

**LA APLICACIÓN DE MÉTODOS MULTIVARIADOS Y ESTRUCTURALES DE ESTADÍSTICA EN LA
INTERPRETACIÓN DE LAS CONDICIONES DE INGRESO PER CÁPITA
(Analysis of the application of Multivariate Methods in Statistics for measuring conditions
Income per capita)**

González Muñoz Oscar*

RESUMEN

Una de las principales herramientas de la Estadística para el análisis directo de datos registrados en una investigación, es el análisis de datos de forma multivariada respecto a los resultados generados por una variable estudiada; académicamente llamado "*análisis multivariado de datos*" es una herramienta que permite el estudio de cada uno de los elementos a estudiar, por medio de los diferentes resultados generados en el objeto de estudio que se trate. Por lo que este trabajo, analiza la importancia de los métodos estadísticos que refieren al análisis *univariante o bivariante*, con el fin de llegar a interpretar los elementos que inciden en el desarrollo de las conclusiones a las que llegan algunos casos particulares.

PALABRAS CLAVE

Estadística, datos, análisis, *univariantes, bivariantes*.

ABSTRACT

One of the main tools of statistics for the direct analysis of data recorded in an investigation is the analysis in a univariate or bivariate variable; academically called "multivariate data analysis" is a mechanism that allows the analysis of each of the elements within a research study through the different results generated in the object of study as research objective in question. So this paper analyzes the importance of this statistical method, in order to get to interpret the elements that directly affect the development of the conclusions we reach some particular cases.

KEYWORDS

Statistical data, analysis, univariate, bivariate.

Classification JEL: I3

* Académico Investigador del Instituto de Investigaciones y Estudios Superiores de las Ciencias Administrativas (IIESCA) de la Universidad Veracruzana; E-mail: oscagonzalez@uv.mx

I. INTRODUCCIÓN

Uno de los sempiternos intereses del hombre ante el entorno que le rodea, radica en la necesidad de responder de manera lógica y científica a las interrogantes que este le plantea. Por lo que mediante la demostración de los hechos empíricos ha tratado de identificar la forma de responder a tales circunstancias. Y con ello, dar salida a los problemas a los que se enfrenta.

No obstante, el proceso investigativo por el cual trata de responder a tales cuestionamientos, implica estudiar la manera en la que se busca dar solución a los obstáculos a que se refieren. Desde el enfoque metodológico y del diagnóstico hasta las herramientas estadísticas utilizadas para tal fin.

Una de las principales aplicaciones estadísticas en cualquier campo, y no sólo en el estudio de las Ciencias Sociales, descansan sobre el hecho de estudiar sobre la posibilidad de realizar observaciones, registros y obtención de información sobre experimentos repetidos. Aceptando esencialmente, que las condiciones son similares a las que prevalecen dentro del hecho a registrar. Lo que de acuerdo con Solanas A.M. (2002) ha dado relevancia a la estadística como instrumento de interpretación de los datos investigados.

De esta manera, es útil mencionar que en algunas áreas de la investigación, los objetos o fenómenos observados, como en las Ciencias Sociales; del comportamiento; de la salud y la mercadotecnia, los estudios de mercado sobre el consumidor, logran generar conclusiones que pueden llegar a variar entre los encuestados, de acuerdo con el momento en el que se obtienen. Es decir, las condiciones en las que se obtienen datos de los elementos estudiados, generan resultados que pueden ser variables entre los encuestados. Por lo que se generan conclusiones que resultan cuestionables.

Una situación más clara es la que resulta en el campo de las Ciencias Químicas, donde los datos obtenidos y registrados por un investigador, pueden ser controlados por medio de las condiciones que privan en el laboratorio y posiblemente semejantes al momento en que se obtienen o registran. De esta manera, la obtención de datos, puede estar basada en condiciones del ambiente en el que se obtiene

la información y de que las aplicaciones estadísticas sean diferentes.

En suma, en el campo de las Ciencias Sociales o de la Conducta, el investigador puede hacer un esfuerzo extralimitado para observar repetidamente las mismas condiciones, aunque consiente de la frecuencia de encontrar diferencias entre las observaciones realizadas y las demostrables empíricamente de acuerdo al momento en el que las realiza; determinadas entonces sus conclusiones de acuerdo a las condiciones del propio ambiente y del entorno que privilegian el desarrollo.

II. EL ANÁLISIS MULTIVARIADO DE DATOS

De acuerdo con Akdogan H. (1997) la Estadística es el estudio de fenómenos donde bajo un mismo conjunto de condiciones, los resultados obtenidos presentan variabilidad, frecuencia distinta y consideraciones variables de acuerdo con la manera de argumentar un posible resultado o conclusión. Por tanto, permite generar resultados impredecibles a priori; es decir, bajo la consideración de la incertidumbre como una condición natural asociada al objeto de estudio y que requiere atenderse.

Considerar que la Estadística trata a la incertidumbre como un factor causal dentro de su naturaleza, tiene lugar en lo inherente a la misma aplicación de sus herramientas y técnicas. Por lo cual, esta condición parte de manera natural a su propia ejecución. La Estadística logra estar determinada por la esencia de la realidad que ante su existencia genera incertidumbre.

En caso representativo dentro de la Estadística, sus herramientas logran servir de análisis de datos por medio del "Análisis Multivariante de Datos" conocido como el conjunto de métodos estadísticos que tienen por objetivo analizar simultáneamente conjuntos de datos multivariantes cuando existen variables distintas y cuantificas para cada individuo u objeto estudiado.

De acuerdo con Martínez Abascal, E. y C. Cabezas (1997) y Hair, J., Anderson, *et al* (1999) el Análisis Multivariante, es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente diferentes conjuntos de datos bajo la consideración de que existen n -ésimas variables medidas para cada individuo u objeto estudiado. Su razón de ser radica en un

mejor entendimiento del fenómeno objeto de estudio obteniendo información que los métodos estadísticos *univariantes* y *bivariantes* son incapaces de conseguir.

Un ejemplo de lo anterior, se logra por medio del resultado de una investigación donde se mide la media, desviación estándar, y coeficiente de variación (en porcentaje) de la renta *per cápita* de 411 Municipios de México. Esto con el fin de conocer el nivel de desigualdad por medio de las condiciones del ingreso. Para ello, los municipios son ordenados en función del valor del ingreso por medio de su desviación estándar, como referente de desigualdad. Es decir, la dispersión de la renta es utilizada como una medida de desigualdad y en cuanto mayor es la dispersión, mayores diferencias se darán en los ingresos de las personas estudiadas.

En un caso propositivo, lo importante será identificar los elementos que inciden en el desarrollo de la desigualdad como referente de desigualdad del ingreso. Y las técnicas utilizadas para el estudio de tales condiciones, implicaría empezar por medio de la distribución de la renta bajo condiciones de espacio geográfico donde se encuentran las personas, acceso a ciertos servicios de salud, educación, nivel de infraestructura desarrollada, etc. Así, la comunidad que representa una mayor o menor dispersión de esa variable puede estar relacionada con las condiciones variables del entorno y la desigualdad entonces deberse a otras condiciones no consideradas.

Considerando un caso representativo de la aplicación multivariada, se puede identificar que el estudio conjunto de datos *multivariantes*, permite el análisis estadístico que las herramientas uni y bidimensional son incapaces de conseguir. Por ello, el analista o investigador debe tomar decisiones óptimas en el contexto en el que se encuentre teniendo en cuenta la información disponible por el conjunto de datos analizado.

Los *Métodos de dependencia* que se han desarrollado con base a la Estadística que bajo el enfoque de análisis multivariado de datos se logran identificar a través de las generalidades de organización de datos en los que se incurre con motivo del objetivo de la investigación. Y suponen que las variables analizadas están divididas en dos grupos: las variables dependientes y las variables independientes.

Por lo que por medio de esta clasificación determinan si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma.

De acuerdo con Afifi A. y Clark V. (1996) y Boyle, G.E. y McCarthy T.G. (1997), los Métodos de dependencia se pueden clasificar en dos grandes subgrupos según que la variable (s) dependiente (s) sea (n) cuantitativas o cualitativas. Si la variable dependiente es cuantitativa algunas de las técnicas que se pueden aplicar son las siguientes:

1) Análisis de Regresión: Es la técnica adecuada si en el análisis hay una o varias variables dependientes métricas cuyo valor depende de una o varias variables independientes métricas. Por ejemplo, intentar predecir el gasto anual en cine de una persona a partir de su nivel de ingresos, nivel educativo, sexo y edad.

2) Análisis de Supervivencia: Es similar al análisis de regresión pero con la diferencia de que la variable independiente es el tiempo de supervivencia de un individuo u objeto. Por ejemplo, intentar predecir el tiempo de permanencia en el desempleo de un individuo a partir de su nivel de estudios y de su edad.

3) Análisis de la Varianza: Se utilizan en situaciones en las que la muestra total está dividida en varios grupos basados en una o varias variables independientes no métricas y las variables dependientes analizadas son métricas. Su objetivo es averiguar si hay diferencias significativas entre dichos grupos en cuanto a las variables dependientes se refiere. Por ejemplo, ¿Hay diferencias en el nivel de colesterol por sexos? ¿Afecta, también, el tipo de ocupación?

4) Correlación Canónica: Su objetivo es relacionar simultáneamente varias variables métricas dependientes e independientes calculando combinaciones lineales de cada conjunto de variables que maximicen la correlación existente entre los dos conjuntos de variables. Por ejemplo, analizar cómo están relacionadas el tiempo dedicado al trabajo y al ocio de una persona con su nivel de ingresos, su edad y su nivel de educación. Si la variable dependiente es cualitativa algunas de las técnicas que se pueden aplicar son las siguientes:

5) **Análisis Discriminante:** Esta técnica proporciona reglas de clasificación óptimas de nuevas observaciones de las que se desconoce su grupo de procedencia basándose en la información proporcionada los valores que en ella toman las variables independientes. Por ejemplo, determinar los ratios financieros que mejor permiten discriminar entre empresas rentables y poco rentables.

6) **Modelos de regresión logística:** Son modelos de regresión en los que la variable dependiente es no métrica. Se utilizan como una alternativa al análisis discriminante cuando no hay normalidad

7) **Análisis Conjoint:** Es una técnica que analiza el efecto de variables independientes no métricas sobre variables métricas o no métricas. La diferencia con el Análisis de la Varianza radica en dos hechos: las variables dependientes pueden ser no métricas y los valores de las variables independientes no métricas son fijadas por el analista. En otras disciplinas se conoce con el nombre de Diseño de Experimentos. Por ejemplo, una empresa quiere diseñar un nuevo producto y para ello necesita especificar la forma del envase, su precio, el contenido por envase y su composición química. Presenta diversas composiciones de estos cuatro factores. 100 clientes proporcionan un ranking de las combinaciones que se le presentan. Se quiere determinar los valores óptimos de estos 4 factores.

Un caso representativo de los “*Métodos de Dependencia*”, es el que se logra en los mercados financieros. De acuerdo con Sharma S. (1998). Cuando se invierte en un activo se valoran varios factores, uno de ellos es la variabilidad que el precio del activo alcanza con respecto al valor medio. Es decir, se busca racionalmente maximizar el beneficio esperado y que este se cumpla. Este concepto es una medida de *riesgo*, puesto que para una misma rentabilidad media esperada será preferible aquella acción que se desvíe lo menos posible de la misma; generando con ello una rentabilidad aceptable a las condiciones de rendimiento. Una aproximación a lo anterior, es el comportamiento de la varianza este resultado. Es decir, el rendimiento obtenido dentro del resultado esperado está determinado por las condiciones de variabilidad dependientes de información conocida o dependiente.

Sin embargo, el cálculo de la varianza con datos dentro de un período de tiempo puede ser considerado una medida de riesgo que se afronta al adquirir una acción. Es decir, el concepto de *volatilidad* es el empleado por los técnicos bursátiles en lugar de la varianza como término asociado a la diferencia existente entre el cumplimiento de mis expectativas y las que realmente se logran en el mercado. Una situación similar ocurrirá cuando el valor esperado en el precio de la acción será más deseable que aquella acción con una menor varianza. De esta manera, la escasa diferencia que existe entre las rentabilidades medias y la desviación estándar de la rentabilidad de la acción será más atractiva si se quiere enfrentar un menor riesgo.

A manera de conclusión Camino (1997) y de los trabajos anteriores de Everitt, B. y Graham D. (1991). concluye que existen ciertas dimensiones que afectan cualquier consideración posible de riesgo en el valor futuro esperado de una acción: Las cotizaciones del valor accionario de cualquier empresa o indicador bursátil, son efectivamente diferentes en los distintos días de la semana. Lo anterior, es un conflicto al cual se enfrentan los métodos de dependencia de variables conocidas en el mercado.

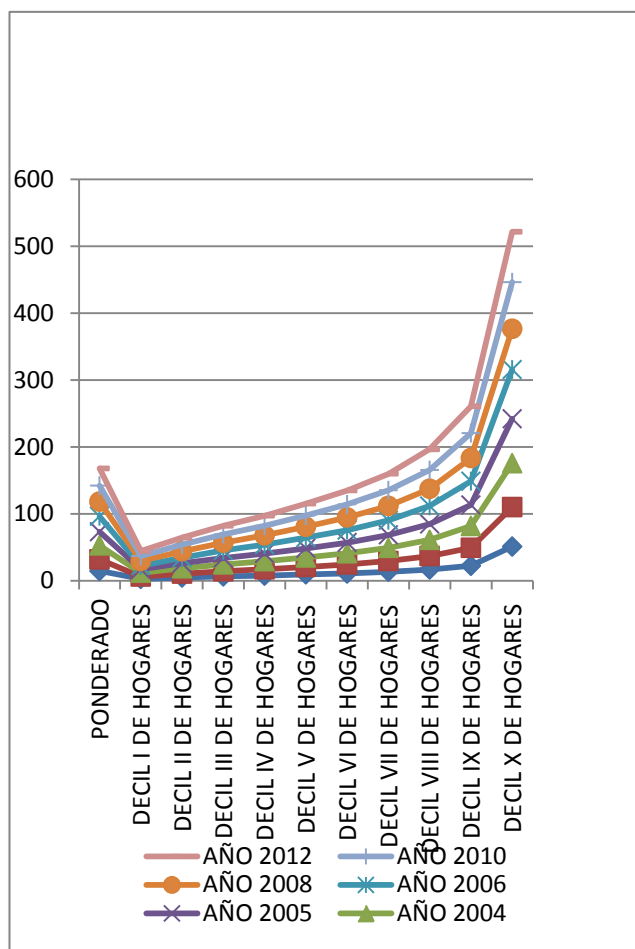
Métodos de interdependencia: Estos métodos no distinguen entre variables dependientes e independientes y su objetivo consiste en identificar qué variables están relacionadas, cómo lo están y por qué. Es una forma de integración de los elementos que inciden directamente en el desarrollo explicativo de una variable resultante.

Un caso de lo anterior, son los estudios basados en los niveles de consumo de los mercados. De acuerdo con González (2012) los bienes alimentarios son un referente de los niveles de Bienestar alcanzados en el mercado. Pues se considera que el nivel de consumo de algunos bienes alimentarios puede estar sujeto a condiciones de capacidad de compra. Y por tanto puede ser la incapacidad de hacerlo, un referente de las condiciones del mercado, sujeta al tiempo o disposición en que se consumen.

De acuerdo con el INEGI (2012) en la Gráfica No. 1, el comportamiento del Gasto en Alimentación en los últimos años por deciles ha cambiado. Y se percibe un incremento en los niveles de gasto por familias en proporción con

el ingreso, lo cual significa un aumento en el costo de los bienes a consumir.

Gráfica 1. Comportamiento de la proporción del gasto en alimentación neto por deciles del año 2000-2012



Fuente: Elaboración propia con datos del INEGI (2012).

De acuerdo con el CONEVAL (2013) el nivel de Bienestar en México, se mide por medio de las líneas de bienestar (LB); y se determinan por medio de la Línea de Bienestar Mínimo (LBM), que considera al valor de la canasta alimentaria por persona al mes; y la Línea de Bienestar (LB), que equivale al valor total de la canasta alimentaria y de la canasta no alimentaria por persona al mes.

El resultado de analizar el gasto en alimentación destinado a cereales y otro tipo de alimentos, permite encontrar que los bienes alimentarios como los cereales en referencia al estudio del Bienestar han aumentado en consumo. Por lo que estos manifiestan cierto resultado en la elección del tipo de alimentos que satisfacen las necesidades nutricionales, y que pueden estar determinados por las calidades, precios y disponibilidad de éstos.

No obstante, de acuerdo con González y Menenes (2013) al establecer posibles relaciones con otras variables, se encuentra que el 52% de los encuestados, considera que el consumo de cereales predomina en el gasto de los alimentos. Así mismo que el nivel de gasto destinado promedio es de \$240.00 por mes entre todos los deciles. La mayor proporción de los hogares estudiados tienen un consumo de otros alimentos como carnes entre el 79% y el 86%. Los hogares mayores de 4 personas, son los que más consumen alimentos cárnicos en general y mantienen un nivel de gasto del 72% del gasto en alimentos dedicado al consumo de estos productos.

Los resultados muestran que el tamaño de los hogares se relaciona con la cantidad de cereales consumidos ($r_s=.218$, $p=.002$) y el consumo sensato ($r_s=.158$, $p=.023$) y una relación inversa significativa del consumo de cereales con el consumo de productos como cigarrillos y bebidas alcohólicas ($r_s= -.118$, $p=.004$). Se observa que el nivel de gasto en cereales mantiene también una relación negativa con el tamaño de la familia ($r_s=-.118$, $p=.029$), y AUDIT ($r_s=.318$, $p=.008$).

II.1 Los Métodos estructurales de evaluación de datos

De acuerdo con Hays W. (1973) y Arnau J. (1996) estos métodos suponen que las variables están divididas en dos grupos: el de las variables dependientes y el de las independientes. El objetivo de estos métodos es analizar, no sólo como las variables independientes afectan a las variables dependientes, sino también cómo están relacionadas las variables de los dos grupos entre sí.

Una de las aplicaciones importantes es la figura f conocida como "Curva de Phillips de macroeconomía". Con base a los datos de la tasa de cambio porcentual de los salarios nominales (Y) como variable dependiente, se identificó la tasa porcentual de desempleo (X) como variable independiente para el Reino Unido durante el periodo 1861 a 1957. De esta manera, se demostró la existencia asimétrica de los cambios salariales en el nivel de desempleo: si la tasa de desempleo está por debajo de U" (denominada por los economistas como tasa natural de desempleo) por cada unidad de cambio en el desempleo, los salarios aumentan con mayor rapidez de lo que caen debido a un

cambio equivalente cuando la tasa de desempleo está por encima del nivel natural. Este hecho particular generó que la curva de Phillip's puede deberse a factores institucionales, tales como el poder de negociación de los sindicatos, los salarios mínimos, compensaciones por desempleo, etc. Es decir variables independientes afectando las dependientes.

III. CONCLUSIONES GENERALES

El "*Análisis Multivariante*" es el medio que presenta las herramientas de interpretación para el análisis de datos dentro del entorno de la investigación y producción de investigaciones. Es considerado la herramienta por la cual se analiza las variables de los métodos estadísticos; y cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes, es decir, dependientes de un conjunto de elementos de investigación conocidas que permitan establecer conclusiones basadas en el objetivo de la investigación. Su razón de ser radica en un mejor entendimiento del fenómeno objeto de estudio obteniendo información que los métodos estadísticos univariantes y bivariantes son incapaces de conseguir.

No obstante, como ha quedado demostrado de manera general, la Estadística aplicada a la comprensión de ciertos problemas sociales como son los que se desarrollan en el marco de la integración de variables económicas, como gasto de los hogares para medir el bienestar. Y en otro, en el caso de las variables que permiten identificar las variables para determinar el nivel de desigualdad social, resulta importante considerar que tales métodos generan una comprensión asíntota del problema estudiado, sujeto a la incertidumbre de los momentos que pueden ser estudiados. Es entonces necesario mencionar el caso de la investigación social de los problemas a solucionar por medio de los esquemas organizados de información.

IV. REFERENCIAS

- Affi, A.A. and Clark, V. (1996) *Computer-Aided Multivariate Analysis*. Third Edition. Texts in Statistical Science. Chapman and Hall.
- Arnau J. (1996) Técnicas de Análisis Avanzadas y Diseño de Investigación: Tendencias Actuales y Líneas Futuras de Desarrollo. En J. Arnau (editor): *Métodos y Técnicas Avanzadas de Análisis de Datos en Ciencias del Comportamiento*. EUB, Barcelona, 1996.
- Solanas A.M. (2002) *La Enseñanza de la Estadística en las Ciencias del Comportamiento a Inicios del Siglo XXI*. Metodología de las Ciencias del Comportamiento 4, no. 2 (2002), 157-183
- Akdogan, H. (1997) "International security selection under segmentation: Theory and application", *The Journal of Portfolio Management*, vol. 24, n. 1, págs. 82-92.
- Barro, R.J., 1991, "Economic growth in a cross section of countries", *Quarterly Journal of Economics*, n.º. vol. 106, n.º. 2, págs. 407-443.
- Boyle, G.E. y T.G. McCarthy, 1997, "A simple measure of b-convergence", *Oxford Bulletin of Economics and Statistics*, vol. 59, n.º. 2, págs. 257-264.
- Camino, D., 1997, "Efectos intradía y día de la semana en la Bolsa de Madrid. Información y Volumen de contratación", *Revista Española de Financiación y Contabilidad*, vol. XXVI, n.º 90, págs. 51-75.
- Doménech, R., M. Fernández y D. Taguas, 1997, "La fiscalidad sobre el trabajo y el desempleo en la OCDE", *Papeles de Economía Española*, n.º. 72, págs. 178-191.
- D. Peña, J. Romo: *Introducción a la Estadística para las Ciencias Sociales*. McGraw-Hill, Madrid, 1997.
- Everitt, B. and Graham, D. (1991). *Applied Multivariate Data Analysis*. Arnold.
- Grim, L. and Yarnold, P.R. (1994). *Reading and understanding multivariate statistics*. American Psychological Association. Washington D.C
- González Muñoz, Oscar (2012) "El Programa Oportunidades y los pobres de México", ISBN 978-3-659-01631-8. Ed. EAE. Madrid, España.
- Hair, J., Anderson, R., Tatham, R. y Black, W. (1999). *Análisis Multivariante*. 5ª Edición. Prentice Hall.
- Jobson, J.D. (1992) *Applied Multivariate Data Analysis. Volume II: Categorical and Multivariate Methods*. Springer-Verlag.
- Martínez Abascal, E. y C. Cabezas (1997) "Correlación entre Bolsas Mundiales. ¿A quién sigue el IBEX?", *Análisis Financiero*, n.º. 73, 3er cuatrimestre, págs. 34-42.
- Michaud, R.O., G.L. Bergstrom, R.D. Frashure y B.K. Wolahan (1996), "Twenty years of International Equity Investing", *The Journal of Portfolio Management*, vol. 23, n. 1, págs. 9-22.
- Moltó, M.L. y E. Uriel (1997) "El problema del paro en la Unión Europea por sexos", *Papeles de Economía Española*, n.º. 72, págs. 122 a 136.
- Norusis, M.J., (1991), *The SPSS Guide to Data Analysis for SPSS/PC+*, 2nd Edition, SPSS Inc., Chicago.
- Martinez Arias, R. (2000). *El Análisis Multivariante en la Investigación Científica*. Cuadernos de Estadística. Editorial La Muralla.

Rute, A., 1997, "Workers or employers: Who is shaping wage inequality?", Oxford Bulletin of Economics and Statistics, vol. 59, nº 4, págs. 523-547.

Sharma, S. (1998). *Applied Multivariate Techniques*. John Wiley and Sons.

Uriel, E. (1995). *Análisis de Datos: Series temporales y Análisis Multivariante*. Colección Plan Nuevo. Editorial AC.

Hays W. (1973) *Statistics for the Social Sciences*. Holt, Rinehart & Winston, London, 1973.