

Departamento de Economía

Documentos Docentes

LA INVESTIGACION CIENTIFICA EN ECONOMIA: ALCANCES METODOLOGICOS

Autor: Mario Gaymer C.

DD 1983 - N° 01

P R E F A C I O

El Departamento de Economía, el de más reciente creación en la Facultad de Administración y Economía de la Universidad de Santiago de Chile, inicia una serie de publicaciones docentes con el trabajo del profesor, Sr. Mario Gaymer Cortés, intitulado "La Investigación Científica en Economía: Alcances Metodológicos".

Las tareas de un Departamento Académico no consisten únicamente en impartir docencia, utilizando textos ya consagrados y de uso muy difundido. Su labor va mucho más allá. Siempre resulta necesario actualizar los concimientos, aclarar conceptos, ampliar el tratamiento de los temas más difíciles, elaborar respuestas para las inquietudes renovadas de los alumnos, modificar aspectos programáticos para adecuarlos a la realidad en continuo rendimiento. De este modo va surgiendo la necesidad de publicar monografías, artículos y notas sobre temas que los textos no dejan lo suficientemente esclarecidos o sólo reseñan brevemente.

Uno de estos temas es la metodología científica usada en Economía, tema soslayado muchas veces en los textos introductorios. ¿Cómo logra la Economía sus resultados?, ¿Tiene validez científica?, ¿Cómo puede llegar a sustentar determinadas afirmaciones, que pueden convertirse en bases de políticas económicas?, ¿Puede dársele categoría de ciencia a la Economía?

El trabajo del señor Gaymer responde estas preguntas con gran claridad y rigurosidad. Es un tema que siempre le ha apasionado y al cual destina un lugar preferente en su propio curso de Introducción a la Economía en nuestra Facultad.

Creemos que esta primera publicación de la serie se constituirá en lectura complementaria indispensable para todos los alumnos de los primeros cursos de Economía.

IVAN YAÑEZ PEREZ
Director

A G R A D E C I M I E N T O S :

El autor agradece los comentarios de sus colegas del departamento de Economía y en particular de los profesores Luis Arturo Fuenzalida Asmussen, Carlos Massad Abud y David Cooper Valenzuela.

También, agradece muy especialmente los comentarios y colaboración del profesor Carlos Gómez Díaz del departamento de Administración y las correcciones del profesor Sergio Ceppi Lupe de Mayol.

La opiniones vertidas son de estricta responsabilidad del autor.

LA INVESTIGACION CIENTIFICA EN ECONOMIA

ALCANCES METODOLOGICOS

MARIO GAYMER C.: Profesor del Departamento de Economía de la Facultad de Administración y Economía de la Universidad de Santiago de Chile.

I INTRODUCCION

En la economía, como en todas las ciencias, es necesario avanzar en el conocimiento. Es permanentemente necesario lograr mejores explicaciones de la realidad que interesa estudiar. Ese avance en el conocimiento debe ser objetivo, es decir, las conclusiones deben tener validez independientemente de la persona que las haya obtenido. Para lograr esa objetividad, y una confiabilidad razonable en las conclusiones, es preciso ceñirse a ciertos esquemas, los que son analizados por la Filosofía de las Ciencias. La investigación científica ha sido definida como un proceso, cuyas etapas se han estandarizado, para la búsqueda de explicaciones a fenómenos. El método científico constituye una lógica de justificación de la validez de las conclusiones obtenidas mediante este proceso. El objetivo de este trabajo es dar una visión general de este proceso y de las características particulares que toman al ser aplicado en la Economía.

A lo largo del desenvolvimiento histórico de la búsqueda del conocimiento ha habido discrepancias con respecto a cuáles deben ser los procedimientos para lograrlo. Si bien en la Edad Media el criterio de autoridad podía ser suficiente para dar validez a una interpretación sobre los fenómenos del mundo que nos rodea, a partir del Renacimiento los hombres de ciencia se rebelan y comienzan a defender nuevas maneras de validar las explicaciones sobre esos fenómenos. Famosa es la afirmación "eppur si muove" atribuida a Galileo.

Como el objetivo de este trabajo no es el de hacer una revisión de la Teoría del Conocimiento, no se mencionará siquiera otras corrientes de pensamiento, pasando directamente a la posición generalmen-

te aceptada por las ciencias nomotéticas (1) modernas, dejando sin embargo, en claro que su aceptación generalizada no constituye necesariamente garantía de definitiva.

El método científico -sustentado en la corriente filosófica del positivismo- es deductivo. Es conveniente detenerse aquí para estudiar el significativo de deducción y de su opuesto, inducción.

La deducción consiste en, a partir de premisas universales, obtener conclusiones particulares, o bien conclusiones referidas a universos menores o subconjuntos del universo inicial. La verdad de las premisas universales garantiza la verdad de las conclusiones obtenidas de éstas, puesto que tales conclusiones están ya contenidas en las premisas universales: son elementos o bien subconjuntos del universo para el cual son válidas las premisas. Si los objetos que tienen masa se atraen mutuamente, entonces al soltar una manzana en el aire, ésta será necesariamente atraída por la tierra (puesto que ambas tienen masa), y consecuentemente la veremos caer hacia ésta.

En la inducción, en cambio, se parte de premisas particulares para llegar a conclusiones universales. Si los objetos que yo he soltado se han caído, entonces todos los objetos libres caerán. La inducción, como método de investigación, arroja conclusiones cuya validez es cierta sólo para los particulares estudiados. La verdad de las premisas no garantiza la verdad de la conclusión y, por lo tanto, las generalizaciones que se pueda elaborar a base de esos particulares son inciertas.

Los satélites, los globos aerostáticos, etc, son objetos libres y, sin embargo, no caen. Si en varios países de alto ingreso per-cápita el nivel educacional es alto, eso no garantiza que, en otros países de ingreso similar, el nivel educacional sea también alto. La generalización inductiva puede ser correcta o puede no serlo, y no incluye sistemas para medir su confiabilidad.

II LA INVESTIGACION CIENTIFICA: UNA VISION GENERAL

Una investigación surge de la inquietud del investigador por la existencia de algún área o algún tipo de situación en que el conocimiento actual es insatisfactorio. El investigador científico define, primero, el problema que va a estudiar. Luego recoge información y revisa las

(1): Es decir, que intentan llegar a establecer "leyes" explicativas, a diferencia de las ciencias históricas, que buscan reconstruir y comprender las manifestaciones de la vida social a través del tiempo, de las ciencias jurídicas o normativas, y de la filosofía.

teorías vigentes que le sean útiles para el estudio de ese problema. A partir de ahí se formula la que él cree es la explicación atingente al problema. Esta explicación tiene generalmente la forma de una asociación causa-efecto. Es del tipo "A es causa de B" (o "B es efecto de A") y, por lo tanto, cuando sucede A sucederá B. Este "Proyecto de Ley General" (2) ha sido obtenido en forma deductiva: el investigador se apoya en el amplio marco de referencia constituido por el cuerpo del saber aceptado comúnmente, le introduce las nuevas ideas que él considera relevantes y, mediante un relacionamiento lógico de estos elementos, genera una explicación válida para la situación específica en estudio, explicación que es una implicancia de los elementos que ha considerado. Este es un primer nivel de deducción en la investigación; a partir de un conjunto de explicaciones generales se ha obtenido una conclusión referida a un universo específico, que está formado por las situaciones de aquel tipo que se desea explicar. Así por ejemplo, del postulado de racionalidad en la conducta económica del hombre, y de ciertos supuestos relativos al grado de satisfacción que éste deriva del consumo, se obtiene como consecuencia lógica la afirmación de que el consumo de un bien cualquiera bajará cuando su precio suba. El "Proyecto de Ley General", no tiene existencia aislada, sino se desprende del marco teórico que constituye el acervo del conocimiento vigente que es relevante al problema en estudio. Luego, tomando este "Proyecto de Ley General", el investigador concluye que, *si éste es realmente una ley general*, tendrá que cumplirse en todas y cada una de las situaciones particulares a las que se aplica; selecciona algunos casos particulares del mundo real y comprueba si efectivamente se cumple en ellos esta hipotética ley general. La lógica subyacente en esta prueba conocida también como test, o contrastación empírica, es que si una relación causa-efecto es válida para un determinado universo, se debe cumplir también en cada uno de los elementos de ese universo. La prueba empírica es el segundo nivel de deducción en la investigación científica.

Si los elementos del universo que fueron escogidos para esa prueba empírica, se comportan de manera contraria o incompatible con lo estipulado por el "Proyecto de Ley General", el investigador concluye que éste era equivocado y lo descarta o estudia las modificaciones que sería necesario hacerle. Si, en cambio, los casos reales contra los cuales comparó el "Proyecto de Ley General" se comportan de acuerdo con éste, el investigador concluye que su explicación ha resistido exitosamente una prueba y, por lo tanto, ha ganado en solidez o confiabilidad.

Prácticamente nunca podrá el investigador, llegar a la conclusión de que su "Proyecto" tiene efectivamente la condición de Ley General. Porque esto implicaría que está seguro de que se cumple en absolutamente todos los casos que pretende cubrir, y eso él no lo puede afirmar, a menos que lo someta a prueba empírica en todos y cada uno de esos casos o elementos del universo. Mientras no lo haga, siempre existirá la posibilidad de que en alguna (elemento) o algunas (subconjunto) de las situaciones (universo)

(2): El nombre de "Proyecto de Ley General" ha sido escogido, a pesar de ser muy poco ortodoxo, por su claridad y comprensibilidad. Será más adelante sustituido por el nombre de hipótesis. Es un Proyecto en el sentido de que se postula como una posible Ley General, pero sin saber aún si es valedera.

so) que el "Proyecto" pretende explicar, éste resulte desmantelado por una realidad contradictoria. Por ello es que se dice que las hipótesis científicas, son validadas y aceptadas sólo provisionalmente: siempre existe la posibilidad de que una prueba posterior lleve a rechazarlas. El no rechazo de una hipótesis no implica su aceptación definitiva, sino sólo provisional.

III PRUEBA DE HIPOTESIS DETERMINISTICAS Y PROBABILISTICAS

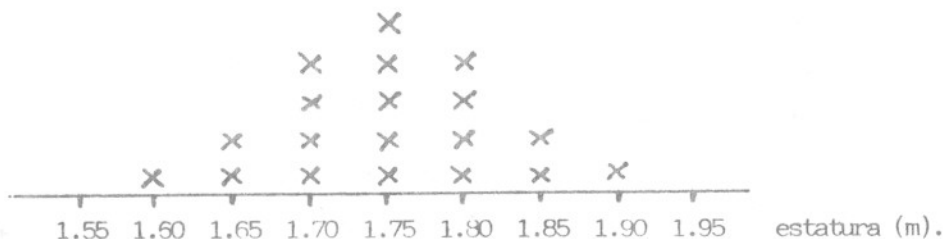
Hasta aquí se ha hablado de "Proyectos de Ley General", o hipótesis, que tienen validez universal dentro del conjunto para el que están definidas, es decir, que si son ciertas se cumplirán siempre, en todos los casos. Estas hipótesis son de tipo determinístico: en todas y cada una de las situaciones en que suceda A, sucederá exactamente B, si es que la hipótesis es verdadera. Pero el mundo que nos rodea no parece obedecer a leyes determinísticas sino, más bien, a leyes probabilísticas. Una ley probabilística postula que, si sucede A, sucederá B o algo o menos parecido, según una distribución de probabilidades centrada en B. Esto queda más claro si se piensa en un ejemplo. Si el problema consiste en explicar de qué depende la estatura de los alumnos de un curso, de tal manera de poder determinar la estatura de un alumno en base a los factores de los cuales ésta depende, es posible formular una hipótesis según la cual la estatura dependa de la edad, sexo, estrato socio económico (E.S.E.) (a través de su influencia en la alimentación en la etapa infantil) y grupo étnico (G.E.). Si esta hipótesis es de carácter determinístico, dirá que un alumno cualquiera del curso, que tenga determinada edad, sexo, E.S.E. y G.E., deberá tener exactamente una cierta estatura, digamos 1,75 cm. Al medir a un alumno de esas características, se deberá encontrar que su estatura es de 1,75 cm. Si resulta tener 1,76 cm. eso significa que la hipótesis era incorrecta. El "Proyecto de Ley General" ha sido rechazado por discordancia con los hechos. Si, en cambio, la hipótesis es de tipo probabilístico, entonces postula que los distintos alumnos que tengan esas mismas características de edad, sexo, E.S.E. y G.E. deberán tener estaturas que se distribuyen probabilísticamente alrededor de 1,75 cm. En caso de que la hipótesis sea correcta, muchos tendrán estaturas cercanas (superiores o inferiores) a 1,75 cm., pocos alumnos de las mismas características tendrán estaturas muy alejadas de 1,75 cm., y el promedio de todos será 1,75 cm. En la figura 1 se muestra un resultado imaginario de la medición de 19 alumnos de 19 años, sexo masculino, E.S.E. medio alto y G.E. nórdico.

ESTATURA (CM.)	Nº DE ALUMNOS
1.60	1
1.65	2
1.70	4
1.75	5
1.80	4
1.85	2
1.90	1

FIGURA Nº 1

Este cuadro se conoce con el nombre de distribución de frecuencias. La información que contiene se puede representar mediante un gráfico conocido como histograma (Figura 2).

FIGURA N°2

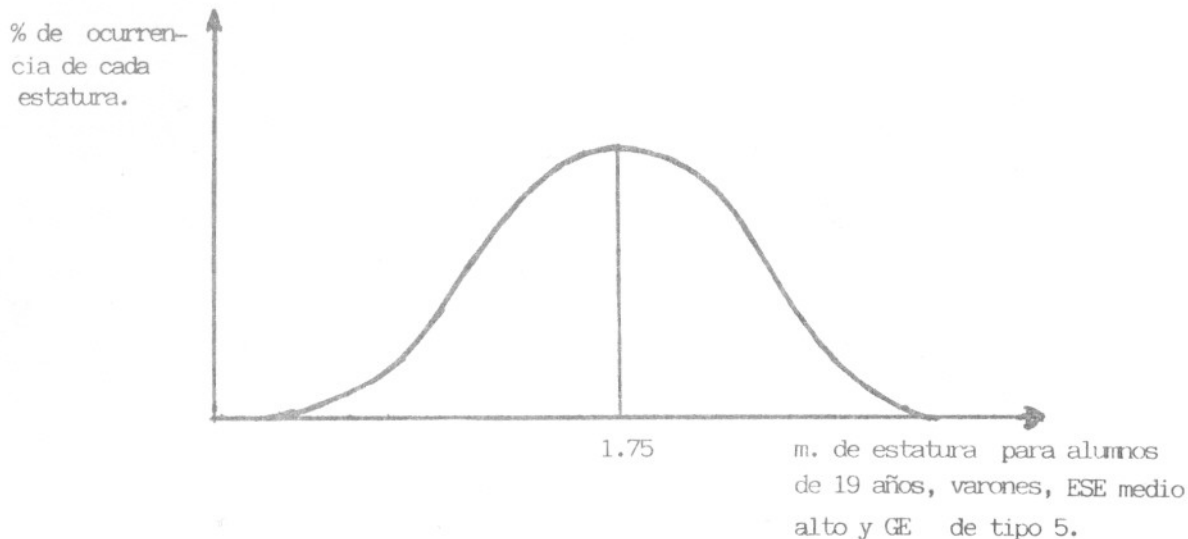


El histograma de la Fig. 2 corresponde a una distribución de frecuencia que podría ser encontrada (aunque rara vez se va a presentar tan "ordenada") en un grupo de 19 alumnos con esas características, extraído de una población mayor.

Pero la hipótesis está definida antes de recoger la información empírica sobre los 19 alumnos. Al ser esta hipótesis probabilística, postula que los alumnos de las características antes mencionadas tendrán estaturas que se distribuyen en forma probabilística, alrededor de 1.75 cm. La hipótesis postula una especie de histograma, pero no empírico sino teórico, y que sigue una curva conocida como distribución de probabilidades. La distribución probabilística más conocida es la llamada "normal", que tiene la forma de una campana simétrica llamada también campana de Gauss, en honor al matemático, físico y astrónomo alemán Karl Friedrich Gauss, quien la desarrolló en gran parte.

En la figura 3, se ve una distribución de probabilidades correspondiente a la hipótesis probabilística que está siendo analizada.

FIGURA N°3



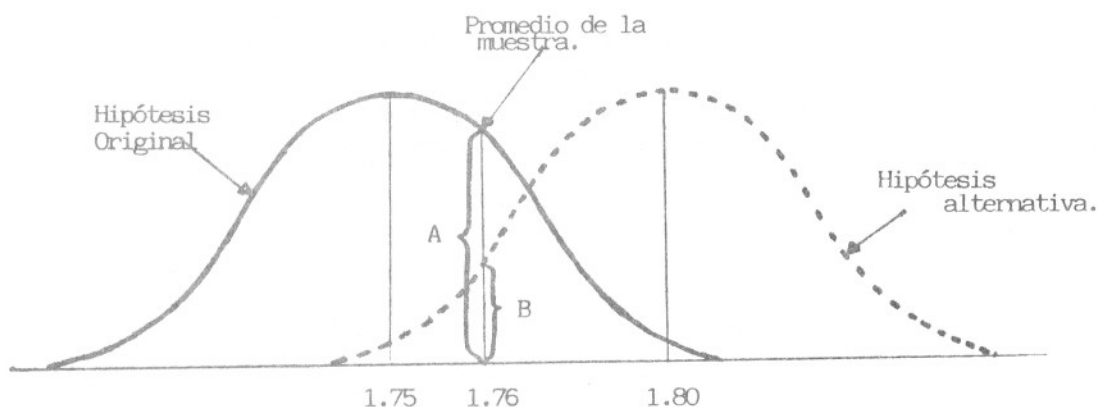
La suma de los porcentajes de ocurrencias teóricas (concepto conocido como "probabilidad") correspondientes a las distintas estaturas tendrá que dar 100, puesto que comprende al total (100%) de los alumnos de las características mencionadas, y el promedio de estaturas teóricas deberá ser 1,75 cm., lo que resulta evidente si se considera que la campana es simétrica (ambos lados son iguales) y su centro está en 1,75 cm.(3).

(3): El análisis gráfico presentado es válido sólo como primera aproximación: la probabilidad total, o 100%, es el área total bajo la curva, y la probabilidad correspondiente a una estatura determinada no es medible, puesto que corresponde al área bajo la curva para esa estatura. Pero esta área es en realidad un trazo (como A y B en las figuras 4 y 5) y un trazo no tiene superficie. No es calculable la probabilidad de un punto (una estatura dada) sino sólo la de un conjunto de estaturas. En la figura 5 la hipótesis se rechaza, porque la probabilidad de encontrar estaturas iguales o superiores a 2.05 cm. es muy baja, y esa probabilidad se mide como el área achurada a la derecha de ese punto y bajo la curva de distribución de probabilidades. El problema se presenta entonces al medir la estatura en forma continua, correspondiendo cada estatura particular a sólo un punto (y no un tramo) de el eje horizontal.

El científico, que ha elaborado su hipótesis probabilística y la ha graficado mediante la figura 3, se encuentra ahora enfrentado a verificar si su hipótesis se cumple en la realidad. Para ello debe obtener información de la realidad y compararla con su hipótesis. Si toma una muestra (subgrupo) al azar de alumnos, y se encuentra con que tiene una estatura promedio de, digamos, 1,76 cm., estará inclinado a considerar que el resultado obtenido es bastante razonable *dada la supuesta veracidad de la hipótesis* y, por lo tanto, aceptar la hipótesis. Este caso se ve en la figura 4.

Naturalmente, él podría equivocarse al aceptarla. La verdadera distribución (que él desconoce, pero acerca de la cual formula su hipótesis) podría estar centrada en 1.80 cm. La probabilidad de encontrar una muestra cuya estatura promedio sea de 1.76 cm., *si el verdadero promedio poblacional es de 1.80 cm.*, es bastante alta. Esta probabilidad mide el riesgo de que el investigador cometa el error (conocido como error β o de tipo II) de aceptar la hipótesis original, cuando la verdadera era una hipótesis alternativa (en este caso, que el centro de la distribución es 1.80cm.). El riesgo --que siempre está presente-- de cometer un error de este tipo es una razón adicional para que la aceptación de cualquier hipótesis sólo pueda ser condicional.

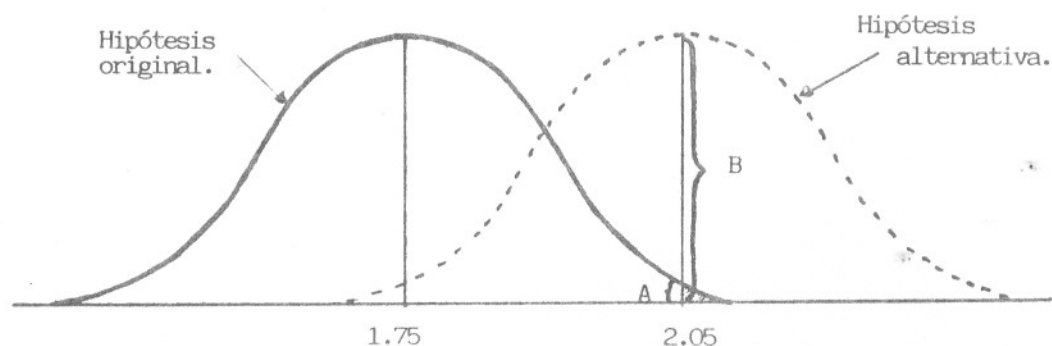
FIGURA N° 4



- A: Probabilidad alta de obtener una muestra de 1.76 m. si la hipótesis original es verdadera. Lleva a aceptar la hipótesis original.
- B: Probabilidad no despreciable de encontrar 1.76 m. si la verdadera es la hipótesis alternativa.

Si en cambio, el investigador encuentra una estatura promedio en su muestra de 2.05 cm., se sentirá inclinado a rechazar la hipótesis; es muy poco probable encontrar una muestra con esa estatura promedio *si la verdadera distribución poblacional tiene un promedio de 1.75 cm.* Pero aquí está también sujeto al riesgo de error: existe una probabilidad (mayor que cero por ciento) de que en una población como la de la figura 3 él extraiga una muestra con promedio 2.05 cm. Y, en ese caso, estaría rechazada indebidamente la hipótesis, como se ve en la figura 5. Este error es conocido como error α o de tipo I.

FIGURA N°5



- A: Probabilidad muy baja de que la muestra tenga promedio 2.05 m., si la hipótesis original es verdadera.
- B: Probabilidad bastante más alta de que la muestra tenga promedio 2.05 m., si la verdadera es una hipótesis alternativa como la representada por la línea punteada. Conduce a rechazar la hipótesis original.

Cuando la forma de la distribución de probabilidades de la población coincide -- por hipótesis o por algún conocimiento previo al respecto-- con alguna de las distribuciones probabilísticas conocidas, tal como la normal, se puede determinar en forma exacta la probabilidad de incurrir en error al aceptar o rechazar la hipótesis. Sin embargo, la forma de la distribución generalmente no es conocida, y se recurre a formular supuestos al respecto, haciéndose entonces menos confiable el test estadístico en sí mismo.

En cambio, cuando la distribución probabi-
lística no es conocida, y también cuando se trabaja con variables no cuantifi-
cables (como personalidad, confianza, etc.), la prueba no podrá ser exacta, y
el investigador deberá evaluar lo más objetivamente posible si los hechos que-
dan o no dentro de un rango de variabilidad que sea compatible con la hipóte-
sis. Esta evaluación, desgraciadamente, nunca estaría completamente exenta de
subjetividad. Sin embargo, este riesgo no invalida la prueba estadística
de las hipótesis, pues ella puede ser mejor en aquellos casos en que una prue-
ba rigurosa requiera transformar las hipótesis a versiones estadísticas que
significarían disvirtuarlas y cuyas conclusiones, por lo tanto, tendrían esca-
sa validez. Es un error pretender descalificar una hipótesis por ser "litera-
ria" (no matemática); muchos de los grandes avances en economía han tenido esa
característica, y sólo algunos estudios posteriores alrededor de esas mismas
hipótesis han permitido la realización de pruebas de rigor estadístico.

Si bien las hipótesis determinísticas son
aceptadas condicionalmente o rechazadas definitivamente, las hipótesis probabi-
lísticas pueden ser también aceptadas, y en tal caso lo serán también condicio-
nalmente, pero si son rechazadas no lo serán definitivamente, sino sólo condi-
cional o provisoriamente.

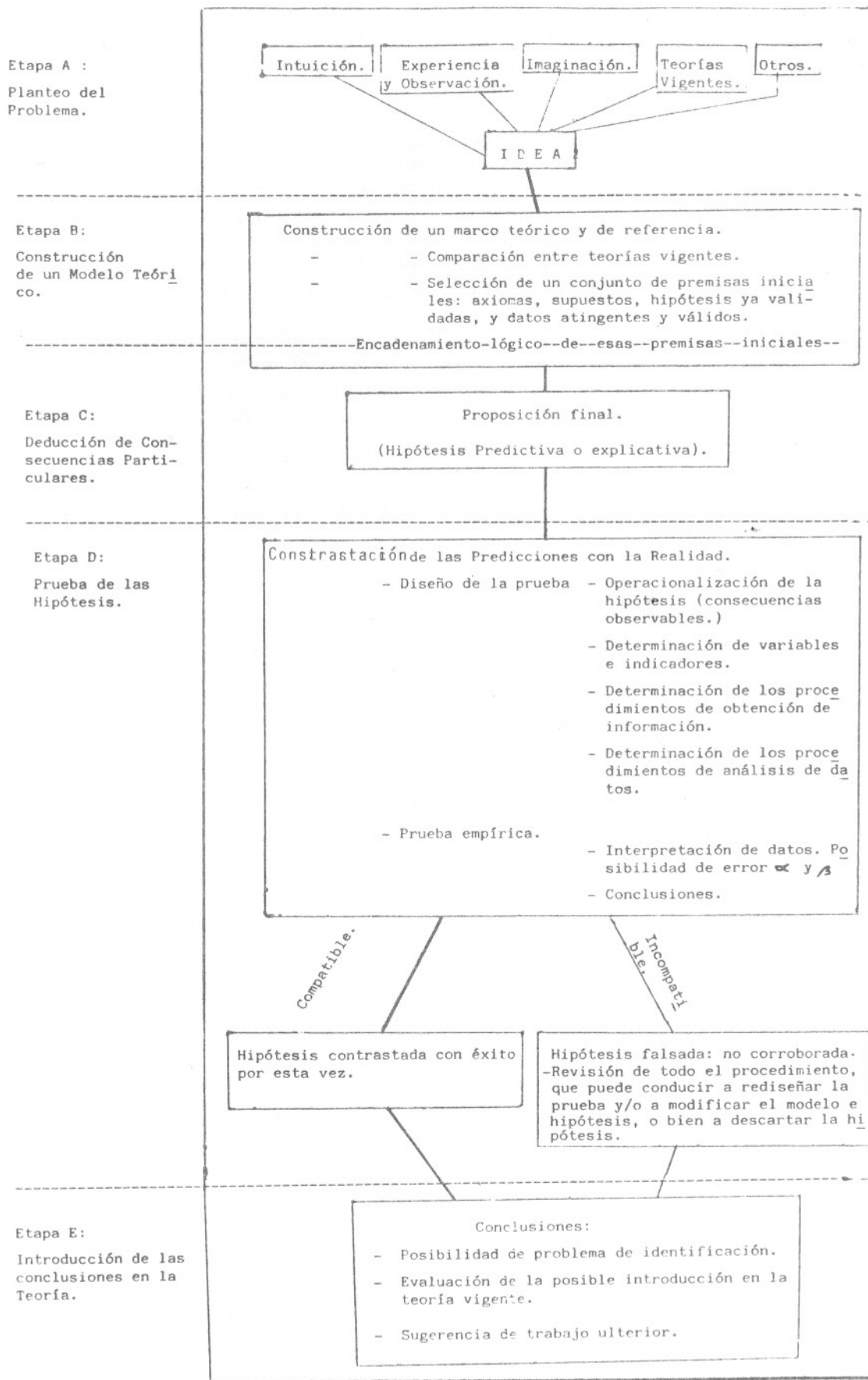
IV CUADRO SINOPTICO DE UNA INVESTIGACION CIENTIFICA

La investigación científica --de la cual se
ha presentado hasta ahora una visión general-- requiere del cumplimiento de ac-
tividades estandarizadas de acuerdo a la lógica que se ha desarrollado hasta a-
quí. Para tener una clara visión de conjunto de esas actividades es útil una
presentación esquemática como la que se muestra en la figura 6, en la próxima pá-
gina.

Acción

	Acepta HN	Rechaza HN
Selección		
Hipótesis	OK	Error 2
ANALISIS	Error 1	OK

INVESTIGACION CIENTIFICA : CUADRO SINOPTICO.



Este cuadro sinóptico es perfectamente compatible con la visión general presentada en las secciones anteriores: primeramente, ante la existencia de una laguna en el saber, el investigador, después de definir el problema a estudiar, echa mano de todos los recursos (intuición, experiencia y observación, imaginación, etc.) para formularse una posible explicación que dé respuesta al problema planteado (idea). En esta etapa es posible, y hasta probable, que se realice un proceso inductivo que, generalmente, no será enteramente consciente, y que consiste en la proposición, como idea de solución, de una generalización de hechos particulares que el investigador ha recogido del mundo que le rodea y ha incorporado como su experiencia. Esta inducción es válida como uno de los caminos para generar la idea de solución, que se postulará como posible respuesta al problema y será sometida a prueba a través de sus conclusiones.

Una vez obtenida esta idea general de solución, es preciso formularla. Para ello se ocupa un conjunto de premisas, que pueden ser: axiomas, es decir, definiciones, y que, por lo tanto, no están sujetos a demostración; supuestos, que son algo que el investigador toma como si se diera en la realidad, que puede que estén bien o mal escogidos en el sentido de que pueden ser consistentes con la realidad o no, y que ayudan a simplificar y hacer manejable el problema; e información no teórica tal como datos atinentes, características institucionales que afectan al problema (por ejemplo, el respeto a la propiedad privada), etc.

A partir de los axiomas y supuestos, el investigador realiza un trabajo lógico que le permite reducir el conjunto de premisas iniciales a una o varias conclusiones finales. Estas (hipótesis) son predicciones del tipo "si sucede A, entonces sucederá B" (y estando toda la relación causal entre A y B estudiada mediante las premisas iniciales y su inter-relación lógica).

Las hipótesis deben tener la característica de ser actual o potencialmente sometibles a prueba y, como se explicó en la sección anterior, son obtenidas deductivamente. Las hipótesis son implicaciones del marco teórico adoptado, que sean verificables y den cuenta del problema en estudio.

Estas hipótesis deben ser sometidas a prueba, es decir, deben ser contrastadas con información empírica que pueda corroborarlas o refutarlas. El resultado de la prueba es la aceptación provisional de la hipótesis, o bien se rechaza. Realizada la prueba y, especialmente, en caso de haber sido rechazada la hipótesis, se debe revisar toda la investigación, a fin de detectar posibles errores en la prueba o en la obtención de la hipótesis o en la construcción del marco teórico, que hubieran podido conducir a aceptar o rechazar indebidamente la hipótesis. En caso de detectar algún error, todos los pasos posteriores a éste deberán ser revisados consecuentemente. Sólo una vez realizadas todas estas tareas puede elaborarse las conclusiones finales del estudio.

V ETAPAS DE LA INVESTIGACION CIENTIFICA

El esquema presentado en el punto anterior, se puede también vertir a un listado de pasos a seguir al realizar una investigación. En este punto se presentará un listado de ese tipo, con cinco etapas de A a E, extractado de Bunge, (4) y a medida que su desarrollo lo vaya haciendo oportuno, se discutirá temas que resulten importantes de considerar. Para la lectura de esta sección es conveniente tener a mano el cuadro sinóptico presentado en la figura 6.

A. Planteo del Problema:

Esta etapa consiste en la detección y organización no rigurosos del problema y su posible solución. Abarca hasta el planteamiento de la idea, en la figura 6.

B. Construcción de un Modelo Teórico:

Esta etapa consiste en la elaboración de un conjunto de premisas (axiomas, supuestos e información no teórica) interrelacionadas que represente lo que el investigador considera medular del problema y de la solución propuesta. Este conjunto constituye un *modelo*, o representación abstracta y simplificada de la realidad.

Para construir el modelo, el investigador debe ocupar supuestos. Debido a que el mundo que nos rodea es demasiado complejo como para poder comprenderlo en su integridad, y existe una pluralidad de interrelaciones que el investigador no es capaz de incluir en su modelo, resulta necesario recurrir a suponer que algunos elementos y algunas relaciones, son especialmente importantes para el problema en estudio, y funcionan de determinada manera. El resto de los hechos que pudieran afectar al problema en estudio se incluye en una especie de supuesto "omnibus" (en el que entran muchas situaciones distintas) conocido como *ceteris paribus*.

Las variables (5) que entran en el *ceteris paribus* se supone que: (a) o no varían, o bien (b) sus variaciones no afec-

-
- (4): Mario Bunge: "La Ciencia, su Método y su Filosofía". Ed. Siglo XX, Buenos Aires, 1972 Págs. 89 a 92. Este listado no debe ser considerado como definitivo y no sujeto a discusión; hay otros autores que presentan una estructuración diferente de etapas. Sin embargo, ha sido considerado un esquema útil para la discusión que se presenta a continuación.
- (5): Al final de esta sección se presenta el concepto de variable y los distintos tipos de variables intervinientes en una investigación.

tan al problema en estudio, o (c) sus variaciones se compensan mutuamente, no produciendo ningún efecto final sobre el problema en estudio.

En el modelo propuesto como ejemplo, acerca de la estatura de los alumnos de un curso, se supone que los alumnos de sexo masculino son más altos que las mujeres, por razones biológicas; que a mayor edad corresponderá una mayor estatura por razones de crecimiento; que la pertenencia a un estrato socio-económico alto significará una mejor nutrición infantil y por ese camino una mayor estatura, y que distintos grupos étnicos tienden a condicionar estaturas distintas. Y se supone que nada más afecta a la estatura. Los supuestos adoptados no son rigurosamente verdaderos; la edad afecta a la estatura sólo hasta determinado límite, y un alumno a los 45 años no será más alto que a los 40. Pero como la abrumadora mayoría de los alumnos de un curso de primer semestre es menor de 25 años, el suponer que existe una relación entre edad y estatura resulta bastante razonable. También hay variables del *ceteris paribus* que pueden estar en realidad afectando a la estatura; por ejemplo: Si se implementa un fuerte programa de becas para basquetbolistas, la estatura promedio es probable que suba por una mayor incorporación de estos deportistas al curso, aún cuando no haya cambios en la edad, sexo, E.-S.E. y G.E.; este sería un caso en que el *ceteris paribus* no se cumple rigurosamente.

Sin embargo, estos supuestos que puede que no sean absolutamente correctos y veraces, son útiles. Sin ellos o algún otro conjunto de supuestos, sería considerablemente más difícil, sino imposible, llegar a conclusiones respecto a la estatura de los alumnos o respecto al problema que esté en estudio.

La selección de los supuestos es crucial en la construcción de un modelo teórico. Una adecuada elección de supuestos permitirá la elaboración de un modelo que represente la realidad en estudio de una forma satisfactoriamente fiel. Existe una larga polémica respecto de los supuestos; muchos autores afirman que éstos deben ser realistas, es decir, que deben asemejarse bastante a la situación real. Milton Friedman tiene una posición esencialmente diferente, afirmando que los supuestos deben ser evaluados únicamente por su capacidad de constituir modelos cuyos resultados sean consistentes con la realidad (es decir, que pasen exitosamente la prueba empírica), y que no tiene ninguna importancia el "realismo" del supuesto considerado aisladamente.

Llevando el argumento al extremo, se podría decir que, si un modelo basado en la influencia de las manchas solares,

ceteris paribus todo lo demás, permite obtener hipótesis que predigan muy bien la estatura de los alumnos, entonces ese conjunto de supuestos queda validado. Un supuesto es bueno cuando permite predecir bien. La posición fridmaniana es sin embargo, discutible, y la selección de supuestos que sean "razonables", en el sentido de ser consistentes con lo que el buen criterio y la experiencia acumulada digan que sucede en la realidad (o mejor aún, que puedan ser y sean, sometidos a test *directo*, y no sólo a través de sus implicaciones, aunque en condiciones normales esta posibilidad se da sólo rara vez), puede dar mayor confiabilidad a las conclusiones. Es cierto que podría elaborarse una argumentación teórica que relacione estatura y manchas solares, pero lo que aparece como más probable es que toda relación entre ambos sea solamente espuria (es decir, sólo casual y no respondiendo a relaciones causa-efecto).

El riesgo de validar -incorrectamente- relaciones espurias sería mucho menor si los supuestos empleados son "razonables".

El modelo constituye una forma estructural, la que debe ser completa y consistente. Es decir, debe conducir a sólo una hipótesis (o conjunto de hipótesis central y auxiliares) o forma reducida.

Un modelo compuesto, por ejemplo, por el siguiente sistema de desigualdades:

$$\begin{array}{rcl} A & > & B \\ B & > & C \\ \hline \therefore C & > & A \end{array} \quad \left. \begin{array}{l} \\ \\ \end{array} \right\} \begin{array}{l} \text{Premisas} \\ \text{Hipótesis} \end{array}$$

es inconsistente: si A es mayor que B y, por lo tanto, que C, eso es incompatible con que C sea mayor que A. El modelo está mal formulado.

Un modelo formado por la siguiente ecuación:

$$A + B = 2$$

está incompleto: los valores de A y B quedan indeterminados y es necesario agregar al menos una relación para completarlo.

Un modelo formado por las siguientes relaciones de implicancia:

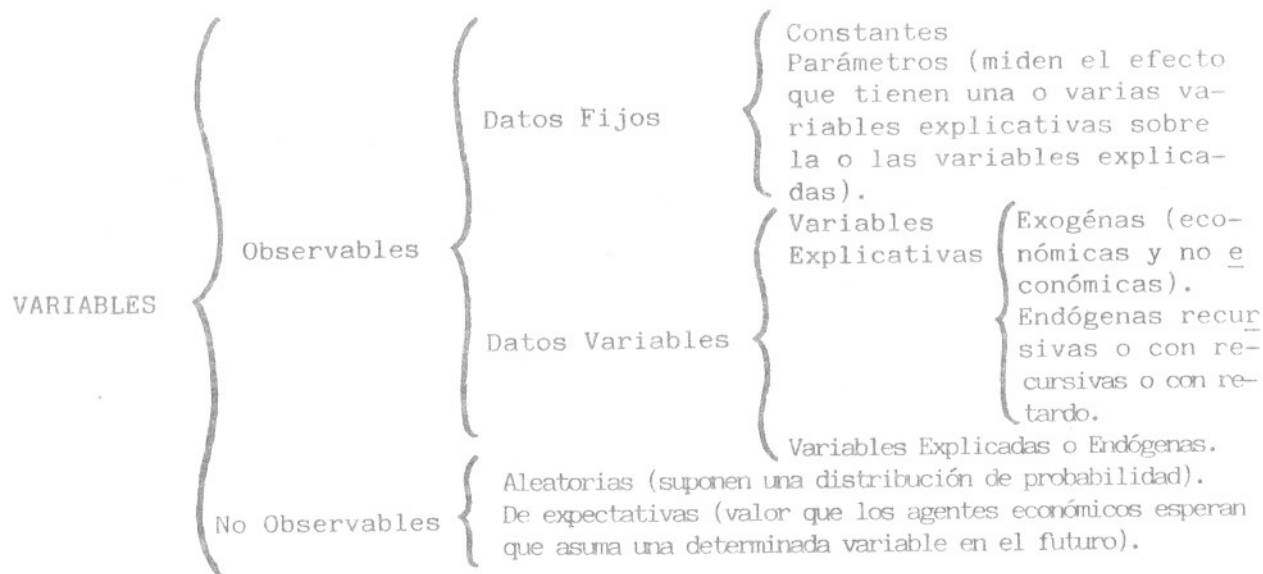
$$\begin{array}{rcl} A & \Rightarrow & B \\ B & \Rightarrow & C \\ \hline \therefore A & \Rightarrow & C \quad (\text{Hipótesis}) \end{array}$$

es en cambio un modelo completo y consistente.

Como se puede observar, el modelo consta de una serie de proposiciones de las cuales se deriva la hipótesis. Esta última también forma parte del modelo. La distinción entre esta etapa de construcción de un modelo teórico y la siguiente, de deducción de consecuencias particulares (o hipótesis), es entonces únicamente analítica, puesto que en el hecho ambas etapas se confunden.

Por último, y siempre dentro de esta etapa de construcción del modelo teórico, el investigador hará un esfuerzo por traducir el modelo un lenguaje formal o matemático. La formulación matemática proporciona una gran ayuda para asegurarse de que el modelo es completo y posee consistencia lógica y, al mismo tiempo, simplifica la obtención de hipótesis cuantitativas, que resultan más fácilmente contrastables con la realidad. Para ello necesitará presentar los sucesos en forma medible. Eso implica definir **variables** (sucesos o eventos que, medidos en distintas circunstancias, registran o pueden registrar distintos valores). La figura 7 muestra una clasificación de estas variables (debe aclararse que, cuando se incluyen los datos fijos, se refiere a que no varían dentro de una aplicación particular del modelo, pero pueden variar entre distintas aplicaciones, por ejemplo en distintos países).

FIGURA N° 7



En el ejemplo sobre estaturas, el modelo es taría constituido por un conjunto de relaciones como:

- 1) Estatura = f (sexo, edad, alimentación que el alumno tuvo en la etapa infantil, herencia).
- 2) Alimentación que el alumno tuvo en la etapa infantil = f (factores culturales y factores económicos).
- 3) Factores culturales = f (estrato socioeconómico pasado).
- 4) Factores económicos = f (estrato socioeconómico pasado).
- 5) Estrato socio económico pasado = estrato socioeconómico actual.
- 6) Herencia = f (grupo étnico).
- 7) (Hipótesis). Por lo tanto, estatura = f (sexo, edad, estrato socioeconómico actual, grupo étnico).
- 8) Dada la forma de las relaciones funcionales del modelo, la hipótesis puede representarse (es decir, no se desvirtúa excesivamente) asignando a la función una forma lineal como:

$$Y = a + b X_1 + c X_2 + d X_3 + e X_4$$

donde:

Y: (= estatura medida en centímetros) es la variable endógena (su valor es generado dentro de, y por, el modelo) o explicada (puesto que su valor resulta explicado por el de las demás variables);

a: es el parámetro de posición de la recta postulada por la hipótesis (y representa la estatura "mínima" no explicada por las variables X_1 a X_4);

b, c, d y e: son los parámetros de relación entre las variables X_1, X_2, X_3 y X_4 , y la variable endógena Y;

X_1, X_2, X_3 y X_4 : son las variables exógenas (cuyo valor es generado fuera del modelo) o explicativas (porque su valor explica o determina el valor que asumirá Y), que corresponde a sexo, edad, E.S.E. y G.E., respectivamente.

Debe hacerse mención aquí a una característica deseable del modelo y de sus conclusiones; debe ser simple y económico.

Una explicación complicada de una realidad tiene un mayor riesgo de incluir errores de una simple y, por otra parte, la inclusión de supuestos más restrictivos de lo necesario le quita generalidad a la teoría.

C Deducción de Consecuencias Particulares:

Una vez construido el modelo teórico, hay que deducir de él, mediante un encadenamiento lógico (representado por el silogismo en la figura 6) un conjunto de consecuencias que puedan ser sometidas a prueba, comparandol~~a~~s con la realidad. Estas son las hipótesis y, normalmente, el investigador deriva de su modelo una consecuencia principal (hipótesis central) y otras auxiliares. Como se explicó en la sección anterior, la o las hipótesis son simultáneas y forman parte del modelo. Se presentan como posteriores sólo para facilitar la diferenciación analítica entre la forma estructural del modelo y su consecuencia o implicación que es la forma reducida (hipótesis).

Las hipótesis deben ser derivadas de la forma estructural teniendo en vista la posibilidad de someterlas a prueba, además de su característica de constituir la conclusión relevante del modelo: respuesta al problema planteado en la etapa A. Deberán ser sometidas a prueba empírica, comparando las predicciones del modelo (hipótesis) con la realidad.

Estas predicciones, que serán del tipo "si sucede A, sucederá B", no siempre pueden ser sometidas a prueba empírica con el carácter riguroso de predicciones; en Economía no se dispone de laboratorios en los que se pueda generar la condición "A" para ver si se presenta la predicción "B". Y si se espera algún momento futuro en que se dé "A", el ritmo de avance de la ciencia puede ser demasiado lento.

Existe una salida para este problema: se puede realizar una especie de "experimento intelectual" que consista en buscar algún momento y lugar en que se haya presentado efectivamente la situación "A" y, sin tener conocimiento previo, verificar si se cumple la "predicción" "B". Esta pasa a ser, entonces, más que una predicción, una retrodicción, y más que afirmar que la hipótesis predice bien o mal (pues es o no consistente con los hechos), se dice que *explica* bien o mal. Se puede *predecir* eventos que no han acaecido aún; los eventos que ya sucedieron no pueden ser predichos sino sólo explicados.

D Prueba de las Hipótesis:

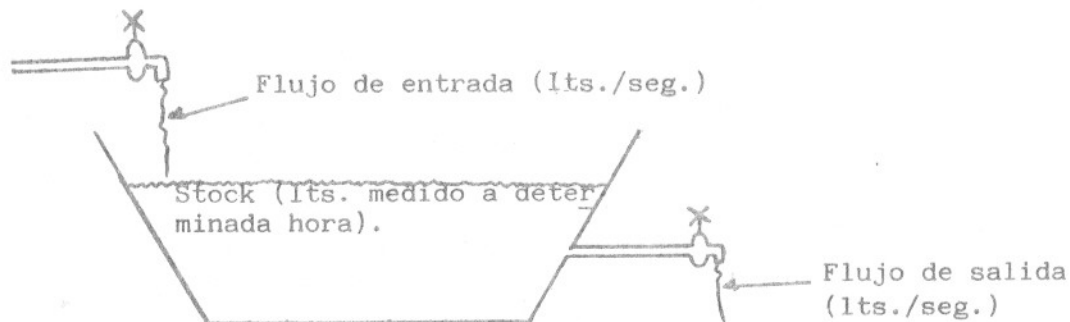
La contrastación de una hipótesis con la evidencia empírica requiere primeramente de un diseño de la prueba (o, del inglés, test).

En ese diseño hay que dar a las variables, una definición operacional. La variable edad, por ejemplo, se medirá en meses completos transcurridos desde el momento del nacimiento, y descartando las fracciones de mes. Esta definición operacional implica casi siempre desvirtuar en cierto grado la variable definida teóricamente. La definición operacional del E.S.E., introducido en el modelo de estaturas de los alumnos, considerará elementos como ingreso familiar, barrio en que vive, colegio en que se educó, etc. Pero con todos estos elementos no se logra una clasificación exacta y precisa el estrato socioeconómico al que pertenece el alumno. Habrá casos, por ejemplo, de alumnos pertenecientes a familias de larga trayectoria social pero con una situación económica sólo regular, y que debieran haber quedado ubicados en un E.S.E., más alto que el que resulta de la aplicación de los criterios objetivos -principalmente económicos- de asignación de valores para esta variable. El caso de esta variable presenta además la característica de no ser cuantitativa (como lo es, por ejemplo, un precio expresado en pesos, o una edad en meses) sino cualitativa. Un valor 2 para esta variable (perteneciente al E.S.E. 2) no significa el doble que un valor 1. Este tipo de variables es manejado mejor (aunque no exclusivamente) mediante una rama de la estadística, la estadística no paramétrica.

Un caso particularmente interesante de dificultades en la definición operacional de las variables se presenta en estudios tales como elaborar una función de producción, en que se postula que la cantidad producida es función de la cantidad de insumos empleados. Entre los insumos para producir un bien está el capital. Por razones prácticas, éste normalmente será medido como un stock o acervo (6): la variable capital estará definida como la cantidad de bienes de capital (máquinas, edificio, etc.) que la empresa posee en un momento dado, valorado a valor libros. Pero no es el stock de capital el que participa en la producción (la cual es un flujo de unidades producidas por día), sino el flujo: horas-máquinas efectivamente ocupadas en la producción. Las máquinas inactivas, que están incluidas en el stock de capital, no contribuyen a la producción. Y una función de producción en que se defina la variable capital como stock dará una mala explicación para la variable producción, a no ser que, por una rara situación, el flujo de capital sea justamente una proporción constante del stock que la empresa posee (por ejemplo, que a lo largo del tiempo que cubra el estudio siempre se haya ocupado el 50% de las máquinas existentes, y durante 35 horas a la semana).

(6):: Existe diferencia entre una variable stock y una variable flujo. La primera se refiere a la cantidad total de lo que se esté midiendo, existente en un período del tiempo. En cambio, una variable flujo mide la cantidad que pasa por un punto determinado de medición *por cada unidad de tiempo*. Esto resulta más claro con el ejemplo clásico del estanque, en la figura 8.

FIGURA N° 8



Además de definir operacionalmente las variables, el investigador deberá dentro del diseño de la prueba, determinar qué tipo de información ocupará para la contratación empírica. Puede recolectar información especialmente para realizar la prueba en cuestión y, en este caso, se dice que ocupa *información primaria*, o bien, puede aprovechar datos recogidos en forma independiente a esta prueba, sea como estadísticas realizadas periódicamente (cálculo de IPC, PNB, índice de desempleo, etc.) o para probar otras hipótesis distintas. Este tipo de información es conocido como *información secundaria*.

La información que se ocupe, sea primaria o secundaria, deberá incluir una cantidad de observaciones (correspondientes a elementos o a subconjuntos del Universo para el cual está definida la hipótesis). Estas observaciones, para que la prueba sea válida, deben ser escogidas de manera de minimizar el riesgo de que correspondan sólo a valores extremos de la respectiva distribución. En el ejemplo de la estatura de los alumnos del curso, no sería correcto escoger a los integrantes del equipo de basquetbol para contrastar la hipótesis. Por ello se define una muestra (subconjunto de la población) *aleatoria*, es decir, escogida al azar, de modo que cada elemento de la población tenga una probabilidad idéntica de aparecer dentro de la muestra. La determinación de esa muestra presenta dificultades, pues requiere, entre otras cosas, contar con un listado completo de la población, dentro del cual seleccionar al azar, o definir algún otro procedimiento que dé igual garantía de aleatoriedad. El tamaño de la muestra depende de la dispersión de las variables que se desea medir; si todos los alumnos son idénticos, una muestra de uno solo representará perfectamente a la población,

mientras que si son muy disímiles se requerirá de una muestra mayor para tener una confianza razonable de que sea representativa de las características poblacionales. En determinadas aplicaciones, es posible reducir el tamaño de la muestra y, por lo tanto, el costo de obtención de la información, utilizando una muestra estratificada.

Cuando se dispone de información recogida mediante un censo, se está trabajando con el universo total y no con una muestra. Sin embargo, debe tenerse en cuenta que, si la hipótesis no está referida al mismo momento y población que el censo, sino por ejemplo, es una hipótesis atemporal, válida para cualquier momento, la información censal representa sólo a un subconjunto del verdadero universo, compuesto por el mismo sujeto censal pero en muchos momentos posibles. En este caso el censo sería sólo un subconjunto del universo al que está referida la hipótesis, asimilable al concepto de una muestra de gran tamaño.

La información ocupada para la contrastación de las hipótesis puede ser obtenida de un subconjunto de la población tomado en un momento dado, de tal manera que las variaciones en los valores de las variables, necesarias para establecer la relación entre ellas (7), resulten de comparar distintos elementos de la población en un mismo momento. En este caso, se habla de un estudio de *cross-section*, palabra inglesa cuya traducción literal de "sección cruzada" arroja quizás menos luzes que si se define como "corte en un momento dado del tiempo" o "corte transversal".

La alternativa opuesta al *cross section* consiste en recoger la información en un grupo o subconjunto de la población, (que podría también ser la población completa), pero en distintos momentos del tiempo. Este sería un estudio de *serie de tiempo*. Ultimamente, además, han aparecido métodos que permiten considerar información provenientes de grupos diferentes de elementos y en momentos distintos del tiempo, formando un *pooling* o "fondo común" de datos.

La selección entre estos tipos de información dependerá básicamente de las características de la hipótesis que se busca contrastar, de la disponibilidad de distintos tipos de datos, y del presupuesto de que se disponga.

(7); Esta variación es necesaria si se piensa en el ejemplo de la estatura de los alumnos, en el cual ésta depende de cuatro variables. Si en la muestra se encuentra sólo alumnos de una misma edad, por ejemplo, resulta imposible detectar el efecto de la edad sobre la estatura (este efecto es medido por el parámetro "c" que equivale a la pendiente $\frac{\Delta Y}{\Delta X_2}$

Para detectar ese efecto es necesario encontrar alumnos con distintas edades y distintas estaturas.

La información recogida debe ser procesada a fin de proceder a la contrastación de las predicciones (o retrodicciones) de las hipótesis con la realidad. Esto implicará preparar tabulaciones, tablas cruzadas, distribuciones de frecuencia y otras formas de presentación de datos.

Además, esa información debe ser evaluada. En el largo y complicado proceso de recolección de la información puede introducirse una multiplicidad de errores, los que se suman a los implicados por las probables diferencias entre la definición teórica y la definición operacional de las variables, todo lo cual puede limitar su capacidad de servir para aceptar o rechazar (provisionalmente) las hipótesis. Entre los errores es dable mencionar fallas de diseño del cuestionario o influencia del encuestador, que tiendan a inducir un determinado tipo de respuesta en el encuestado o a producir faltas de respuesta con sesgo (es decir, que no sean neutras, sino tiendan a corresponder a alguna de las alternativas de respuesta. Por ejemplo ante la pregunta ¿sostiene Ud. relaciones extramaritales? podría esperarse que muchas "no corresponde" ocultarán un "si"). Además, en el proceso de codificación y tabulación puede cometerse una cantidad importante de errores humanos.

Aceptando que la información recogida sea de buena calidad, y corresponda a los conceptos teóricos que desea medir, procede realizar la prueba misma, es decir, comparar las hipótesis con los hechos reales. Aquí también existe un riesgo de error; la prueba estadística para aceptar la hipótesis propuesta, o rechazarla (según los criterios mostrados en el punto III de este trabajo), ocupa ciertos supuestos sobre el comportamiento de las variables, como por ejemplo, el de que la población se distribuye de acuerdo a determinada función de distribución probabilística (que puede ser la normal u otra), y dados esos supuestos determina cuán probable es el haber obtenido los valores que se encontró en la muestra si es que es verdadera la hipótesis propuesta, de manera aceptar o rechazar esa hipótesis según si esa probabilidad sea alta o baja. Si el supuesto de normalidad de la población, o algún otro supuesto del test estadístico, no se cumple, la probabilidad de haber obtenido esa muestra estará mal calculada, y no servirá como criterio de aceptación o rechazo. Este tema se vuelve a tocar en la sección VI de este trabajo.

E. Introducción de las Conclusiones en la Teoría.

Si una hipótesis es aceptada, el modelo gana en confiabilidad, y si esta aceptación se produce en pruebas reiteradas, va siendo incorporada a la teoría vigente, generalmente aceptada, en tanto que, en caso de haber sido rechazada, se hace necesario revisar todo el proceso.

En caso de que a raíz de la revisión se efectúen modificaciones al modelo, las nuevas hipótesis resultantes deben ser sometidas a todo el proceso de comprobación, teniendo un tratamiento similar

al de una hipótesis absolutamente nueva. Esta validación no puede realizarse con la misma información ocupada para la primera prueba. La razón es que, si los resultados del test llevan a un ajuste en el modelo, el investigador introducirá en éste cambios que lo hagan consistente con los datos empíricos hallados. Si se vuelve a someter a prueba con esos mismos datos que fueron considerados para modificar el modelo, evidentemente éste será consistente, *por construcción*, con los datos empíricos. Esa no es, obviamente, una prueba válida, siendo necesario recopilar nueva información empírica para el test.

El hecho de que una hipótesis pase exitosamente el test no necesariamente implica que la forma estructural de la cual fue derivada haya quedado validada (8). Ello es debido a la posible existencia de un *problema de identificación*. Este consiste en que es posible que más de una forma estructural (o modelo) arroje una misma forma reducida (hipótesis). Cuando ello sucede, el investigador no tiene seguridad de si está validando la forma estructural que él propuso y que generó la hipótesis que sometió a test, o está validando una forma estructural distinta que genera la misma hipótesis. La relación entre la forma reducida y las formas estructurales de las cuales fue derivada se muestra en la figura 9.

FIGURA N°9



(8); El argumento que aquí se presenta es independiente del carácter de condicional de la aceptación, y consiste en que no hay seguridad de que, por el hecho de haber pasado la prueba empírica, la teoría deba ser aceptada condicionalmente.

Si la forma reducida no pudo haber sido obtenida de ninguna forma estructural, entonces resulta subidentificada: hay un error lógico que consiste en que la hipótesis no proviene de ningún conjunto de premisas. Si la forma reducida pudo haber sido obtenida sólo de una forma estructural, resulta estar exactamente identificada: sólo un modelo es capaz de generar esa hipótesis, y aceptar la capacidad predictiva o explicativa de ésta equivale a aceptar ese modelo. El problema de identificación se presenta cuando la forma reducida puede provenir de más de una forma estructural: hay más de un modelo que arroje la misma predicción; cuando se somete a prueba esta última, no se sabe cuál es el modelo correcto, de entre los que generan la misma forma reducida.

Un ejemplo ilustrador de un problema de identificación es el de dos investigadores, A y B, que tratan de elaborar un modelo explicativo respecto del suicidio. El investigador A postula que la gente se suicida por razones económicas, y como las crisis de los problemas de negocios progresan durante los fines de semana y culminan el primer día hábil ante el "no" del banco, entonces habrá más suicidios los días lunes que el resto de la semana. El modelo del investigador B, en cambio, afirma que los suicidios son por razones sentimentales, y que los conflictos de pareja se desarrollan principalmente durante los días en que no se trabaja, de tal manera que los enamorados se ven enfrentados a su dura realidad al terminar el fin de semana y no pudiendo soportarla se suicidan el día lunes en un porcentaje mayor que los demás días. Si, al someter sus modelos a prueba, verifican que efectivamente el porcentaje de suicidios de los días lunes es mayor, tanto A como B, independientemente, creerán ver validado su modelo. Pero no está claro cuál de ellos es el verdaderamente correcto (o si lo son ambos, o ninguno de los dos y el verdadero es un tercero que, por ejemplo, postula que los suicidios son fruto de desórdenes mentales, que se presentan indistintamente en cualquier día de la semana, y que los lunes aparecerán registrados los suicidios del mismo día y además algunos del sábado y domingo que no habían sido descubiertos antes).

Existiendo el riesgo de un problema de identificación, y ya que no es factible explorar todas las formas estructurales posibles que hubieran generado una forma reducida igual a la que se está sometiendo a prueba, es importante tratar de probar el mayor número posible de implicancias del modelo, es decir, el mayor número posible de hipótesis laterales. En el ejemplo visto, el investigador A postularía además que la mayoría de los suicidios corresponderán a hombres (que participan más activamente que las mujeres en el mundo de los negocios), mientras que B afirmará que habrá aproximadamente tantos suicidios masculinos como femeninos, y muchos de ellos serán conjuntos. Así como las hipótesis laterales ayudan a discriminar entre teorías alternativas con una misma hipótesis, y a dar carácter más probable de única a cada teoría, también una revisión de cuán razonables sean los supuestos en sí mismos (o, mejor aún, una prueba empírica de éstos) ayuda a evitar el problema de identificación. Si resulta que los bancos estudian la si-

tuación de sus clientes en reuniones que se efectúan sólo los días martes, el modelo de A se verá debilitado respecto del de B. Este es un aspecto adicional que debe ser considerado en la discusión sobre la posición friedmaniana respecto de los supuestos.

Un punto final de la investigación consiste en la elaboración de las conclusiones finales, teniendo en cuenta todo lo anterior, y de sugerencias acerca del trabajo ulterior que puede ser realizado a base de las conclusiones obtenidas.

Una consideración general sobre la filosofía de todo el método, es que el investigador debe realizar un serio esfuerzo por concluir que su modelo está errado, y sólo después de someterlo a prueba "dura" puede afirmar que le ha resultado imposible destruirlo, y que por lo tanto, es probable que sea verdadero. Esta actitud es difícil de adoptar, pues un investigador tiende a encariñarse con su modelo explicativo, pero es la única verdaderamente científica.

Cuando la hipótesis ha sido sometida a prueba dura repetidas veces y ha resultado exitosa, y el riesgo de que esté presente un problema de identificación parece ser muy bajo, se justifica incluir dicha hipótesis, y el modelo del cual se desprende, en el marco de la Teoría comúnmente aceptada. Si el modelo validado se interrelaciona con otros, formando un conjunto de proposiciones explicativas, lógicamente articuladas, y que pueden ser validadas empíricamente, constituyen una teoría. De una teoría se desprende una cantidad de hipótesis, y la validación de algunas de ellas va dando confiabilidad a la teoría, aún cuando no todos los aspectos de ésta hayan sido sometidos a prueba.

VI EL ANALISIS DE REGRESION

Existe una técnica matemático - estadística que es ampliamente utilizada en Economía, y que merece una consideración aparte.

Debe aclararse que esta técnica, bastante poderosa desde un punto de vista cuantitativo, no es la única técnica usada en economía para la verificación de hipótesis sobre relaciones entre variables, y que tiene limitaciones impuestas por la necesidad de llevar las hipótesis a la forma de relaciones funcionales entre variables cuantificables, lo que en muchos casos implica desvirtuar esas hipótesis a un punto tal que lo que se somete a prueba es sólo una especie de caricatura de ellas. Eso hace que sea una técnica bastante discutida entre los economistas. Sin embargo, y a pesar de sus fallas, es una de las mejores técnicas de prueba disponibles, y una de las más frecuentemente empleadas.

cas son del tipo:

$$Y = f (X_1, X_2, \dots X_n)$$

con Y como variable dependiente de las variables exógenas X_1 a X_n .

Para poder hacer una representación gráfica se considerará solamente la exógena X_1 , haciendo abstracción de las demás. Sin embargo, las conclusiones que se obtenga gráficamente pueden ser generalizadas algebraicamente -aunque sean difíciles o imposibles de dibujar por el número de ejes- a más variables exógenas.

La hipótesis mostrada en las secciones anteriores podría quedar reducida a:

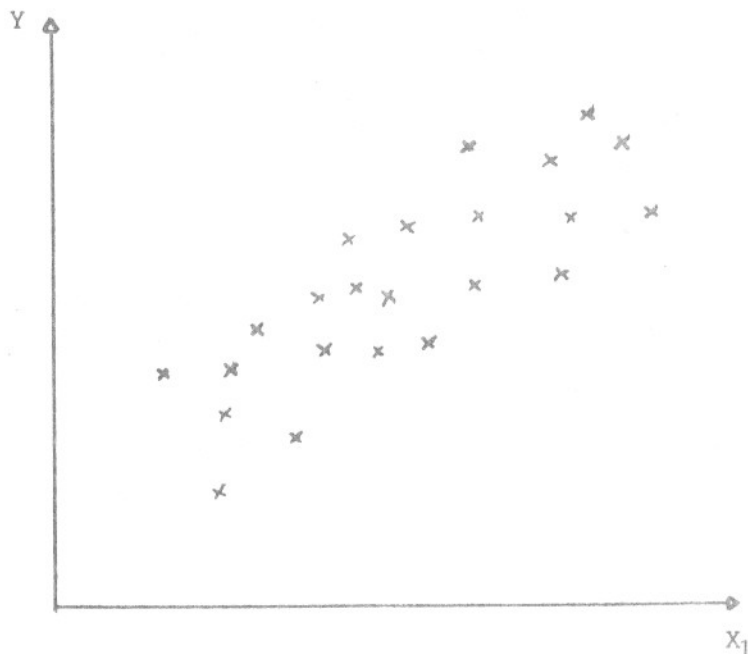
$$Y = f (X_1) \quad \text{con } Y = \text{estatura.}$$

$$X_1 = \text{edad.}$$

y con una f tal que, a mayor X_1 mayor Y.

Las medidas efectuadas en un grupo de alumnos se pueden representar gráficamente, correspondiendo cada punto al par de valores estatura-edad de un alumno en particular (fig. 10), obteniéndose una "nube de puntos".

FIGURA N°10



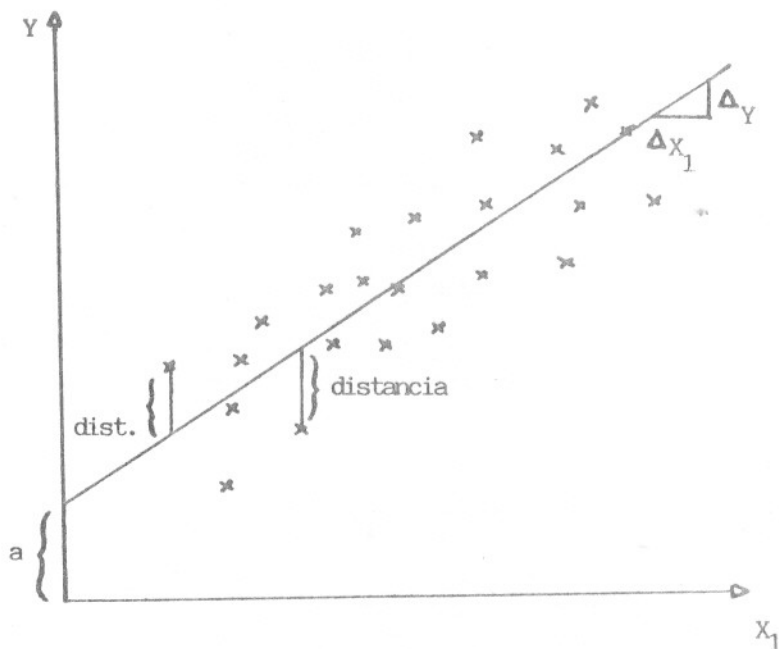
Esta nube de puntos, si la hipótesis es correcta, es una aproximación a la función $Y = f(X_1)$. Forzando un poco el ejemplo (puesto que el crecimiento o ΔY por cada año de edad es distinto según la edad, de manera que $\frac{\Delta Y}{\Delta X}$ no es constante a lo largo de la función)

supondremos que la función hipotetizada es una recta:

$$Y = a + b X_1 \quad \left| \quad \text{con} \quad b = \frac{\Delta Y}{\Delta X_1} \quad \text{constante} \right.$$

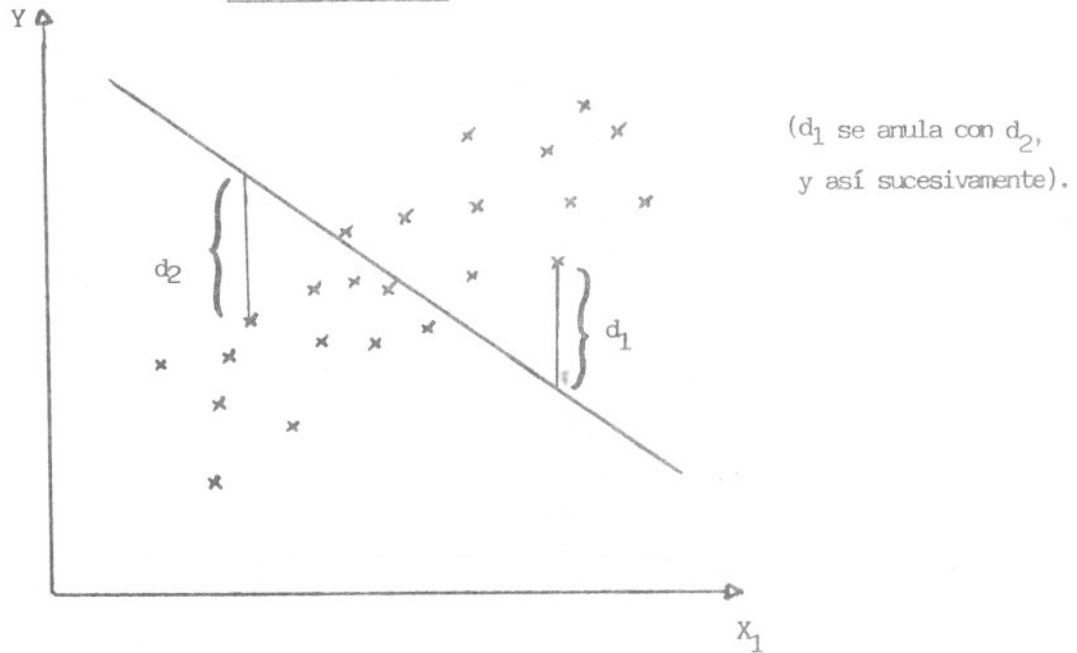
En esta situación es posible encontrar aquella recta que mejor represente a la nube de puntos, en el sentido de que las distintas verticales entre los puntos y la recta sean mínimas. (Fig. 11).

FIGURA N°11



Si se consideran únicamente las distancias verticales, se tendría que, para cualquier recta que pasara por el centro de la nube de puntos, las distancias positivas se anularían con las negativas, resultando que la suma de las distancias sería cero (Fig. 12).

FIGURA N°12



Pero la recta dibujada en la figura 12 indudablemente no es la que mejor representa la nube de puntos. Por ello lo que se minimiza no es el conjunto (o suma) de las distancias, sino la suma de estas distancias elevadas al cuadrado. Por eso el método es conocido con el nombre de "método de los mínimos cuadrados", aunque más comúnmente se le da el nombre de *regresión* (9).

La recta de regresión, entonces, será aquella que minimice la suma de los cuadrados de las distancias de los puntos a la recta.

Pero, como cada punto no se encuentra precisamente en la recta, sino que estará a alguna distancia de ella, si se pretende representar cualquier punto mediante la línea de regresión, será necesario agregarle un término de error, el que represente esa distancia. La línea de regresión quedará entonces como:

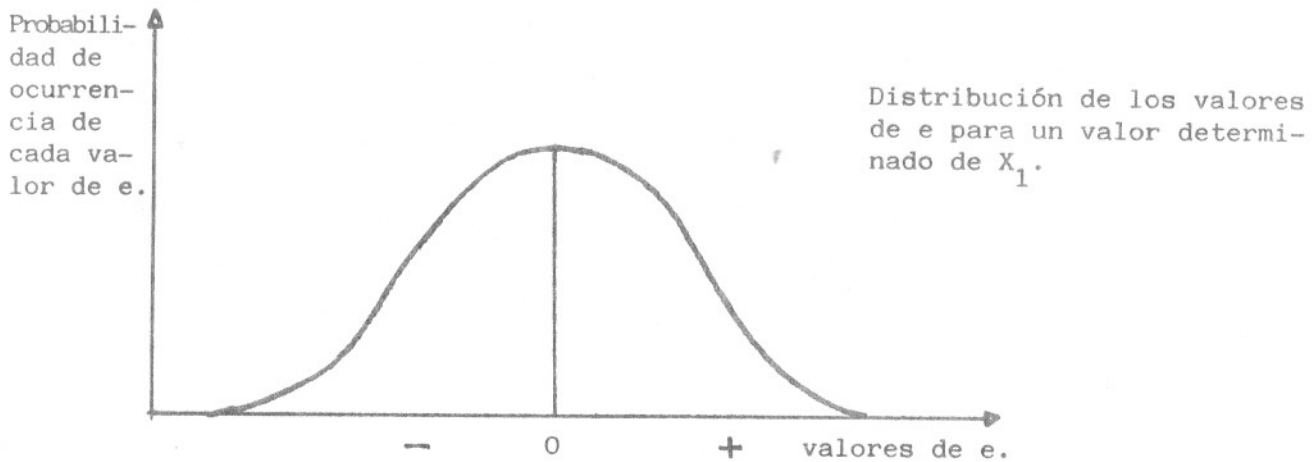
$$Y = a + bX_1 + e$$

con e = distancia entre cada punto y la recta.

La variable " e " es una variable aleatoria (ver la fig. 7). Eso quiere decir que los errores pueden tomar distintos valores, y que esos valores pueden ser caracterizados mediante una distribución de probabilidades (Fig. 13).

- (9): Si bien el modelo de análisis de regresión más comúnmente utilizado está definido sólo para rectas, es posible ocuparlo en una cierta cantidad de curvas, mediante el truco de transformar estas curvas en rectas, principalmente mediante la aplicación de logaritmo: como ejemplo, es posible transformar: $Y = X_1^2$ en: $\log Y = 2 \log X_1$, lo que es una recta, si se representa en los ejes $\log Y$ y $\log X_1$.

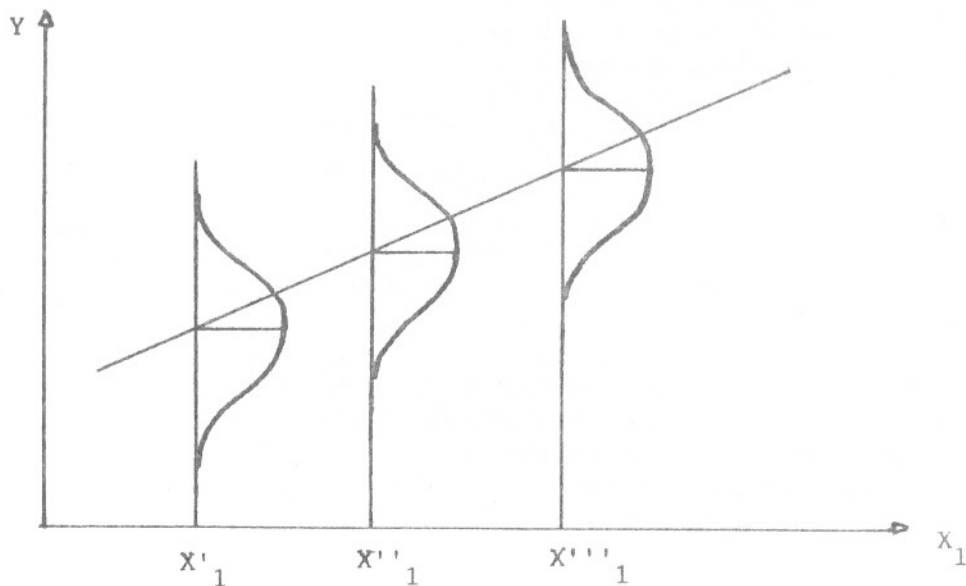
FIGURA N°13



Los distintos valores que asume el término de error para un valor dado de X_1 siguen una distribución de probabilidades normal centrada en cero (es decir, el promedio de los errores será cero), y con la característica de que resulta más probable encontrar errores pequeños (valores cercanos a 0, positivos o negativos) que grandes.

Al mostrar la distribución de los errores para distintos puntos de la recta de regresión, es decir, para distintos valores de X_1 se obtiene un gráfico como el de la figura 14, en que la altura de un punto de la campana representa la probabilidad de obtener ese valor del error (ver nuevamente la nota 3).

FIGURA N°14



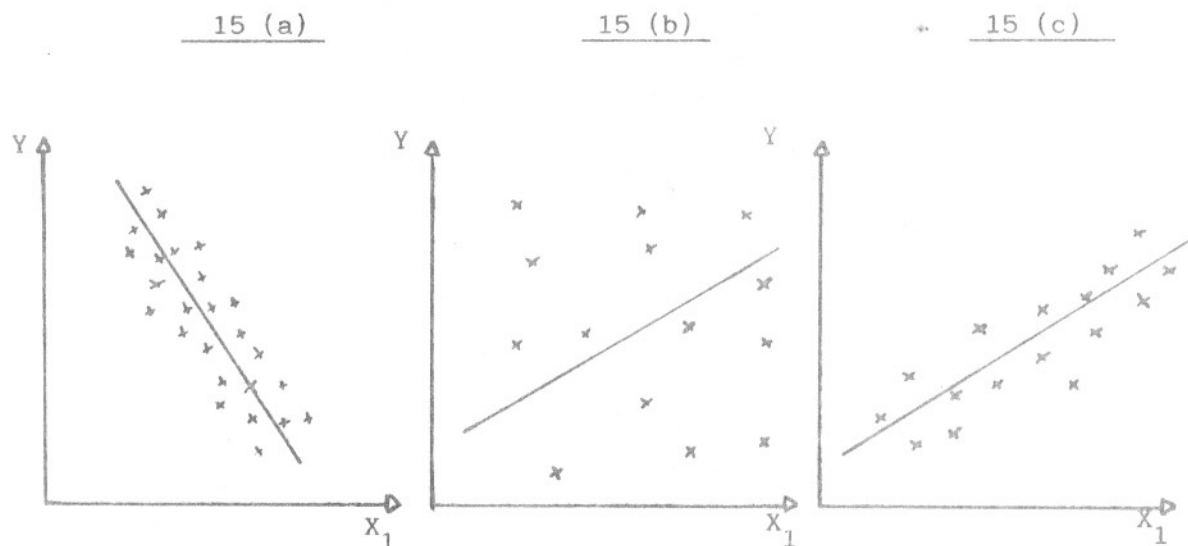
Volviendo a la hipótesis, ésta postula que:

$$Y = a + b X_1 \quad \text{con} \quad b > 0$$

es la verdadera característica poblacional, y que los puntos encontrados empíricamente podrán discrepar de esta recta, por la variabilidad normal de los hechos económicos (10), pero que esos errores no serán demasiado grandes (puesto que, en caso de ser así, significaría que la recta hipotetizada no da cuenta satisfactoria de la situación, existiendo discrepancias entre la recta y la realidad, o *residuos no explicados* por la recta, demasiado grandes).

Para aceptar la hipótesis, es necesario que los parámetros a y b de la recta ajustada a la nube de puntos empíricos (parámetros cuyo valor resulta de encontrar aquella recta que minimice los cuadrados de las distancias) tengan valores compatibles con la hipótesis (en este caso, $b > 0$), y además que el ajuste obtenido sea bueno, es decir, que los errores o residuo no explicado por la función, sean pequeños. En la figura 15 (a) se muestra una nube de puntos que lleva a rechazar la hipótesis por que el valor de " b " es distinto del hipotetizado, en 15 b se observa una nube de puntos en que los errores son demasiados grandes, llevando a rechazar la hipótesis, y en 15 c se ve una situación en que la hipótesis es aceptada.

FIGURA N°15

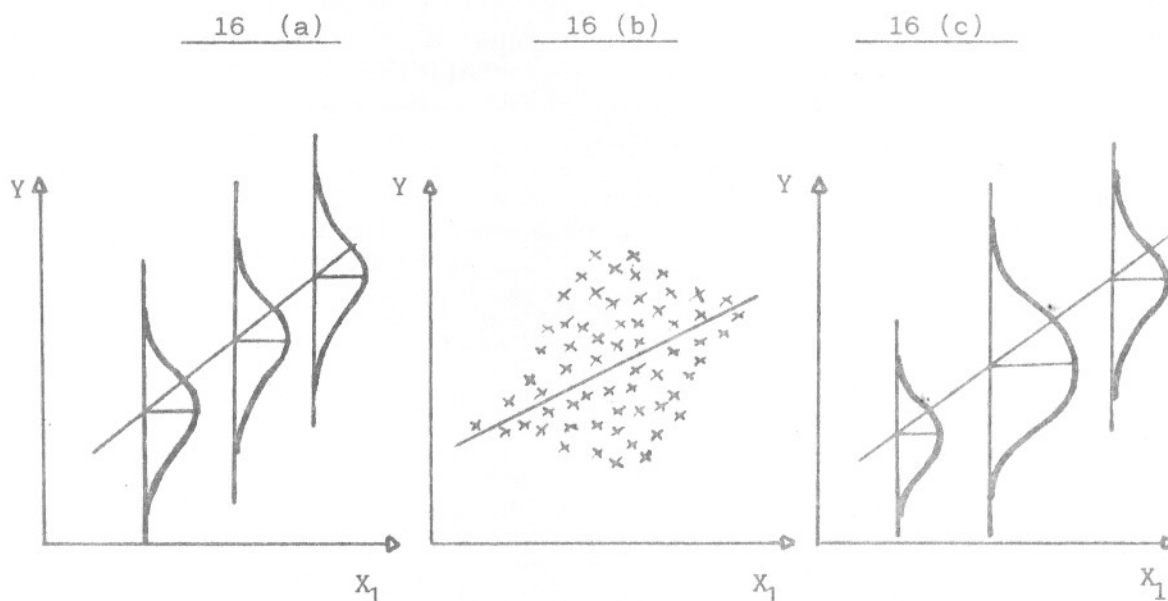


(10); La inclusión del término de error es una aceptación de que el supuesto del *ceteris paribus* no se cumple plenamente, de tal manera que el valor de Y es el que postula la ecuación para un X_1 dado, más/menos un residuo que se debe a la acción de las variables no controladas, es decir, no incluidas en el modelo. El supuesto del *ceteris paribus* quedaría entonces transformado a: "suponiendo que las demás variables, o permanecen constantes, o no afectan a la variable dependiente, o su efecto combinado resulta en un error aleatorio que sigue una distribución normal con media (o promedio) cero".

Retomando el punto referente a los supuestos de los test estadísticos, debe mencionarse algunos de los supuestos del análisis de regresión.

El supuesto de homoscedasticidad implica que la distribución de los errores es constante a lo largo de la función. La figura 16 (a) muestra un caso en que este supuesto se cumple, mientras que en 16 (b) y (c) se grafica una situación en la que no hay homoscedasticidad: la dispersión de los errores varía a lo largo de la línea de regresión.

FIGURA N°16



El supuesto de independencia lineal de las variables exógenas significa que, cuando hay más de una variable independiente, ninguna de ellas puede ser expresada como una función lineal de otra u otras. Si tal situación existiera (por ejemplo, $X_1 = 2 X_2$) resultaría imposible aislar el parámetro de inclinación correspondiente a cada una de ellas $\left(b = \frac{\Delta Y}{\Delta X_1} \text{ ; y } c = \frac{\Delta Y}{\Delta X_2} \right)$, y que mide el efecto de cada una sobre la variable dependiente.

Cuando estos, u otros, de los supuestos del modelo no se cumplen, no resultan válidos los test estadísticos que se ha diseñado para medir si los errores son demasiado grandes como para poder aceptar la hipótesis, o si el efecto de cada una de las variables considerada por separado es significativo (es decir, su parámetro de inclinación es distinto de cero más allá de lo que podría explicarse por simple azar), y es necesario buscar soluciones -que no siempre existirán- en el campo de la Econometría, que se ocupa de las aplicaciones del análisis de regresión.

APENDICE: EJERCICIOS

- 1) El proceso de contrastación de las teorías en Economía, consiste en verificar la consistencia de la teorías entre sí. Comente.
(Prueba global de repetición, 23/12/82).
- 2) Las teorías económicas no ofrecen ninguna garantía de representar adecuadamente la realidad, porque ocupan supuestos los cuales no son sometidos a prueba. Comente.
(Prueba global, 09/12/82).
- 3) Un conocido economista realizó un estudio en que pretende medir el impacto del crédito en la producción de cacahuates.
Para ello hipotetizó que la producción de cacahuates (Q) depende de una serie de variables:

$$Q = f (X_1, X_2, \dots X_n)$$

Como las variables X_1 a X_n son fenómenos sociales que se caracterizan por su fluctuabilidad en el tiempo (t), postuló que:

$$X_1 = g_1 (t).$$

$$X_2 = g_2 (t).$$

\vdots

$$X_n = g_n (t).$$

por lo que:
$$\left[Q = f (g_1 (t), g_2 (t), \dots g_n (t)) \right]$$

lo que se puede reducir a:

$$Q = G (t).$$

Postuló que el mundo debería dedicarse a la producción de cacahuates por su alto contenido proteico, por lo cual la función $G(t)$ será uniformemente creciente.

$$Q = a + b t \text{ con } a > 0, \quad b > 0, \text{ donde } t = \text{tiempo medido en años.}$$

Q = Producción anual de cacahuates, en toneladas, medida mediante los registros de exportación de Chicania, país que en 1981 produjo más del 80% del total mundial de cacahuates.

Luego recogió información empírica y estimó la función como:

$$Q = a + b t + e \quad \text{con } e = \text{término de error (aleatorio)}.$$

Encontró que la estimación empírica de los parámetros es:

$$a = 2.000.000$$

$$b = 0$$

Encontró además que las discrepancias entre los datos empíricos y los puntos de la recta eran muy pequeñas, por lo que consideró aceptada su hipótesis en forma definitiva, formulando una recomendación de que se aumente el crédito concedido a los plantadores de cacahuates a fin de incrementar la producción de éstos.

Emita un juicio crítico sobre el estudio y sobre sus conclusiones. Indique además, en no más de media página, cómo habría encarado Ud. ese mismo estudio. (Primera Prueba Parcial, 27/09/1982.

) A fin de determinar la conveniencia de realizar publicidad en una tienda de artículos de regalo, un investigador estudió las ventas mensuales durante los últimos 8 años y las relacionó con las variables de las cuales consideró que dependían esas ventas, en la siguiente relación funcional:
 $Q = f(X, Y, Z).$

Donde Q = Venta mensual total, en pesos.

X = Precio promedio mensual, en pesos.

Y = Crédito medido como total de valores adeudados por los clientes, al día 1º de cada mes.

Z = Valor gastado mensualmente en publicidad.

Se observa los siguientes hechos de importancia:

- a) La publicidad se realiza principalmente en los meses de noviembre y diciembre.
- b) Los pagos a plazo se realizaron, durante los primeros 4 años, con vencimiento los días 5 de cada mes; desde entonces, los nuevos créditos han tenido vencimiento en el día del mes que el cliente elija, habiéndose observado una preferencia de los clientes por pagar los días 30.

El investigador representó los datos obtenidos (nube de puntos) mediante una recta estimada con criterios estadísticos, de la forma:

$$Q = a + bX + cY + dZ.$$

Como las distancias de los puntos a la recta le resultaron excesivamente grandes, analizó cada punto que le resultara muy distante y encontró que la mayoría de ellos correspondía a los meses de verano, por lo que optó por agregar una cuarta variable W, que toma valor 0 para todos los meses, excepto enero y febrero, en que toma valor 1; luego volvió a estimar la recta ahora de la forma:

$$Q = a + bX + cY + dZ + eW$$

y obtuvo un mejor ajuste (distancias más pequeñas, debido a la inclusión de la cuarta variable (como un factor explicativo de Q), por lo que aceptó la hipótesis representada por la función.

Como "d" representa el impacto de la publicidad en las ventas, comparó ese impacto con el costo de la publicidad, y llegó a la conclusión de que la publicidad era conveniente para la firma, aunque por un margen muy pequeño (es decir, el aumento en ventas atribuido a la publicidad deja una utilidad mayor, pero por poco, que el costo de esa publicidad).

- Se pide:
- identificar, en la relación funcional, los distintos tipos de variables y parámetros.
 - evaluar el trabajo realizado, indicando los errores cometidos por el investigador en cuestión.
 - Si usted es el dueño de la firma ¿mantendría o suspendería la publicidad? (Tenga en cuenta sólo su impacto directo en las ventas, y no consideraciones como imagen de la empresa, etc.). (Primera Prueba Parcial, 18/05/1981).