LA REVISTA DE LA CARRERA DE ESTADISTICA

VARIANZA

Universidad Mayor de San Andres

Facultad de Ciencias Puras y Naturales

Instituto de Estadistica Teorica aplicada

Campus Cota Cota

AÑO 5

N° 5

Octubre 2006

REVISTA VARIANZA
Número 5, Año 05
Octubre del 2006

PRESENTACIÓN

El material de la quinta Versión de la revista "Varianza" publicación anual de la Carrera de Estadística de la Facultad de Ciencias Puras y Naturales, muestra el interés de docentes y estadísticos asociados a esta revista para hacer conocer el aporte ante la comunidad y por ende a la sociedad, temas de actualidad del ámbito estadístico.

Esperando que los lectores de la presente publicación nos honren con sus comentarios y aportes a los temas expuestos, felicito a los lectores como al personal que hace posible la presente edición.

Lic. Raúl Delgado Alvarez
JEFE CARRERA DE ESTADÍSTICA

CARRERA ESTADÍSTICA
INSTITUTO DE ESTADÍSTICA TEÓRICA Y APLICADA
FACULTAD DE CIENCIAS PURAS Y NATURALES
UNIVERSIDAD MAYOR DE SAN ANDRÉS

La Paz - Bolivia
Edificio Antiguo. Planta Baja PB 05
Telefax: 2442100

INDICE DE CONTENIDOS

¿QUÉ APORTAN LOS ORDENADORES A LA ENSEÑANZA Y APRENDIZAJE DE LA ESTADÍSTICA?

Juan D. Godino

1.-INTRODUCCIÓN

En la actualidad se reconoce la importancia de la enseñanza de la estadística en los diferentes centros educativos a los estudiantes sobre ideas probabilísticas y estadísticas y su mutua interdependencia.

La interpretación, representación y tratamiento de la información, tratamiento del azar en la formación profesional y en las universidades los cursos sobre estadística aplicada se incluyen prácticamente en todas las especialidades.

Este fenómeno no es exclusivo de los países desarrollados. La enseñanza de los contenidos referidos a la estadística y probabilidades se incrementa en los nuevos planes de estudio de diferentes países. Un indicador significativo del interés por la formación estadística a todos los niveles es el hecho de que el ultimo congreso internacional sobre enseñanza de la estadística, hubo un grupo de trabajo especial dedicado a "El currículo estadístico"

Este interés se explica por la importancia que la estadística ha alcanzado en nuestros días, tanto como cultura básica, como en el trabajo profesional y en la investigación, debido a la abundancia de información a la que el ciudadano, el técnico y el científico deben enfrentarse en su trabajo diario. El rápido desarrollo de la estadística y su difusión en los últimos años se ha debido a la influencia de los ordenadores, que también han contribuido a la acelerada cuantificación de nuestra sociedad y al modo en que los datos son recogidos y procesados.

Hasta hace pocos años, el análisis de datos reales estaba reservado a estadísticos profesionales, quienes debían escribir sus propios programas de ordenador para realizar los cálculos. Posteriormente, el uso de los paquetes potentes de análisis de datos requería el conocimiento de los comandos y sintaxis de los mismos. Esta situación, aparentemente, ha sido superada

Por un lado aparecen los entornos operativos" amistosos", que permiten acceder directamente al manejo de cualquier de los módulos de un paquete estadístico y, con la ayuda del "ratón", explorar sus posibilidades. Por otro lado, existen programas"de consulta" a los cuales se pude recurrir para obtener un "consejo" sobre el método de análisis que se debe aplicar en función del tipo de datos y nuestra hipótesis sobre los mismos. ¿Quiere esto decir que hemos resuelto definitivamente el problema de la estadística? ¿se debe reducir esta enseñanza a los alumnos el usos de este tipos de programas informáticos? Si no es así, ¿Cómo debemos reconsiderar los contenidos, objetivos y metodología de aprendizaje en función de las nuevas tecnologías?

En este trabajo discutimos estas cuestiones, aportando nuestra respuesta sobre las mismas: por un lado, la capacitación estadística incluye hoy en día el conocimiento del modo de procesar datos mediante un programa estadístico, por los que deberíamos, en la medida de los posible, ofrecer a nuestros alumnos un primer contacto con este tipo de programas.

Por otro, el ordenador no es sólo un recurso de cálculo, sino un potente útil didáctico, que nos permite conseguir una aproximación más exploratoria y significativa en la enseñanza de la estadística

2.- ORDENADORES Y OBJETIVOS EDUCATIVOS

Como hemos razonado, la creciente disponibilidad de programas de ordenador para el análisis de datos nos obliga a una reflexión sobre sus implicaciones en la enseñanza de esta materia.

En primer lugar, el ordenador puede y debe usarse en la enseñanza como instrumento de cálculo y representación gráfica para analizar datos recogidos por el alumno o proporcionados por el profesor. Nos enfrentamos a diario a la necesidad de recoger, organizar e interpretar sistemas complejos de datos y esta necesidad aumentará en el futuro, debido al desarrollo de los sistemas de comunicación y las bases de datos. Uno de los objetivos que debiera incluirse en un curso de estadística es capacitar al alumno para recoger, organizar, depurar, almacenar, representar y analizar sistemas de datos de complejidad accesibles para él. Este objetivo comienza por la comprensión de la idea básica de sistema de datos.

Este término es más adecuado que el conjunto de datos, para describir las estructuras de datos en las aplicaciones reales. Un conjunto no tiene por qué se ordenado, mientras que un sistema de datos ha de organizarse para poder ser procesado. Organizamos un sistema de datos al identificar en el mismo la unidad de análisis, las variables y las categorías de las mismas. Un conjunto no tienen elementos repetidos, mientras que una de las características de las variables de un sistema de datos es cada uno de sus valores se darán con una cierta frecuencia. No tendría ningún interés estadístico un sistema de datos en que todos sus elementos fuesen diferentes. Es precisamente las regularidades globales, dentro de la variabilidad individual el objeto de estudio de la estadística

En la mayor parte de los sistemas de datos hay al menos tres componentes: la descripción de las variables, los valores de la variable, que es el cuerpo principal de los datos, y los resúmenes estadísticos de cada variable. Los campos pueden ser de longitud fija o variables, y puede haber campos vacíos. Asimismo, clasificamos las variables según diversas tipologías cualitativas o cuantitativas; discretas, continuas; nominales, datos de intervalo, de razón.

Sobre cada una de estas componentes pueden realizarse operaciones o transformaciones internas (clasificación remodificación, agrupamiento) y externas (insertar, borrar, seleccionar...). Podemos clasificar variables, clasificar los casos dentro de una variable o clasificar los resúmenes estadísticos, por ejemplo, por su magnitud, podemos seleccionar casos por los valores de una variable, o seleccionar variables porque sus valores coinciden en una serie de casos. También es posible determinar relaciones entre estos componentes, por ejemplo, de dependencia, implicación, similaridad (dependencia entre variables; similaridad de sujetos; similaridad de variables,...Estos tipos de operaciones deben ser presentados para casos sencillos a los estudiantes, de modo que sean comprendidas. Aunque parezca muy simple, nuestra experiencia personal en el trabajo de análisis de datos no ha mostrado que la principal dificultad de muchos investigadores es precisamente el definir de una forma adecuada sus unidades de análisis y variables.

.El punto de comienzo de la estadística debería ser el encuentro de los alumnos con sistemas de datos reales: resultados deportivos de sus equipos favoritos, precios de las golosinas que compran en el recreo, medios de transporte usados para ir a sus estudios, temperaturas máximas t mínimas a lo largo de un mes; color o tipo de vehículo que pasa por delante de la ventana etc.

De este modo podrán ver que construir un sistema de datos propio y analizarlo no es lo mismo que resolver un problema de cálculo rutinario tomado de un libro de texto. Si quieren que el sistema de datos sea real, tendrán que buscar información cuando les falte, comprobar y depurar los errores que comenten el recoger los datos, añadir nueva información a la base de datos cuando se tenga disponible, aprenderán a comprender y apreciar más el trabajo de los que realizan las estadísticas para el gobierno y los medios de comunicación. Si comprenden la importancia de la información fiable, se mostrarán más dispuestos a colaborar cuando se les solicite colaboración en encuestas y censos.

Estos sistemas de datos pueden se la base de trabajo interdisciplinares en geografía, ciencias sociales, historia, deportes, etc.

En el caso de que los datos se tomen de los resultados de experimentos aleatorios realizados en la clase, estaremos integrando el estudio de la estadística y probabilidad. Una vez construido un sistema de datos el siguiente paso sería analizar con ayuda del ordenador. El manejo de un paquete es un objetivo importante ya que, en la actualidad, el uso de las técnicas estadísticas está ligado a los ordenadores. Un problema tradicional en la enseñanza de la estadística ha sido la existencia de un desfase entre la comprensión de los conceptos y los medios técnicos de cálculo para poder aplicarlos. La solución de los problemas dependía en gran medida de la habilidad de cálculo de los usuarios, que con frecuencia no tenían una formación específica en matemáticas.

Hoy en día la existencia de programas fácilmente manejables permite salvar este desfase. Esta mayor facilidad actual de empleo de procedimientos estadísticos, implica, sin embargo, el peligro de los usos no adecuado de la estadística. En el trabajo de consultoría estadística no es difícil encontrar a investigadores que, habiendo recogido un conjunto de datos sin ningún tipo de consulta con un estadístico profesional en la etapa del diseño de la investigación, piensan que el análisis consiste simplemente e la elección de un programa adecuado que automáticamente dará una interpretación a sus investigaciones.

Acostumbremos, pues, a los alumnos a planificar el análisis que quieren realizar incluso antes de finalizar la construcción de su sistema de datos. Si por ejemplo, quieren hacer un estudio en su instituto para comparar la intención de voto de jóvenes y señoritas en las próximas elecciones al centro de estudiantes, deben recoger una muestra suficientemente representativa de ambos sexos en los diferentes cursos y deben recoger datos sobre las principales variables que influyen en esta intención de voto.

De otro modo, sus conclusiones pudieran estar sesgadas o ser poco explicativas. Debemos también hacer conscientes a los alumnos de que un mismo problema estadístico puede se resuelto por diferentes procedimientos y las respuestas que se obtienen pueden ser complementarias y a veces poco adecuadas. No todos los procedimientos estadísticos se adaptan bien para todos los problemas. Por ejemplo, la media aritmética no sería un representante adecuado de un conjunto de datos bimodal o con valores atípicos mus acusados.

Finalmente está el problema de la interpretación de los resultados y la generación de hipótesis sobre el problema investigado a partir de los resultados de los análisis. Esta cuestión se incluye dentro de la filosofía del análisis exploratorio de datos.

3.-ANALISIS EXPLORATORIO DE DATOS

Los ordenadores actuales, con sus posibilidades interactivas, favorecen la introducción, desde los primeros niveles de enseñanza, de una nueva filosofía en los estudios estadísticos introducida por Tukey (19779: el análisis exploratorio de datos.
Esta filosofía no solo se aplica a nivel de estadística elemental. Por el contrario, muchos de los métodos de análisis de datos multivariados son empleados en la actualidad dentro de esta misma filosofía, para analizar fenómenos físicos o sociales complejos.

El análisis de datos tradicional tenía como función principal la confirmación de hipótesis, que se establecían de antemano a al recogida de datos. Estaba basado fundamentalmente en el cálculo de estadísticos en muestras recogidas con el fin de poner las hipótesis a prueba y su comparación con los valores supuestos para los parámetros de la población.

La representación gráfica de los datos no era muy importante, porque la mayor parte de las veces se suponía que los datos provenián de una distribución normal o aproximadamente normal.

En el análisis exploratorio de datos se concede una importancia similar a los dos componentes que componen los datos estadísticos: la “regularidad” o “tendencia” y las “desviaciones” o “variabilidad”. La regularidad es la estructura simplificada del conjunto de observaciones: la media o mediana en su distribución, la línea de regresión en una nube de puntos, Las diferencias de los datos con respecto a esta estructura (diferencia respecto a la madia, respecto a la línea de regresión, etc., son las desviaciones o residuos de los datos.

En la inferencia clásica se supone que estas desviaciones siguen un patrón aleatorio. De acuerdo con la teoría de errores, la distribución de estos residuos sería normal con media cero.

El estudio se centraba en buscar un modelo, dentro de una colección dada, para expresar la regularidad de las observaciones. Por ejemplo, en un estudio de regresión lineal se trata de elegir la recta que representa lo mejor posible la nube de puntos. Asimismo, se definen unos ciertos coeficientes cuyo fin es probar la “bondad” de ajuste del modelo mediante un contraste de hipótesis, en este caso, sobre el coeficiente de correlación

Por el contrario, en el análisis exploratorio de datos se concede una importancia similar a los dos componentes de los datos que hemos citado. En lugar de imponer un modelo a las observaciones, se genera dicho modelo desde las mismas.

Por ejemplo, cuando se estudian las relaciones entre dos variables, el investigador no solamente necesita ajustar los puntos a una línea, sino que estudia los estadísticos, compara la línea con los residuos, estudia la significación estadística de la razón de correlación u otros parámetros para descubrir si la relación entre las variables se debe o no al azar. Si se piensa que es posible extraer nueva información de los residuos, se rea analizan éstos, tratando de relacionarlos con otras variables

Un punto importante en el análisis exploratorio de datos es que no se trata de un conjunto de métodos aunque se han creado algunas técnicas, especialmente gráficas, asociadas a ella -sino de una filosofía de aplicación de la estadística. Esto lo podemos ver en el ejemplo anterior en el que una misma técnica la regresión lineal la podemos aplicar con enfoque exploratorio o confirmatorio. Esta filosofía consiste en el estudio de los datos desde todas las perspectivas y con todas las nuevas, en el sentido de conjeturar sobre las observaciones de las que disponemos. Como contrapartida, tales “hipótesis” no quedan contrastadas en el sentido estadístico del término al finalizar el análisis, por lo que sería preciso la toma de nuevos datos (una replicación) sobre fenómenos y efectuar sobre ellos un análisis estadísticos confirmatorio con el fina de contrastarlas.

En la educación secundaria, la enseñanza de la estadística descriptiva debe hacerse desde la perspectiva del análisis exploratorio de datos. A titulo de ejemplo, describimos, a continuación, una situación problemática que podría proponerse a los alumnos para que tengan la oportunidad de aplicar esta filosofía y aprender de las técnicas gráficas asociadas al análisis exploratorio de datos.

Ejemplo ¿Cuáles son las principales diferencias entre los chicos y chicas en la clase de educación física?

Un proyecto fácil de plantear a los alumnos en clase consiste en proponerles la investigación de la influencia del sexo en la clase de educación física. Una tarea similar puede proponerse eligiendo otro tema, como los rasgos físicos, las preferencias

sobre empleo del tiempo libre, el tipo de música a lectura que prefieren, etc. En el ejemplo podemos recoger datos sobre rendimiento en pruebas de salto de altura y longitud, flexibilidad, número de abdominales por minuto, tiempo en recorrer 100 metros y 500 metros, etc. Estas marcas podrían recogerse a principios y finales de un año, con el objeto de comprobar cuál de los dos sexos han mejorado más sus marcas iniciales. También podríamos recoger datos de algunas variables físicas como el peso y la altura. Si los alumnos tienen acceso a los usos de un paquete de programas para el análisis de datos el profesor puede proponer la grabación de un fichero con estas variables y la realización de los análisis estadísticos necesarios para dar respuesta a las siguientes interrogantes.

a) ¿Existe diferencias entre hombres y mujeres en las características estadísticas de alguna de estas variables? ¿en cuál es esta diferencia más o menos acusada?

b) ¿Cuáles de estas características son independientes del sexo y en cuáles existe asociación estadística con dicha variable? ¿Cuál es la intensidad de la asociación?

Una respuesta concluyente a estas preguntas no es posible darla en el marco de una actividad estudiantil en cursos de iniciación a la estadística, ya que precisaría la recogida de datos representativos de la población y la aplicación de procedimientos de contraste de tipo confirmatorio.

Pero sí es posible e instructivo que los alumnos traten de formular hipótesis plausibles apoyadas en los datos empíricos disponibles y utilizando diversas técnicas de análisis exploratorio de datos. A titulo de ejemplo citamos algunos procedimientos que los estudiantes puedan aplicar usando un paquete de programas y que el profesor puede sugerir.

Un primer nivel de análisis se referirá al estudio de las distribuciones de frecuencias y resúmenes estadísticos de cada una de las variables para todo el conjunto de datos. Esto puede hacerse mediante opciones del paquete que permita calcular los estadísticos elementales (promedios, dispersión y medida de simetría) y las tablas y gráficas de distribución de frecuencias como histogramas y polígonos de frecuencia, gráficos del “tronco” o la “caja”. La representación gráfica de estas distribuciones, con la posible presencia de varias modas y de asimetría mas o menos acusada, nos ayudará también a decidir, para cada variable, cuál es el promedio media, mediana o moda y medida de dispersión más adecuada como base para la comparación.

El disponer de ordenadores hace posible una actitud más crítica y analítica hacia los datos recogidos. Si se detectan algunos valores aislados, que se desvían demasiado de los restantes, podemos optar, tras una indagación de las posibles causas, por suprimirlos y repetir el estudio uní variante. Asimismo, la selección de una amplitud de intervalos de clase en los histogramas, que se adapte bien a los datos, pueden hacerse probando con distintos valores y viendo su efecto en la pantalla.

También se pueden comparar las características de estas distribuciones con otras disponibles, procedentes de tablas de medidas físicas correspondientes a la, población de la cual se ha extraído la muestra, en nuestro caso de una población de jóvenes de edades similares a la de nuestros estudiantes, para juzgar la posible representatividad de la misma

Comparación de subpoblaciones La opción de selección de casos permite hacer los análisis estadísticos para distintas submuestras, en nuestro caso, por ejemplo, para hombres y mujeres. Los que hace posible la comparación de los estadísticos y las distribuciones de frecuencias. Los “gráficos de cajas paralelas” permiten una

comparación simultánea y visual de los valores de las medianas, cuarteles, recorrido intercuartílicos, asimetría y valore atípicos.

En algunos casos, las diferencias entre grupos serán menores para otras variables, llegando incluso el caso en que tengamos duda en la interpretación. Esto nos llevará a dos cuestiones: por un lado, nos interesará dar un coeficiente o medida de la intensidad de la relación entre variables, lo que nos conduce a la idea de asociación. Además, necesitaríamos la herramienta del contraste estadístico de hipótesis, para poder tomar una decisión sobre la existencia de diferencias significativas en las poblaciones. La introducción de estas técnicas para alumnos podría ser motivada con este ejemplo, en las poblaciones. La introducción de estas técnicas para alumnos mayores también podría ser motivada con este ejemplo

Estudio de la asociación estadística: tablas de contingencia, medidas de asociación y regresión el estudio de las tablas de contingencia, de las variables agrupadas en intervalo, frente al sexo, nos permite comparar las distribuciones condicionales y la diferencia entre frecuencias observadas y esperadas y estudiar los conceptos de asociación e independencia entre variables estadísticas.

Además, puede ser necesario el estudio de la nube de puntos de pares de variables tales como tiempo en recorrer 100 y 500 metros, salto de longitud y altura, que nos da una primera visión gráfica de la existencia o no de asociación entre variables numéricas y de la intensidad de la misma. El estudio del coeficiente de correlación junto con el de la recta de regresión puede precisar aún más esta idea de asociación.

Puesto que algunas de las variables que queríamos relacionar con el sexo están relacionadas entre sí, surgirá la idea de hasta qué punto la asociación detectada entre, por ejemplo, el salto de longitud y la altura depende del sexo. Podemos discutir aquí la idea de cuando una variable debe o no jugar el papel de dependiente o independiente, o si sus papeles son intercambiables. También se verá la necesidad de buscar modelos más complejos que incluyan las relaciones entre efectuado, servirá como base intuitiva para la comprensión posterior de estos conceptos

4.- PROGRAMAS UTILIZABLES EN LA ENSEÑANZA

Actualmente existe una gran variedad de programas estadísticos, tanto los de tipo profesional, como los desarrollados especialmente con fines educativos. A continuación describimos brevement e los principales tipos de sftware utilizable:

Paquetes estadísticos profesionales como B.M.D.P., S.P.S.S. , SYSTAT, SATVIEW, STATGRAPHICS, especialmente las versiones para entorno Windows o Mac Intosh, que no requieren el aprendizaje de los comandos, también los desarrollados especialmente para ser usados en la enseñanza, como MINÍTAB. La principal finalidad es el Cálculo y representación gráfica. Son también un recurso profesional y permiten el aprendizaje a diversos niveles de complejidad.

Hojas electrónicas, disponible en diferentes paquetes integrados. Aunque más incompletas, permiten comprender los algoritmos de cálculo y pueden servir para otras aplicaciones diferentes de la estadística, concretamente el Microsoft Office Excel, en la tabla de herramientas tiene el comando Análisis de Datos.

Software didáctico para fines especiales, como los siguientes: Statlab (Inferencia), Gasp (Procesos Estocásticos) Tabletop(Exploración de contextos multivariantes para alumnos muy jóvenes), Quercus(Curs o de autoaprendizaje de bioestadística)

5.- Bibliografía

Batanero, C., Estepa, A. y Godino, J. D. (1991a). Análisis exploratorio de datos: Sus posibilidades en la enseñanza secundaria. Suma, 9, pp. 25-31.

Batanero, C., Godino, J. D. y Estepa, A. (1991b). Laboratorio de estadística. Uso del paquete de programas PRODEST. Granada: Dpto de Didáctica de la Matemática. Universidad de Granada.

Estepa, A. (1994). Concepciones iniciales sobre la asociación estadística y su evolución como consecuencia de una enseñanza basada en el uso de ordenadores. Tesis Doctoral. Universidad de Granada.

Godino, J. D. y Batanero, C. (1993). Ordenadores y enseñanza de la estadística. Revista de Educación de la Universidad de Granada , 7, 173-186.

Godino, J. D. y Batanero, C. (1994). Enfoque exploratorio en el análisis multivariante de los datos educativos. Epsilon, 10 (2), pp. 11-22.

Batanero (De.): Research on the Role of Technology in Teaching and Learning Statistics, 1996 IASE Round Table Conferenc Papers (pp. 255-264). Universidad de Granada.

APLICACIONES DE MODELOS DE SUPERPOBLACIÓN A LA ASIGNACIÓN ÓPTIMA

Rubén Belmonte Coloma

Un método muy conocido en el muestreo de poblaciones finitas estratificado para asignar tamaños de muestra a los estratos es el método de Neyman. Este método esta destinado a poblaciones fijas, en la actualidad existe muchos estudios sobre muestreo en superpoblaciones, en este artículo se encara este método.

1.- MODELO SUPERPOBLACIONAL

El uso de modelos superpoblacionales se remonta a trabajos desarrollados por los fundadores de la Teoría del Muestreo. Cochran (1939) la utilizó, Godambe (1955) recomendó su uso para hacer inferencias al considerar que Y es generado por un mecanismo aleatorio denominado superpoblación. Este enfoque es considerado como parcialmente Bayesiano pues a pesar de utilizar como principio el inferir a partir de una distribución a priori no busca la minimización del riesgo aposteriori. Un modelo popular es el dado por asumir que una variable conocida X permite establecer la familia a la que pertenece la distribución de Y que a su vez es caracterizada por un modelo superpoblacional. El modelo para la estratificación es:

$$Y_{hj} = \alpha_{h0} + \alpha_{h1} X_{hj} + \varepsilon_{hj}, j = 1, \dots, N_h, h = 1, \dots, H$$

Donde $E_{\mathfrak{P}}(\varepsilon_{hj}) = 0$

$$Cov_{\mathfrak{P}}(\varepsilon_{hj}, \varepsilon_{h'j'}) = \begin{cases} \sigma_h^2 g(X_{hj}) & si \quad h \neq h' \quad y \quad j \neq j' \\ \rho_{\prime h} \sigma_h^2 \sqrt{g(X_{hj}) g(X_{hj'})} & si \quad h = h' \quad y \quad j \neq j' \\ 0 \; en \; otro \; caso \end{cases}$$

α_{ho} α_h y g (X_{hj})son estrictamente positivos y ρ es el coeficiente de correlación entre X y Y.
Simplificaremos este modelo haciendo $\rho_h = \alpha_{ho}=0$ y $g(X_{hj}) = 1$. Esto permite modelar
problemas en los que se espera una expansión de Y como función de X.
En estudios del crecimiento en que X mide los resultados 'antes' y la Y 'después' por ejemplo.
Se espera suavizar la restricción de insesgamiento de mu por la de que esta se cumpla respecto al modelo. Entonces tendremos el problema de optimización

$$Min \left\{ \sum_{h=1}^{H} \beta_h^2 \frac{\sigma_h^2}{n_h} \middle| \sum_{h=1}^{H} \beta_h \mu_{X(h)} - \mu_X = 0, \; \beta_h > 0, h = 1, \dots, H \right\}$$

Los pesos óptimos se definen convenientemente a partir de los resultados de X y

$$\beta_h^* = \frac{Q_h^*}{\sum_{h=1}^{H} Q_h^*}$$

donde

$$Q_h^* = \frac{n_h \mu_{X(h)} / \sigma_h^2}{\sum_{h=1}^{H} n_h \mu_{X(h)} / \sigma_h^2}$$

y el error está dado por

$$V(m_\beta) = \sum \frac{\beta_h^{*2}\sigma_h^2}{n_h}$$

Si se acepta que E(A/B)≡E(A)/E(B) (linearización de Taylor) tendremos que

$E_{\mathfrak{P}}(Q_h) \cong Q_h^*$ y $E_{\mathfrak{P}}(\beta_h) \cong \beta_h^*$

De ahí que

$$E(E(m_\beta)) \cong E\left(\sum_{h=1}^{H} \beta_h \mu_h\right) \cong \sum_{h=1}^{H} \beta_h^* \mu_h$$

por lo que

$$\sum_{h=1}^{H} \beta_h^* m_h$$

es aproximadamente insesgado

2.- COMPARACIÓN DE LOS PESOS OPTIMOS

En muchas aplicaciones los estratos están definidos a partir de intervalos de una variable continua. Tal es el caso cuando se usan variables como el área de parcelas, los ingresos de las familias, etc. En ellos es muy común que la media y la varianza crezcan en forma no proporcional y que una medida relativa sea más adecuada para medir la precisión de los estimadores.

Población fija

Analizamos algunos ejemplos de los libros de texto para fijar el comportamiento de este método. En algunso casos lo datos considerados como meustrales son de una población artificial.

Se analiza la ganancia en precisión debida al uso del método propuesto respecto al uso de la Asignación de Neymann, y la proporcional $n_h^p = nW_h$. Para cualcular las ponderaciones óptimas son usadas n_h^0 y n_h^P

Ejemplo

Tabla 1 Parámetros del Ejemplo

h	w_h	s_h^2	c_h	n_h^0	μ_h	Q_h^0	β_h^0
1	0,5	3,2	1	2	5	3,1	0,6
2	0,5	14,8	4	2	15	2,39	0,4

Tabla 2 Eficiencia de las Ponderaciones Óptimas en el Ejemplo

Criterio	Varianza	V(mβ)/Varianza
Asignación de Neymann	1,50	1,50
Asignación Proporcional	1,50	1,50
Ponderación Optima	1,76	1,00

En este caso el suso del método prouesto de las ponderaciones óptimas (PO) tiene un comportamiento peor que la asignación de Neymann (AN) y que el de la proporcional (AP). Si tuviéramos que μ1=15 y μ2=5 los resultados cambian y por tanto son otros los valores de los parámetros.

Tabla 3 . Parámetros al variar los valores de la media en el Ejemplo.

H	W_h	s_h^2	μ_h	Q_h^0	β_h^0
1	0,5	3,2	15	9,40	0,92
2	0,5	14,8	5	0,78	0,08

Ahora hay una mayor eficiencia al tener una relación diferente entre varianza y media: el método es mejor que la AP.

Tabla 1.4. Eficiencia del método de las proporciones óptimas con los parámetros variados para el Ejemplo:

Criterio	Varianza	$V(m_\beta)$/Varianza
Asignación de Neymann	1,50	1,17
Asignación Proporcional	2,75	0,64
Ponderación Optima	1,40	1,00

BIBLIOGRAFIA

ALLENDE,S. and C. BOUZA (1993): "Stochastic Programming approaches for the estiamtion of the mean in stratified populations", Inv. Operacional, **13,109-118.**

COCHRAN, W. (1977) : **Sampling Techniques,** Willey, New York.

Índice

DISTRIBUCIÓN DE ITEMS VS. CALIDAD

David Barrera

1.-INTRODUCCION

En los momentos de profunda crisis en al educación en todos sus niveles, de grandes desgarros sociales producto de la desigualdad existente, de pérdida de identidad y del atraso científico, todos coinciden en que la solución reside en la educación.

Educación para la participación de los ciudadanos en la toma de decisiones, para consolidar un sistema basado en la interacción, en el diálogo, tanto a escala local, municipal, departamental, como supranacional. La ciudadanía plena implica no ser espectador y receptor solamente sino figurar, desempeñando cada uno su papel adecuadamente, en el escenario de la comunidad o municipio a la que pertenecemos.

En definitiva para ser más libres o menos oprimidos, como se quiera, la clave sigue siendo la educación. Ante las demandas de mayor atención, se escucha con frecuencia tanto de autoridades departamentales, como sindicales, donde se da por PROBADA la hipótesis que a un mayor número de profesores es sinónimo de mejor educación, suena muy ligero, a simple vista hasta parece correcto el razonamiento. El gran defecto de éste planteamiento, es que muy poco o casi nada se habla de calidad, es que la lógica no es lineal, es decir que mayor número de profesores no implica mejor calidad, como tampoco a menor número de profesores no necesariamente implica menor calidad.

El Presente trabajo es un trabajo empírico, es decir no existe el diseño de investigación formal, sin embargo con la información disponible que existe en el Ministerio de Educación, intenta mostrar que existen herramientas estadísticas que nos permitirían optimizar los escasos recursos destinados a la educación.

¿Cómo medir la calidad? Primero tendríamos la necesidad de dar una definición de calidad, en pedagogía existe una infinidad de trabajos al respecto, sin embargo en el país no se existe muchos avances al respecto, se hizo algo con los datos de SIMECAL, digo algo porqué es posible realizar análisis muchos más profundos de la realidad de nuestros educandos con pruebas objetivas, sin embargo existe un celo exagerado de los amos y dueños de los datos, es algo que no se entiende, tendrán una buena razón y/o justificación.

En el presente trabajo, en esta oportunidad utilizó como una variable proxy del indicador de la calidad, los años de escolaridad en los 314 municipios.

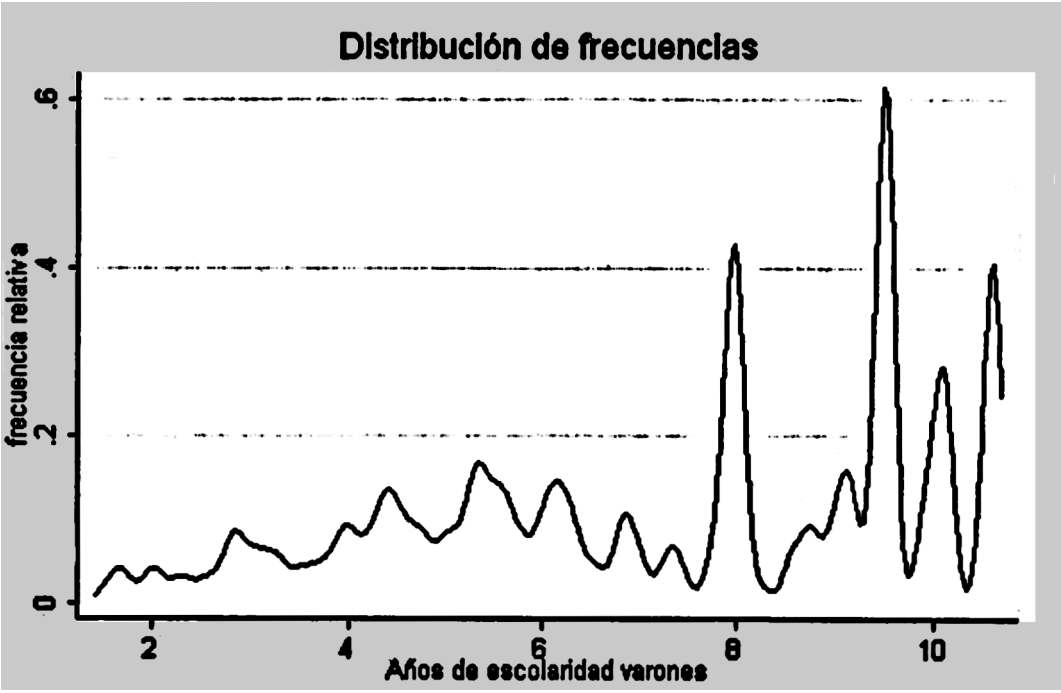
MODELO ECONOMETRICO

Para definir de manera analítica el modelo de **inversión en capital humano**, se plantea la siguiente relación teórica:

$$Y_i = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon_i$$

Donde Yi es una variable observable que representa los años de escolaridad en promedio en el i-ésimo municipio, asumiendo un criterio legítimo de maximización de los años de escolaridad; **X1** representa la proporción de profesores normalistas(normalista); **X2** representa la proporción de profesores interinos(interino); **X3** recoge la proporción de personas que sólo hablan castellano (solocast); **X4** incluye a la proporción de personas que sólo hablan lenguaje castellano y nativo (castnat) y **X5** el pib per cápita. Donde **?¡** es la disturbancia, es decir aquí están representadas el resto de las variables que tienen efecto sobre Yi, que sin embargo no existen o si existen no pueden ser medidas por muchas razones; esta se distribuye de manera aleatoria entre los municipios.

2.-DISTRIBUCIÓN DE LOS AÑOS DE ESCOLARIDAD



Este gráfico(KERNEL) que es una versión mejorada del histograma, muestra que asumir que los años de escolaridad provengan de una distribución normal es muy remoto, además es notable la dispersión que existe, en decir para minimizar la heterocedasticidad una de las alternativas es particionarla por género.

3.-¿PORQUÉ EXPLICAR LOS AÑOS DE ESCOLARIDAD POR GENERO?
PRUEBA DE DIFERENCIA DE MEDIAS APAREJADA

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
escoh	314	6.08051	.0975903	1.729304	5.888494	6.272525
escom	314	4.198662	.0978835	1.734499	4.00607	4.391255
diff	314	1.881847	.045124	.7995992	1.793062	1.970632

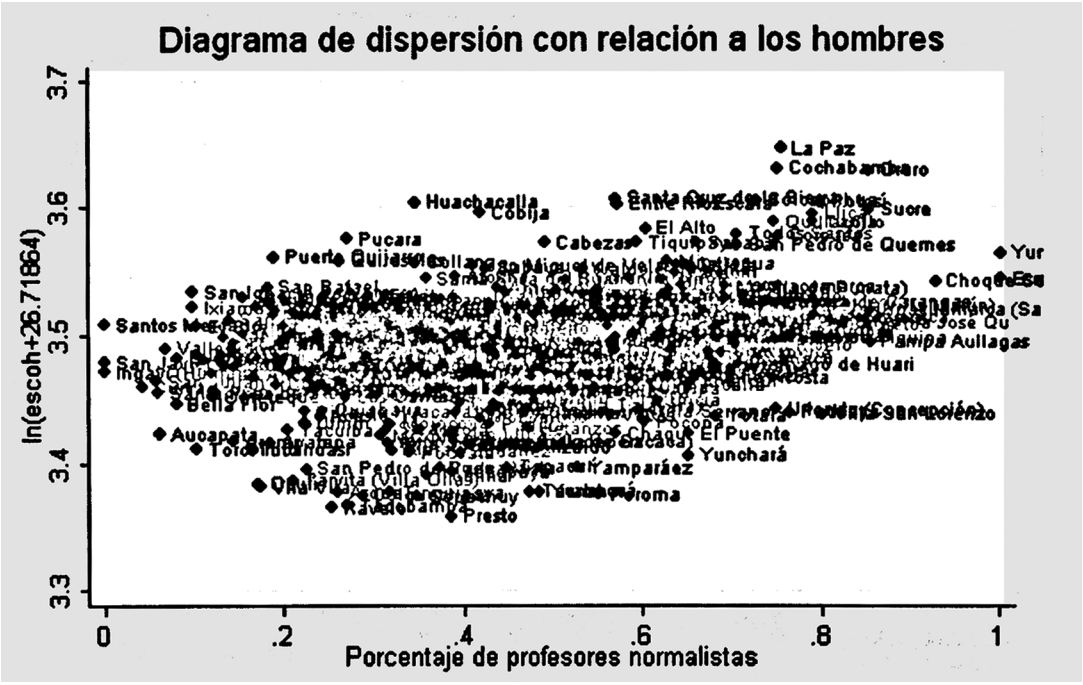
mean(diff) = mean(escoh - escom) t = 41.7039
Ho: mean(diff) = 0 degrees of freedom = 313

Ha: mean(diff) < 0 Ha: mean(diff) != 0 Ha: mean(diff) > 0
Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 0.0000

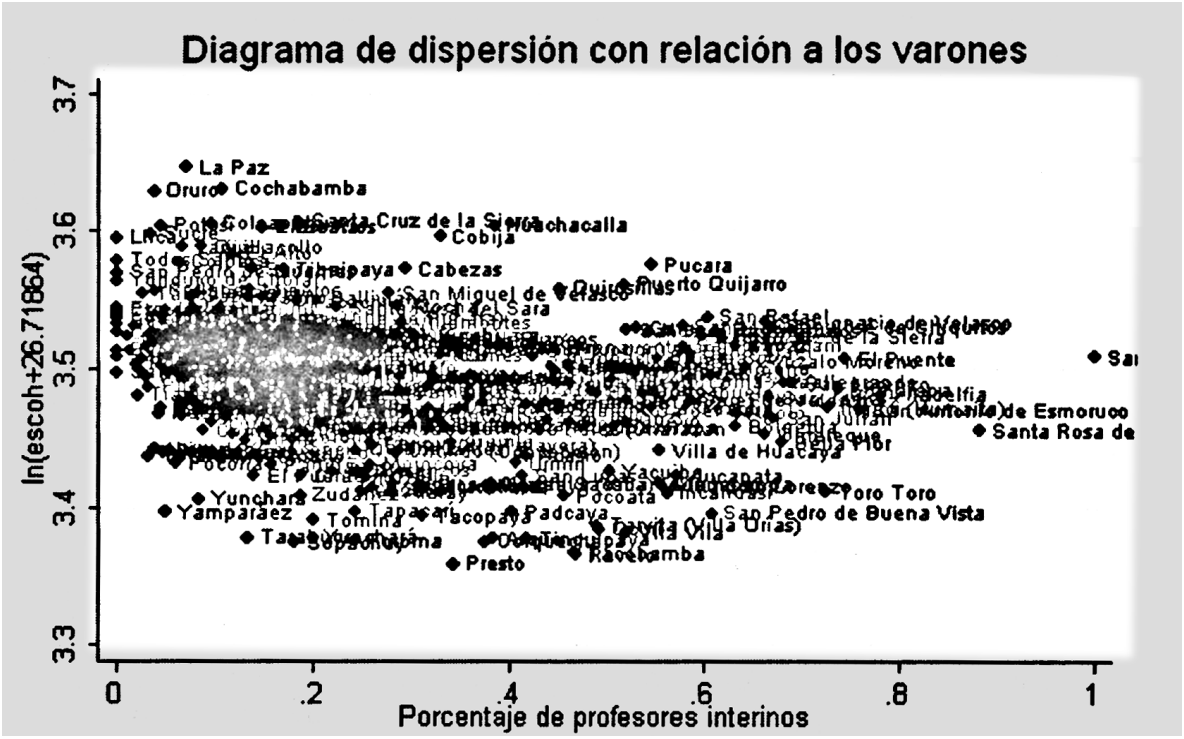
Los resultados de la prueba señalan que efectivamente los años de escolaridad en promedio son diferentes de los hombres con respecto a los años de escolaridad de las mujeres.

¿ALGUNAS VARIABLES EXOGENAS?

DIAGRAMA DE DISPERSIÓN
Box-Cox de los años de escolaridad vs. normalistas
Varones

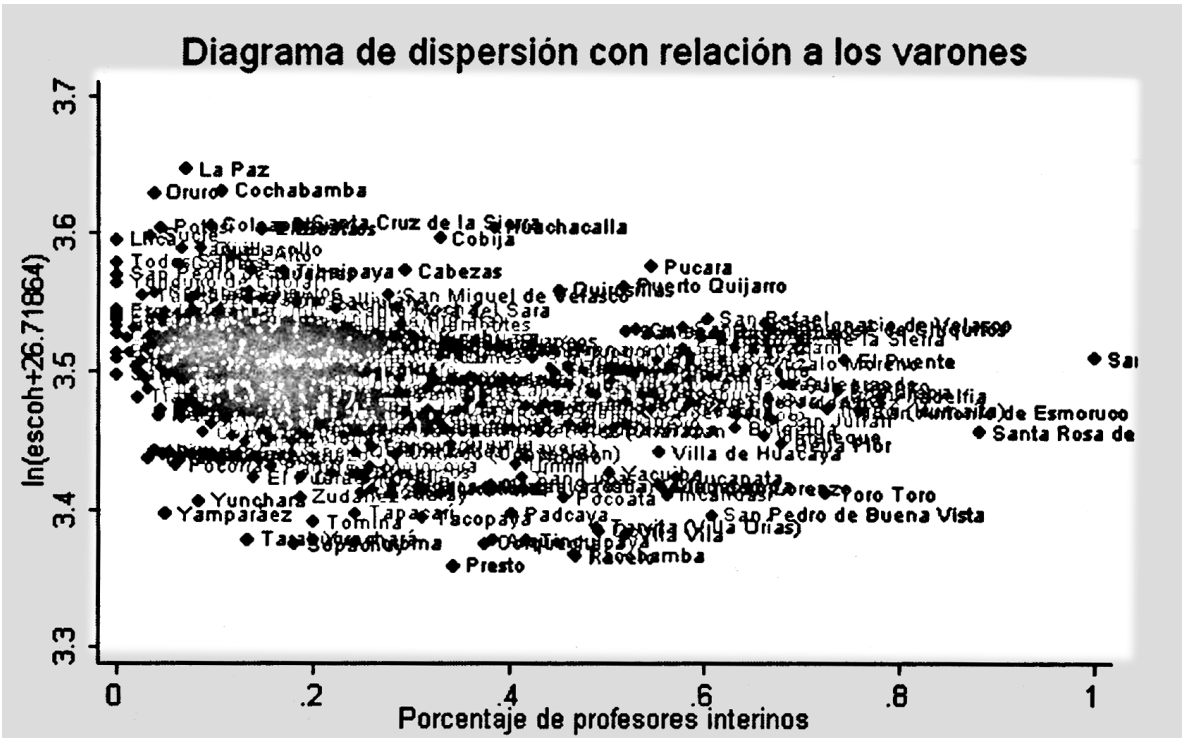


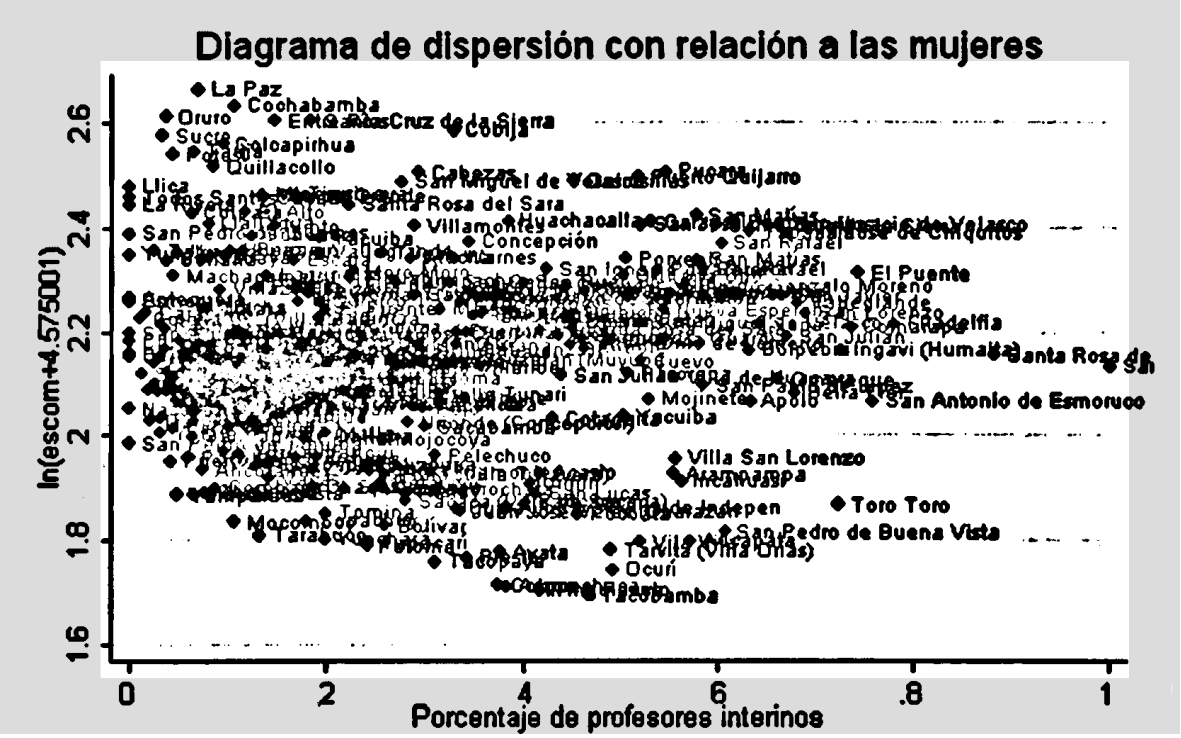
Mujeres



El gráfico de los varones, señala la marcada realción de los años de escolaridad respecto a la proporción de profesores normalistas, en tanto de las mujeres no es tan marcado, si bien existe pero no es visible.

Box-Cox de años de escolaridad vs. interinos
Varones





Los gráficos muestra nuevamente la tendencia lineal entre la proporción de profesores interinos y los años de escolaridad; es notable la dispersión de los años de escolaridad de las mujeres.

Nota.- Es importante indicar que para el kernel, el diagrama de dispersión, como para el modelo estimado, se ponderó por el número de habitantes en edad escolar, esto representa el peso de cada municipio.

Con seguridad que existe otras variables exógenas, por falta de espacio señalaremos estos mediante la correlación parcial.

CORRELACION PARCIAL

	VARONES		MUJERES	
Regresores	Corr.	Prob.	Corr.	Prob.
Normalista	0.374	0.000	0.374	0.000
Interno	0.199	0.000	0.263	0.000
Solocast	0.718	0.000	0.795	0.000
Castnat	0.665	0.000	0.658	0.000
Lpibperca	0.528	0.000	0.596	0.000

Los regresores señalados, muestran claramente que los regresores son significativos, tanto para los años de escolaridad de los varones y de las mujeres. Es decir son candidatos para ser tomados en cuenta como regresores.

VALORES INFLACTORES DE LA VARIANZA

REGRESORES	VIF	Indice de Condición
NORMALISTA	5.500	0.182
INTERINO	4.940	0.202
SOLOCAST	4.040	0.247
CASTNAT	3.420	0.293
LPIBPERCA	2.230	0.448

Los valores inflactores de la varianza y los índices de condición, reflejan claramente que no existe problemas de cuasi-multicolinealidad.

ESTIMACIÓN DEL MODELO

	Varones			Mujeres		
Lnescoh	Coef.	Std. Err.	P>t	Coef.	Std. Err.	P>t
Normalista	0.146	0.032	0.000	0.426	0.087	0.000
Interino	0.082	0.037	0.027	0.313	0.102	0.002
Solocast	0.269	0.059	0.000	1.111	0.148	0.000
solocast2	-0.054	0.068	0.428	-0.366	0.177	0.040
Castnat	0.272	0.057	0.000	0.714	0.148	0.000
lpibperca	0.047	0.013	0.000	0.164	0.033	0.000

Con un nivel de significación del 5%, casi todas las variables en ambos modelos son significativas.

ELASTICIDADES

	Hombres	**Mujeres**
Regresores	Ey/ex	Eey/ex
Normalista	0.022	0.099
Interino	0.005	0.026
Solocast	0.031	0.195
solocast2	-0.004	-0.037
Castnat	0.024	0.096
Lpibperca	0.095	0.495

4.- Conclusiones.-

1.- El efecto de la proporción de los profesores normalistas es desproporcionado el de los profesores interinos; en los años de escolaridad de los hombres como de las mujeres. Este resultado es importante, que a pesar de las deficiencias de las normales, un normalista en promedio es mejor que un interino.

2.- El efecto de la proporción de personas que sólo hablan castellano en los años de educación en los hombres es lineal, en tanto en el de las mujeres no es lineal. Este factor es importante, cuando hay propuestas de implementar como materia un idioma nativo, luego un idioma extranjero, que no es lo mismo que una educación bilingüe.

3.- En tanto en los municipios donde se habla castellano y un idioma nativo el efecto es casi la misma tanto en los hombres como en las mujeres; El efecto sin embargo tiene mayor impacto que los municipios donde se habla solo castellano. Esto parece sugerir que para más años de escolaridad sería más conveniente castellano e idioma nativo, pero no tanto como asignatura; sugiere analizar con más cuidado antes de implementar reformas que a la larga no tengan el efecto esperado.

4.- El pib per cápita, definitivamente era de esperarse que tenga su efecto en los años de escolaridad. Si se incrementa en un 1% el pib per capita, en promedio los años de escolaridad mejora en un 50%.

5.- Sería interesante que los planteamientos para una reforma estructural se realicen con análisis objetivos, es decir con fundamento estadístico. Porque de lo contrario existe se corre el riesgo de invertir muchos dinero con poco efecto.

Sugerencias.-

1.- Sería más adecuado un estudio con modelos econométricos espaciales por la inminente correlación espacial existente entre un municipio y sus vecinos, en trabajos posteriores trabajos es probable éste análisis.

2.- También sería importante abordar la calidad de la enseñanza del conjunto de municipios y las materias impartidas, como por ejemplo con modelos jerárquicos, es una técnica muy interesante.

5.- Bibliografía

1.- Análisis econométrico, William Greene, edit Prenticel Hall, año 2002
2.- Introducción a la Econometría, G.S. Maddala, edit Prentice Hall, año 1996

La Estadística en Ciencias de la Salud

Edgar Osorio Brun

1.- INTRODUCCION

En el estudio de la Epidemiología se encuentran las medidas de asociación e impacto de variables que cuantifican la relación existente entre las variables. La estadística contribuye con los indicadores de asociación e intervalos de confianza para garantizar la confiabilidad y validez del estudio clínico, pero para una mejor comprensión del cálculo de indicadores es necesario conocer algunos conceptos importantes que nos guía a la aplicación de la teoría estadística en Ciencias de la salud.

2.- Estudios de cohortes

Los estudios de cohorte son diseños de observación con seguimiento y sentido hacia delante, partiendo de la exposición se estudia el efecto (una enfermedad, una medición como el colesterol). Son estudios analíticos, que comprueban hipótesis de asociación.

Medidas a las que dan origen los estudios de cohorte

Un estudio de cohorte permite obtener información sobre incidencia y a partir de ésta, indicadores de riesgos absoluto y relativo.

Tipos de cohorte

Se denomina **cohorte cerrada** a aquella cuyos miembros son reclutados en el mismo periodo de tiempo y a la cual no ingresan personas durante el periodo de seguimiento. En consecuencia, en esta modalidad el total de miembros de la cohorte tiene periodos de seguimiento que comienzan al mismo tiempo.

Cohorte abierta o dinámica es aquella en la cual sus integrantes pueden ingresar a seguimiento en diferente momento durante el periodo que este dure. Por tanto, los miembros de esta cohorte pueden tener tiempos de exposición heterogéneos

3.- Incidencia

El seguimiento de individuos sanos por un período determinado de tiempo permite medir el número de casos de una enfermedad que aparecen en dicho período. Esta cifra constituye la tasa de incidencia de la enfermedad en estudio que puede ser medida para la cohorte expuesta (Ti exp), la no expuesta (Ti noexp) y para ambas en conjunto (Ti).

La incidencia acumulada se calcula considerando todos los sujetos que presentaron el outcome en estudio independientemente del momento en el cual lo presentaron (*cumulative risk*). Su cálculo se aplica cuando se trata de una cohorte cerrada.

Para el caso particular de un diseño de cohorte en que se permita eliminar o ingresar individuos a las cohortes después de haber iniciado el seguimiento (cohortes abiertas) seguimiento), se prefiere el término densidad de incidencia. (*incidence rate*)

La densidad de incidencia suma todos los tiempos con que efectivamente contribuyeron los individuos estudiados. El indicador se construye dividiendo el total de enfermos encontrados a lo largo del estudio por el total del tiempo de seguimiento (tiempo - persona) y amplificando según corresponda.

4.- Riesgo

El cálculo de incidencia de la enfermedad en expuestos y no expuestos permite evaluar riesgo asociado a la condición de exposición. La relación matemática que se establezca entre estas dos medidas permite el cálculo de a lo menos seis expresiones de riesgo:

- Riesgo Relativo (en la literatura anglosajona el término Risk Ratio corresponde al cálculo utilizando incidencia acumulada)
- El término Rate ratio se utiliza cuando se utiliza densidad de incidencia en el cálculo,
- Riesgo Atribuible,
- Riesgo Atribuible Porcentual (fracción etiológica)
- Riesgo Atribuible Poblacional.
- Riesgo Atribuible Poblacional Porcentual

Para explicar el sentido de cada una de estas medidas se puede recurrir a la tabla tetracórica o de doble entrada, en este caso, siendo la situación más simple: relacionar dos variable binarias (si/no), una de las cuales es la exposición (EX) y otra el efecto (E), la información se recoge basándose en unidades de personas.

Distribución de los datos entre las variables

Exposición	Efecto		**Total**
	Si	No	
Si	a	b	$\mathbf{n_1}$
No	c	d	$\mathbf{n_0}$
Total	$\mathbf{m_1}$	$\mathbf{M_0}$	**n**

Se puede establecer si hay o no asociación estadística y para esto se aplica la Ji de Mantel-Haenszel (χ_{MH}) que tiene la siguiente expresión:

$$\chi_{MH} = \sqrt{\frac{(a \times d - b \times c)^2}{n_1 \times n_0 \times m_1 \times m_0}(n-1)}$$

Se busca su valor en una tabla de distribución normal para ver si es o no es significativo indicándonos si hay o no hay asociación de las variables, pero esto no nos indica la magnitud de la asociación que es lo más importante en los estudios de epidemiología. Para medir esta magnitud se tiene que calcular el riesgo del efecto en los individuos expuestos o no expuestos y el total a través de:

$$R_1 = \frac{a}{n_1}; \qquad R_0 = \frac{a}{n_0}; \quad R = \frac{m_1}{n}$$

Riesgo Relativo: Es el cociente entre la tasa de incidencia de la enfermedad en expuestos y la incidencia en no expuestos. Permite conocer la magnitud de riesgo o protección asociada a la exposición estudiada. Carece de unidades de medida y se calcula con la siguiente formula:

$$RR = \frac{R_1}{R_0}$$

En un estudio realizado para ver si existe alguna relación entre el habito de fumar en el embarazo conduce a tener un recién nacido de bajo peso. Los datos se dan en la tabla siguiente (datos ficticios)

Fumar en el embarazo y tener un recién nacido de bajo peso (Datos Ficticios)

Exposición	Efecto		**Total**
	Si	No	
Si	2	18	**20**
No	4	76	**80**
Total	**6**	**94**	**100**

$$\chi_{MH} = \sqrt{\frac{(a \times d - b \times c)^2}{n_1 \times n_0 \times m_1 \times m_0}(n-1)} = \sqrt{\frac{(20 \times 760 - 40 \times 180)^2}{200 \times 800 \times 60 \times 940}(1000-1)} =$$

$$\chi_{MH} = 2.662$$

El valor calculado se busca en la tabla de la distribución normal y este nos da p=0.008 que es inferior a 0.05 es un resultado significativo por lo que se puede decir que hay asociación entre el consumo de cigarrillos y tener recién nacidos con bajos pesos.

Cual es la asociación que hay entre las variables para ello calculamos el Riesgo Relativo (RR)

$$RR = \frac{R_1}{R_0} = \frac{0{,}1}{0{,}05} = 2$$

Índice

LOS SONDEOS ELECTORALES

Aníbal Angulo A.

1.-INTRODUCCION

Los sondeos electorales son una de las aplicaciones estadísticas de la que más se habla. Además, este tipo de estudios resulta singular desde varios puntos de vista, por ejemplo. Es un tema que despierta gran interés, incluso pasión, en gran parte de la población. A diferencia de otros estudios estadísticos, en este caso se acaba sabiendo el verdadero valor de los parámetros estimados, cosa que no ocurre si, por ejemplo, hacemos una encuesta para conocer el porcentaje de estudiantes que tienen conexión a Internet en su casa.

Los sondeos electorales fallan con frecuencia (aunque no siempre), y como esta es la única relación de muchas personas con la estadística, se abona la impresión de que esta es una ciencia poca seria.
Siempre que se estiman las características de una población a partir de una muestra se corre un cierto riesgo de error, pero sabemos que este riesgo es controlable mediante la selección de un adecuado tamaño de la muestra. Ahora bien, cuando los errores son de bulto y se producen de forma repetitiva, es que fallan otras cosas. No es que no se cumpla la teoría estadística, lo que suele ocurrir es que no se han aplicado bien los principios en que se basa, ya sea por falta de recursos o porque dadas las circunstancias en que se realizan, es muy difícil de cumplirlos.

2.- La muestra no es representativa

Para la correcta predicción de lo que pasa en la población a partir de una muestra es fundamental que la muestra sea representativa. La representatividad de la muestra se consigue mediante una adecuada selección de sus componentes, de forma que todos los individuos de la población tengan una probabilidad conocida de ser incluidos. Pero esto no es fácil. Ni barato.

Tal vez el fiasco más famoso de un sondeo electoral se produjo en las elecciones presidenciales de los Estados Unidos en 1936, cuando se enfrentaron los candidatos Roosevelt (demócrata) y Landon (republicano). La revista Literary Digest realizó un sondeo mediante el envió por correo de 10 millones de cuestionarios, de los cuales recibió complementados2,4 millones. Pero a pesar de lo generoso del tamaño muestral, la revista predijo una clara victoria de Landon (con 57% de los votos) cuando en realidad ocurrió todo lo contrario(gano Roosevelt con 61% frente al 37% de Landon). El problema estaba en la selección de la muestra, ya que las direcciones a las que se enviaron los cuestionarios se obtuvieron de la guías telefónica y del registro de propietarios de automóviles, este fue un gran error, pues en aquella época tener teléfono o coche estaba muy relacionado con la clase socioeconómica a la que se pertenecía, de forma que los partidarios de Landon estaban sobredimensionados en la muestra, en detrimento de los de Roosevelt. Hoy en día, seleccionar a los integrantes de la muestra a través de la guía telefónica es mucho menos problemático porque el uso del teléfono se ha extendido a todas las capas de la población en los países desarrollados tecnológicamente. Pero en el caso de realizar la entrevista por teléfono(caso frecuente) sí tiene mucha importancia a qué hora se llama, por quién se pregunta, o cómo se sustituye a los que no desean contestar. Descuidar estos aspectos puede conducir a graves errores en las predicciones por sesgo en la selección de la muestra.

3.- La intención de voto va cambiando

Los sondeos electorales se basan en encuestas realizadas varios días. O incluso varías semanas, antes de las elecciones. En algunos países está prohibido publicar resultados de sondeos electorales durante cierto período de tiempo antes de las elecciones. Por tanto, nos encontramos con dos tipos de extrapolación:
La que se hace desde la muestra hacia la población, siendo esta la que trata la teoría estadística del muestreo.
La que generaliza los resultados de las fechas en que se ha hecho el sondeo, al día de las elecciones. Pero los partidos se emplean a fondo en sus campañas electorales, se producen debates entre candidatos, pueden ocurrir sucesos sobre los que se posicionan los candidatos, y todo esto puede afectar al voto decidido o a la decisión finad de los que en el momento de la encuesta estaban indecisos.

Un caso paradigmático de cambio en la intensión de voto es el que se produjo en las elecciones españolas de 2004. la gestión informática que hizo el gobierno del Partido Popular sobre el atentado del 11 de marzo(solo 3 días antes de las elecciones) indignó y movilizó a muchos electores de forma que, en contra de los pronósticos y sondeos que se venían realizando, ganó el Partido socialista con una clara mayoría. En definitiva, los cambios en la intensión de voto, si se producen, no se pueden predecir estadísticamente con base en las encuestas realizadas.
¿A quién votarán los indecisos?.
Los indecisos son un verdadero dolor de cabeza para los encargados de realizar sondeos electorales, especialmente en los casos en que este grupo representa un porcentaje grande y cuando no hay un candidato que lleve una clara ventaja sobre el resto. El problema está en que las personas que no responden no son una muestra al azar de la población. Si las razones de no respuesta fueran estadísticamente independientes de las preferencias de voto, el problema tendría solución desde la estadística. Sin embargo, la práctica ha demostrado que esto no es así, y el problema se complica

Fermín Bouza, en un interesante artículo publicado en la revista 'Praxis Sociológica? Escribe "Como quiera que el grupo de NS/NR (No sabe /No responde) suele ser muy amplio en vísperas electorales no muy inmediatas, y suele ir de un20% a un 50% de los encuestados, el sociólogo presionado por medios y partidos, tiene que hacer una estimación del voto, es decir, imputar a los indecisos un voto y predecir lo que ocurriría hoy sise celebran las elecciones. Para hacer esa imputación a los indecisos cuenta con varios procedimientos, siendo los más frecuentes la atribución por simpatía ("¿con qué partido simpatiza usted más" o ¿De qué partido se siente usted más cercano?", o cualquier otra formula) o por recuerdo de voto (¿A qué partido votó usted en las últimas elecciones?"). No son las únicas fórmulas: los sociólogos pueden improvisar otras preguntas para conseguir la intención más probable de voto, y aquí ponen en juego su conocimiento, su intuición, o cualquier otra virtual adivinación. Los que el sociólogo tiene delante es un 'menú' de estimaciones entre las que va a elegir una para dar a la opinión pública" Está claro que la asignación de los votos de los indecisos a uno u otro partido es una tarea de importancia crítica, y su éxito responde más a conocimientos relacionados con la Sociología y con la política que con la Estadística.

4.- La falta de sinceridad en las respuestas

La redacción de las preguntas y el orden en que se realizan son también aspectos críticos. Escribir preguntas claras, que no induzcan respuestas, no es tarea fácil y requiere conocer bien la técnica de cómo plantear las preguntas y también requiere entrevistadores bien entrenados y motivados.
Puede preguntarse dando el nombre del candidato y el partido al que pertenece o dando solamente el nombre. Y esta pregunta puede hacerse antes o después de preguntas relacionadas con la situación del país, o con la valoración que se da a los candidatos, y las respuestas puede n variar dependiend o de cóm o se haga n la pregunta. Aveces existen condiciones de libertad de expresión particulares, que hacen más o menos creíbles las respuestas de los ciudadanos y que harán que el volumen de los llamados "indecisos" sea mayor o menos, ya que quizá en vez de indecisos lo que se tiene son "deciso s cautos", que prefieren mantener reservad a su opinión.

Del porcentaje de votos al número de escaños

En muchas ocasiones lo verdaderamente relevante, más que el porcentaje de votos que va a obtener cada partido, es el número de escaños, y los sistemas que se utilizan para distribuirlas cosas. Por ejemplo, para una determinada circunscripción electoral en la que hay 5 escaños en juego se pude predecir con una confianza del 95% que un determinado partido obtendrá un 32% de los votos con un margen de error del 3%. El problema está en que si obtiene un 31% le corresponderá un escaño, mientras que si obtiene el 33% le corresponderán 2. Y esta es una diferencia importante pero no sabemos por cuál opción decantarnos con la información disponible.
Otro problema es que algunas legislaciones electorales exigen un mínimo porcentaje de votos (por ejemplo, el 5%) para entrar en el reparto. Si un partido está rozando este porcentaje (por ejemplo, si su estimación de voto es del % no se puede saber si llegará o no, y el hecho de que ocurra una u otra afectará también al numero de escaños del resto de partidos.

5.-A modo de resumen

Cuando se realizan sondeos electorales existen muchas dificultades para lograr buenas predicciones, dificultades que van más allá de aquellas que se refieren al ámbito de la teoría del muestreo. A pesar de ello también se producen notables éxitos en las predicciones, o en el adelanto de los resultados muy poco tiempo después de serrarse las urnas.
Sería conveniente tener medida la frecuencia y la magnitud con que fallan los sondeos electorales serios, después de la misma manera que las malas noticias son las que no invaden a través de los medios informativos, también las pifias en las predicciones son las más destacadas, incluso en el ambiente académico, pues es más sensacional y a veces más pedagógico ilustrar lo que no debe hacerse, que mostrar ejemplos donde las predicciones han funcionado bien.
Debe quedar claro que en este caso de los sondeos electorales hay varios tipos de procesos involucrados, y en muchos de ellos la sociología, la psicología o la polito logia juegan un papel más protagonista que la propia estadística. En cualquier caso, el uso apropiado de las herramientas estadísticas, la seriedad con que se aborde el trabajo de campo, la supervisión, el conocimiento sociológico del grupo humano de interés, son factores claves en el éxito de las predicciones.

También pueden existir, y de hecho existen, encuestas que son el resultado de consultas interesadas que pretenden influir sobre opinión de los electores. El ciudadano, tendrá que aprender a distinguirlas, aunque a veces es difícil por la proliferación y bombardeo de resultados de encuestas y sondeos. Indagar sobre el patrocinador y responsable del estudio puede dar cuenta del interés por la divulgación de determinados resultados, sin que ello obligue a sospechar de sesgo o falsedad. La experiencia y seriedad de la agencia responsable del estudio, así como del medio en que se publica, también son un buen indicador de la confianza que merecen los sondeos, además de aquella que se indica en la ficha técnica.

Bibliografía

José M Bernardo “Monitoring the 1982 Spanish Socialist Victory
A Bayesian Analysis” American Statistical Association Vol, 79, Núm. 387(1984)
Bouza Álvarez, F.: Comunicación política: encuestas, agendas y procesos cognitivos eletorales.
Revista ‘Praxis Sociológica’ 1998 número 3.
Estimación de la opinión pública y previsión electoral
Quantum. Vol1, núm. 3 Montevideo, invierno de 1994. págs 111 -122
Jorge Blanco.

Índice

Herramientas para el Control de Calidad

Dindo Valdez Blanco

1.-INTRODUCCION

En general en cualquier negocio o emprendimiento la mejora continua es crítica para la supervivencia. En todo el mundo, los estándares de calidad y productividad son
fundamentales, por tanto la supervivencia de las empresas depende en parte del mejoramiento continuo de la calidad.

Según la filosofía japonesa la clave para ser el mejor de su industria es adelantarse a los cambios de los costos y la demanda. En nuestro país esta filosofía se esta comenzando a aplicar por ejemplo en la banca, pues no cabe duda que con la globalización y el crecimiento de los mercados en todos los rubros, la calidad en los procesos será cada vez más importante. Es de esperar entonces que las empresas dediquen una parte de sus recursos a invertir en la mejora en la calidad de sus procesos. Por esta razón es importante estudiar más a fondo los diversos aspectos que engloba el estudio de la “calidad”.

2.- Fases de un proyecto de control de calidad

Para implementar la mejora de un proceso, se deben realizar cinco fases:

1. Fase de Definición
2. Fase de Medición
3. Fase de Análisis
4. Fase de Mejora
5. Fase de Control

La fase definición trata de la definición de objetivos, los instrumentos y alcances del proyecto. En la fase de medición se identifican y miden los factores que influyen en el proceso que se desea mejorar, el cual no se encuentra en los niveles deseados. En la fase de análisis se aplican las herramientas estadísticas para identificar los niveles críticos que explican la mayor parte de la variación del proceso. En la fase de mejora se utilizan los niveles críticos para llevar el proceso al nivel de desempeño deseado. Por último la fase de control se utiliza para aplicar un plan de control para mantener los factores dentro de los rangos requeridos

En las cinco fases del control de calidad se utilizan técnicas estadísticas desde las más simples hasta las más avanzadas. Es por esta razón que no hay una receta general de cómo encarar cada fase, lo que está claro es que se necesita de un sólido conocimiento de la estadística

En general un proyecto de control de calidad persigue cumplir las siguientes metas:

- Reducción de defectos
  - Menores costos de producción
  - Tiempos de ciclo más cortos
  - Mayor satisfacción del cliente
- Cambio de la cultura operacional
  - Enfoque a la calidad al cliente y a hacer las cosas bien
  - Orgullo de ser el mejor
  - Estándares para solucionar problemas
- Personal altamente entrenado y consciente de su rol en la empresa
  - Todos hablan un lenguaje común
  - Existe un compromiso con la institución

Esta metodología del control de calidad es un enfoque del método japonés. La filosofía de esta metodología del control de calidad comenzó después de la segunda guerra mundial con la llamada “Revolución japonesa de la Calidad”. El éxito de este país en

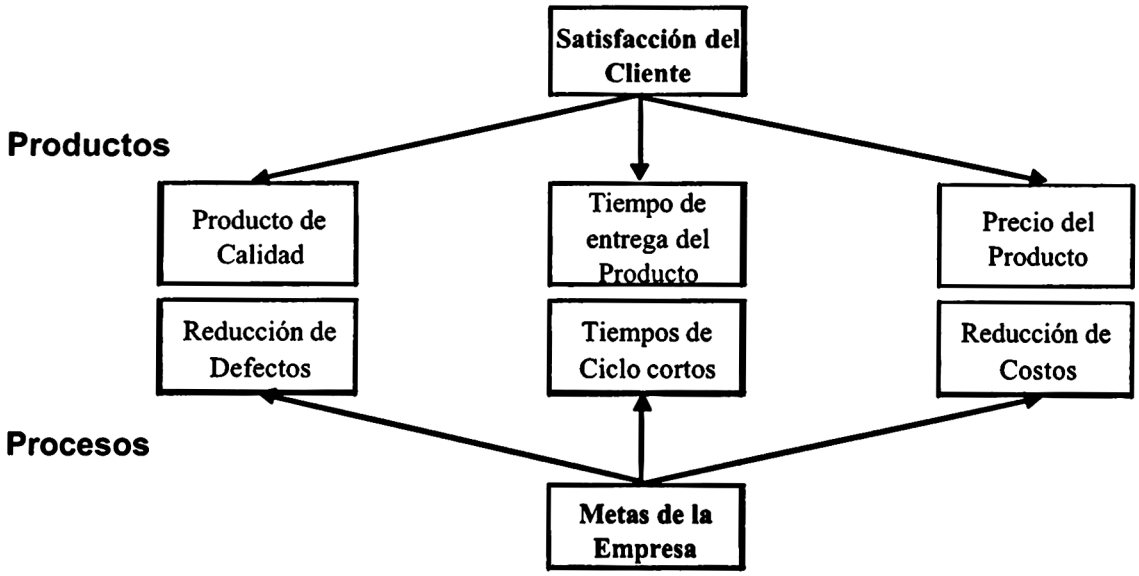
parte ha sido por el enfoque que le han dado a los procesos los cuales siempre están enfocados a la satisfacción del cliente, ya que los estándares de calidad que el cliente espera son cada vez mayores ha medida que pasa el tiempo. Por ejemplo Henry Ford acostumbraba incluir una herramienta especial y un manual de procedimiento para ajustar las válvulas del famoso modelo "A".

La Pirámide de las Metas

Muchas veces los propietarios de los negocios esperan ver al final del mes únicamente los rendimientos económicos dejando de lado aspectos de la calidad. Es ahí donde se debe cambiar la cultura, más bien la ganancia vendrá como consecuencia de la calidad con la que se entreguen los productos y servicios, este aspecto tal vez es la razón por la que un negocio no rinde como se quisiera. Si se ilustra lo anterior como una pirámide de metas, está claro que la primera meta debe ser la reducción de defectos en todo el proceso, seguido de la mejora de los rendimientos (costos, tiempos, etc.). Estos aspectos mejorarán la calidad del producto y se llegará a la satisfacción del cliente que se traducirá es una mayor demanda y esto generará las ganancias esperadas. El siguiente diagrama ilustra todo lo dicho:

La Pirámide de Metas del Negocio

En la mayoría de las ocasiones lo que es importante para el cliente no lo es para la empresa. Por ejemplo en el siguiente diagrama se puede observar cómo se puede enfocar los objetivos de la empresa a las necesidades del cliente.



De esta forma se puede decir que cualquier cosa que sea crítica para el cliente debeser critica también para el negocio. Claro está que es parte de la calidad el encontrar que cosa es crítica para el usuario final del producto, algo no evidente y que debe ser constante preocupación del proveedor del producto o servicio.

Herramientas para la definición del problema

La primera fase en la implementación de un proyecto de control de calidad es la definición del problema, en esta fase se debe realizar un diagnóstico de la situación y determinar los problemas y su magnitud. Existen diversas herramientas para esta tarea dentro de las cuales podemos mencionar:

- Los diagramas de afinidad
- Matrices de Causa y Efecto
- Mapas de los procesos
- Diagramas de Espina de Pescado
- Diagramas de Pareto
- Diagramas de árbol de necesidades

Como una aplicación del paquete estadística Minitab versión 14 en inglés vamos a explicar la construcción del Diagrama de Espina de Pescado.

4.-El Diagrama de Espina de Pescado

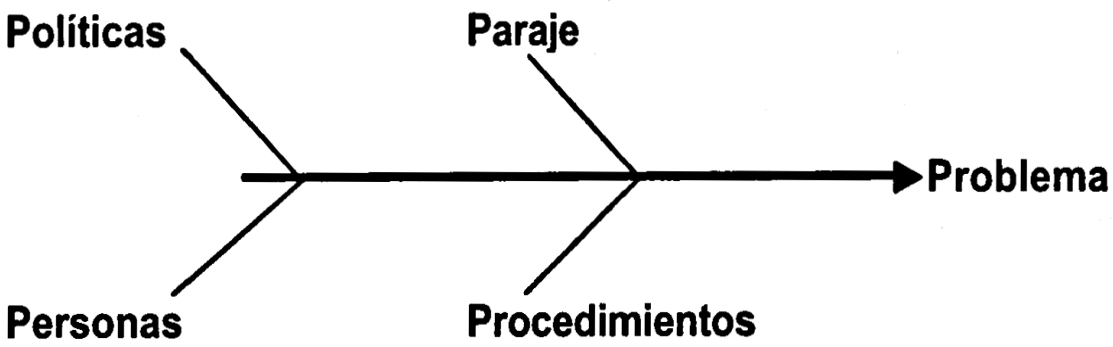
Los diagramas de espina de pescado primero fueron desarrollados por Ishikawa, un ingeniero japonés en los años 40's. El buscaba una manera simple y gráfica de mostrar la relación entre las causas (espinas) probables de un problema de un proceso

El diagrama de Espina de Pescado es un método para ir identificando sistemáticamente todas las causas potenciales que pueden estar contribuyendo a un problema (efecto).

Los diagramas más usados comúnmente son dos: El diagrama 4P, y el diagrama 6M.

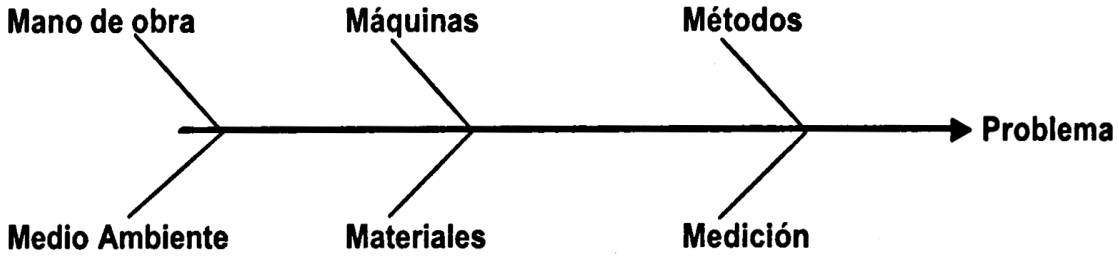
El diagrama 4P es utilizado cuando el problema es administrativo, algo común en las empresas de servicios. Se trata de identificar las causas asignables de un problema a través de una mesa redonda donde todos los miembros del personal contribuyen a identificarlos, la existencia del problema se puede deber a la ubicación del negocio (Paraje), a las políticas de la empresa, a los procedimientos de operación establecidos o debido al personal de la empresa.

Diagrama 4P



En cambio el diagrama 6M es aplicado en la industria, en la producción de sus artículos. Claramente un problema en la producción se puede asignar a: la mano de obra (personal), a la maquinaria, a los métodos utilizados, al medio ambiente (por ejemplo la temperatura), a los materiales utilizados en la producción o por último a los efectos de la mala medición y/o calibración.

Diagrama 6M



La principal utilidad de un diagrama de espina de pescado es que no solo sirve para identificar las causas, sino que como se basa en una lluvia de ideas a partir de todo el personal que interviene en el proceso, esto coadyuva para que todos conozcan como se desarrolla el proceso y de esta forma se sientan comprometidos con la calidad del producto.

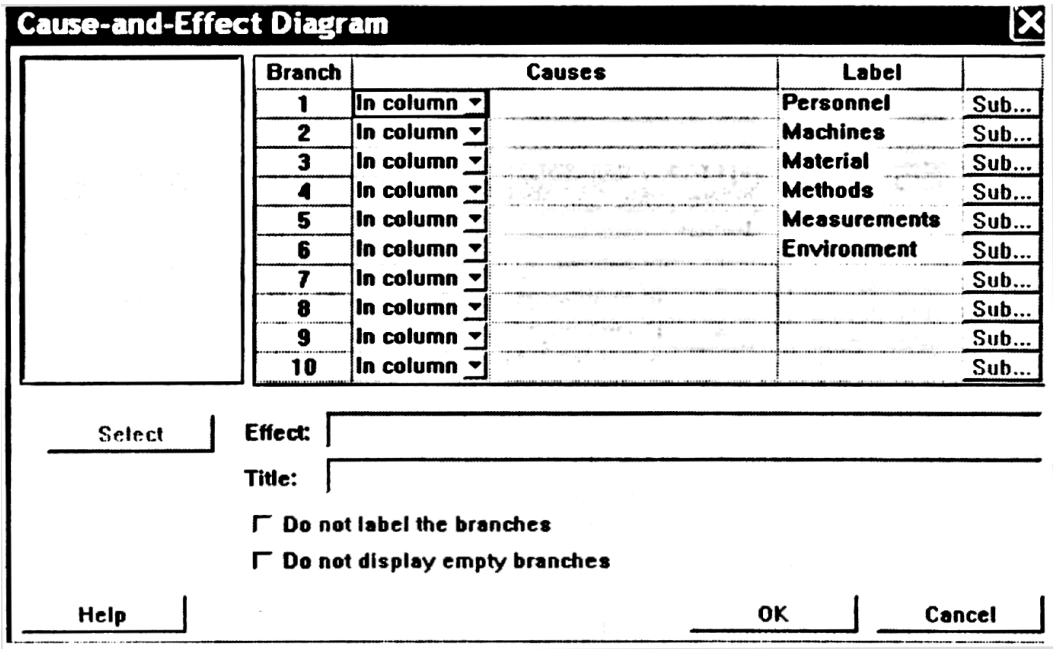
5.- Creación del diagrama de Espina de Pescado en Minitab

Minitab es un paquete estadístico que tiene muchas herramientas para el control de calidad, entre ellos brinda la posibilidad de elaborar el diagrama de Espina de pescado de una forma sistemática y sencilla, veamos el siguiente ejemplo de aplicación: Imagine que a través de un diagrama de Pareto ha descubierto que las partes de un producto fueron rechazadas en su mayoría debido a desperfectos en su superficie. Usted se reúne con representantes de varios departamentos para generar una lluvia de ideas de las causas potenciales para tales desperfectos. Usted decide utilizar el diagrama de Espina de Pescado 6M elaborando una lista para cada espina:

Mano de Obra	Maquina	Material	Método	Medición	Medio Ambiente
Muy indecisos	Fallas en los enchufes	Mala calidad en el material	Mal empleo del Ángulo de ruptura	Falta de precición del micrómetro	Mucha Humedad
Mal Supervisor	Existencia de Basura	Lubricante de mala calidad	Fallas al poner los frenos	Micrómetros no acondicionados	Mucha condensación
Poco entrenamiento	Falla del torno	Distribuidor Malo			
	Velocidad muy lenta				

Una vez definidas las causas, se procede a la elaboración del diagrama:

- Abra Minitab
- Seleccione STAT>Quality Tools> Causse and Effect



Minitab tiene una utilidad que automáticamente genera diagramas simples de espina de pescado. Tiene preprogramado en las etiquetas las &M's.

Las entradas a las espinas del pescado se almacenan en seis columnas en una hoja de trabajo de Minitab. Las columnas entonces se asocian a las etiquetas de esta ventana.

- Ingrese el título de las 6 columnas en la hoja de trabajo de Minitab
- Ingrese las causas debajo del título apropiado

MINITAB - Untitled - [Ejemplo.MTW ***]

File Edit Data Calc Stat Graph Editor Tools Window Help

	C1-T	C2-T	C3-T	C4-T	C5-T
	Mano de Obra	**Maquina**	**Material**	**Método**	**Medición**
1	Muy indecisos	Fallas en los enchufes	Mala calidad en el material	Mal empleo del Ángulo de ruptura	Falta de precición del r
2	Mal Supervisor	Existencia de Basura	Lubricante de mala calidad	Fallas al poner los frenos	Micrómetros no acondi
3	Poco entrenamiento	Falla del torno	Distribuidor Malo		
4		Velocidad muy lenta			
5					

- Lleve la columna correspondiente a cada etiqueta del menú
- Ingrese un nombre para el problema en Effect

Cause-and-Effect Diagram

C1	Mano de Obra
C2	Maquina
C3	Material
C4	Método
C5	Medición
C6	Medio Ambier
C7	Training
C8	Speed
C9	Micrometers

Branch	Causes		Label	
1	In column	'Mano de Obra'	Personnel	Sub...
2	In column	Maquina	Machines	Sub...
3	In column	Material	Material	Sub...
4	In column	Método	Methods	Sub...
5	In column	Medición	Measurements	Sub...
6	In column	'Medio Ambiente'	Environment	Sub...
7	In column			Sub...
8	In column			Sub...
9	In column			Sub...
10	In column			Sub...

Select

Effect: Desperfectos en la Superficie

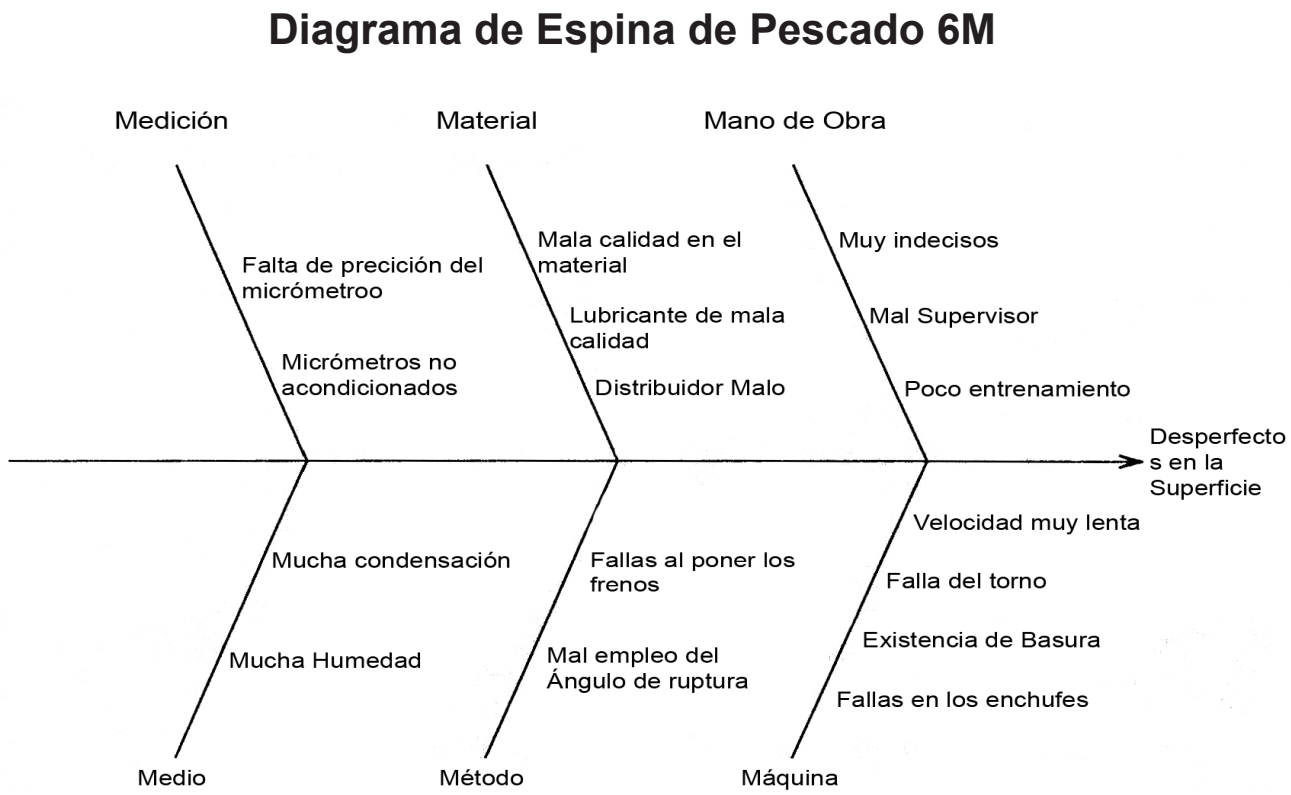
Title: Diagrama de Espina de Pescado 6M

☐ Do not label the branches

☐ Do not display empty branches

Help OK Cancel

❖ Cree el diagrama simple haciendo un clic en el botón OK, y listo.



No existe una receta general para elaborar los diagramas de espina de pescado, pero se podría resumir en los siguientes pasos:

- Identificar el Proceso
- Construir el diagrama de espina de pescado
  - o Definir claramente el problema
  - o Realizar una lluvia de ideas para identificar las ramas de cada espina
  - o Trabajar en grupo
  - o Elaborar el diagrama en Minitab
  - o Presentarlo a todo el grupo
  - o Discutir y tomar decisiones

El último punto es muy importante, el diagrama de espina de pescado no solo debe servir para mostrar de una manera agradable las causas de un problema, sino fundamentalmente el objetivo tomar decisiones que ayuden a mejorar los procesos. Si al final de todo el trabajo no se toma ninguna acción, entonces no tendrá sentido.

Índice

CLASIFICACIÓN DE MÉTODOS MULTIVARIADOS

Juan Carlos Flores L.

1.-INTRODUCCION

La investigación de la gestión 2005 sobre la ***Clasificación de métodos del Análisis Multivariante***, es con el fin de contribuir con la comunidad estudiantil y todos lo que tengan interés en conocer estos métodos que es de gran interés en las aplicaciones reales en el análisis de datos en las distintas disciplinas, estos métodos pueden ser aplicados en ciencias, administración de empresas, ingeniería, psicología, área salud, en el área sociológico, económico, la industria, centros de investigación de ámbito universitario, etc.

El trabajo de investigación consta de Capitulo I que es la introducción, el Capitulo II desarrolla los métodos de dependencia, el Capitulo III desarrolla métodos de Interdependencia, luego las Conclusiones, Recomendaciones y la Bibliografía.

El Análisis Multivariante es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que existen varias variables medidas para cada individuo ú 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 no lo podrían conseguir. Estos métodos hacen posible plantear preguntas específicas y precisas de considerable complejidad en marcos idóneos, lo que posibilita llevar a cabo investigaciones teóricamente significativas y evaluar los efectos de las variaciones paramétricas ocurridas de forma natural en el contexto en que normalmente ocurren. De esta forma, se pueden preservar las correlaciones naturales entre las múltiples influencias sobre el comportamiento y se pueden estudiar estadísticamente los efectos aislados de esas influencias sin provocar el típico aislamiento de esos individuos o variables.

2.- Definición del análisis multivariante

El análisis multivariante es un conjunto de técnicas de análisis de datos en expansión. Entre las técnicas más conocidas se tiene a (1) regresión múltiple y correlación múltiple; (2) análisis discriminante múltiple; (3) componentes principales y análisis factorial común; (4) análisis multivariante de varianza y covarianza; (5) correlación canónica; (6) análisis cluster; (7) análisis multidimensional y (8) análisis conjunto. Entre las técnicas emergentes también incluidas están (9) análisis de correspondencias; (10) modelos de probabilidad lineal como logit y probit; y (11) modelos de ecuaciones

simultáneas/estructurales. (*En este resumen se, desarrolla cada una de las técnicas multivariantes, definiendo brevemente la técnica y el objetivo de su aplicación.*)

Los objetivos del Análisis Multivariante pueden sintetizarse en dos:

i) Proporcionar métodos cuya finalidad es el estudio conjunto de datos multivariantes que el análisis estadístico uni y bidimensional es incapaz de conseguir.

ii) Ayudar al analista o investigador a tomar decisiones óptimas en el contexto en el que se encuentre teniendo en cuenta la información disponible por el conjunto de datos analizado.

3.-Tipos de Técnicas Multivariantes

Se pueden clasificar en tres grandes grupos (ver grafica 1):

Métodos de dependencia

Suponen que las variables analizadas están divididas en dos grupos: las variables dependientes y las variables independientes. El objetivo de los métodos de dependencia consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma.

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.

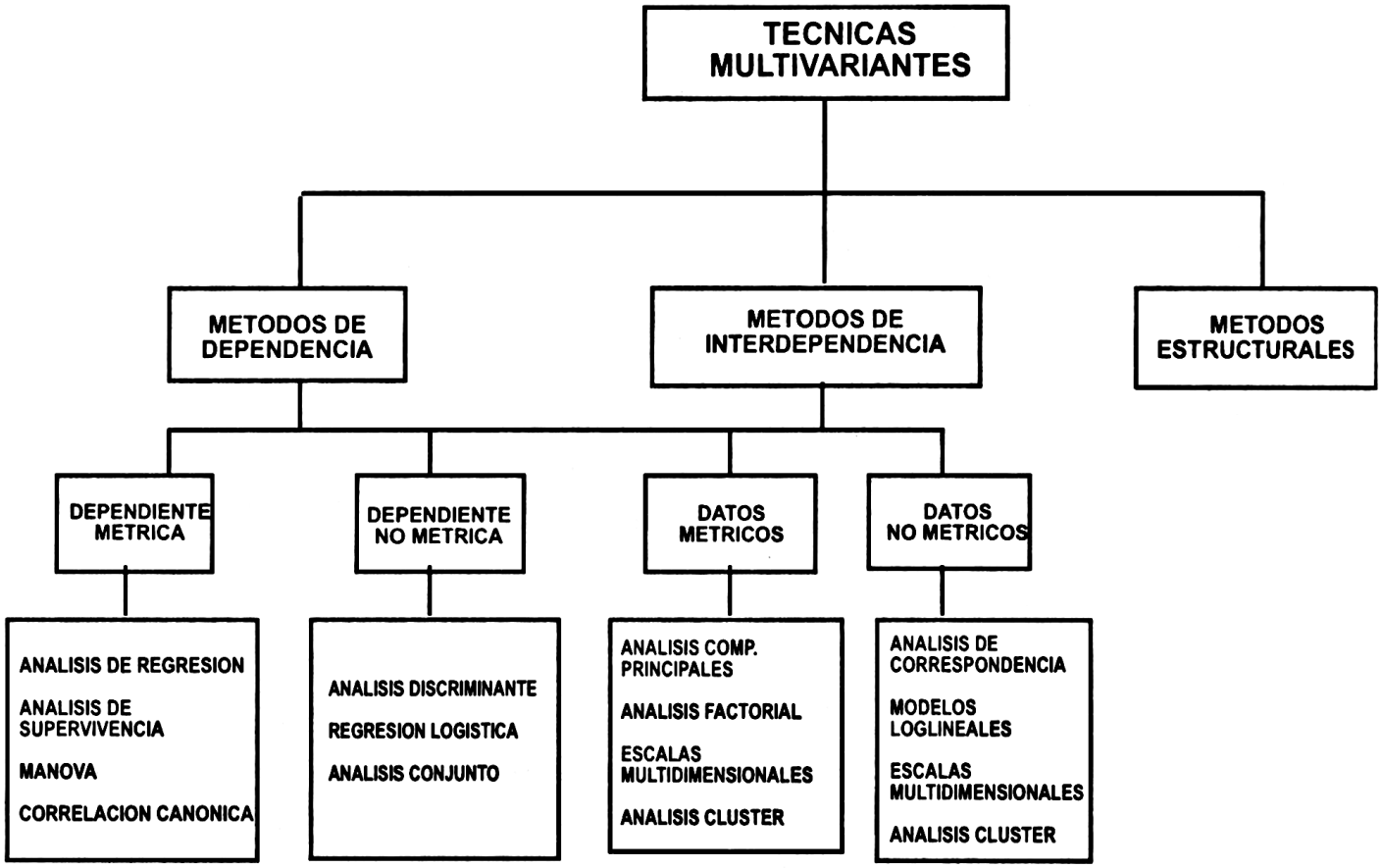
Métodos de dependencia

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:

Métodos estructurales

Suponen que las variables están divididas en dos grupos: el de las variables dependientes yelde 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. En resumen los mostramos en la siguiente grafica 1

Grafica 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, 1 intentar predecir el gasto anual en cine de una persona a partir de su nivel de ingresos, nivel educativo, sexo y edad.

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 ú 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.

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?

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 está relacionado 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, las técnicas que se pueden aplicar son las siguientes:

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.

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.

Análisis Conjunto (Conjoint Analysis)

Es una técnica que analiza el efecto de variables independientes no métricas sobre variables métricas o no métricas. La diferencia de Conjoint con el Análisis de la Varianza radica en que las variables dependientes pueden ser no métricas y los valores de las variables independientes no métricas son fijadas por el analista.

En otras disciplinas, a la técnica Conjoint se la 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.

Métodos de Interdependencia

Se pueden clasificar en dos grandes grupos según que el tipo de datos que analicen sean métricos o no métricos.

Si los datos son métricos, se pueden utilizar, entre otras, las siguientes técnicas:

Análisis Factorial (AF) y Análisis de Componentes Principales (ACP)

Se utiliza para analizar interrelaciones entre un número elevado de variables métricas explicando dichas interrelaciones en términos de un número menor de variables denominadas factores (si son inobservares) o componentes principales (si son observables).

Así, por ejemplo, si un analista financiero quiere determinar cuál es el estado de salud financiero de una empresa a partir del conocimiento de un número de ratios financieros, construyendo varios índices numéricos que definan su situación, el problema se resolvería mediante un ACP. Si un psicólogo quiere determinar los

factores que caracterizan la inteligencia de un individuo a partir de sus respuestas a un test de inteligencia, utilizaría para resolver este problema un AF.

Escalas Multidimensionales

Su objetivo es transformar juicios de semejanza o preferencia en distancias representadas en un espacio multidimensional. Como consecuencia, se construye un mapa en el que se dibujan las posiciones de los objetos comparados de forma que aquellos percibidos como similares están cercanos unos de otros y alejados de objetos percibidos como distintos. Por ejemplo, analizar, en el mercado de refrescos, las percepciones que un grupo de consumidores tiene acerca de una lista de refrescos y marcas con el fin de estudiar qué factores subjetivos utiliza un consumidor a la hora de clasificar dichos productos.

Análisis Cluster

Su objetivo es clasificar una muestra de entidades (individuos o variables) en un número pequeño de grupos de forma que las observaciones pertenecientes a un grupo sean muy similares entre sí y muy disimilares del resto. A diferencia del Análisis Discriminante, en el Análisis Cluster se desconoce el número y la composición de dichos grupos. Por ejemplo, clasificar grupos de alimentos (pescados, carnes, vegetales y leche) en función de sus valores nutritivos.
Si los datos fuesen no métricos, se podrían utilizar, además de las Escalas Multidimensionales y el Análisis Cluster, las siguientes técnicas:

Análisis de Correspondencias

Se aplica a tablas de contingencia multidimensionales y persigue un objetivo similar al de las escalas multidimensionales pero representando simultáneamente las filas y columnas de las tablas de contingencia. Por ejemplo, analizar el paro teniendo en cuenta la provincia, sexo, edad y nivel de estudios del parado.

Modelos log-lineales

Se aplican a tablas de contingencia multidimensionales y modelizan relaciones de dependencia multidimensional de las variables observadas que buscan explicar las frecuencias observadas.

Métodos o Modelos Estructurales

Analizan las relaciones existentes entre un grupo de variables representadas por sistemas de ecuaciones simultáneas en las que se suponen que algunas de ellas (denominadas constructos) se miden con error a partir de otras variables observables denominadas indicadores.

Los modelos utilizados constan, por lo tanto, de dos partes:

i) Un modelo estructural, que especifica las relaciones de dependencia existente entre los constructos latentes.

ii) y un modelo de medida, que especifica como los indicadores se relacionan con sus correspondientes constructos.

Por ejemplo, los Modelos Estructurales permiten analizar cómo se relacionan los niveles de utilización de los servicios de una empresa con las percepciones que sus clientes tienen de ella.

4.-Etapas de un análisis multivariante

Con el fin de guiar el desarrollo de la aplicación de un análisis multivariante, podemos decir que se tienen las siguientes etapas:

1.-Objetivos del análisis

Se define el problema especificando los objetivos y las técnicas multivariantes que se van a utilizar. El investigador debe establecer el problema en términos conceptuales, definiendo los conceptos y las relaciones fundamentales que se van a investigar. Se deben establecer si dichas relaciones van a ser relaciones de dependencia o de interdependencia. Con todo esto se determinan las variables a observar.

2.- Diseño del análisis

Se determina el tamaño muestral, las ecuaciones a estimar (si procede), las distancias a calcular (si procede) y las técnicas de estimación a emplear. Una vez determinado todo esto, se proceden a observar los datos.

3.-Hipótesis del análisis

Se evalúan las hipótesis subyacentes a la técnica multivariante. Dichas hipótesis pueden ser de normalidad, linealidad, independencia, homocedasticidad, etc. También se debe decidir qué hacer con los datos missing.

4.-Realización del análisis

Se estima el modelo y se evalúa el ajuste a los datos. En este paso pueden aparecer observaciones atípicas (outliers) o influyentes cuya influencia sobre las estimaciones y la bondad de ajuste se debe analizar.

5.-Interpretación de los resultados

Dichas interpretaciones pueden llevar a reespecificaciones adicionales de las variables o del modelo con lo cual se puede volver de nuevo a los pasos 3 y 4.

5.-Validación del análisis

Consiste en establecer la validez de los resultados obtenidos analizando sí los resultados obtenidos con la muestra se generalizar a la población de la que procede. Para ello se puede dividir la muestra en varias partes en las que el modelo se vuelve a estimar y se comparan los resultados. Otras técnicas que se pueden utilizar aquí son las técnicas de remuestreo (jacknife y bootstrap).

6.- BIBLIOGRAFIA

HAIR, J., ANDERSON, R,, TATHAM, R. y BLACK, W. (1999). Análisis Multivariante. 5ª Edición. Prentice Hall.

T.W. ANDERSON. An Introduction to Multivariate Statistical Analysis. 2ª Edición. Wiley.

S. JAMES PRESS . Applied Multivariat e Análisis . 2ª Edición . Wiley.

JHONSON WICHERN. Aplplied Method of Multivariate Statisitical Analysis. 3ª Edición. Prentice Hall.

EVERITT, B. And GRAHAM, D. (1991). Applied Multivariate Data Analysis. Arnold.

SHARMA, S. (1998). Applied Multivariate Techiques. John Wiley and Sons.

MARDIA, K.V., KENT, J.T. y BIBBY, J.M. (1994). Multivariate Analysis. Academic Press.

FERRAN, M. (1997). SPSS para Windows. Programación y Análisis Estadístico. Mc.Graw Hill.

MEDICIÓN DE LA POBREZA ASPECTOS TEORICOS Y METODOLÓGICOS

Carlos Fernando Silva Viamonte

1. Introducción[1].

En la literatura económica, siempre ha sido un problema de indudable interés el de la identificación y la cuantificación de la pobreza, teniendo en cuenta sus consecuencias y su amplia repercusión, ya sea en su vertiente económica, en la social ó en otras.

En los últimos treinta y cinco años, desde el trabajo de Atkinson (1970), se ha desarrollado un cúmulo importante de aportaciones sobre la medición de diferentes aspectos distributivos: desigualdad, polarización, movilidad, pobreza, etc. Se ha ido abriendo camino una literatura extensa sobre los conceptos y la metodología adecuada para medirlos. Se ha hecho igualmente un esfuerzo en encuadrarlos en marcos normativos que enlacen con el resto de las áreas de la economía y de la política económica. Para ello se han ido dotando de marcos axiomáticos generalmente aceptados y de unos métodos de medición coherentes con esos marcos. El estudio de la desigualdad económica ha centrado principalmente el análisis. No obstante, otros temas distributivos, como la pobreza, han ido llamando la atención, sobre todo a partir de la publicación, en 1976, en la revista Econometrica, del artículo de Amartya Sen en el que se sientan las bases para el estudio de lo que ha venido en denominarse pobreza económica, en el sentido de utilizar un indicador de la posición económica de los hogares, como la renta, el ingreso ó el gasto familiares, como punto de partida para el diseño y la construcción de medidas de pobreza. Es así, que a partir de este trabajo de Sen son temas conectados entre sí, pero esencialmente diferentes.

"Pobreza" es un término complejo puesto que sobre él caben muchas realidades sociales, así como, muy diferentes interpretaciones. Esta dificultad de sistematizar el término ha dado lugar al establecimiento de una serie de axiomas. Todo lo cual hace muy difícil la unanimidad y, en consecuencia, hace que unos sean más aceptados y otros más conflictivos. A este respecto, un nuevo término, "exclusión social", relacionado pero no equivalente a la pobreza, se abre camino. Parece ser un concepto más adecuado en relación a la política económica. Es claramente un concepto multidimensional, pero aún no está ni definido formalmente ni establecido. Este trabajo revisará en términos generales sobre la medición tradicional de la pobreza. Solamente comentaremos posibles líneas de avance y conexiones con la exclusión social cuando sea apropiado. Al fin y al cabo, cualquier desarrollo "multidimensional" se tiene que basar en los pilares de la medición clásica unidimensional. Veamos sus fundamentos. El trabajo aborda en primer lugar la medición de la pobreza. En la segunda sección se repasan los principales índices de pobreza propuestos y, al final, en la sección cuarta, se presentan tests de dominancia para la comparación robusta de la pobreza.

[1] FERES, Juan Carlos; MANCERO, Xavier (2001). "Enfoques para la medición de la pobreza. Breve revisión de la literatura". CEPAL, ECLAC Estudios estadísticos y prosp ectivos, serie 4. páginas 5-7 .

2. Resumen histórico

Haciendo un análisis histórico, el estudio científico de la pobreza se remonta a comienzos del siglo XX. Atkinson (1987) señala que antes de esa fecha se habían realizado algunas estimaciones sobre pobreza, pero que fue Booth entre 1892 y 1897 “el primero en combinar la observación con un intento sistemático de medición de la extensión del problema”, elaborando un mapa de pobreza de Londres.

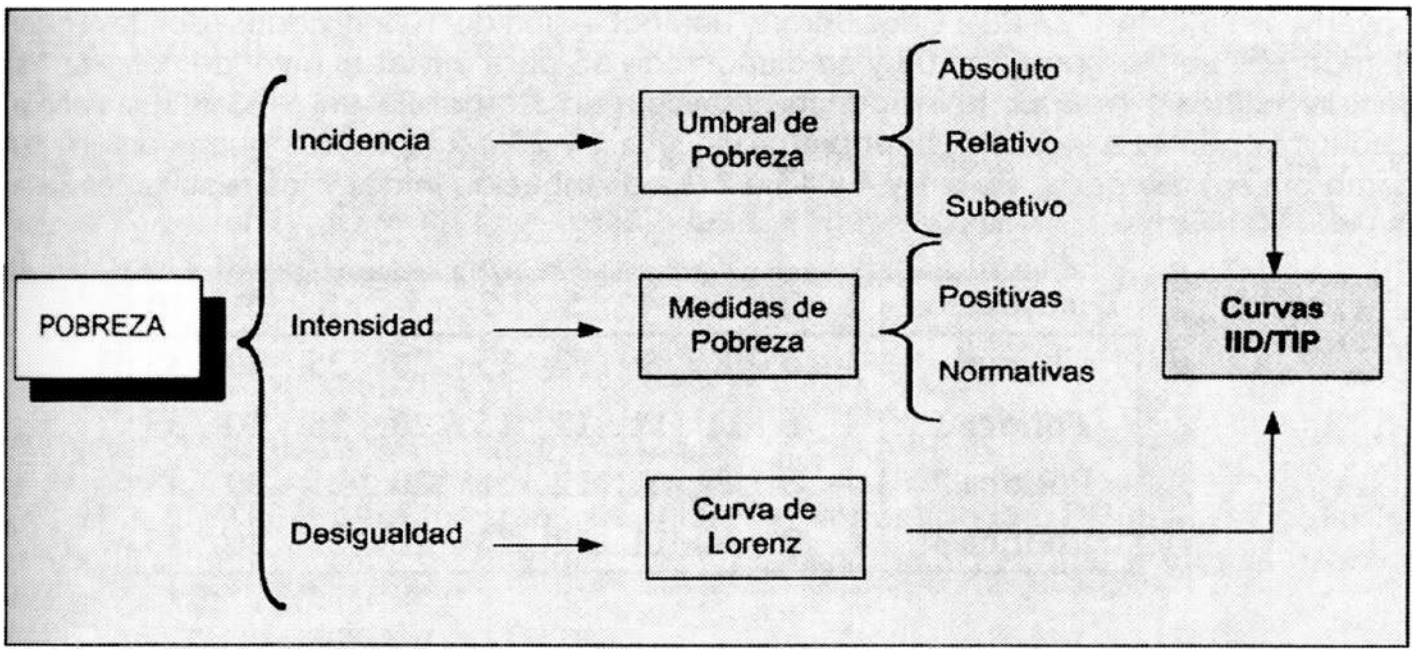
Posteriormente, Rowntree (1901) realizó un estudio para medir la pobreza en York, y utilizó un estándar de pobreza basado en requerimientos nutricionales. A partir de entonces se han desarrollado nuevos conceptos sobre la medición del bienestar y nuevas metodologías para medir la pobreza.

Cuando hablamos de pobreza en términos generales, se refiere a la incapacidad de las personas de vivir una vida tolerable (PNUD, 1997). Entre los aspectos que la componen se menciona llevar una vida larga y saludable, tener educación y disfrutar de un nivel de vida decente, además de otros elementos como la libertad política, el respeto de los derechos humanos, la seguridad personal, el acceso al trabajo productivo y bien remunerado y la participación en la vida comunitaria. No obstante, dada la natural dificultad de medir algunos elementos constituyentes de la “calidad de vida”, el estudio de la pobreza se ha restringido a los aspectos cuantificables (y generalmente materiales) de la misma, usualmente relacionados con el concepto de “nivel de vida”.

Como se sabe, para analizar la pobreza primero que nada es necesario definirla. Una vez establecidos los aspectos que abarca el término “pobreza”, su medición requiere de indicadores cuantificables, que guarden relación con la definición elegida. Sea cual fuere ésta y el o los indicadores utilizados, el proceso de medición comporta dos elementos: la identificación de las personas que se considere pobres y la agregación del bienestar de esos individuos en una medida de pobreza.

3.- Elementos del análisis estadístico de la Pobreza.

En el siguiente esquema se puede observar cada unos de los elementos que involucran la medición de la Pobreza desde un punto de vista estadístico.



4-Cuantificación

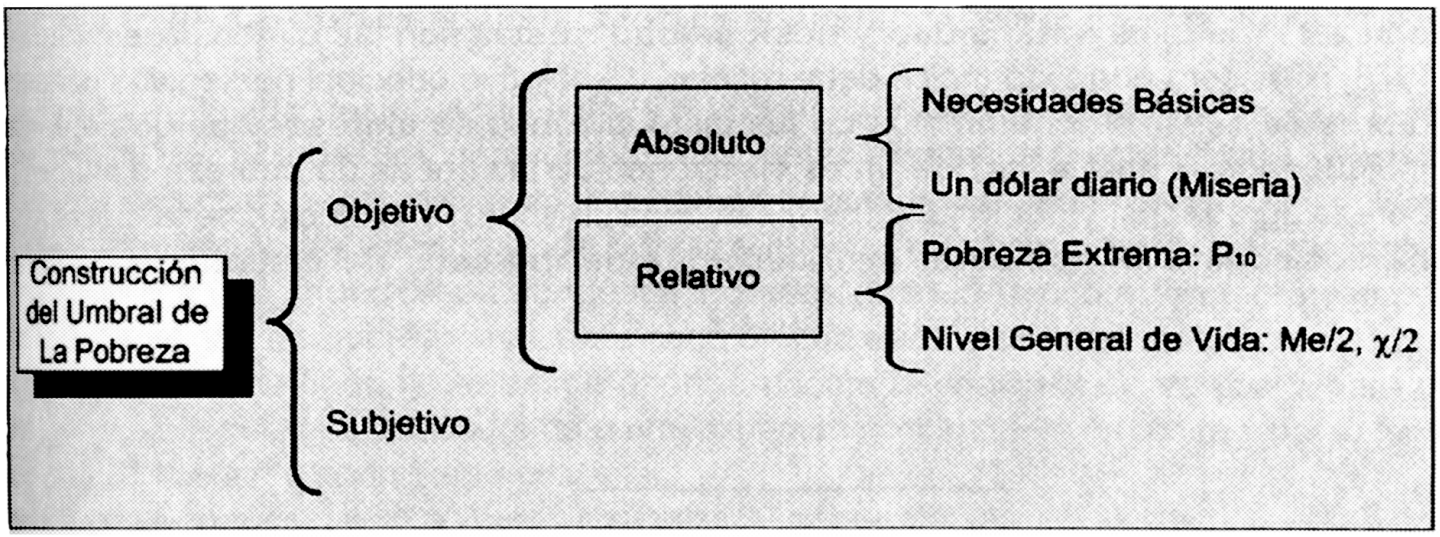
4.1 Umbral o Línea de La Pobreza.

Dada una sociedad con N individuos, cuyo vector de rentas no negativas es , entonces el umbral de pobreza es un valor z>0, tal que:

Xi < z Û'3f i es pobre

El conjunto de pobres será T(x,z) = {i: xi<z} y su cardinal(q) será el número de pobres.

La construcción del umbral de la pobreza se puede apreciar a través del siguiente esquema:



4.1 Incidencia e Intensidad de La Pobreza.

Utilicemos el siguiente ejemplo extraído de Creedy (The dynamics of inequality and poverty, 1998, págs. 27-28): Considérese una población de 10 individuos, donde el umbral de pobreza se ha fijado en 10$ y se dispone de 5$ para aliviar el nivel de pobreza. Para ello, la política 1 (reducir la incidencia) consiste en otorgar 2$ a 4 y 3$ a 3, la política 2 (reducir la pobreza extrema) podría otorgar 3$ a 1 y 2$ a 2, mientras que la política 3 (de compromiso) otorgaría: 2$ a 1 y 4 y 1$ a 2. La distribución inicial y las resultantes serían:

Distribución	1	2	3	4	5	6	7	8	9	10
Inicial	4	6	8	9	12	15	20	25	30	35
Política 1	4	6	11	11	12	15	20	25	30	35
Política 2	7	8	8	9	12	15	20	25	30	35
Política 3	6	7	8	11	12	15	20	25	30	35

¿Qué política resulta la más correcta?

Justamente para tener una visión clara de la mejor política es que a continuación se desarrolla la metodología de las curvas IID o curvas TIP.

5.- Curvas IID/TIP (Jenkins y Lambert).

En un intento por encontrar un instrumento gráfico que permita analizar la pobreza en forma similar a lo que se hace a través de las funciones de densidad y las curvas de Lorenz en el caso de la distribución de ingresos, Jenkins y Lambert (1997) desarrollan las curvas **IID/TIP**[2].
La denominación TIP's obedece a :Three l's of Poverty: incidence, intensity, inequality..

Estas curvas permiten visualizar gráficamente tres dimensiones fundamentales de la pobreza: incidencia, **intensidad y desigualdad**[3]. Estas son las dimensiones que Sen (1976) considera que todo índice debe reflejar.. El objetivo principal perseguido por estos autores es lograr ordenamientos de las distribuciones de ingresos que no presenten ambigüedades frente a las diferentes elecciones de las líneas de pobreza e índices de pobreza agregados, y por lo tanto permitan concluir sobre la evolución de la pobreza en una sociedad o al comparar distribuciones de ingresos de distintas economías.

[2]La denominación TIP's obedece a :Three l's of Poverty: incidence, intensity, inequality.
[3]Estas son las dimensiones que Sen (1976) considera que todo índice debe reflejar.

A continuación se describen los fundamentos de esta metodologías, siguiendo a Jenkinsy Lambert (1997).Se considera un vector $x = (x_p x_2, x_p, ..., x_n)$ que representa una distribución de ingresos entre n unidades perceptoras de ingresos (hogares o personas), donde x_i denota el ingreso de la unidad i y se cumple que $0 < x_l \leq x_2 \leq x_i \leq ... \leq x_n$ El vector g_x refleja las brechas o gaps de pobreza asociadas con el vector de ingresos x y la línea de pobreza z:

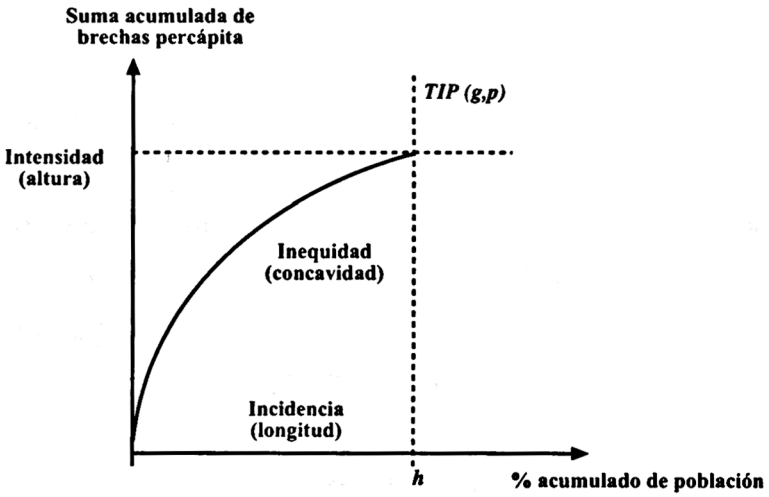
$$g_{xi} = \max\{z - x_i, o\}$$

Las curvas IID/TIP's para los gaps de pobreza, denominadas TIP(g;p), se obtienen ordenando a las personas en orden creciente de ingresos y acumulando sus gaps de pobreza, de forma tal que:

$$TIP(g;p) = \frac{\sum_{i=1}^{q} g_i}{n}$$

Donde hace referencia al total de hogares pobres. Para cada valor de p, TIP(g; p) representa el gap acumulado por ese porcentaje de unidades perceptoras de ingresos dividido el total de unidades. Se acumulan los gaps de pobreza solamente para las unidades pobres, para el resto estos gaps valen cero. Este instrumento gráfico permite visualizar la incidencia de la pobreza, que está dada por el percentil para el cual la TIP se vuelve horizontal, es decir h=q/n. También permite visualizar la intensidad o severidad déla pobreza, que está dada por la altura máxima alcanzada por curva, ya que refleja la suma total de gaps dividida el total de unidades perceptoras de ingresos.

Finalmente refleja la desigualdad entre los pobres, a través de la concavidad de parte
curva de la TIP Si todos los gaps de pobreza fueran iguales, esta parte de la TIP sería una línea recta con pendiente igual a . Análogamente con las medidas de desigualdad, existen situaciones de referencia de pobreza mínima y máxima. La máxima pobreza corresponde a una situación en donde los ingresos de toda la población fueran cero, y por lo tanto el gap de pobreza es z para cada unidad perceptora de ingresos, por lo que la curva TIP es una línea recta desde el origen hasta la intersección vertical z para un p=1. La mínima pobreza corresponde a una situación donde nadie es pobre y por lo tanto la curva TIPcoincide con el eje horizontal.



Dadas dos distribuciones de ingresos x e y, y una línea de pobreza común z, se dice que la distribución g_y domina en el sentido TIP a g_x cuando la TIP g_y no se sitúa por debajo de la TIP g_x en ningún punto, es decir:

$$TIP(g_v;p) \text{ domina a } TIP(g_x;p) \Leftrightarrow TIP(g_v;p) \geq TIPg_x(g_x;p) \text{ para todo } p \in [0,1]^4$$

Para cualquier línea de pobreza en común, la dominancia de g_y sobre es g_x condición necesaria y suficiente para asegurar que el nivel de pobreza en *x* no es superior al nivel de pobreza en *y*. Los autores señalan que esto se sigue manteniendo para líneas de pobreza menores, cualquiera sea el índice de pobreza elegido perteneciente a una amplia familia de índices *P*. Este primer resultado refiere a la comparación de dos distribuciones de ingresos con líneas de pobreza comunes, pero puede resultar de interés comparar dos distribuciones considerando las diferencias en los niveles de vida entre ambas distribuciones, es decir incorporar la dimensión relativa de la pobreza.

Para ello los autores desarrollan las curvas *TIP* normalizadas, que se construyen a partir de la acumulación de brechas de pobreza normalizadas. Las brechas de pobreza normalizadas, Γ_x , se construyen a partir de:

$$\Gamma_x = \frac{g_{xi}}{z} = \max\left[\frac{z - x_i}{z},\right.$$

6. Axiomática.

Se considera el sistema básico de axiomas que cualquier indicador de pobreza debe satisfacer.

- ☑ Axioma Focal: El indicador de pobreza está determinado por las rentas de los pobres. Es decir, sean x e y dos vectores ordenados de rentas, y sean x(z) e y(z) los subvectores formados por las rentas de los pobres. Entonces:
  x(z) = y(z) => P(x,z) = P(y,z)

[4]La dominancia estricta requiere q u e la desigualda d sea estricta para algún *p*,

- ☑ Axioma de Monotonía: Sean $x = (x_1,x_2,...,x_n)$, $x^* = (x_1,x_2,...,x_i - \alpha,...,x_n)$, siendo z>0 la línea de pobreza y a >0. Entonces:
  $x_i < z \Rightarrow P(x,z) < P(x^*,z)$
- ☑ Axioma de Transferencia Débil: Sean $x = (x_1,x_2,...,x_n)$, $x^* = (x_1,x_2,...,x_i - \alpha,..., x_j+\alpha,..., x_n)$, con a >0.
  $x_i < z \Rightarrow P(x,z) < P(x^*,z)$

7.- Otros Axiomas.

- ☑ Axioma de Simetría: Si y se obtiene de x mediante una permutación, la medida de la pobreza no varía: P(x,z) = P(y,z), ? z>0.
- ☑ Axioma de Sensibilidad frente a incrementos del umbral de pobreza: Dado un vector x, ordenado de rentas, P(x,z')>P(x,z), ? z.z'XD: z'>z
- ☑ Axioma de Normalización: Si no existen individuos cuya renta se sitúe por debajo del umbral de pobreza, la medida de la pobreza es cero.
- ☑ Axioma de Continuidad: Para un umbral de pobreza (z) determinado, a medida P(x,z) es una función continua de la renta (x).
- ☑ Axioma de Descomponibilidad Aditiva: Sea x el vector de rentas de una población y supongamos que ésta se particiona en m subgrupos: $x = (x^{(1)},x^{(2)},...,x^{(n)})$. En este caso:

$$P(x,z) = \sum_{i=1}^{m} \frac{N_i}{N} P(x^{(i)},z)$$

8.-Indicadores Simples de Pobreza

En todos los casos, se supone una población de n individuos, cuyo vector ordenado de rentas es $x = (x_1,x_2,...,x_n)$, siendo z>0, el umbral de pobreza. El número de pobres (q) es el cardinal de T=T(x,z)= {i: x_i<z}.

☑ Proporción de Pobres: $H(x,z) = \frac{q}{N}$

☑ Ratio de Pobreza: $I(x,z) = \frac{1}{q^z}\sum_{i\in T} g(x_i,z) = \frac{1}{q^z}\sum_{i\in T}(z - x_i) = 1 - \frac{\mu_q}{z}$

☑ Ratio Combinado de Pobreza:

$$H(x,z) = H(x,z).I(x,z) = \frac{1}{n_z}\sum_{i=1}^{q}(z - x_i) = \frac{q}{n} - \frac{q\mu_q}{nz}$$

9.-Medidas Basadas en Déficits de Pobreza.

$$P(x,z) = A(q,n,z)\sum_{i\in T} g(x_i,z)V_i(x,z)$$

☑ Medida de Sen:

$$V_i(x,z) = q + 1 - i \Rightarrow S(x,z) = \frac{2}{(q+1)nz}\sum_{i\in T}(z - x_i)(q + 1 - i)) = H[I + (1 - I)G_q]$$

☑ Medida de Thon:

$$V_i(x,z) = n+1-i \Rightarrow T(x,z) = \frac{2}{n(n+1)z}\sum_{i\in T}(z-x_i)(n+1-i)) = \frac{q+1}{n+1}S(x,z) + \frac{2(n-q)}{n+1}HI(x,z)$$

☑ Medida Exponencial

$$V_i(x,z) = e^{z-x_i} \Rightarrow E(x,z) = \frac{qk}{nz}\sum_{i\in T}(z-x_i)e^{z-x_i} = \frac{1}{nz}\sum_{i=1}^{q}(z-x_i)\,e^{\frac{-x_i}{z}}$$

10.- Familias de Medidas de Pobreza

☑ $V_i(x,z) = (q+1-i)^{\alpha} \Rightarrow K(x,z,\alpha) = \frac{q}{nz}\left(\sum_{i=1}^{q} i^{\alpha}\right)\sum_{i=1}^{q}(z-x_i)(q+1-i)^{\alpha}, \alpha \geq 0$

$\alpha = 0 \Rightarrow K(x,z,0) = HI(x,z)$

$\alpha = 1 \Rightarrow K(x,z,1) = S(x,z)$

☑ $V_i(x,z) = (z-x_i)^{\alpha-1} \Rightarrow F(x,z,\alpha) = \frac{1}{nz^{\alpha}}\sum_{i=1}^{q}(z-x_i)^{\alpha}, \alpha \geq 0$

$\alpha = 0 \Rightarrow F(x,z,0) = H(x,z)$

$\alpha = 1 \Rightarrow F(x,z,1) = HI(x,z)$

$\alpha = 2 \Rightarrow F(x,z,2) = H(I^2 + (1+I)^2 + CV^2 q)$

11.- Cumplimiento de Axiomas

Axiomas	H	I	HI	Kα	Fα	T	E
Focal	SI	SI	SI	SI	SI	SI	**SI**
Monotonía	NO	SI	SI	SI	SI	SI	**SI**
Transf.. Débil	NO	NO	NO	SI	SI	SI	**SI**
Simetría	SI	SI	SI	SI	SI	SI	**SI**
Incr. De la línea de pob.	NO	NO	NO	SI	SI	SI	**SI**
Normalización	SI	SI	SI	SI	SI	SI	**SI**
Continuidad	NO	NO	SI	NO	SI	SI	**SI**
Descomponibilidad	SI	NO	**SI**	NO	**SI**	NO	NO

12.- Bibliografía.

BARTELS, C.P.A. (1977). Economics Aspects of Regional Welfare. Martinus Nijhoff Sciences División.

CALLEALTA, F.J.; CASAS, J.M.; NÚÑEZ, J.J. (1995). "Un modelo para la distribución de ingresos en España: Ajuste y evolución de la desigualdad". Actas de la IX Reunión Asepelt-España. Volumen III, págs. 219-232. Santiago de Compostela.

CALLEALTA, F.J.; CASAS, J.M.; NÚÑEZ, J.J. (1996). "Distribución de la Renta per Cápita Disponible en España: Descripción, Desigualdad y Modelización". En Distribución Personal de la Renta en España, Cap. 5, J.B. Pena (Director). Ed. Pirámide. Madrid.

CASAS, J.M.; NÚÑEZ, J.J. (1991). Sobre la medición de la desigualdad y conceptos afines. Actas de la V Reunión Anual de ASEPELT-España. Libro 2. Las Palmas de Gran Canaria.

FOSTER, J.E.; SHORROCKS, A.F. (1988). Inequality and Poverty Orderings. European Economic Review, 32, págs. 654-662.

JOHNSON, R.A.; WICHERN, D. W. (1992). Applied Multivariate Statistical Analysis. Prentice-Hall.

NYGARD, F.; SANDSTROM, A. (1981). Measuring Income Inequality. Amqvist & Wiksell International.
Stockholm.
PENA, J.B. (1977). Problemas de la medición del bienestar y conceptos afines. INE. Madrid.

PENA, J.B. (Director); CALLEALTA, F.J.; CASAS, J.M.; MEREDIZ, A.; NÚÑEZ, J.J. (1996). Distribución Personal de la Renta en España. Pirámide. Madrid.

RUIZ-CASTILLO, J. (1986). Problemas Conceptuales en la Medición de la Desigualdad. Hacienda Pública Española, 101, págs. 17-31.

SEN, A. (1997). On Economic Inequality. Clarendon Press, Paperbacks. Oxford.
SHORROCKS, A.F. (1983). Ranking Income Distributions. Economica, 50, págs. 3-17.
URIEL, E. (1995). Análisis de Datos. Series Temporales y Análisis Multivariante. Ac, Madrid.

ZARZOSA, P. (1996). Aproximación a la medición del bienestar social. Universidad de Valladolid. Valladolid.

Índice

REGRESIÓN APARENTEMENTE NO RELACIONADOS

Verónica Cuenca Ramallo

Desarrolla una técnica econométrica conocida como Regresión Aparentemente No Relacionados o modelos SUR, la cual constituye un caso muy especifico de un sistema de ecuaciones simultaneas en el que la correlación entre las ecuaciones se origina entre los errores de éstas y no la incorporación de variables endógenas (dependientes) como variables predeterminadas en otras ecuaciones del sistema.

1.- INTRODUCCION.-

Usualmente los modelos de ecuaciones simultáneas, se piensa de inmediato en aquellos sistemas en los que se especifican variables endógenas (dependientes) en algunas ecuaciones como predeterminadas en otras ecuaciones del mismo modelo.

Bajo esta especificación existe una correlación identificable de los términos de error entre las ecuaciones del sistema donde no debe confundirse con el problema de autocorrelación, el cual toma lugar cuando existe correlación de los términos de error dentro de cada ecuación.

Para las ecuaciones simultaneas, la aplicación de mínimos cuadrados ordinarios(MCO) resulta en estimadores sesgados y con errores cuadrados medios que pueden ser bastante elevados, especialmente en muestras pequeñas, se utilizan procedimientos tales como los mínimos cuadrado s ordinarios y la estimació n máximo verosímil.

2.- MODELO SUR.-

Las siglas **SUR** provienen del nombre que recibe en inglés este sistema de ecuaciones (Seemingly Unrelated Regressions) que en castellano significa Regresiones aparentemente no relacionados.

Sin embargo, considérese un conjunto de ecuaciones de regresión de la siguiente forma.

$$y_{1i} = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{\ i} + \cdots + \beta_k x_{Ki} + u_i$$
$$y_{2i} = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{Ki} + u_{2i}$$
$$y_{3i} = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{Ki} + u_{3i}$$
$$\cdots$$
$$y_{Mi} = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{Ki} + u_{Mi}$$

Donde i= 1,2,3 ..... N

En este sistema existen **M** variables endógenas (dependientes) denotadas como $\mathbf{y_{Mi}}$, cada una asociada con un término de error $\mathbf{u_{Mi}}$, así como con un conjunto de **K** variables exógenos (independientes) $\mathbf{X_{Ki}}$.

La relación entre las variables endógenas y las exógenos está dada por los coeficientes b_k,el número de observaciones ***i*** para cada variable (endógenas, exógenas y términos de error) es de **N**.

En principio se podría razonar que, al no observarse variables endógenas $\mathbf{y_{Mi}}$ como variables predeterminadas en otras ecuaciones del sistema, cada una de las ecuaciones podría ser estimada con el uso de mínimos cuadrados ordinarios.

Esto sería posible si las ecuaciones fueran completamente independientes en el sentido de que la variabilidad de alguna de las variables endógenas no afectara el comportamiento de alguna otra ecuación.

En el vocabulario econométrico ello sería equivalente a decir que la matriz de varianzas ycovarianzas del sistema de ecuaciones tiene triángulos (inferior y superior) iguales a cero.

En otras palabras, sería una matriz con una diagonal diferente de cero y cuyas entradas serían las variancias de los términos de error de cada ecuación.

Esta conjetura, sin embargo, podría ser incorrecta si se detectara algún tipo de movimiento simultáneo de todas las ecuaciones originado por una supuesta relación contemporánea entre los términos de error que no se origina por la presencia de variables endógenas como variables predeterminadas en las ecuaciones.

Es decir, las regresiones que no están aparentemente correlacionadas, sí lo estarían por medio de correlaciones implícitas, sin modelar específicamente, entre los términos de error.

3.- LA ESPECIFICACIÓN DEL MODELO GENERAL.-

Suponga que la representación matricial de la m-ésima ecuación del sistema (1) es de la siguiente forma:

$$Y_m = X_m \beta_m + U_m \quad m=1, 2,3,....., M$$

Donde los vectores **Y** y **U** son de orden (N x 1), **X** es una matriz de orden
(N x Km), donde Km es el número de variables exógenas en la m-ésima ecuación, el numero de variable exógena no tienen que ser el mismo para cada una de las ecuaciones, es más, la eficiencia del método SUR aumenta en el tanto que las variables X muestren una menor asociación entre ecuaciones y **βm** es un vector de parámetros de orden (Km x 1). Al considerar las **M** ecuaciones de forma matricial se tiene la siguiente representación:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 & 0 \\ 0 & 0 & . & 0 & 0 \\ 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_M \end{bmatrix} + \begin{bmatrix} U_1 \\ U_2 \\ . \\ . \\ U_M \end{bmatrix}$$

La cual se puede expresar, en forma matricial, como:

$$Y = XB + U \qquad (4)$$

Donde la dimensión de Y es (MN x 1), la de X es (MN x K), la de b es

(K x 1) y la de U es (MN x ¡) en este caso, $K = \sum_{m=1}^{M} K_m$

Dado que U_{mi} es el valor observado del término de error de la m-ésima ecuación en el i-ésimo período, el supuesto de **correlación contemporánea** de los errores, pero no correlación serial, implica que, E [Umi, Ujs] = s_{mj} si i=s, pero es igual a cero si i#s, en otras palabras, cuando los períodos **i** y **s** coinciden existe una covariancia diferente de cero entre los errores de las ecuaciones **m** y **j**.
En forma matricial se puede especificar una matriz de variancias y covariancias para las ecuaciones **m** y **j** como E [Um U_j^t]=s_{mj} I_N, donde **I** es la matriz identidad de orden **N**.

Para los M vectores de términos de errores existe una matriz de varianzas y covarianzas que asume la siguiente forma:

$$\Phi = E[UU^t] = \begin{bmatrix} \sigma_{11}I_N & \sigma_{12}I_N & \dots & \sigma_{1M}I_N \\ \sigma_{21}I_N & \sigma_{22}I_N & \dots & \sigma_{2M}I_N \\ . & & & \\ . & & & \\ \sigma_{M1}I_N & \sigma_{M1}I_N & \dots & \sigma_{MM}I_N \end{bmatrix} = \Sigma \otimes I_N$$

Donde $\otimes$ es el producto Kronecker y Σ la matriz de varianzas y covarianzas de la forma:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2M} \\ . & & & \\ . & & & \\ \sigma_{M1} & \sigma_{M1} & \dots & \sigma_{MM} \end{bmatrix}$$

La matriz I es simétrica y se supone que es definida positiva y que no es singular, el producto Kronecker es la expansión de cada una de las entradas de una matriz (en este caso las variancias y covariancias de la matriz I) por una matriz, se dice que una matriz es simétrica cuando las entradas de sus triángulos superior e inferior son iguales.

En una estructura matricial el producto $S_{mj}I_N$ denota los términos de error dentro de cada ecuación son homocedásticos (varianza del error constante) y que no tienen autocorrelación en el tiempo (existe casos que se trata de datos de corte transversal).

4.- ESTIMACIÓN CON MATRIZ Σ CONOCIDA.-

Dado que el modelo no involucra la simultaneidad de las variables endógenas en el sentido de los sistemas de ecuaciones sim ultáneas, el procedimiento de estimaciones de un modelo SUR, cuando son conocidas las varianzas y covarianzas de la matriz Σ , es el correspondiente al de mínimos cuadrados generalizados. En este caso, el estimador toma la forma:

$$\hat{\beta} = [X^1\Phi^{-1}X]^{-1}X^1\Phi^{-1}Y$$

La cual es equivalente a:

$$\hat{\beta} = [X^1(\Sigma^{-1}\otimes I)^{-1}X]^{-1}X^1(\Sigma^{-1}\otimes I)Y$$

Es decir, el mejor estimador lineal insesgado. El estimador expresado detalladamente es de la forma:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_M \end{bmatrix} = \begin{bmatrix} \sigma^{11}X_1^1X_1 & \sigma^{12}X_1^1X_2 & \dots & \sigma^{1M}X_1^1X_M \\ \sigma^{21}X_2^1X_1 & \sigma^{22}X_2^1X_2 & \dots & \sigma^{2M}X_2^1X_M \\ \vdots & \vdots & \vdots & \vdots \\ \sigma^{M1}X_M^1X_1 & \sigma^{M2}X_M^1X_2 & \dots & \sigma^{MM}X_M^1X_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{m=1}^{M} \sigma^{1m}X_1^1y_m \\ \sum_{m=1}^{M} \sigma^{2m}X_2^1y_m \\ \vdots \\ \sum_{m=1}^{M} \sigma^{Mm}X_M^1y_m \end{bmatrix}$$

Donde los escalare σ^{mj} corresponden al elemento (m,j) de la matriz de? .
La matriz de varianzas y covarianzas del estimador mínimo cuadrático generalizado de β es:

$$VAR(\hat{\beta}) = [X^1 \Phi^{-1} X]^{-1} = [X^1 (\Sigma^{-1} \otimes I)^{-1} X]^{-1}$$

En la discusión de los estimadores SUR hay tres aspectos que destacar sobre la eficiencia (entendido como varianza mínima).

1) Cuanto más elevada la correlación contemporánea de los términos de error entre ecuaciones mayor será la ganancia en eficiencia del estimador generalizado.
2) Si la correlación contemporánea es muy baja (los triángulos de la matriz E se aproxima a cero) no hay una ganancia importante por aplicar la regresión SUR a las ecuaciones en vez de utilizar los mínimos cuadrados ordinarios en cada ecuación.
3) Si cada una de las ecuaciones del sistema tiene las rrismas variables exógenas, entonces los estimadores SUR son equivalentes a los mínimos cuadrados ordinarios.

En general, cualquier ganancia en eficiencia (mínima varianza) tiende a ser mayor cuando las variables explicativas en las diferentes ecuaciones no están altamente correlacionadas.

5.- ESTIMACION CON MATRIZ Σ DESCONOCIDA.-

En la práctica es poco probable contar con los verdaderos valores de la varianzas y covarianzas (σ_{mj}) de los errores, por lo que es necesario recurrir a una estimación preliminar de los errores.

Este procedimiento se logra mediante la aplicación del método de mínimos cuadrados ordinarios a cada una de las ecuaciones del sistema.

El estimador es de la forma $b = (X_M^1 X)^{-1} X_M^1 y_m$, el cual permite obtener una primera estimación de errores mínimo cuadráticos $\hat{u}_m = y_m - X_m b_m$ para cada ecuación.

Si bien las variancias y covariancias calculadas con estos errores mínimos cuadráticos son sesgadas en muestras pequeñas , tienen la propiedad de consistencia que permite continuar con el procedimiento SUR.

La forma genérica de cálculo es la siguiente:

$$\hat{\sigma}_m = \frac{1}{N} \hat{u}_m^1 \hat{u}_j = \frac{1}{N} \sum_{i=1}^{N} \hat{u}_{mi} \hat{u}_{ji}$$

Si se definiera $\hat{\Sigma}$ como la matriz de variancias y covariancias conformada por las estimaciones $\hat{\sigma}_{mj}$; , el correspondiente estimador mínimo cuadrático generalizado asume la forma:

$$\tilde{\beta} = \left[X^1 \left(\hat{\Sigma}^{-1} \otimes I\right)\right]^{-1} X^1 (\hat{\Sigma}^{-1} \otimes I) Y$$

Este estimador es de uso generalizado y en la literatura se le conoce como el estimador mínimo cuadrático generalizado de Séller.

Sin embargo, otro estimador con propiedades asintóticas más deseables es el carácter iterativo, es decir, una vez que se cuenta con los estimadores de la fórmula $\tilde{\beta} = \left[X^1 \left(\hat{\Sigma}^{-1} \otimes I\right)\right]^{-1} X^1 (\hat{\Sigma}^{-1} \otimes I) Y$, se calculan los errores y las varianzas a partir de ello, los cuales se utilizan en un nuevo cálculo de los mínimos cuadrados generalizados, el procedimiento se repite hasta que la función de verosimilitud alcance un máximo.

En tanto que las varianzas y covarianzas tiendan a permanecer constantes, ambos procedimientos tiene la misma distribución límite, esta distribución es aproximadamente normal con una media β y con una matriz de varianzas y covarianzas consistentemente estimada por:

$$\left[X^1\left(\hat{\Sigma}^{-1} \otimes I\right)X\right]^{-1}$$

Por consiguiente, el estimador de Zellner y el estimador iterado serán asintoticamente más eficiente que el estimador mínimo cuadrático.

En muestras finitas, sin embargo, existe una región en el espacio paramétrico de Σ , en laque el estimador mínimos cuadrático es más eficiente.

Es el caso cuando la pérdida de eficiencia originada por el uso de una estimación de Σ vez de la verdadera matriz, es mayor que la ganancia obtenida por utilizar un estimador mínimo cuadrático generalizado (especialmente si las correlaciones contemporáneas entre los errores son pequeñas), el caso extremo es cuando Σ es diagonal y el estimador mínimo cuadrático es el relevante.

6 -EJEMPLO DE LA APLICACIÓN DEL MODELO SUR.-

Como ejemplo de aplicación, puede ser los cuestionarios del Curso Prefacultativo, variables que se puede considerar son.

Donde vive, procedencia del colegio, edad, sexo, si trabaja o no trabaja, el factor de la economía de los alumnos y a la vez de los padres, que les afecta para que reprueben o aprueben, etc., estudiar si existe un impacto del conjunto de regresores sobre las puntuaciones en Matemáticas, Informática, Química y Física.

7-CONCLUSIÓN.-

Lo interesante de los modelos **SUR** (Modelo Aparentemente no Relacionados), es que se logra evaluar simultáneamente el efecto de las variables independientes sobre el conjunto de las variables dependientes, tomando en cuenta la correlación que existe entre las ecuaciones.

8.- Bibliografía

[1] Agresti Alan Wiley Jhon (1938), Análisis de Datos Categóricos, New Cork.
[2] J.Aníbal Angulo (2003), Estudio Socioeconómico de la Población Estudiantil Biblioteca de la Carrera de Estadística de la Facultad de Ciencias Puras y Naturales, La Paz- Bolivia.
[3] Dhrymes Phoebus J., Me Gran Hill (1995) Económetria, México.
[4] Greene William H. (1999), Análisis Econométrico Tercera edición, Madrid España
[5] Gujarati Damodar N. (1997), Econometría Básica, Santafé de Bogotá, Colombia.
[6] Hair Jr.Joseph F., Ralph E. Anderson (1999), Análisis Multivariante, Madrid España.

[7] Araya Monge Rigoberto, Muñoz Giro Juan E., Regresiones que Aparentemente no están Relacionados, disponible en internet.

“El éxito de la vida: si piensas que estas vencido, vencido estarás, Si piensas que no te atreves no lo harás, Si piensas que te gustaría ganar, pero no lo puedes no lo lograras, Si piensas que perderás, ya has perdido, **PORQUE EN EL MUNDO EL ÉXITO COMIENZA CON LA VOLUNTAD DEL HOMBRE**, Todo esto esta en el estado mental porque muchas carreras se ha perdido entes de haberse corrido y muchos cobardes han fracasado antes de haber su trabajo comenzado”.

Calle 27 de Cota Cota
Bloque F.C.P.N. - Primer Piso

La Paz - Bolivia