una distribución normal estándar y donde se tiene que es equivalente a:

$$W = \frac{(\hat{\beta}_i - \beta_i)^2}{\widehat{Var}(\hat{\beta}_i)}$$

con una distribución ji-cuadrada con un grado de libertad.

Por otro lado, recordando que el intervalo de confianza para los coeficientes β_i está determinado por $\{\beta | W(\beta) < q_{1-\alpha}\}$, se tiene que:

- ★ Si $W \leq q_{1-\alpha}$ el modelo ajustado es adecuado. En otro caso,
- ★ Si $W > q_{1-\alpha}$ el modelo ajustado no es adecuado.

INTERVALOS DE CONFIANZA

Dada la búsqueda del mejor modelo de regresión, existe otra forma en la cual se puede evaluar el modelo, la cual, consiste en encontrar intervalos dentro de los cuales con cierta confiabilidad el valor buscado se encuentre.

En el caso de la regresión logística, dichos intervalos pueden ser construidos para los elementos siguientes:

PARA β_1

Considerando el modelo $Y = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x$ de la regresión logística, el intervalo de confianza asociado con la variable β_i es el determinado por:

$$\hat{\beta}_i \pm z_{\alpha/2} \hat{E}E(\hat{\beta}_i)$$

Donde:

- ★ $\hat{\beta}_i$ es el estimado de β_i .
- ★ $z_{\alpha/2}$ es el punto porcentual de una distribución normal (0,1).
- ★ $\hat{E}E$ es la estimación del error estándar que puede ser visto como $\widehat{Var}(\hat{\beta}_i)$.

PARA EL CAMBIO DE MOMIOS

Dentro de la regresión logística se tienen elementos importantes, los cuales forman parte fundamental de este tipo de regresión, conocidos como “momios”, por lo cual es substancial conocer los intervalos asociados a estos conceptos.

Por medio de la construcción de los momios y del modelo de regresión logística $\left(\left(\frac{\pi}{1-\pi}\right) = \exp[\beta_0 + \beta_1 x]\right)$, y debido a que $\exp(\beta_1)$ representa el cambio en los momios

derivado de las variaciones por unidad de x , es considerable la afirmación de que los intervalos para los momios están descritos como:

$$\exp[\hat{\beta} \pm z_{\alpha/2} \hat{E}E(\hat{\beta})]$$

PARA π

Derivado de la búsqueda de un estimador para la probabilidad π , se tiene la inquietud de construir su intervalo.

Por lo anterior, para el caso en el que la muestra es lo suficientemente grande se puede considerar el intervalo como:

$$\hat{y} \pm t_{(n-2)}^{1-\alpha/2} \sqrt{\hat{\pi}^2 \left(\frac{1}{n} + \frac{(x - \bar{X})^2}{S_{xx}} \right)}$$

Sin embargo, una estructura usada en los softwares estadísticos, para la construcción de dichos intervalos, es la definida por⁹:

$$\frac{\exp[\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{(n-2)}^{1-\alpha/2} EE(\hat{\beta})]}{1 + \exp[\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{(n-2)}^{1-\alpha/2} EE(\hat{\beta})]}$$

⁹ Agresti, Alan. An introduction to categorical data analysis. 1996

**2.2.2. EJEMPLO DEL MODELO SIMPLE DE REGRESION LOGÍSTICA CON
 VARIABLE DE RESPUESTA BINARIA**

Sea un estudio realizado a 260 individuos, con la finalidad de ser clasificados en consumidores o no consumidores de drogas y en mujeres y hombres.

Resumiendo los datos obtenidos aleatoriamente y asignando los valores de 3 y 4 si es hombre o si es mujer respectivamente, de igual manera el valor de 1 y 2 si es que la persona consume drogas o no. Cabe señalar que el software estadístico SPSS recodifica esta última variable con los valores 0 y 1 respectivamente.

Consumidor de drogas	Código	masculino	femenino
		3	4
sí	1	47	23
no	2	98	92

Codificación de la variable dependiente

Valor original	Valor interno
1.00	0
2.00	1

Por otro lado, suponiendo:

- ★ **Variable dependiente:** la variable “*sn*”, que representará la afirmación o negación de que el individuo haya consumido drogas.
 - ☆ Que obtiene el valor codificado de 1 en el caso de que el individuo no haya consumido drogas y de 0 en el caso contrario.
- ★ **Variable independiente:** la variable “*fm*”, corresponderá al género de la persona, indicando con valor de 3 si es masculino y 4 en el caso femenino.

Considerando las significancias arrojados por el software en la prueba de Ji-cuadrada y para la variable independiente, se puede apreciar que este último valor es muy cercano al primero.

Pruebas omnibus sobre los coeficientes del modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	5.118	1	.024
	Bloque	5.118	1	.024
	Modelo	5.118	1	.024

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95.0% para EXP(B)	
								Inferior	Superior
Paso 1(a)	fm	.651	.293	4.945	1	.026	1.918	1.080	3.406
	Constante	-1.220	.996	1.498	1	.221	.295		

a. Variable(s) introducida(s) en el paso 1: *fm*.

Cabe señalar, que la asociación entre el género del individuo y el consumo de drogas viene citado por el cociente de momios correspondiente en la tabla anterior al “exponencial del coeficiente asociado a la variable *fm*”.

2.2.3. MODELO DE REGRESIÓN LOGÍSTICA MÚLTIPLE CON VARIABLE DE RESPUESTA BINARIA

Derivado del objetivo de modelar en la regresión, el tipo logístico se enfrenta al caso en el que una variable independiente no es suficiente para explicar dicho evento, por tal razón, se busca el uso del modelo de regresión logística múltiple, en el cual se relaciona la variable respuesta con k variables independientes, es decir, donde existen k coeficientes de regresión.

Este modelo, de igual forma que el modelo de regresión lineal múltiple, representa un hiperespacio, correspondiente a las k dimensiones de las variables regresoras x_j con $j \in [1, k]$.

AJUSTE DEL MODELO

Suponiendo que se tienen $n > k$ observaciones, y_i la i – ésima respuesta observada, x_{ij} la i – ésima observación y siguiendo el esquema de la regresión logística con una sola variable independiente, se forma para cada observación i con $i \in [1, n]$ las ecuaciones siguientes:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \dots + \beta_k x_{1k} + \varepsilon_1 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{n1} + \dots + \beta_k x_{nk} + \varepsilon_n \end{aligned}$$

Que de otra manera puede ser visto como:

$$y_i = \text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \sum_j^k \beta_j x_{ij} + \varepsilon_i$$

De donde se tiene que: $\pi_i = \frac{\exp[\beta_0 + \sum_j^k \beta_j x_{ij}]}{1 + \exp[\beta_0 + \sum_j^k \beta_j x_{ij}]}$

O bien:

$$\pi_i = \frac{1}{1 + \exp[-\beta_0 - \sum_j^k \beta_j x_{ij}]}$$

En esta opción de modelo, otra representación útil, es aquella en la que intervienen matrices.

$$\underline{Y} = \underline{\beta}\underline{X} + \underline{\varepsilon}$$

Donde:

- ★ \underline{Y} es un vector de $n \times 1$, correspondiente a los valores de la variable respuesta, los cuales estarán en el conjunto $\{0,1\}$.

$$\underline{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

- ★ \underline{X} es una matriz de $n \times (k + 1)$

$$\underline{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}$$

- ★ $\underline{\beta}$ es un vector de $k \times 1$, el cual contiene los coeficientes de regresión.

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$$

- ★ $\underline{\varepsilon}$ es el vector de $n \times 1$ de los errores asociados.

$$\underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

ESTIMACIÓN DE PARÁMETROS POR MÁXIMA VEROSIMILITUD

Considerando el modelo de regresión logística múltiple:

$$\text{logit}\left(\frac{\pi_i}{1 - \pi_i}\right) = \underline{\beta}\underline{X} + \underline{\varepsilon}$$

Y que la variable dependiente Y es binaria con posibles valores cero o uno, las observaciones en la muestra se tendrían en la clasificación: dentro del conjunto (A) de elementos cuyo valor asociado a la variable independiente es uno o dentro del conjunto complemento (B), es decir, donde $Y = 0$.

Suponiendo que la cardinalidad del conjunto A es d y por consiguiente para el conjunto B es $n - d$, por otro lado, sea $\hat{\pi}_i$ la probabilidad estimada por el modelo de que el i -ésimo valor de Y_i sea uno.

Se define como la función de verosimilitud como:

$$V = \left(\prod_{i=1}^d \pi_i \right) \left(\prod_{i=d+1}^{n-d} (1 - \pi_i) \right)$$

Donde los mejores valores de los parámetros serán aquellos que maximicen esta función.

Por otra parte, para simplificar, sea $\lambda_i = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$, con lo que la función de verosimilitud queda definida como:

$$V = \prod_{i=1}^n \lambda_i = \prod_{i=1}^n \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$$

De donde resulta que:

$$\begin{aligned} \prod_{i=1}^n \left(\frac{\pi_i}{1-\pi_i} \right)^{y_i} (1-\pi_i) &= \prod_{i=1}^n (1-\pi_i) \exp \left[y_i \left(\frac{\pi_i}{1-\pi_i} \right) \right] \\ &= \prod_{i=1}^n (1-\pi_i) \prod_{i=1}^n \exp \left[y_i \left(\frac{\pi_i}{1-\pi_i} \right) \right] \\ &= \left[\prod_{i=1}^n (1-\pi_i) \right] \exp \left[\sum_{i=1}^n y_i \ln \left(\frac{\pi_i}{1-\pi_i} \right) \right] \end{aligned}$$

Y que recordando la ecuación $\ln \left(\frac{\pi_i}{1-\pi_i} \right) = \beta_0 + \sum_j^k \beta_j x_{ij} + \varepsilon_i$, que muestra la definición del modelo de regresión logística múltiple; el último producto anterior queda definido como:

$$\left[\prod_{i=1}^n (1-\pi_i) \right] \exp \left[\sum_{i=1}^n y_i \sum_{j=0}^k \beta_j x_{ij} \right]$$

De igual manera, con la ecuación $\pi_i = \frac{1}{1+\exp[-\beta_0 - \sum_j^k \beta_j x_{ij}]}$, que muestra una expresión de π , dicho término queda como:

$$\left\{ \prod_{i=1}^n \left[1 + \exp \left(\sum_{j=0}^k \beta_j x_{ij} \right) \right]^{-1} \right\} \exp \left[\sum_{i=1}^n y_i \sum_{j=0}^k \beta_j x_{ij} \right]$$

Prosiguiendo con el método y con apoyo de la expresión de la función de verosimilitud antes encontrada, la aplicación del logaritmo en dicha función, se concreta como:

$$\begin{aligned}
 \ln(V) &= \sum_{i=1}^n y_i \sum_{j=0}^k \beta_j x_{ij} - \sum_{i=1}^n \ln \left[1 + \exp \left(\sum_{j=0}^k \beta_j x_{ij} \right) \right] \\
 &= \sum_{i=1}^n y_i \ln \left(\frac{\pi_i}{1 - \pi_i} \right) - \sum_{i=1}^n \ln \left[1 + \exp \left(\sum_{j=0}^k \beta_j x_{ij} \right) \right] \\
 &= \sum_{i=1}^n y_i \ln(\pi_i) - \sum_{i=1}^n y_i \ln(1 - \pi_i) - \sum_{i=1}^n \ln \left[1 + \left(\frac{\pi_i}{1 - \pi_i} \right) \right] \\
 &= \sum_{i=1}^n y_i \ln(\pi_i) - \sum_{i=1}^n y_i \ln(1 - \pi_i) - \sum_{i=1}^n \ln(1 - \pi_i)^{-1} \\
 &= \sum_{i=1}^n y_i \ln(\pi_i) + \sum_{i=1}^n \ln(1 - \pi_i) - \sum_{i=1}^n y_i \ln(1 - \pi_i) \\
 &= \sum_{i=1}^n y_i \ln(\pi_i) + \sum_{i=1}^n \{ \ln(1 - \pi_i) - y_i \ln(1 - \pi_i) \} \\
 &= \sum_{i=1}^n y_i \ln(\pi_i) - \sum_{i=1}^n (1 - y_i) \ln(1 - \pi_i)
 \end{aligned}$$

Por otro lado al aplicar la derivada y buscar igualarla a cero se tiene que:

$$\frac{\partial \ln(V)}{\partial \beta_j} = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \left[\frac{\exp(\sum_{j=0}^k \beta_j x_{ij})}{1 + \exp(\sum_{j=0}^k \beta_j x_{ij})} \right] = 0 \quad j = 0, \dots, k$$

Del resultado anterior se obtiene la matriz H en ocasiones conocida como la matriz de información, cuyas entradas son de la forma:

$$\frac{\partial^2 \ln(V)}{\partial \beta_j \partial \beta_l} = \sum_{i=1}^n \frac{x_{ij} x_{il} \exp(\sum_{j=0}^k \beta_j x_{ij})}{\left(1 + \exp(\sum_{j=0}^k \beta_j x_{ij}) \right)^2} = - \sum_{j=0}^k x_{ij} x_{il} \pi_i (1 - \pi_i) < 0 \quad j, l = 0, \dots, k$$

Los elementos anteriores constituyen una matriz de segundas derivadas o hessiana, la cual debe ser definida negativa y así maximizar la función de verosimilitud.

Una forma completa y opcional de ver esta matriz es:

$$-\left(\frac{\partial^2 \ln(V)}{\partial \underline{\beta} \underline{\beta}'}\right) = \sum_{i=1}^n \left(\frac{e^{\underline{X}_i \underline{\beta}}}{1 + e^{\underline{X}_i \underline{\beta}}}\right) \left(1 - \frac{e^{\underline{X}_i \underline{\beta}}}{1 + e^{\underline{X}_i \underline{\beta}}}\right) (x_i x_i')$$

A partir de lo anterior y por medio de métodos iterativos se resuelve el sistema y se obtienen los estimadores de β .

Para el caso de la obtención de la matriz de la covarianza se considera la inversa de la matriz de información, $(I(\hat{\underline{\beta}}))^{-1}$, siendo el resultado:

$$\widehat{Cov}(\hat{\underline{\beta}}) = \{\underline{X}'[\hat{\pi}(1 - \hat{\pi})]\underline{X}\}^{-1}$$

Y de la varianza:

$$\widehat{Var}(\hat{\underline{\beta}}) = \text{Diag}\left(I(\hat{\underline{\beta}})\right)^{-1} = \{\underline{X}'[\hat{\pi}(1 - \hat{\pi})]\underline{X}\}^{-1}$$

La raíz cuadrada de cada uno de estos últimos elementos corresponde a los valores de los errores estándar estimados del modelo.

Con base en el método arriba descrito, para una muestra de tamaño considerable, los valores estimados de β se comportan con una distribución normal estándar, mostrada por:

$$\frac{\hat{\underline{\beta}}}{\text{Diag}\left(I(\hat{\underline{\beta}})\right)^{-1}}$$

Esta última definición es utilizada en las pruebas individuales para los parámetros.

INTERPRETACIÓN DE PARÁMETROS

Por un lado, el coeficiente β_0 representa el logaritmo del cociente de momios cuando todas las variables independientes tienen el valor de cero.

En el caso de la regresión logística múltiple, al igual que en el caso simple, el parámetro β_i mide el incremento en una unidad existente en la variable independiente x_i ; sin embargo, en el modelo múltiple es de suma importancia la consideración de que existen otras variables independientes involucradas, y que para llevar a cabo la medición de dicho incremento, todas aquellas variables que en ese momento no son consideradas deben de controlarse.

Considerando los conjuntos $\{x_1^*, \dots, x_k^*\}$ y $\{x_1^{**}, \dots, x_k^{**}\}$, que evaluándolos en el modelo múltiple $\ln\left(\frac{\pi_i}{1-\pi_i}\right) = y_i = \beta_0 + \sum_j^k \beta_j x_{ij} + \varepsilon_i$ y recordando la definición de momio, se pueden obtener los correspondientes $O(X^*)$ y $O(X^{**})$ que al relacionarlos por medio del cociente de momios se tiene lo siguiente:

$$\frac{O(X^*)}{O(X^{**})} = \exp\left[\sum_j^k \beta_j (X_j^* - X_j^{**})\right]$$

Expresión que al considerar que los valores de las variables fuesen iguales $x_1^* = x_1^{**}, \dots, x_{i-1}^* = x_{i-1}^{**}, x_{i+1}^* = x_{i+1}^{**}, \dots, x_k^* = x_k^{**}$ excepto uno, suponiendo esa variable $x_i^* \neq x_i^{**}$, quedaría como:

$$\frac{O(X^*)}{O(X^{**})} = \exp[\beta_i (X_i^* - X_i^{**})]$$

Adicionalmente considerando que la diferencia entre ambas, sea tan sólo de una unidad $X_i^* = X_i^{**} + 1$, la expresión queda como:

$$\frac{O(X^*)}{O(X^{**})} = \exp(\beta_i)$$

De modo tal, que mientras las otras variables sean iguales o estén controladas, si el valor de la variable X cambia de x_0 a $x_0 + C$, la variación existente será de $C\beta_i$.

VALORACIÓN DEL MODELO ESTIMADO POR MEDIO DE MEDIDAS GLOBALES DE BONDAD DE AJUSTE

En el modelo de regresión logística ya se vieron algunas pruebas de significancia, pero dada la adición de variables independientes al modelo, algunas varían y otras cambian, por lo que a continuación se presentan las más conocidas para este modelo en especial.

DEVIANZA

Como antes se señaló, derivado de la función de verosimilitud, se define el estadístico $D = -2\ln L$, el cual evalúa que tan drásticos son los cambios que se producen al agregar o quitar variables al modelo.

Cabe destacar que la característica del estadístico anterior en relación a la función de verosimilitud es; que al dicha función adquirir el valor de uno, la devianza tendrá por valor un número muy cercano a cero o en un caso especial cero.

Considerando el modelo logístico múltiple con n parámetros y probabilidades π_i ; mismo que alcanza su punto máximo para la función de verosimilitud, dado que en ocasiones es denominado modelo saturado, puesto que, incluye todas las variables

relativas al evento observado, es de utilidad contemplar que el valor máximo es la unidad.

Por otro lado, en el modelo logístico ajustado o aquel en el que están consideradas un cierto número de variables, con parámetros estimados $\hat{\beta}$, al aplicarle la función logaritmo de verosimilitud, se observa que alcanza su valor máximo en:

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

La devianza del modelo que realiza la comparación entre los logaritmos de verosimilitud de los modelos saturado y ajustado, queda definida como:

$$D = -2 \sum \left\{ y_i \ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1 - y_i) \ln\left(\frac{1 - \hat{\pi}_i}{1 - y_i}\right) \right\}$$

De donde, con la correcta función de regresión logística y una muestra de tamaño n lo suficientemente grande, la devianza del modelo tendría una distribución *ji-cuadrada* con $n - p$ grados de libertad y determinaría que:

- ★ Si $\lambda(\beta) \leq \chi_{\alpha, n-p}^2$, el modelo ajustado puede ser considerado como adecuado. Es decir, aún con menor cantidad de parámetros que un modelo saturado, se ajusta casi igual que éste.
- ★ En otro caso, si $\lambda(\beta) > \chi_{\alpha, n-p}^2$, el modelo ajustado no puede ser considerado como adecuado.

ESTADISTICO LOG-VEROSIMILITUD O RAZÓN DE VEROSIMILITUD

Este método se basa en la comparación de la devianza de dos modelos, con una distribución *ji-cuadrada*.

PRUEBA DE HIPÓTESIS PARA SUBCONJUNTOS DE VARIABLES

Basándose en el modelo $\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = y_i = \beta_0 + \sum_j^k \beta_j x_{ij} + \varepsilon_i$ y separando las variables independientes en dos conjuntos y de igual manera sus respectivos parámetros.

Sea el anterior modelo logístico escrito de la forma:

$$\eta = x\beta = x_1\beta_1 + x_2\beta_2$$

Con:

β_1 : vector de $p - r$ parámetros

β_2 : vector de r parámetros

x_i : variables asociadas a los vectores de parámetros β_1 y β_2

Considerando:

- ★ Hipótesis: $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$
- ★ Suponiendo cierta la hipótesis H_0 , el modelo reducido es: $\eta_i = x_1'\beta_1$.

Sea $\lambda(\beta_1)$ devianza para el modelo reducido, lo que implica que la diferencia en la devianza es:

$$\lambda(\beta_2|\beta_1) = \lambda(\beta_1) - \lambda(\beta)$$

La diferencia anterior, en algunas ocasiones recibe el nombre de “desviación parcial”, que es una prueba del cociente de verosimilitudes.

El cociente anterior, puede visualizarse como $\frac{L(\widehat{\beta}_1)}{L(\widehat{\beta})}$, donde: $L(\widehat{\beta})$ y $L(\widehat{\beta}_1)$ son los valores máximos de las funciones de verosimilitud de los modelos completo y reducido respectivamente.

Por otra parte, el estadístico para la prueba del cociente de verosimilitud es:

$$\chi^2 = -2\ln\left(\frac{L(\widehat{\beta}_1)}{L(\widehat{\beta})}\right)$$

Que concuerda con la diferencia de la devianza, es decir:

$$\begin{aligned}\chi^2 &= -2\ln\left(\frac{L(\widehat{\beta}_1)}{L(\widehat{\beta})}\right) \\ &= -2\ln(L(\widehat{\beta}_1)) + 2\ln(L(\widehat{\beta})) \\ &= -2\ln(L(\widehat{\beta}_1)) + 2\ln(L(\widehat{\beta})) + 2\ln(L(ms)) - 2\ln(L(ms)) \\ &= \lambda(\widehat{\beta}_1) - \lambda(\widehat{\beta})\end{aligned}$$

Donde “ms” denota el modelo saturado.

Por lo que, suponiendo cierta la hipótesis $H_0: \beta_2 = 0$ y una muestra grande, la diferencia se distribuye como una $\chi^2_{n-(p-r)-(n-p)} = \chi^2_r$.

Lo que implica, que los criterios de decisión quedan determinados como:

- ★ Si $\lambda(\beta_2|\beta_1) \geq \chi^2_{a,r}$, se rechaza H_0 .

En este criterio se muestra la presencia de por lo menos un parámetro distinto de cero en β_2 .

- ★ Si $\lambda(\beta_2|\beta_1) < \chi^2_{a,r}$, no se rechaza H_0 .

Lo que simboliza que el modelo reducido tiene un ajuste similar al modelo completo.

PSEUDO-ESTADÍSTICO R^2

En la búsqueda por encontrar la mejor manera de evaluar los modelos logísticos, se han construido diversos estadísticos análogos a la función de R^2 en la regresión lineal.

Estos pseudo-estadísticos a menudo son rechazados, sin embargo, algunos textos los confrontan con los métodos reconocidos y los apoyan cuando se trata de elegir entre dos modelos ajustados.

Entre las sugerencias de dichos estadísticos; existe uno y el más conocido, que indica que proporción de la varianza de la variable dependiente es explicada por el modelo; ésta aproximación, se basa en la comparación de la verosimilitud del modelo nulo con la verosimilitud del modelo saturado, denotado por la ecuación:

$$Pseudo - R^2 = 1 - \left[\frac{\ln L_1}{\ln L_0} \right]$$

CRITERIO DE INFORMACIÓN DE AKAIKE

Otro de los métodos desarrollados dentro del análisis de la regresión logística, se encuentra el propuesto por Hirotugu Akaike en 1974, el cual se basa en un concepto de entropía, llevado a la búsqueda de un equilibrio entre la precisión del modelo ajustado y la construcción del mismo.

Este criterio no es considerado como prueba de significancia, pero, de igual manera que el *pseudo – estadístico* R^2 puede ser utilizado al momento de elegir modelos.

La evaluación se realiza por medio de la ecuación:

$$AIC = D + 2(p + 1)$$

Donde: D es la devianza, p es el número de variables predictoras y se elige el modelo cuyo valor de AIC sea el más pequeño.

VALORACIÓN DEL MODELO ESTIMADO POR MEDIO DE LA PRUEBA HOSMER-LEMESHOW

Similar al aprendizaje natural de un niño, surgen ideas de la manera en la cual se pueden comparar y evaluar los modelos ajustados y los datos reales, es sencillo pensar que se pueden ir revisando pareja por pareja, sin embargo, al tener una muestra de tamaño significativo, dicha idea resulta un tanto complicada.

De esta forma, partiendo de ideas sencillas es que la prueba de Hosmer-Lemeshow explica una técnica para la evaluación de los modelos de regresión.

La secuencia de las ideas en las que se postula este método son:

- ★ Calcular las n probabilidades relativas al modelo ajustado.
- ★ Ordenar dichos n valores de manera ascendente.
- ★ Realizar una separación en varios conjuntos, teniendo distintas posibilidades para dichas clasificaciones, por ejemplo:
 - ☆ Conjuntos de un número determinado de elementos, para la investigación realizada por Hosmer y Lemeshow, los grupos fueron considerados de 10 elementos cada uno.
 - ☆ Otra opción es de acuerdo a los valores de las probabilidades, es decir, juntar aquellas probabilidades que estén dentro de determinados intervalos.

Considérense n_i ; $i = 1, \dots, k$ las frecuencias respectivas.

- ★ Sumar los valores de las probabilidades dentro de cada grupo, para así constituir los valores esperados: $E_i; i = 1, \dots, k$.
- ★ Por último se cuenta el número de valores observados, denominados así los elementos para los cuales $Y = 1$ y sean: $O_i; i = 1, \dots, k$.

Con los pasos anteriores, es constituido el “estadígrafo de Hosmer y Lemeshow” como:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} + \sum_{i=1}^k \frac{(O_i^* - E_i^*)^2}{E_i^*}$$

Donde:

- ★ $E_i^* = n_i - E_i$
- ★ $O_i^* = n_i - O_i$

Con una distribución *ji – cuadrada* de $k - 2$ grados de libertad. Es conveniente acentuar que una de las restricciones de este modelo, es que los valores $E_i; i = 1, \dots, k$ no deben de ser cercanos a cero o nulos.

INTERVALOS DE CONFIANZA

PARA β_1

De igual forma que para el modelo simple, el intervalo de confianza de un coeficiente es conformado por la estimación del parámetro más o menos el punto porcentual de una distribución normal y multiplicado por la estimación del error estándar.

Donde por ejemplo, si el modelo es conformado por $Y_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$, el intervalo para β_3 sería:

$$\beta_3 \pm t_{(n-2)}^{1-\alpha/2} \hat{E}E(\widehat{\beta}_3)$$

Intervalo que busca el 95% de confianza con un $\alpha = 5\%$ por lo que $t_{(n-2)}^{1-\alpha/2} = 1.96$ es el valor obtenido de tablas.

2.2.4. EJEMPLO DEL MODELO DE REGRESION LOGÍSTICA CON VARIABLE DE RESPUESTA BINARIA

Con base en los conceptos arriba citados, es posible concretar un ejemplo que aplique el modelo de regresión logística con más de una variable independiente.

Considerando un estudio¹⁰ realizado en Estados Unidos, acerca de las condenas por el delito de asesinato, se constituye la hipótesis de que dependiendo del color de la piel de la persona se dicta la condena de muerte. Los datos de dicha muestra pueden ser vistos de la manera siguiente:

		Raza del acusado		Raza de la Víctima	
		Blanco	Negro	Negro	Blanco
Pena de Muerte	No	432	178	157	453
	Sí	54	15	4	65

Raza de la Víctima		Blanco		Negro	
Raza del acusado		Blanco	Negro	Blanco	Negro
Pena de Muerte	No	415	38	17	140
	Sí	54	11	0	4

¹⁰ <http://www.bioestadistica.uma.es/baron>

Para este análisis, son consideradas como variables:

★ **Variable dependiente:** la variable “*pena de muerte*”

☆ Que obtiene el valor codificado de 0 en el caso de que el individuo no tenga condena y de 1 en el caso contrario.

★ **Variables independientes:**

☆ la variable “*raza del acusado*” (asignando el valor 0 al color blanco en la piel y el valor 1 a la piel oscura) y

☆ la variable “*raza de la víctima*” (asignando el valor 1 al color blanco en la piel y el valor 0 a la piel oscura)

En este caso el ejemplo determina un modelo constituido por:

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95.0% para EXP(B)	
								Inferior	Superior
Paso 1(a)	rzAcusado	.827	.365	5.129	1	.024	2.286	1.118	4.677
	rzVictima	2.391	.599	15.937	1	.000	10.928	3.378	35.355
	Constante	-4.435	.614	52.198	1	.000	.012		

a. Variable(s) introducida(s) en el paso 1: rzAcusado, rzVictima.

Donde la columna denominada *B*, muestra los valores estimados de los coeficientes de regresión. De igual manera en la tabla anterior, se puede observar que de acuerdo a los cocientes de momios ($exp(B)$) son significativamente mayores que uno para un nivel de significancia de 0.05 y los intervalos de confianza no contienen a la unidad.

Con lo anterior, la hipótesis de que el color de la piel de los acusados y víctimas influyen en la condena resultaría afirmativa, siempre que se siga considerando la opción de que en un determinado caso, pueden existir otras variables que aporten mayor efecto sobre la condena.

Lo anterior nos lleva a las conclusiones siguientes:

- ★ Si la condena dependiera del color de piel del acusado, es 2.286 veces más probable que al acusado le dicten sentencia de muerte si es negro.
- ★ Por otro lado, si la raza de la víctima define el castigo y ésta fuese blanca, es 10.928 veces más probable que la pena de muerte que si fuese negra.

2.2.5. MODELO POLITÓMICO DE REGRESIÓN LOGÍSTICA

En el proceso de dar explicación a un evento, se encuentran condiciones, las cuales son relacionadas con la variable independiente, sin embargo, dicha variable a diferencia del caso binario tiene la posibilidad de elegir sus valores no sólo en un conjunto de dos elementos sino de más.

MODELO POLITÓMICO

Para este tipo de regresión, donde a la variable dependiente no le es asignado un valor de un conjunto con más de dos elementos, el procedimiento a seguir es la consideración de una categoría de referencia enfrentada contra las otras.

Considerando el conjunto $C = \{1, 2, \dots, r\}$ correspondiente a las categorías o valores asignados a la variable dependiente Y ; sea $w \in C$ el elemento que será nombrado como categoría de referencia, para el cual existirá $r - 1$ categorías restantes con las cuales se contrastará, es decir:

$$\frac{P(Y = s)}{P(Y = w)}; s \in C - \{w\}$$

Suponiendo que para cada Y existen k variables independientes X_1, \dots, X_k , la probabilidad de que Y asuma el valor w es descrita en función de las $r - 1$ variables independientes, de la manera siguiente:

$$P(Y = s) = \frac{\exp[\beta_0 + \sum_j^k \beta_j x_{ij}]}{1 + \exp[\beta_0 + \sum_j^k \beta_j x_{ij}]}$$

que visto de otra forma:

$$\ln \frac{P(Y = s)}{P(Y = w)} = \beta_0 + \sum_j^k \beta_j x_{ij}$$

donde para cada $s \in C$, se tiene el conjunto de parámetros $B_s = \{\beta_{0s}, \beta_{1s}, \dots, \beta_{ks}\}$ relacionados a dicha categoría.

Por otra parte, dado que las probabilidades involucradas son relacionadas con los posibles valores de la variable independiente, se tiene la condición de que la suma de todas ellas es uno.

Con respecto al punto anterior, si la variable independiente tiene en consideración tres categorías, se tendría que:

$$P_1 + P_2 + P_3 = 1$$

Con lo cual, el logaritmo de las probabilidades define sólo un análisis entre dos de las probabilidades asociadas.

Haciendo una especificación adicional, asociando a la variable independiente los valores de X_1^* y X_1^{**} , la ecuación para el cociente de momios en este tipo de regresión queda determinado por:

$$CM_r = \exp[\beta_{r1}(X_1^{**} - X_1^*)],$$

con r como la categoría de la variable Y .

INTERVALO DE CONFIANZA

De forma análoga al modelo de regresión logística estándar, el intervalo de confianza asociado con esta regresión, está definido como:

$$\exp[\hat{\beta}_{r1}(X_1^{**} - X_1^*) \pm 1.96(X_1^{**} - X_1^*)ASE_{\hat{\beta}_{r1}}]$$

3. APLICACIÓN DE UN MODELO DE REGRESIÓN LOGÍSTICA EN AMÉRICA LATINA A LOS PROCESOS DE GLOBALIZACIÓN

En la presente sección, se tiene como objetivo la aplicación de la teoría antes vista en el análisis de la conformación de los Regímenes del Bienestar en América Latina.

Dentro de este estudio, se tienen los considerados por la doctora en sociología Juliana Martínez Franzoni, los cuales son denotados por:

1. Estado del Bienestar estatal-proteccionista
2. Estado del Bienestar productivista-informal
3. Estado del Bienestar familiarista
4. Estado del Bienestar altamente familiarista

Asociados a estos estados, se encuentran los siguientes países: Argentina y Chile para el primer estado; Brasil, Costa Rica, México, Panamá y Uruguay para el segundo; Colombia, Ecuador, El Salvador, Guatemala, Perú, República Dominicana y Venezuela para el tercero y por último, Bolivia, Honduras, Nicaragua y Paraguay para el cuarto.

De los cuales, en este trabajo se consideran únicamente catorce países: Argentina, Bolivia, Brasil, Chile, Costa Rica, Ecuador, El Salvador, Guatemala, México, Nicaragua, Paraguay, Perú, Uruguay y Venezuela.

Asimismo, para la realización del análisis, se tendrá la base de datos elaborada con los indicadores siguientes:

1. Tasa de variación del PIB.
2. Estructura de la población ocupada total por sector de actividad económica (Agricultura, Industria y Servicios)
3. Salario mínimo real.
4. Tasa de participación en la actividad económica.

5. Tasa media anual de desempleo
6. Gasto público social en salud como porcentaje del PIB.
7. Gasto público social como porcentaje del PIB.
8. Tasa de mortalidad infantil.
9. Índice de desarrollo humano.
10. Esperanza de vida.

Cabe señalar, que se considera el periodo comprendido entre el año 2000 y 2005; y que en los casos en los que los indicadores carecían de información (Estructura de la población ocupada total por sector de actividad económica (Agricultura, Industria y Servicios), Tasa de participación en la actividad económica, Tasa media anual de desempleo, Índice de desarrollo humano, Esperanza de vida), se construyeron modelos de regresión lineal para aproximar las referencias faltantes.

Para conceptualizar dicho procedimiento, se puede ver el ejemplo elaborado con los datos encontrados del indicador: “Esperanza de vida”.

Inicialmente, el informe presentado por el Banco Interamericano de Desarrollo, se constituía por:

	2000	2001	2002	2003	2004	2005
Argentina	74		74	74	75	75
Bolivia	63		64	64	65	65
Brasil	70		70	71	71	71
Chile	77		78	78	78	78
Costa Rica	78		79	79	79	79
Ecuador	73		74	74	75	75
El Salvador	70		71	71	71	71
Guatemala	66		67	67	68	68
México	74		75	75	75	75
Nicaragua	69		69	70	70	70
Paraguay	70		71	71	71	71
Perú	69		70	70	70	71

Uruguay	75	75	75	75	76
Venezuela	73	74	74	74	74

Donde el año 2001 está carente de la información solicitada; por lo cual, por medio del software SPSS, se obtuvieron los modelos de regresión lineal correspondientes a cada uno de los países, por ejemplo para el caso de Argentina, el modelo fue:

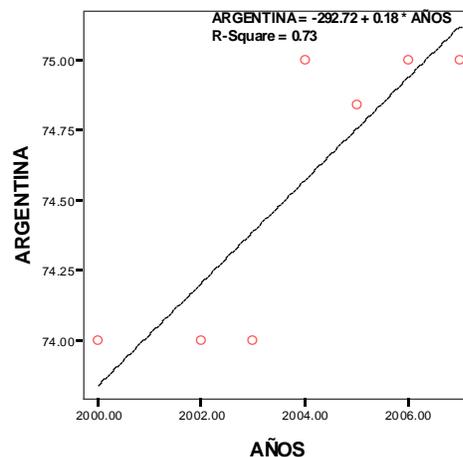
$$Y = -292.716 + 0.183x$$

Donde:

- ★ la variable dependiente “Y” corresponde a la esperanza de vida
- ★ y la variable independiente “x” a los años.

Por otro lado, al realizar un análisis gráfico exploratorio, es posible apreciar que al ser una cantidad no muy significativa de puntos, los datos no se encuentran tan dispersos.

Otro aspecto importante por considerar es el valor R^2 arrojado por el programa, que para el ejemplo de Argentina es: $R^2 = 0.732$, valor que describe como el modelo explica el 73.2% de los datos examinados; lo cual refleja que el valor pronosticado para el año faltante podrá ser considerado como aceptable.



De esta manera, al realizar el análisis de regresión lineal correspondiente a los países restantes, se tiene que la tabla completa que contiene los datos de la esperanza de vida es:

	2000	2001	2002	2003	2004	2005
Argentina	74	74	74	74	75	75
Bolivia	63	63	64	64	65	65
Brasil	70	70	70	71	71	71
Chile	77	78	78	78	78	78
Costa Rica	78	78	79	79	79	79
Ecuador	73	74	74	74	75	75
El Salvador	70	71	71	71	71	71
Guatemala	66	66	67	67	68	68
México	74	74	75	75	75	75
Nicaragua	69	69	69	70	70	70
Paraguay	70	70	71	71	71	71
Perú	69	69	70	70	70	71
Uruguay	75	75	75	75	75	76
Venezuela	73	73	74	74	74	74

Por otro lado, buscando cumplir el objetivo de la presente sección, el cual se centra en encontrar cuál o cuáles Estados del Bienestar en América Latina son los dominantes, lo anterior, de acuerdo a algunas características demográficas, económicas o relacionadas con la Salud.

Para tal efecto, se consideran los regímenes antes citados, codificados como:

- 1 Estado del Bienestar productivista-informal (categoría que comprende a Brasil, Costa Rica, México y Uruguay).
- 2 Estado del Bienestar estatal-proteccionista (categoría que contiene a Argentina y Chile).
- 3 Estado del Bienestar familiarista (categoría que engloba a Ecuador, El Salvador, Guatemala, Perú y Venezuela)
- 4 Estado del Bienestar altamente familiarista (categoría que incluye a Bolivia, Nicaragua y Paraguay)

Siendo la escala nominal, la correspondiente a esta variable, la cual de ahora en adelante, se denominará “Régimen” y será considerada como variable dependiente.

A manera de evaluar la información obtenida para los países, se clasifican las variables por medio de sus características en demográficas, económicas y relacionadas con la salud, quedando de la siguiente manera:

★ Demográficas:

- ☆ TMI: tasa de mortalidad infantil
- ☆ IDI: índice de desarrollo infantil
- ☆ EV: esperanza de vida

★ Económicas:

- ☆ PIB: tasa de variación del PIB
- ☆ Agricultura: estructura de la población dedicada a la actividad agrícola.
- ☆ Industria: estructura de la población dedicada a la actividad industrial.
- ☆ Servicio: estructura de la población dedicada a la actividad de servicios.
- ☆ Salario: Salario mínimo real.
- ☆ PAE: tasa de participación de la población en la actividad económica
- ☆ Desempleo: tasa media anual de desempleo.

★ De Salud:

- ☆ GPSS_PIB: gasto público social en salud como porcentaje del PIB.
- ☆ GPS_PIB: gasto público social como porcentaje del PIB.

Una vez organizada la información contenida en las variables, los procesos aplicados de aquí en adelante, se realizaron en el software Stata.

Con base en la clasificación anterior y realizando un primer análisis a cada grupo de variables (Demográficas, Económicas y de Salud), se obtuvieron las correlaciones internas. En relación a dichos resultados, se corrieron las regresiones logísticas con las variables que tenían un aporte significativo para el modelo.

Las tablas siguientes muestran los resultados arrojados por Stata; en el Anexo A, se presenta una breve explicación de los datos contenidos en dicho reporte.

★ Para las variables demográficas:

AÑO		MODELO					
2000	Multinomial logistic regression						Number of obs = 14
	Log likelihood = -11.27476						LR chi2(6) = 14.80
							Prob > chi2 = 0.0219
							Pseudo R2 = 0.3962
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	TMI	-.1641438	.3067716	-0.54	0.593	-.7654052	.4371175
	EV	.1960008	.81066	0.24	0.809	-1.392863	1.784865
	_cons	-10.12742	65.60528	-0.15	0.877	-138.7114	118.4566
	2						
TMI	-1.299139	1.122801	-1.16	0.247	-3.499787	.90151	
EV	-2.269118	2.778269	-0.82	0.414	-7.714425	3.176189	
_cons	193.0746	226.1024	0.85	0.393	-250.0779	636.2271	
4							
TMI	.0946653	.1710912	0.55	0.580	-.2406672	.4299979	
EV	.0102072	.6144313	0.02	0.987	-1.194056	1.21447	
_cons	-4.646551	47.69134	-0.10	0.922	-98.11986	88.82676	
(RÉGIMEN==3 is the base outcome)							
2001	Multinomial logistic regression						Number of obs = 14
	Log likelihood = -11.981229						LR chi2(6) = 13.38
							Prob > chi2 = 0.0374
							Pseudo R2 = 0.3583
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	TMI	-.3019567	.3007234	-1.00	0.315	-.8913637	.2874503
	EV	-.1269116	.7173934	-0.18	0.860	-1.532977	1.279154
	_cons	16.38325	58.24571	0.28	0.778	-97.77624	130.5428
	2						
TMI	-.7632674	.6996518	-1.09	0.275	-2.13456	.6080248	
EV	-.8323563	1.533193	-0.54	0.587	-3.837359	2.172647	
_cons	76.25519	125.8703	0.61	0.545	-170.4461	322.9565	
4							
TMI	.0611861	.1614946	0.38	0.705	-.2553374	.3777097	
EV	-.13431	.5564642	-0.24	0.809	-1.22496	.9563398	
_cons	6.646569	43.25501	0.15	0.878	-78.1317	91.42484	
(RÉGIMEN==3 is the base outcome)							

REGRESIÓN LOGÍSTICA APLICADA EN LA CONFIGURACIÓN DE ESTADOS EN
ALGUNOS PAÍSES DE AMÉRICA LATINA

Diana Citlalli Márquez Palacios

78

2002	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -11.048717				LR chi2(6)	=	15.25
					Prob > chi2	=	0.0184
					Pseudo R2	=	0.4083
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	TMI	-.3844647	.3496906	-1.10	0.272	-1.069846	.3009163
	EV	-.3090281	.7205701	-0.43	0.668	-1.721319	1.103263
	_cons	31.36697	59.85467	0.52	0.600	-85.94603	148.68
	2						
TMI	-1.284246	.8807535	-1.46	0.145	-3.010491	.4419991	
EV	-1.834246	1.835297	-1.00	0.318	-5.431361	1.762869	
_cons	159.8814	151.261	1.06	0.291	-136.5847	456.3475	
4							
TMI	.0898506	.18728	0.48	0.631	-.2772115	.4569127	
EV	-.0829282	.5987552	-0.14	0.890	-1.256467	1.09061	
_cons	2.286293	47.00816	0.05	0.961	-89.84801	94.4206	
(RÉGIMEN==3 is the base outcome)							
2003	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -11.461055				LR chi2(6)	=	14.42
					Prob > chi2	=	0.0253
					Pseudo R2	=	0.3862
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	TMI	-.3622876	.3728315	-0.97	0.331	-1.093024	.3684487
	EV	-.0840074	.7811784	-0.11	0.914	-1.615089	1.447074
	_cons	14.0434	64.03975	0.22	0.826	-111.4722	139.559
	2						
TMI	-.9391497	.6680819	-1.41	0.160	-2.248566	.3702668	
EV	-1.001582	1.322953	-0.76	0.449	-3.594521	1.591357	
_cons	91.4989	109.1747	0.84	0.402	-122.4796	305.4774	
4							
TMI	.1475287	.2265517	0.65	0.515	-.2965044	.5915618	
EV	.1145493	.6404284	0.18	0.858	-1.140667	1.369766	
_cons	-13.19262	51.02555	-0.26	0.796	-113.2009	86.81562	
(RÉGIMEN==3 is the base outcome)							
2004	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -11.528711				LR chi2(6)	=	14.29
					Prob > chi2	=	0.0266
					Pseudo R2	=	0.3826
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	TMI	-.3780046	.3418259	-1.11	0.269	-1.047971	.2919618
	EV	-.1054856	.6739146	-0.16	0.876	-1.426334	1.215363
	_cons	15.62048	55.26661	0.28	0.777	-92.70008	123.941
	2						
TMI	-.7736725	.580461	-1.33	0.183	-1.911355	.3640101	
EV	-.5498085	1.103892	-0.50	0.618	-2.713397	1.61378	
_cons	54.39948	91.09052	0.60	0.550	-124.1347	232.9336	
4							
TMI	.1187148	.2153066	0.55	0.581	-.3032784	.540708	
EV	-.060787	.6166342	-0.10	0.921	-1.269368	1.147794	
_cons	.1964344	48.60278	0.00	0.997	-95.06326	95.45613	
(RÉGIMEN==3 is the base outcome)							

RÉGIMEN		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1	TMI	-.3945756	.3506985	-1.13	0.261	-1.081932	.2927809
	EV	-.0421344	.6233638	-0.07	0.946	-1.263905	1.179636
	_cons	10.94585	51.12696	0.21	0.830	-89.26115	111.1528
2	TMI	-.8211776	.6166412	-1.33	0.183	-2.029772	.3874169
	EV	-.5546397	1.059245	-0.52	0.601	-2.630721	1.521442
	_cons	55.26856	88.05941	0.63	0.530	-117.3247	227.8618
4	TMI	.0821561	.2072081	0.40	0.692	-.3239643	.4882764
	EV	-.2057631	.6553237	-0.31	0.754	-1.490174	1.078648
	_cons	11.5903	51.03372	0.23	0.820	-88.43395	111.6145

(RÉGIMEN==3 is the base outcome)

2005

★ Para las variables económicas:

RÉGIMEN		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1	PIB	-.1530242	.3866084	-0.40	0.692	-.9107627	.6047142
	PAE	.0252443	.1646845	0.15	0.878	-.2975313	.34802
	DESEMPLEO	-.1319468	.2636132	-0.50	0.617	-.6486192	.3847257
	_cons	-.330704	10.81996	-0.03	0.976	-21.53743	20.87603
2	PIB	-2.260915	4.367014	-0.52	0.605	-10.82011	6.298275
	PAE	-2.473951	5.169129	-0.48	0.632	-12.60526	7.657356
	DESEMPLEO	-.5339938	1.475651	-0.36	0.717	-3.426217	2.358229
	_cons	153.1676	321.9148	0.48	0.634	-477.7739	784.1091
4	PIB	-.5798414	.5252713	-1.10	0.270	-1.609354	.4496714
	PAE	.3110925	.2369802	1.31	0.189	-.15338	.7755651
	DESEMPLEO	-.2980792	.4619904	-0.65	0.519	-1.203564	.6074053
	_cons	-16.93535	14.86433	-1.14	0.255	-46.0689	12.19821

(RÉGIMEN==3 is the base outcome)

2000

2001	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -9.9559316				LR chi2(9)	=	17.43
					Prob > chi2	=	0.0424
					Pseudo R2	=	0.4668
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	PIB	-11.39671	8.712997	-1.31	0.191	-28.47387	5.68045
	SALARIO	.2306426	.4580211	0.50	0.615	-.6670622	1.128347
	DESEMPLEO	-4.493756	3.988462	-1.13	0.260	-12.311	3.323486
	_cons	18.51542	52.95887	0.35	0.727	-85.28206	122.3129
2							
PIB	-1.246932	.894632	-1.39	0.163	-3.000378	.5065145	
SALARIO	.3673357	.3613766	1.02	0.309	-.3409494	1.075621	
DESEMPLEO	1.220519	.7802735	1.56	0.118	-.3087894	2.749826	
_cons	-49.08967	42.22953	-1.16	0.245	-131.858	33.67869	
4							
PIB	-1.057937	.8458089	-1.25	0.211	-2.715693	.5998176	
SALARIO	.3388701	.3058041	1.11	0.268	-.260495	.9382352	
DESEMPLEO	.8140411	.6859204	1.19	0.235	-.5303382	2.15842	
_cons	-41.31311	35.89751	-1.15	0.250	-111.6709	29.04471	
(RÉGIMEN==3 is the base outcome)							
2002	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -13.331184				LR chi2(9)	=	10.68
					Prob > chi2	=	0.2981
					Pseudo R2	=	0.2860
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	SALARIO	-.0266464	.1015317	-0.26	0.793	-.2256448	.172352
	PAE	.0691757	.2190478	0.32	0.752	-.36015	.4985014
	DESEMPLEO	.0006856	.1913478	0.00	0.997	-.3743491	.3757203
	_cons	-1.813693	13.17031	-0.14	0.890	-27.62703	23.99964
2							
SALARIO	-.0836874	.3231875	-0.26	0.796	-.7171232	.5497484	
PAE	-.8880948	1.016727	-0.87	0.382	-2.880842	1.104653	
DESEMPLEO	.2043862	.480007	0.43	0.670	-.7364102	1.145183	
_cons	57.27243	88.6809	0.65	0.518	-116.539	231.0838	
4							
SALARIO	.1063768	.1719294	0.62	0.536	-.2305986	.4433522	
PAE	.2488574	.3339061	0.75	0.456	-.4055866	.9033014	
DESEMPLEO	.3506793	.407375	0.86	0.389	-.447761	1.14912	
_cons	-31.54855	21.40417	-1.47	0.140	-73.49995	10.40284	
(RÉGIMEN==3 is the base outcome)							

REGRESIÓN LOGÍSTICA APLICADA EN LA CONFIGURACIÓN DE ESTADOS EN
ALGUNOS PAÍSES DE AMÉRICA LATINA

Diana Citlalli Márquez Palacios

81

2003	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -14.295541				LR chi2(9)	=	8.75
					Prob > chi2	=	0.4603
					Pseudo R2	=	0.2344
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	PIB	.3014299	.2922861	1.03	0.302	-.2714403	.8743001
	SALARIO	-.065805	.0750276	-0.88	0.380	-.2128564	.0812463
	DESEMPLEO	.0277356	.2271537	0.12	0.903	-.4174774	.4729487
	_cons	5.62637	8.670922	0.65	0.516	-11.36832	22.62106
2							
PIB	1.149303	.822463	1.40	0.162	-.4626953	2.7613	
SALARIO	.0469236	.136108	0.34	0.730	-.2198433	.3136905	
DESEMPLEO	.3842782	.4409702	0.87	0.384	-.4800075	1.248564	
_cons	-14.06392	18.5337	-0.76	0.448	-50.38931	22.26147	
4							
PIB	.3672985	.4910325	0.75	0.454	-.5951075	1.329704	
SALARIO	.0479371	.0859619	0.56	0.577	-.1205452	.2164193	
DESEMPLEO	.1663737	.2865775	0.58	0.562	-.3953079	.7280553	
_cons	-8.165766	10.64329	-0.77	0.443	-29.02622	12.69469	
(RÉGIMEN==3 is the base outcome)							
2004	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -13.429427				LR chi2(9)	=	10.49
					Prob > chi2	=	0.3126
					Pseudo R2	=	0.2808
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	SERVICIO	.0186259	.0894213	0.21	0.835	-.1566367	.1938885
	PAE	.0466257	.2136009	0.22	0.827	-.3720243	.4652757
	DESEMPLEO	-.0091589	.2338437	-0.04	0.969	-.4674841	.4491662
	_cons	-4.065241	15.02793	-0.27	0.787	-33.51944	25.38896
2							
SERVICIO	.7211568	2.483836	0.29	0.772	-4.147073	5.589386	
PAE	-.7338826	2.118635	-0.35	0.729	-4.886331	3.418566	
DESEMPLEO	-.4652658	2.260411	-0.21	0.837	-4.895589	3.965058	
_cons	2.04666	40.27039	0.05	0.959	-76.88185	80.97517	
4							
SERVICIO	-.1624163	.1996834	-0.81	0.416	-.5537885	.228956	
PAE	.5492693	.4431719	1.24	0.215	-.3193316	1.41787	
DESEMPLEO	-.2194015	.4193629	-0.52	0.601	-1.041338	.6025347	
_cons	-25.826	22.13874	-1.17	0.243	-69.21713	17.56513	
(RÉGIMEN==3 is the base outcome)							

		Multinomial logistic regression				Number of obs = 14		
		Log likelihood = -15.13743				LR chi2(9) = 7.07		
						Prob > chi2 = 0.6299		
						Pseudo R2 = 0.1893		
RÉGIMEN		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
2005	1							
		PIB	-.4892435	.5895007	-0.83	0.407	-1.644644	.6661567
		SALARIO	.0336117	.0616045	0.55	0.585	-.087131	.1543544
		DESEMPLEO	.1949966	.5670538	0.34	0.731	-.9164085	1.306402
		_cons	-3.058299	6.035103	-0.51	0.612	-14.88688	8.770285
		2						
		PIB	.0435512	.8010285	0.05	0.957	-1.526436	1.613538
		SALARIO	.0795876	.0677564	1.17	0.240	-.0532126	.2123877
		DESEMPLEO	-.0074645	.7901111	-0.01	0.992	-1.556054	1.541125
		_cons	-10.69333	7.608107	-1.41	0.160	-25.60494	4.218287
		4						
		PIB	-.8461605	.7257155	-1.17	0.244	-2.268537	.5762157
	SALARIO	.0191418	.0748825	0.26	0.798	-.1276251	.1659087	
	DESEMPLEO	.1806082	.6134706	0.29	0.768	-1.021772	1.382988	
	_cons	-.1020097	7.148099	-0.01	0.989	-14.11203	13.90801	

(RÉGIMEN==3 is the base outcome)

★ Para las variables de salud:

AÑO	MODELO							
2000	Multinomial logistic regression				Number of obs = 14			
	Log likelihood = -10.535295				LR chi2(6) = 16.27			
					Prob > chi2 = 0.0124			
					Pseudo R2 = 0.4358			
	RÉGIMEN		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1							
		GPSS_PIB	3.189705	3.023404	1.06	0.291	-2.736058	9.115468
		GPS_PIB	.7538159	.6689526	1.13	0.260	-.5573071	2.064939
		_cons	-13.55732	8.046676	-1.68	0.092	-29.32851	2.213879
		2						
		GPSS_PIB	3.612728	3.095391	1.17	0.243	-2.454127	9.679582
		GPS_PIB	.7443679	.6892462	1.08	0.280	-.6065299	2.095266
	_cons	-15.63732	8.668308	-1.80	0.071	-32.62689	1.352255	
	4							
	GPSS_PIB	3.108285	2.949778	1.05	0.292	-2.673174	8.889744	
	GPS_PIB	.5182332	.6415525	0.81	0.419	-.7391867	1.775653	
	_cons	-10.15434	7.542734	-1.35	0.178	-24.93783	4.629143	

(RÉGIMEN==3 is the base outcome)

2001	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -11.039789				LR chi2(6)	=	15.27
					Prob > chi2	=	0.0183
					Pseudo R2	=	0.4088
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	GPSS_PIB	4.613689	3.996912	1.15	0.248	-3.220115	12.44749
	GPS_PIB	-.0774257	.4064594	-0.19	0.849	-.8740714	.71922
	_cons	-8.364244	4.159978	-2.01	0.044	-16.51765	-.2108363
	2						
GPSS_PIB	5.057808	4.052727	1.25	0.212	-2.88539	13.00101	
GPS_PIB	-.1084539	.4459395	-0.24	0.808	-.9824793	.7655715	
_cons	-10.20848	5.157331	-1.98	0.048	-20.31666	-.1002979	
4							
GPSS_PIB	4.621158	3.973371	1.16	0.245	-3.166507	12.40882	
GPS_PIB	-.3308614	.4392234	-0.75	0.451	-1.191723	.5300066	
_cons	-5.008559	3.518964	-1.42	0.155	-11.9056	1.888484	
(RÉGIMEN==3 is the base outcome)							
2002	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -11.42325				LR chi2(6)	=	14.50
					Prob > chi2	=	0.0245
					Pseudo R2	=	0.3882
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	GPSS_PIB	3.221647	3.036424	1.06	0.289	-2.729634	9.172928
	GPS_PIB	.6207253	.7163615	0.87	0.386	-.7833173	2.024768
	_cons	-13.24744	7.961839	-1.66	0.096	-28.85236	2.357478
	2						
GPSS_PIB	3.392165	3.065416	1.11	0.268	-2.61594	9.400269	
GPS_PIB	.5645576	.7239571	0.78	0.435	-.8543722	1.983487	
_cons	-13.54988	8.163377	-1.66	0.097	-29.54981	2.450042	
4							
GPSS_PIB	3.145849	2.990894	1.05	0.293	-2.716195	9.007894	
GPS_PIB	.4084478	.698229	0.58	0.559	-.9600559	1.776951	
_cons	-9.996733	7.501841	-1.33	0.183	-24.70007	4.706606	
(RÉGIMEN==3 is the base outcome)							
2003	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -12.125675				LR chi2(6)	=	13.09
					Prob > chi2	=	0.0416
					Pseudo R2	=	0.3506
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	GPSS_PIB	2.50449	2.103231	1.19	0.234	-1.617767	6.626747
	GPS_PIB	.32897	.3356545	0.98	0.327	-.3289007	.9868407
	_cons	-9.342314	4.400806	-2.12	0.034	-17.96774	-.7168918
	2						
GPSS_PIB	2.525695	2.144784	1.18	0.239	-1.678005	6.729394	
GPS_PIB	.2786326	.3564684	0.78	0.434	-.4200327	.9772979	
_cons	-9.248237	4.672297	-1.98	0.048	-18.40577	-.090702	
4							
GPSS_PIB	2.391073	2.062319	1.16	0.246	-1.650998	6.433144	
GPS_PIB	.0860959	.3209116	0.27	0.788	-.5428792	.7150711	
_cons	-5.609237	3.489959	-1.61	0.108	-12.44943	1.230957	
(RÉGIMEN==3 is the base outcome)							

		Multinomial logistic regression				Number of obs = 14		
		Log likelihood = -12.710753				LR chi2(6) = 11.92		
						Prob > chi2 = 0.0637		
						Pseudo R2 = 0.3193		
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
2004	1							
		GPSS_PIB	2.053542	2.068167	0.99	0.321	-1.999991	6.107075
		GPS_PIB	.4077671	.3556965	1.15	0.252	-.2893852	1.104919
		_cons	-8.920017	4.24153	-2.10	0.035	-17.23326	-.606771
	2							
		GPSS_PIB	2.082295	2.130331	0.98	0.328	-2.093078	6.257668
		GPS_PIB	.365921	.3839959	0.95	0.341	-.3866971	1.118539
		_cons	-9.01902	4.539369	-1.99	0.047	-17.91602	-.1220203
	4							
		GPSS_PIB	2.192745	2.056381	1.07	0.286	-1.837687	6.223178
		GPS_PIB	.1418597	.3455134	0.41	0.681	-.535334	.8190534
		_cons	-5.753446	3.517453	-1.64	0.102	-12.64753	1.140636
(RÉGIMEN==3 is the base outcome)								
		Multinomial logistic regression				Number of obs = 14		
		Log likelihood = -12.575873				LR chi2(6) = 12.19		
						Prob > chi2 = 0.0578		
						Pseudo R2 = 0.3265		
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
2005	1							
		GPSS_PIB	1.845529	1.997504	0.92	0.356	-2.069507	5.760565
		GPS_PIB	.4841301	.4082336	1.19	0.236	-.3159932	1.284253
		_cons	-9.4885	4.456572	-2.13	0.033	-18.22322	-.753779
	2							
		GPSS_PIB	1.857496	2.058949	0.90	0.367	-2.17797	5.892962
		GPS_PIB	.4470456	.4307196	1.04	0.299	-.3971493	1.291241
		_cons	-9.608123	4.758673	-2.02	0.043	-18.93495	-.2812954
	4							
		GPSS_PIB	1.790732	1.965299	0.91	0.362	-2.061183	5.642646
		GPS_PIB	.2573413	.38957	0.66	0.509	-.506202	1.020884
		_cons	-6.271275	3.694248	-1.70	0.090	-13.51187	.9693186
(RÉGIMEN==3 is the base outcome)								

De los modelos arriba descritos, es posible apreciar que en algunos la veracidad para la explicación por modelo es aparentemente buena, como se puede observar en el ejemplo de las variables demográficas año 2004, en el cual, el modelo tiene una probabilidad de 0.0266 de ser mayor que el estadístico ji cuadrada.

Sin embargo, de manera individual se entra en una conjetura al encontrar que la significancia por variables es mínima, puesto que por ejemplo la variable “Tasa de mortalidad infantil” en el valor 4 de la variable régimen, cuenta con una probabilidad de 0.921 de ser mayor que el estadístico.

Dando otra opción a la respuesta buscada, se formuló un modelo completo de regresión logística en el cual se incluyan todas las categorías de las variables independientes (Demográficas, Económicas y de Salud).

De igual manera que para los modelos de cada categoría de variables independientes, se consideró el análisis de las matrices de correlaciones conjuntas por año de todas las variables regresoras, para así lograr quitar del modelo conjunto aquellas que pudieran afectar negativamente el resultado, debido a su estrecha relación. De tal manera que los modelos quedaron como sigue:

AÑO	MODELO						
2000	Multinomial logistic regression			Number of obs = 14			
	Log likelihood = -10.535295			LR chi2(6) = 16.27			
				Prob > chi2 = 0.0124			
				Pseudo R2 = 0.4358			
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	GPSS_PIB	3.189705	3.023404	1.06	0.291	-2.736058	9.115468
	GPS_PIB	.7538159	.6689526	1.13	0.260	-.5573071	2.064939
	_cons	-13.55732	8.046676	-1.68	0.092	-29.32851	2.213879
	2						
GPSS_PIB	3.612728	3.095391	1.17	0.243	-2.454127	9.679582	
GPS_PIB	.7443679	.6892462	1.08	0.280	-.6065299	2.095266	
_cons	-15.63732	8.668308	-1.80	0.071	-32.62689	1.352255	
4							
GPSS_PIB	3.108285	2.949778	1.05	0.292	-2.673174	8.889744	
GPS_PIB	.5182332	.6415525	0.81	0.419	-.7391867	1.775653	
_cons	-10.15434	7.542734	-1.35	0.178	-24.93783	4.629143	
(RÉGIMEN==3 is the base outcome)							
2001	Multinomial logistic regression			Number of obs = 15			
	Log likelihood = -8.6849179			LR chi2(6) = 23.50			
				Prob > chi2 = 0.0006			
				Pseudo R2 = 0.5750			
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	GPSS_PIB	6.726609	6.696975	1.00	0.315	-6.39922	19.85244
	SERVICIO	.2778917	.5476923	0.51	0.612	-.7955655	1.351349
	_cons	-30.19578	44.80179	-0.67	0.500	-118.0057	57.61412
	2						
GPSS_PIB	7.361528	6.742554	1.09	0.275	-5.853635	20.57669	
SERVICIO	.4116866	.5852809	0.70	0.482	-.7354428	1.558816	
_cons	-41.29341	47.15035	-0.88	0.381	-133.7064	51.11959	
4							
GPSS_PIB	3.824279	2.819985	1.36	0.175	-1.702789	9.351348	
SERVICIO	-.2476557	.2061462	-1.20	0.230	-.6516948	.1563834	
_cons	5.93851	9.003804	0.66	0.510	-11.70862	23.58564	
(RÉGIMEN==3 is the base outcome)							

REGRESIÓN LOGÍSTICA APLICADA EN LA CONFIGURACIÓN DE ESTADOS EN
ALGUNOS PAÍSES DE AMÉRICA LATINA

Diana Citlalli Márquez Palacios

86

2002	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -11.42325				LR chi2(6)	=	14.50
					Prob > chi2	=	0.0245
					Pseudo R2	=	0.3882
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	GPSS_PIB	3.221647	3.036424	1.06	0.289	-2.729634	9.172928
	GPS_PIB	.6207253	.7163615	0.87	0.386	-.7833173	2.024768
	_cons	-13.24744	7.961839	-1.66	0.096	-28.85236	2.357478
	2						
GPSS_PIB	3.392165	3.065416	1.11	0.268	-2.61594	9.400269	
GPS_PIB	.5645576	.7239571	0.78	0.435	-.8543722	1.983487	
_cons	-13.54988	8.163377	-1.66	0.097	-29.54981	2.450042	
4							
GPSS_PIB	3.145849	2.990894	1.05	0.293	-2.716195	9.007894	
GPS_PIB	.4084478	.698229	0.58	0.559	-.9600559	1.776951	
_cons	-9.996733	7.501841	-1.33	0.183	-24.70007	4.706606	
(RÉGIMEN==3 is the base outcome)							
2003	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -9.5403532				LR chi2(9)	=	18.26
					Prob > chi2	=	0.0322
					Pseudo R2	=	0.4891
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	GPSS_PIB	6.078863	4.394196	1.38	0.167	-2.533604	14.69133
	PIB	.5045566	.6953044	0.73	0.468	-.858215	1.867328
	DESEMPLEO	.5078555	.678754	0.75	0.454	-.8224778	1.838189
	_cons	-17.80894	14.23674	-1.25	0.211	-45.71244	10.09456
2							
GPSS_PIB	4.948433	4.391272	1.13	0.260	-3.658301	13.55517	
PIB	1.658159	1.07105	1.55	0.122	-.4410603	3.757379	
DESEMPLEO	.5227031	.7497274	0.70	0.486	-.9467356	1.992142	
_cons	-19.73252	14.70714	-1.34	0.180	-48.55799	9.092946	
4							
GPSS_PIB	5.192466	4.308889	1.21	0.228	-3.252801	13.63773	
PIB	.8012706	.8099055	0.99	0.322	-.7861149	2.388656	
DESEMPLEO	.4149986	.6702554	0.62	0.536	-.8986778	1.728675	
_cons	-15.20094	13.90413	-1.09	0.274	-42.45253	12.05064	
(RÉGIMEN==3 is the base outcome)							

2004	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -9.3498481				LR chi2(9)	=	18.64
					Prob > chi2	=	0.0284
					Pseudo R2	=	0.4993
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	GPSS_PIB	18.79598	14.03857	1.34	0.181	-8.719107	46.31107
	PIB	-1.604998	1.484567	-1.08	0.280	-4.514695	1.304699
	DESEMPLEO	3.500013	2.948614	1.19	0.235	-2.279165	9.279191
	_cons	-56.74033	43.38069	-1.31	0.191	-141.7649	28.28426
2							
GPSS_PIB	17.3034	12.97839	1.33	0.182	-8.133775	42.74058	
PIB	-3.158409	2.618317	-1.21	0.228	-8.290216	1.973398	
DESEMPLEO	5.738259	4.680689	1.23	0.220	-3.435724	14.91224	
_cons	-66.09189	49.43281	-1.34	0.181	-162.9784	30.79463	
4							
GPSS_PIB	17.78124	13.55088	1.31	0.189	-8.778	44.34047	
PIB	-4.592713	3.674299	-1.25	0.211	-11.79421	2.608781	
DESEMPLEO	4.775295	3.697903	1.29	0.197	-2.472462	12.02305	
_cons	-49.27381	37.40117	-1.32	0.188	-122.5788	24.03113	
(RÉGIMEN==3 is the base outcome)							
2005	Multinomial logistic regression				Number of obs	=	14
	Log likelihood = -8.8244211				LR chi2(9)	=	19.70
					Prob > chi2	=	0.0199
					Pseudo R2	=	0.5274
	RÉGIMEN	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	1						
	GPSS_PIB	10.22679	6.567173	1.56	0.119	-2.644637	23.09821
	PIB	-5.525788	4.169766	-1.33	0.185	-13.69838	2.646804
	DESEMPLEO	3.999818	2.891006	1.38	0.166	-1.66645	9.666085
	_cons	-28.19109	18.59074	-1.52	0.129	-64.62827	8.246089
2							
GPSS_PIB	6.603431	4.05451	1.63	0.103	-1.343263	14.55012	
PIB	-1.074348	1.509254	-0.71	0.477	-4.032431	1.883735	
DESEMPLEO	1.532651	1.466535	1.05	0.296	-1.341705	4.407008	
_cons	-21.65074	13.87587	-1.56	0.119	-48.84694	5.545462	
4							
GPSS_PIB	9.672857	6.521111	1.48	0.138	-3.108285	22.454	
PIB	-5.952187	4.230668	-1.41	0.159	-14.24414	2.33977	
DESEMPLEO	3.871551	2.874456	1.35	0.178	-1.762279	9.505382	
_cons	-23.91502	18.01098	-1.33	0.184	-59.21589	11.38586	
(RÉGIMEN==3 is the base outcome)							

Analizando las significancias obtenidas tanto de manera grupal como individual (cada variable), se observa cierta consistencia en los resultados por lo que se elegirán los modelos anteriores, como modelos base para dar la respuesta a que régimen es el dominante en América Latina.

Para todos los modelos anteriores la categoría de referencia seleccionada por el software la categoría 3 correspondiente al Régimen de bienestar familiarista.

Por otro lado, cada regresión refleja la ocurrencia de cada Estado del Bienestar de acuerdo con las variables independientes involucradas.

Recordando un poco, en la sección teórica referente a la regresión logística politómica, la ocurrencia antes reportada, queda definida por:

$$P(Y = s) = \frac{1}{1 + (\sum_j^k \eta_{ij})} \quad P(Y = j) = \frac{\sum_j^k \eta_{ij}}{1 + (\sum_j^k \eta_{ij})}$$

Donde:

$$\eta_{ij} = \exp\left(\beta_{i0} + \sum_j^k \beta_{ij}x_{ij}\right)$$

Para este ejercicio, se tiene considerado que $k = 4$, es decir, las categorías de la variable dependiente o los Regímenes de Bienestar.

Derivando que para cada año, los modelos determinarán las probabilidades siguientes:

Para el año 2000:

$$\eta_{i1} = \exp(-13.56 + 3.19 * GPSS_PIB + 0.75 * GPS_PIB)$$

$$\eta_{i2} = \exp(-15.64 + 3.21 * GPSS_PIB + 0.74 * GPS_PIB)$$

$$\eta_{i4} = \exp(-10.15 + 3.11 * GPSS_PIB + 0.52 * GPS_PIB)$$

De donde es posible determinar la variación de la probabilidad conforme al incremento en una unidad de los valores de cada variable independiente, el cual se manifiesta como:

RÉGIMEN	PROBABILIDAD
1	0.00007
2	0.00001
3	0.99846
4	0.00146

Para el año 2001:

$$\eta_{i1} = \exp(-30.20 + 6.73 * GPSS_PIB + 0.28 * SERVICIO)$$

$$\eta_{i2} = \exp(-41.29 + 7.36 * GPSS_PIB + 0.41 * SERVICIO)$$

$$\eta_{i4} = \exp(-5.94 + 3.82 * GPS_PIB - 0.25 * SERVICIO)$$

Con probabilidades:

RÉGIMEN	PROBABILIDAD
1	0.00000
2	0.00000
3	0.91387
4	0.08613

Para el año 2002:

$$\eta_{i1} = \exp(-13.25 + 3.22 * GPSS_PIB + 0.62 * GPS_PIB)$$

$$\eta_{i2} = \exp(-13.55 + 3.39 * GPSS_PIB + 0.56 * GPS_PIB)$$

$$\eta_{i4} = \exp(-10.00 + 3.15 * GPSS_PIB + 0.41 * GPS_PIB)$$

Con probabilidades:

RÉGIMEN	PROBABILIDAD
1	0.00008
2	0.00007
3	0.99826
4	0.00159

Para el año 2003:

$$\eta_{i1} = \exp(-17.81 + 6.08 * GPSS_PIB + 0.50 * PIB + 0.51 * DESEMPLEO)$$

$$\eta_{i2} = \exp(-19.73 + 4.95 * GPSS_PIB + 1.66 * PIB + 0.52 * DESEMPLEO)$$

$$\eta_{i4} = \exp(-15.20 + 5.19 * GPSS_PIB + 0.80 * PIB + 0.41 * DESEMPLEO)$$

Con probabilidades:

RÉGIMEN	PROBABILIDAD
1	0.00002
2	0.00000
3	0.99982
4	0.00015

Para el año 2004:

$$\eta_{i1} = \exp(-56.74 + 18.80 * GPS_PIB - 1.60 * PIB + 3.50 * DESEMPLEO)$$

$$\eta_{i2} = \exp(-66.09 + 17.30 * GPS_PIB - 3.16 * PIB + 5.74 * DESEMPLEO)$$

$$\eta_{i4} = \exp(-49.27 + 17.78 * GPS_PIB - 4.59 * PIB + 4.78 * DESEMPLEO)$$

Con probabilidades:

RÉGIMEN	PROBABILIDAD
1	0.00000
2	0.00000
3	1.00000
4	0.00000

Para el año 2005:

$$\eta_{i1} = \exp(-28.19 + 10.23 * GPS_PIB - 5.53 * PIB + 4.00 * DESEMPLEO)$$

$$\eta_{i2} = \exp(-21.65 + 6.60 * GPS_PIB - 1.07 * PIB + 1.53 * DESEMPLEO)$$

$$\eta_{i4} = \exp(-23.92 + 9.67 * GPS_PIB - 5.95 * PIB + 3.87 * DESEMPLEO)$$

Con probabilidades:

RÉGIMEN	PROBABILIDAD
1	0.00000
2	0.00000
3	1.00000
4	0.00000

Con la aplicación del modelo politómico se obtuvo una respuesta a la probabilidad de aparición de cada estado de bienestar en América Latina, según cambian las variables relacionadas con el PIB y desempleo.

Derivado de lo anterior, el régimen que en el transcurso de los años tiene dominancia es el tercero, puesto que la probabilidad que lo define es bastante alta.

CONCLUSIONES

En la amplia extensión del continente americano es posible encontrar una variedad muy amplia de características económicas por países o regiones; sin embargo, los estudios estadísticos logran concentrar parte de ese total de información para lograr presentar una característica uniforme y representativa con la cual sea posible no sólo ubicar a América como una región fragmentada sino como la unión de países que además de compartir una extensión territorial, también comparte características sociales y económicas, las cuales pueden ser referencia para el bienestar conjunto.

Manteniendo ese ideal, el presente trabajo evidencia los patrones de bienestar entendido como la manera de manejar los riesgos inherentes a ciertos aspectos de la Seguridad Social y que pueden ser encontrados en la región de América Latina.

Como precedentes de esta investigación es posible ubicar diversas fuentes que desde años atrás se han interesado en este estudio, cabe señalar a al autor Gosta Esping-Andersen, sociólogo cuyo principal interés es el estudio de los Estados del Bienestar.

Así mismo, de manera direccionada hacia América Latina, es posible encontrar los estudios publicados por la Fundación Carolina CeALCI, destacando el estudio realizado por la Doctora Juliana Martínez Franzoni, en cual se analiza y proponen tres tipos de regímenes del Bienestar para dicha región de América.

A partir de ese análisis, los regímenes propuestos son:

- ★ Estado del Bienestar estatal-proteccionista,
- ★ Estado del Bienestar productivista-informal y
- ★ Estado del Bienestar familiarista.

Sin embargo, en el presente trabajo se abordó el último estado propuesto como dos diferentes, debido al grado de familiarización contemplado en el estudio de la Doctora Martínez; quedando determinados de la manera siguiente:

- ★ El estado del Bienestar estatal-proteccionista, se denota de manera principal el papel del estado en este sistema y la búsqueda del deslinde de su control en la prestación de servicios a la sociedad.
- ★ En el estado del Bienestar productivista-informal, el sector privado es el principal elemento de este régimen, con la diferencia de un sistema privado en cuanto a que los recursos utilizados son públicos.
- ★ En los estado del Bienestar familiarista y altamente familiarista, como su nombre lo dice, tiene su base en la familia, es decir la población en general depende de los arreglos económicos familiares que se establezcan. Cabe mencionar que, entre estos dos estados, la diferencia simplemente radica en la magnitud de familiarización existente.

De esta manera, el análisis basado en la regresión logística tuvo como objetivo la comprobación de la existencia de los anteriores regímenes en América Latina.

Una vez aplicados los modelos basados en la regresión logística múltiple, es posible concluir que el régimen encontrado para la región estudiada de América Latina, determinada por los países cuyas bases de datos fueron investigadas es el Estado del Bienestar familiarista.

Por lo tanto, para los países analizados, es viable decir que se aplica un sistema en el cual las obligaciones propias de la distribución de bienestar a las familias; aunado a ello, la aplicación de un régimen de este tipo, provoca que se aprecie de una manera más clara la división entre la población que se encuentra dentro y fuera del mercado laboral.

Con la revisión anterior, se observa que los regímenes encontrados por medio de conglomerados elaborados por la Doctora Martínez Franzoni se presentan de manera más heterogénea, y que por el contrario, según las condiciones que se tuvieron para este estudio, la región del continente americano es homogénea y con el sistema antes reportado.

Una primera conjetura que surge se presenta cuando al querer encontrar la diferencia entre los resultados observados en la investigación base y el presente trabajo es la cantidad de variables independientes encontradas; y asimismo la cantidad de países integrantes de la muestra, puesto que, en comparación del análisis base, esta investigación aborda más países que sólo un representante de cada régimen de bienestar.

Cabe señalar que el análisis estadístico utilizado se basó en el modelo de regresión logística politómico por cumplir con el objetivo de encontrar un mejor ajuste de los datos.

Para la realización del análisis anterior, se utilizaron diversas series históricas; sin embargo, la obtención de ellas, reflejó las restricciones existentes para el acceso a información referente a algunos países, destacando que si se contara con toda la información estadística de los países estudiados, cualquier estudio realizado podría verse mejorado o simplemente ajustado a los valores más reales.

Por otro lado, aún con los valores conseguidos, fue posible realizar diversas estimaciones por medio de regresiones lineales, para así completar algunas series históricas que requerían información.

De manera similar, para la elaboración de los modelos logísticos, la elección de las variables a utilizar se realizó por medio de las correlaciones existentes entre dichas variables.

Por lo anterior, es conveniente recordar que para estudios realizados sobre este tema y en general para aquellos referentes a características económicas; en especial cuando se busca encontrar un punto de referencia con el cual realizar una comparación, lo más idóneo es la elección semejante de variables como en el caso de los representantes de cada Estado de Bienestar.

Finalmente, el punto más importante a destacar de la presente investigación, es que el resultado obtenido en el cual se hace referencia a que el estado del bienestar existente en América Latina es el familiarista, depende en gran sentido de las variables independientes que se trabajaron.

ANEXO A

Al realizar la instrucción del modelo de regresión en el software Stata, los resultados obtenidos en dicho proceso se muestran en un formato como el siguiente:

```
Multinomial logistic regression          Number of obs =      14
                                         LR chi2(6)       =     14.80
                                         Prob > chi2      =     0.0219
Log likelihood = -11.27476              Pseudo R2       =     0.3962
```

RÉGIMEN		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1	TMI	-.1641438	.3067716	-0.54	0.593	-.7654052	.4371175
	EV	.1960008	.81066	0.24	0.809	-1.392863	1.784865
	_cons	-10.12742	65.60528	-0.15	0.877	-138.7114	118.4566
2	TMI	-1.299139	1.122801	-1.16	0.247	-3.499787	.90151
	EV	-2.269118	2.778269	-0.82	0.414	-7.714425	3.176189
	_cons	193.0746	226.1024	0.85	0.393	-250.0779	636.2271
4	TMI	.0946653	.1710912	0.55	0.580	-.2406672	.4299979
	EV	.0102072	.6144313	0.02	0.987	-1.194056	1.21447
	_cons	-4.646551	47.69134	-0.10	0.922	-98.11986	88.82676

(RÉGIMEN==3 is the base outcome)

Donde las variables que se han especificado como independientes son: TMI (tasa de mortalidad infantil) y EV (esperanza de vida); así como, RÉGIMEN la dependiente.

Lo primero denotado en el lado izquierdo superior a la tabla denominado “Log likelihood”, muestra el valor estimado de la función de verosimilitud alcanzado cuando el mecanismo iterativo realizado por Stata, llega al punto de tope; es decir, cuando la diferencia entre la iteración última i y la anterior $i - 1$ es mínima.

```
Multinomial logistic regression          Number of obs =      14
                                         LR chi2(6)       =     14.80
                                         Prob > chi2      =     0.0219
Log likelihood = -11.27476              Pseudo R2       =     0.3962
```

Arriba a la derecha aparece el número de observaciones, seguido de la razón de verosimilitud con distribución Ji-cuadrada con (i) grados de libertad ($LR\ chi2(i)$). El siguiente punto en la lista es una prueba estadística de significación del modelo

basada nuevamente en el estadístico Ji-cuadrada; con un nivel de confianza del 95%, con lo que se determina que si la probabilidad que aparece es menor a 0.05 , el modelo es significativo.

Por último en esa columna, se encuentra el estadístico “Pseudo R2”, mismo que indica la bondad de ajuste del modelo a los datos.

Adentrando en la información contenida en la tabla, en la segunda columna titulada “Coef.” se presentan los coeficientes correspondientes al modelo, ordenados por grupos definidos de acuerdo a los valores de la variable independiente, mismos que son {1,2,4 } en la imagen.

RÉGIMEN	Coef.
1	TMI
	EV
	_cons
2	TMI
	EV
	_cons
4	TMI
	EV
	_cons

En la misma columna, al final de los coeficientes de las variables independientes se encuentran aquellos asociados a las constantes de los modelos.

Avanzando en las columnas, la tercer columna (“Std. Err.”) se presentan los errores estándar de los coeficientes de regresión individual para los modelos estimados respectivos.

RÉGIMEN		Coef.	Std. Err.
1	TMI	-.1641438	.3067716
	EV	.1960008	.81066
	_cons	-10.12742	65.60528
2	TMI	-1.299139	1.122801
	EV	-2.269118	2.778269
	_cons	193.0746	226.1024
4	TMI	.0946653	.1710912
	EV	.0102072	.6144313
	_cons	-4.646551	47.69134

Respecto a las columnas “z” y “P>|z|” presentan los valores de los estadísticos de una distribución normal estándar utilizados en la prueba y los de las probabilidades asociadas, que indican si el coeficiente es estadísticamente significativo o no; en el caso de los números en “z”, se tiene la relación existente entre los coeficientes y los errores estándar (*Coef./Std. Err.*).

RÉGIMEN		Coef.	Std. Err.	z	P> z
1	TMI	-.1641438	.3067716	-0.54	0.593
	EV	.1960008	.81066	0.24	0.809
	_cons	-10.12742	65.60528	-0.15	0.877
2	TMI	-1.299139	1.122801	-1.16	0.247
	EV	-2.269118	2.778269	-0.82	0.414
	_cons	193.0746	226.1024	0.85	0.393
4	TMI	.0946653	.1710912	0.55	0.580
	EV	.0102072	.6144313	0.02	0.987
	_cons	-4.646551	47.69134	-0.10	0.922

Por último, en la tabla se presentan los intervalos de confianza correspondientes a los coeficientes, con un nivel de confianza del 95%, establecidos por $Coef. \pm [(z_{\alpha/2})(Std. Err.)]$, donde $z_{\alpha/2}$ es el valor de la distribución normal estándar.

RÉGIMEN		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
1	TMI	-.1641438	.3067716	-0.54	0.593	-.7654052 .4371175
	EV	.1960008	.81066	0.24	0.809	-1.392863 1.784865
	_cons	-10.12742	65.60528	-0.15	0.877	-138.7114 118.4566

REFERENCIAS

- ☆ Adelantado, José; Calderón, Eduardo. Globalización y estados del bienestar: ¿respuestas semejantes a problemas parecidos?. Cuadernos de Relaciones Laborales, 2005,23,Vol. 2
<http://revistas.ucm.es/rla/11318635/articulos/CRLA0505220015A.PDF>
- ☆ Agresti, Alan. An introduction to categorical data analysis. JohnWiley & Sons, Estados Unidos de América, 2007.
- ☆ Arriba González de Durana, Ana. El estado de bienestar: Modelos, regímenes y objetivos. Universidad de Alcalá. http://www2.uah.es/ana_arriba/
- ☆ Banco Interamericano de Desarrollo www.iadb.org
- ☆ Barba Solano, Carlos. Régimen de bienestar y reforma social en México. CEPAL, Comisión Económica para América Latina y el Caribe. Santiago de Chile, 2004.
- ☆ CEPAL: Comisión Económica para América Latina y el Caribe: Sobre la base de información oficial. CEPALSTAT www.eclac.org/estadisticas , www.risalc.org
- ☆ Chomsky, Noam. Los Límites de la Globalización. Ariel. Barcelona, 2002.
- ☆ Dietz; Gail; Krickeberg; Tsiatis; Samet. Statistics for biology and health. Springer-Verlag. New York, 1995.
- ☆ Draibe, Sonia; Resco, Manuel. Estado de bienestar, desarrollo económico y ciudadanía: algunas lecciones de la literatura contemporánea. CEPAL, Comisión Económica para América Latina y el Caribe. México, 2006.

- ☆ Gómez Olivares, Mario. Globalización, Inversiones Extranjeras y Desarrollo en América Latina. <http://eumed.net/>
- ☆ Gosta Esping, Andersen. Fundamentos sociales de las Economías Pos-industriales. Ariel. Barcelona, 2000.
- ☆ H. Del Valle, Alejandro. Los Estudios sobre regímenes de bienestar en América Latina. Artículo de la Revista A-MÉRICA. Volumen 2 Número 3, 2009. Universidad de Deusto, España. www.revistaamerika.deusto.es
- ☆ Hosmer Jr, David W; Lemeshow, Stanley. Applied Logistic Regression. WILEY, Estados Unidos de América, 1989.
- ☆ Kemeny , Jim. Theories of Power in “The Three Worlds of Welfare capitalism”. Journal of European Social Policy, 1995 Vol. 5 No. 22.
- ☆ Kleinbaum, David G.; Klein, Michael. Logistic regression : a self-learning text. Springer Verlag, New York, 2002.
- ☆ Manual de Stata. Demos Apoyo a la Docencia. España. <http://demos.usal.es/>
- ☆ Martínez Franzoni, Juliana. ¿Arañando bienestar? Trabajo remunerado, protección social y familias en América Central. CLACSO, Consejo Latinoamericano de Ciencias Sociales. Buenos Aires, 2008. <http://bibliotecavirtual.clacso.org.ar>
- ☆ Montgomery, Peck, Vining, Introducción al Análisis de Regresión Lineal. 3ª edición, CECSA, 2002.

- ☆ Moral Peláez, Irene. Modelos de Regresión: Lineal Simple y Regresión Logística. Revista de la Sociedad Española de Enfermería Nefrológica <http://www.revistaseden.org/files/14-CAP%2014.pdf>
- ☆ Moreno Mínguez, Almudena. Análisis de la reducida fecundidad y el reducido empleo femenino en los regímenes de bienestar del sur de Europa. Universidad de Valladolid, Departamento de Sociología y Trabajo Social. Diciembre, 2006.
- ☆ Regresión Logística. Fundación Andaluza Beturia, Investigación en Salud, España. <http://www.fabis.org/html/>
- ☆ Regresión Logística Asociación de la sociedad española de hipertension, liga española para la lucha contra la hipertensión arterial, <http://www.seh-lelha.org/rlogis1.htm>
- ☆ Silva Aycaguer, Luis Carlos. Regresión logística. La muralla, Madrid, 2004
- ☆ Silva Aycaguer, Luis Carlos. Excursión a la Regresión Logística en las ciencias de la Salud. La muralla, Madrid, 2004
- ☆ Simonoff, Jeffrey. Analysing Categorical Data. Springer Texts in Statistics, New York, 2003.
- ☆ Stata Annotated Outpu, Multinomial Logistic Regression. UCLA: Academic Technology Services, Statistical Consulting Group. <http://www.ats.ucla.edu/stat/stata/>.
- ☆ Ulrich, Beck. ¿Qué es la globalización? Falacias del globalismo respuestas a la globalización. PAIDÓS, España, 1998.