



Escuela de Estadística

# VII Coloquio Regional de Estadística

Búsqueda del Conocimiento y Pensamiento  
Estadístico

*Roberto Behar Gutiérrez.*  
Profesor Titular.  
Universidad del Valle

Medellín, 20 al 23 de Julio de 2010

# PROLOGO

Revisando las estadísticas nos damos cuenta del impacto de los esfuerzos por erradicar el analfabetismo. Hoy los índices de analfabetismo en Colombia, son bastante menores que hace 50 años, no obstante que siempre hay un espacio para mejorar sobre todo en algunas regiones del país bastante olvidadas por el estado.

Los avances tecnológicos en estas últimas décadas han sido abrumadores sobre todo en las comunicaciones. En este mundo globalizado cada vez nos vemos impactados por volúmenes más grandes de información y la capacidad de síntesis cobra especial importancia. En no pocas ocasiones esta síntesis debe realizarse con procesos estadísticos y los ciudadanos debemos estar preparados para la adecuada interpretación de cifras.

Los políticos y los gobiernos expresan sus intenciones o sus logros con base en indicadores numéricos y si la población no está preparada para su correcta valoración, no podrá ejercer su derecho de control, ni podrá ejercer con responsabilidad su participación democrática, además de que será presa fácil de charlatanes de dudosa ética.

Esto también es verdad en todas las facetas de la vida cotidiana, empezando con la publicidad y los indicadores de eficiencia de un cierto producto o con los resultados de investigaciones sobre productos que tienen que ver con cosas tan delicadas como su salud.

Me atrevo a decir que no podrá existir democracia completa y verdadera mientras exista analfabetismo numérico en general y analfabetismo estadístico en particular.

*“La Búsqueda del conocimiento y el Pensamiento Estadístico”* pretenden poner en relieve, entre otros, aquel proceso oculto que sucede antes de los datos.

Muchos textos guía para el aprendizaje de la probabilidad y Estadística empiezan con los datos, como si antes de tenerlos no ocurriera nada.

Haciendo alusión al origen de la palabra “dato”, el diccionario de María Moliner nos dice que viene del latín “datum” y significa “dado” ( de dar). Todos sabemos que lo menos dado son los datos. Nadie nos da los datos, estos hay que producirlos y el proceso de generación de los mismos es determinante en la búsqueda del conocimiento en cualquier campo.

Todos los procesos de análisis de los datos, por sofisticados que sean, pierden valor si los datos no son fiables. Por esta razón considero pertinente que este cursillo se convierta en un espacio de reflexión sobre esos procesos que se realizan con el propósito de obtener los datos, el proceso de análisis y de interpretación y comunicación de los resultados.

Discutir aspectos relacionados con la validez de las conclusiones de un estudio, es algo que siempre estará en el orden del día de todo aquel quiera investigar o que sea consumidor de información producida por otros.

Poder leer críticamente los resultados de una investigación debe ser un objetivo primordial de todo curso que aborde la metodología estadística como un objeto de aprendizaje.

En la secundaria, no son muchos los profesores que se arriesgan a trabajar con sus estudiantes el tema de estadística, no obstante que aparece considerado en los lineamientos curriculares del Ministerio de Educación. En ocasiones cuando se aborda, el curso se convierte en un curso de fórmulas para calcular algunos indicadores, y hacer algunos gráficos fuera de contexto.

Los libros de estadística que sirven de guía, generalmente traen una breve introducción y luego se concentran en el desarrollo de los temas específicos que corresponden a un curso introductorio. En los últimos años, se ha sentido la necesidad de incluir en todos los cursos un tema que permite contextualizar la metodología estadística y su enseñanza. Este tema es la relación de la estadística con aquello que se ha denominado "Método Científico", que intenta ser una guía para una buena búsqueda del conocimiento.

Además del efecto contextualizador, incluir este tópico, podría servir como un elemento de motivación en un curso, que no pocas veces, genera sentimientos y actitudes negativas de parte del estudiante.

En algunas ocasiones recibimos de los medios de comunicación o en publicaciones científicas o en las noticias, información aparentemente contradictoria sobre temas diversos. Cosas o hábitos que ayer eran consideradas patológicas por la ciencia, en nuevas publicaciones se plantean como inocuas o como virtuosas. Los ejemplos son abundantes.

Nos preguntamos: ¿Las investigaciones se hicieron mal? ¿Por qué estos bandazos? ¿Si esta no es segura, cómo confiar entonces en los resultados de la investigación científica? ¿Cómo juzgar la validez de una investigación? ¿Qué tiene que ver la estadística en todo esto?

El Cursillo desarrollará este tema que se propone se incluya en todo cursos de estadística, con base en el análisis de recortes de periódicos y revistas.

Este cursillo no se ofrece solo a profesores de matemáticas, pues no exige ningún prerrequisito, pueden aprovecharlo los profesores de todas las disciplinas, incluyendo a los de ciencias sociales y también estudiantes de estadística y de licenciatura.

En la actualidad intento escribir un libro sobre este tema, basado en numerosos recortes de revistas y periódicos. El libro se encuentra en una etapa muy preliminar, razón por la cual tuve duda de entregarles este material bastante inacabado. Sin embargo me convencí de que es una buena y significativa oportunidad para compartir con ustedes este naciente esfuerzo, que espero muy pronto mejorar con el apoyo de todos.

Espero que sea de utilidad y con este material poner un granito de arena para el gran proyecto que es edificar para la erradicación del alfabetismos estadístico.

*Roberto Behar Gutiérrez.*

*Medellín, Colombia, , Julio de 2010.*



# Búsqueda Científica del Conocimiento y Pensamiento Estadístico

## **Introducción.**

Cuando nos referimos a la búsqueda científica del conocimiento, no estamos pensando solamente en grandes investigaciones científicas con impactos muy grandes para la humanidad o para una comunidad específica, sino que también enmarcamos como búsqueda científica aquella que debemos hacer para resolver un problema concreto en un ámbito empresarial o de un entorno social particular o inclusive familiar.

Podríamos en principio definir “problema” como la ausencia de conocimiento que nos impide, por ejemplo, tomar decisiones con racionalidad. Esa falta de conocimiento, puede ser de muy variadas clase; puede tratarse de buscar explicación para cierto comportamiento

de la naturaleza, buscamos “causas”, buscamos relaciones que pueden existir entre características de un fenómeno, buscamos conocer una cierta cantidad o cuantificar la magnitud de un parámetro.

Desde siempre el hombre ha investigado, siempre ha estado interesado en buscar explicaciones a las cosas que suceden a su alrededor, ¿Por qué “sale” el sol? ¿Por qué llueve? ¿Qué son los rayos? ¿Qué hay más allá de la muerte? ¿Por qué los cuerpos caen? ¿Hay vida inteligente en otros planetas? ¿La luz son ondas o partículas? ¿Puede calcularse la velocidad y la posición de un electrón en un instante dado? Otras más prosaicas: ¿El reo es culpable o inocente? Frente a una demanda por discriminación racial, el juez se pregunta: ¿Hay evidencia “suficiente” para soportar la hipótesis de discriminación? ¿El proceso de producción está bajo control? ¿Qué tan efectiva es la vacuna V?

¿Cómo resolver “el problema”? o ¿Cómo llenar ese vacío de conocimiento?

La historia de la humanidad va marcada por el desarrollo de estrategias que han dominado grandes periodos y que ha ido evolucionado y flexibilizando paradigmas de búsqueda de acuerdo a campos de conocimiento y a situaciones particulares.

Los griegos, quienes son referencia obligada cuando se hable de filosofía y de epistemología tenían al pensamiento deductivo de la matemática, como su método por excelencia para llegar a la “verdad” sin riesgos. El método de la geometría como su modelo máspreciado.

Las cuentas iban más o menos de la siguiente manera:

Si se acepta que la proposición

(Si ocurre P entonces ocurre Q) es verdadera

.. Y usted descubrió que en su situación de interés ocurrió P

Entonces, sin duda, puede asegurar que Q ha ocurrido.

La lógica de la matemática se basa en este paradigma deductivo, que produce conclusiones

libres de riesgo, por eso se oye decir que “la matemática no falla”. (Descontando a Gödel).

Afirmaciones resultado de la inducción tenían baja jerarquía para los griegos; conclusiones obtenidas de esta manera eran en el mejor de los casos “probables”, en el sentido de “aprobables” o plausibles mas bien por algún criterio de autoridad.

La evolución de los procesos de búsqueda del conocimiento, tiene que ver con la concepción filosófica del mundo, con la teología, pues no fueron pocos los años del dominio de un paradigma mecanicista del mundo, un modelo en el cual el mundo es un máquina con piezas que se comportan de manera determinística y solo hay que descubrir los engranajes y la relación entre los “piñones” para hacer predecibles todos los sucesos de la naturaleza. Por eso el gran interés de las culturas antiguas, en especial de los griegos, por la mecánica celeste, allí estaba lo verdaderamente digno de conocer, nada más estable, nada más constante que la “salida” del sol y la llegada cíclica de las estaciones.

Lo variable, lo impredecible, es fruto de la ignorancia, se alberga el sentimiento de que es posible conocer con exactitud las condiciones iniciales y una vez conocidas es posible la predicción sin error del comportamiento de la naturaleza.

La religión por otro lado, en principio era determinista: todos tenemos escrito el destino que dios nos ha deparado y nada puede hacerse para cambiarlo. Esta postura, es a juicio de Hacking (1975), una de las posibles razones del estancamiento del desarrollo del pensamiento probabilístico, pero aun con el advenimiento del catolicismo, cuya doctrina admitía que nuestra suerte en el más allá depende de nuestro comportamiento en la vida, aparece como impío intentar predecir la voluntad de dios.

Hoy, se acepta que la aleatoriedad existe como una característica inherente a los fenómenos, que es real y que no necesariamente está asociada con la ignorancia. Los aportes del científico Alemán Heisemberg, que dan origen al principio de indeterminación que lleva su nombre, es un puntal clave para la inclusión de la incertidumbre que genera la aleatoriedad como parte de los modelos que se construyen para explicar los fenómenos.

¿El cigarrillo “da” cáncer?.

Preguntas de este tipo, no pueden ser resueltas directamente con el método deductivo de la lógica matemática, pues para falsearlo, bastará con encontrar un contraejemplo en el cual se presente una persona que fue fumador empedernido durante toda su vida, murió a los 90 años y nunca padeció de cáncer.

Sin embargo hoy en día, hay cierto consenso en que la intensidad de consumo del cigarrillo si está relacionado con la mayor o menor “probabilidad” de desarrollar cáncer.

Observe que esta frase encierra la esencia del nuevo paradigma, no se garantiza que a un fumador habitual se le desarrolle la enfermedad, pero si se establece que sus probabilidades son mucho mayores que en los no fumadores. Esta afirmación como ésta no aplica a los individuos, sino a los colectivos. La Figura 1, ilustra distintas posibilidades de una compleja relación de causa a efecto, entre un factor X y un evento Y.

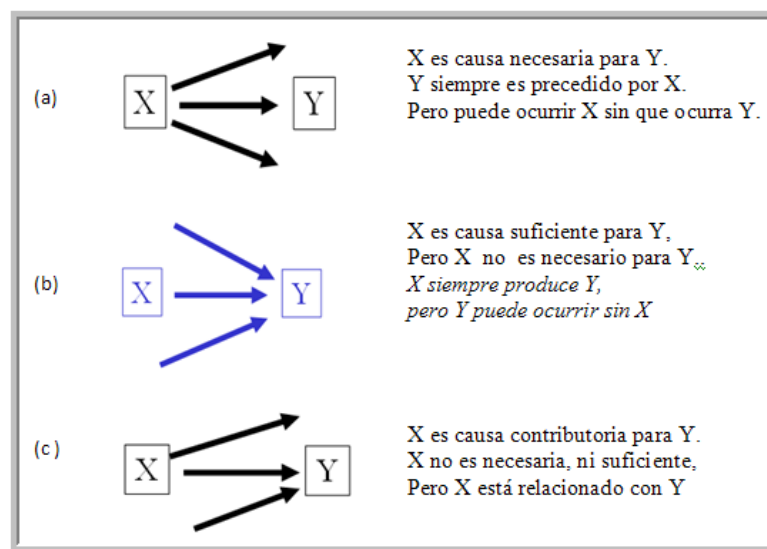


Figura 1. Algunas Formas posibles de relación causal entre X y Y (Tomado de Transparencias de Ignacio Méndez Ramírez, UNAM, México)

Nuestra situación del cigarrillo y el cáncer, estaría representada por la situación c), pues la presencia de consumo habitual de cigarrillo (X) no garantiza la presencia de la enfermedad Y. La primera flecha estaría indicando que el cigarrillo en alguna situación particular puede que no conduzca al cáncer, pero podría generar otra patología, digamos Z. (X no es

causa suficiente para Y), la segunda flecha estaría indicando que en algunos casos el hábito del cigarrillo puede contribuir a la presencia del cáncer (X está relacionado con Y) y en la tercera flecha estaríamos diciendo que el cigarrillo no es causa única de Y, (X no es necesario para Y), pues pueden haber otras causas distintas que generen la presencia de Y.

El esquema de la Figura 1, muestra en parte la complejidad de los fenómenos en el mundo real, que se hace aun más grande en presencia de la llamada interacción, fenómeno que se presenta cuando un factor X puede cambiar la probabilidad de generar Y, de acuerdo con la intensidad con que se presente otro factor W, como lo detallaremos más adelante.

En el intento por llenar un vacío de conocimiento, personas diferentes podrían tener propuestas distintas sobre estrategias para llenar el mencionado vacío, es decir, maneras variadas de ir a la búsqueda del conocimiento, distintas opciones de investigar el mismo problema. Por esta razón, entre muchas otras, no hay que confundirse cuando se habla de “*El Método Científico*”, pues pareciera que para cada problema de investigación existiera una única manera (método) de aproximarse a “la verdad”.

Si consultamos en el diccionario<sup>1</sup> el significado de la palabra “Método”:

*“Manera sistemática de hacer una cosa. Procedimiento, sistema. Se aplica al conjunto de reglas, lecciones o ejercicios para enseñar o aprender algo. Particularmente, procedimiento que se sigue en la investigación científica para descubrir o demostrar algo”*

Esta definición sugiere además que existen una serie de pasos, una especie de receta que hay que seguir al momento de investigar, si se quiere llegar al conocimiento de “la verdad”.

Hemos puesto “la verdad” (entre comillas), con la intención de poner en duda su carácter absoluto.

---

<sup>1</sup> María Moliner (1998). “Diccionario de Uso del Español”. Segunda edición. Madrid, España.

Existe de verdad “el método” en el sentido descrito? o es una forma de expresar que no puede ser de cualquier manera y que existen unas condiciones o características del proceso de búsqueda para considerarlo “valido”?, es decir, para hacer “creíbles” (fiables?) los hallazgos que resultaren? (¿Creíbles para quién?)

Si en realidad existen distintas posibilidades o estrategias de búsqueda, todas ellas conducirán a dar respuestas “válidas” a nuestros interrogantes?

¿Qué características debe tener el procedimiento de búsqueda para producir “confianza” en los resultados que se encuentren?

¿Si las estrategias de búsqueda cumplieren con las presuntas condiciones, se garantiza siempre que se llega a los mismos hallazgos? o por lo menos a resultados ciertos aunque pudieran ser distintos?

¿Qué significan los términos: “valido”, “confianza”, “creíble”, en el contexto mencionado?

Todas estas inquietudes están relacionadas con la llamada “Epistemología”, la cual es ineludible cuando se habla de búsqueda del conocimiento. El diccionario de María Moliner, la define como:

*“Tratado de los métodos del conocimiento científico, en general o de determinada ciencia”*

Es decir que, ¿existe una ciencia que “rige” a las ciencias o a los procesos de búsqueda del conocimiento en general?

*"La epistemología es a mi criterio, ese punto de vista desde el cual me relaciono con las cosas, con los fenómenos, con los hombres y eventualmente con lo trascendente. Esto, que se produce en el ámbito personal y cotidiano también ocurre en el ámbito científico, donde proliferan distintas corrientes y sistemas de pensamiento que resultan ser, en definitiva, formas de ver el mundo..... Podríamos comparar la Epistemología con un mirador u observatorio de la realidad que, con lentes más o menos agudos, la miran críticamente, dándose cuenta que no se halla*

*fuera de ella, sino contribuyendo a constituirla; es decir, no es meramente un proceso de observación, sino una participación activa en el hecho o fenómeno estudiado" Ricci (1999)<sup>2</sup>.*

Jaramillo (2003)<sup>3</sup>, en su artículo "¿Qué es la Epistemología?" Escribe:

*Para Ceberio y Watzlawick (1998)<sup>4</sup>, "el término epistemología deriva del griego episteme que significa conocimiento, y es una rama de la filosofía que se ocupa de todos los elementos que procuran la adquisición de conocimiento e investiga los fundamentos, límites, métodos y validez del mismo".*

*Ahora bien, la adquisición de conocimiento se fundamenta en vivencias otorgadas por el mundo de la vida, en la cotidianidad del sujeto; pero son las constantes que se verifican en esas vivencias, en la adecuación y relación sujeto –objeto - sujeto, en la validez de los conceptos que surjan de dicha adecuación, y en la posibilidad de predecir o interpretar acciones estableciendo causas o comprensiones sobre lo que realmente la epistemología legisla. Se puede esbozar entonces que la epistemología tiene por objeto ese conocimiento que se soporta en sí mismo o que soporta alguna disciplina en su especificidad; lo que la sustenta como tal, su esencia, sus alcances y límites en su acepción interna (propia de la disciplina) y externa (su influencia en el contexto social).*

*Para otros autores, la epistemología es aquella parte de la ciencia que tiene como objeto (no el único) hacer un recorrido por la historia del sujeto respecto a la construcción del conocimiento científico; es decir, la forma cómo éste ha objetivado, especializado y otorgado un status de científicidad al mismo; pero a su vez, el reconocimiento que goza este tipo de conocimiento por parte de la*

---

2 Ricci, Ricardo Teodoro. 1999. Acerca de una Epistemología Integradora. En: Revista Cinta de Moebio No.5. Abril. Facultad de Ciencias Sociales. Universidad de Chile

3 Cinta de Moebio No. 18. Diciembre 2003. Facultad de Ciencias Sociales. Universidad de Chile <http://www.moebio.uchile.cl/18/jaramillo1.htm>

4 Ceberio, Marcelo y Paul Watzlawick. 1998. La Construcción del Universo. Herder. Barcelona.

*comunidad científica. Es aquella epistemología que estudia la génesis de las ciencias; que escudriña cómo el ser humano ha transformado o comprendido su entorno por la vía de métodos experimentales o hermenéuticos en el deseo o necesidad de explicar fenómenos en sus causas y en sus esencias.*

*Concepción esta última asumida por Thuillier (citado por Mardones<sup>5</sup> 1991) al considerar la noción de epistemología como aquella "ciencia o filosofía de la ciencia que no impone dogmas a los científicos sino que estudia la génesis y la estructura de los conocimientos científicos". Para el autor, la epistemología no es un sistema dogmático conformado por leyes inmutables e impuestas; sino que mas bien, es ese trasegar por el conocimiento científico que se mueve en el imaginario de la época; las reflexiones sobre el mismo, y el quebranto o "crisis" de las normas que sustentan un paradigma en particular propio de una comunidad científica (Ver Kuhn 2001).*

*Ahora bien, relacionar la epistemología con la génesis de los conocimientos científicos, permite reconocer en ella los diferentes alcances que tienen este tipo de conocimientos en las instituciones de una sociedad, los saberes ideológicos de la época, y el impacto y transformación cultural causado por un conocimiento objetivo que a veces se abstrae de la realidad.*

*Desde el reconocimiento del entorno y la emergencia (emerger) del conocimiento en medio de la historia, Osorio<sup>6</sup> (1998) percibe dos visiones en la epistemología: una externa y una interna; yo entiendo la visión externa, comenta el autor, "como la consideración del contexto social y cultural en la actividad reflexiva acerca del conocimiento; y por la visión interna, los aspectos epistemológicos sin dicho contexto." Visión esta que se aleja un tanto del término formal de epistemología*

---

5 Mardones, José M. 1991. Filosofía de las Ciencias Humanas y Sociales. Materiales para una fundamentación científica. Barcelona: Anthropos.

6 Osorio, Francisco. 1998. Qué es Epistemología. Conferencia dictada para el Magíster en Antropología Mención Desarrollo. Departamento de Antropología. Universidad de Chile. Abril 7 de 1998.

*como objeto de estudio respecto a la esencia de algo, e invita más bien, a tener una mirada ecológica del mundo real en relación con un saber disciplinar.*

*...El Pensar reflexivo debe ser parte de nuestra mirada epistemológica, de nuestros progresos científicos al tratar de comprender o explicar un fenómeno natural, cultural o social. La epistemología debe llevar consigo un pensamiento crítico del conocimiento en confrontación consigo mismo, con sus pares e impares (otros) y con su mundo (lo otro); debe ser una epistemología proliferante movida por el viento, la cual reacciona frente a aquellos adelantos científicos que el sujeto se niega aceptar ciegamente y sin ningún juicio. Lo producido "como científico" por el que hace ciencia, debe ser continuamente depurado en procesos subjetivos e intersubjetivos de reflexión "como fuente de respiración, expiración - inspiración que lleva consigo gérmenes fecundantes de una animación capaz de resistir a largo término la pesadez mortífera que tiende a la esclerosis....un viento así, es la metáfora por excelencia de la circulación sin freno" Maffesoli<sup>7</sup> (1999).*

Como puede deducirse del fragmento del escrito de Jaramillo (2003) que se acaba de citar, esto de “normar” los procesos de búsqueda del conocimiento es muchísimo más complejo que intentar plantear un método, es el objeto de una metadisciplina que nunca estará acabada, que vive y se desarrolla en una dinámica continua y permanente, que ha sido una preocupación del hombre desde siempre.

Aunque este no es un libro de filosofía, conviene dar una ida de la complejidad de los procesos de investigación científica que en no pocas ocasiones se presentan en la metodología estadística con una sobre simplificación que podría dar una idea equivocada a las personas sobre este importante asunto.

A manera de colofón en la temática filosófica que aborda la epistemología, quisiera rescatar de los planteamientos de Jaramillo (2003) algunas expresiones que conviene matizar para ponerlas en el contexto de lo que se pretende en este escrito.

---

7 Maffesoli, Michel. 1999. El Nomadismo fundador. En: Revista Nómadas. No 10. Santafé de Bogotá. Departamento de Investigaciones Universidad Central. Abril. Pág. 126 -142.

Empecemos por la definición: “..es una rama de la filosofía que se ocupa de todos los elementos que procuran la adquisición de conocimiento e investiga los fundamentos, límites, métodos y validez del mismo”.

Se revela aquí la existencia de unos “elementos” que se presentan en esa pretensión de adquirir el conocimiento y que están relacionados con unos “límites” (no hay verdades absolutas) que muy seguramente están definidos por un contexto, por unos “métodos” y que también se relacionan con la validez del resultado.

Es el sujeto que investiga que en dicho proceso establece una relación con el objeto que pretende conocer y que esta mediado por características de experiencias, de vivencias, percepción del sujeto, destacándose entonces una componente subjetiva en el proceso que en principio puede significar una restricción inevitable. Sin embargo, se habla de la “construcción del conocimiento científico” refiriéndose a “la forma como éste ha objetivado, especializado y otorgado un status de cientificidad al mismo; pero a su vez, el reconocimiento que goza este tipo de conocimiento por parte de la comunidad científica”. En este sentido se estaría dando un paso de lo subjetivo a lo objetivo, al someterlo a la crítica de la comunidad académica, a través por ejemplo de la publicación en revistas científicas generando lo que se ha dado en llamar la objetividad intersubjetiva, que se relaciona con el status de cientificidad.

La epistemología debe ser dinámica, en permanente cuestionamiento, y las verdades que se obtienen con la investigación científica son verdades relativas, fruto de una mayor o menor simplificación de una realidad compleja, por medio de procesos siempre sujetos a error. Desde esa óptica las verdades en la ciencia son transitorias, al menos en lo que respecta a sus límites. Un ejemplo muy útil para ilustrar este punto, es el desarrollo del modelo de átomo. Muy seguramente el átomo que estudiaron nuestros abuelos, está muy desactualizado con respecto al modelo que la comunidad científica acepta como válido hoy en día. Observe que no decimos que el modelo de átomo actual, es el verdadero, sino el que es aceptado hoy por la comunidad de académicos. Muy seguramente nuestros abuelos, no tuvieron noticias de los mesones, de los neutrinos y de otras partículas que ha sido necesario colgarle al modelo de átomo para ponerlo a tono con los resultados de nuevos

comportamientos de la naturaleza que tan solo registramos ahora gracias al avance de la tecnología disponible.

Las Leyes de Newton, siguen vigentes y con base en ellas se sigue diseñando edificios, automóviles, cohetes, etc., estas leyes, tuvieron el carácter de universales por más de dos siglos. Hoy la comunidad científica acepta que para velocidades cercanas a la de la luz, se requiere de nuevos modelos para explicar el comportamiento de la naturaleza, pues las leyes de Newton se quedan cortas. Descubrimos restricciones o límites a las leyes de Newton. Para estas situaciones aparecen nuevos modelos como los planteados por Einstein, sobre la teoría de la relatividad, que estará vigente hasta nueva orden, es decir, a la espera de nuevos hallazgos que no encuentren explicación en el modelo y que hagan necesario su replanteamiento incluyendo las nuevas observaciones.

Una reflexión que pone en evidencia el carácter dinámico de la epistemología, es que una buena proporción de las personas cuando escuchan la palabra “investigación” traen a su mente una persona de bata blanca en un laboratorio. Paradójicamente el concepto de experimentación como un medio para obtener conocimiento válido, o aceptado por la comunidad académica es relativamente reciente, toda vez que se asocia al famoso Galileo Galilei, (Siglo XVI) con el inicio de la llamada “ciencia experimental”, no obstante que en la antigüedad, Arquímedes y el propio Aristóteles pueden haber sido precursores del “método experimental”.

De hecho la ciencia o episteme para los antiguos griegos, según Platón, estaba ligada a la idea del bien, y según Aristóteles, a la *deducción* de los atributos de un fenómeno, a partir de la definición de su esencia.

La dinámica de la ciencia, se ha movido con el desarrollo tecnológico, cultural, filosófico y económico del medio donde tiene su desarrollo.

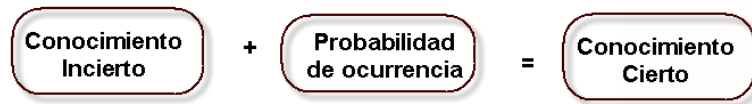
Hitos que marcan rupturas y se constituyen en mojones, pueden ser en la edad moderna la “teoría científica” de Galileo, en siglo XVI y de Newton en el siglo XVII. El Racionalismo de Descartes en el siglo XVIII. Descartes es el fundador del Racionalismo, creador de la teoría de las ideas innatas, y defensor del criterio de considerar a la

matemática como la ciencia pura, independiente de la experiencia humana, que encarnaba el ideal mismo de la Ciencia. El Racionalismo representa la concepción opuesta del Empirismo, no obstante que tienen sus puntos de contacto.

El empirismo en el siglo XIX, y luego en el siglo XX y en el XXI, el surgimiento de teorías y modelos asociados a situaciones específicas, desarrollan formas muy particulares de búsqueda del conocimiento, y han definido estándares de validez en sus propias comunidades académicas. En este último siglo, tenemos por ejemplo: la teoría de la relatividad, de la mecánica ondulatoria, de la mecánica cuántica, de los quarks, de la electrodinámica cuántica. Por otro lado, la teoría de la creación del universo y su permanente expansión, el Big Bang, la biología molecular, la química de los seres vivos, la ingeniería genética (genoma humano), etc.

En el marco de complejidad de la situación causal, que pretende ilustrar la Figura 1, que cobija una gama amplia de fenómenos en los cuales la presencia de factores aleatorios o de ruido, hacen difícil encontrar patrones de naturaleza matemática (determinística), un verdadero hito que ha permitido un avance vertiginoso de la ciencia moderna, es la consolidación de la teoría estadística en el siglo XX, con los aportes de Karl Pearson y su pionera distribución Ji-cuadrado, como referencia para la contrastación empírica de hipótesis estadísticas, Kolmogorov y la axiomatización de la teoría de la probabilidad, Sir Ronald Fisher, con la instrumentación del método experimental, la introducción de la selección aleatoria y el desarrollo de la matemática, para realizar la contrastación de hipótesis a partir de distintos modelos de diseños experimentales, la creación del famoso *P-value*, como indicador a partir del cual se obtiene información acerca de la compatibilidad de los resultados experimentales con una hipótesis predefinida. Luego Los aportes de Jerzy Neyman y Egon Pearson, con la construcción de pruebas de contraste de hipótesis, con criterios de optimización matemática como la verosimilitud, la definición de la potencia de una prueba, que sirve para comparar familias de contrastes de hipótesis, los errores tipo I y Tipo II, que se han constituido en verdaderos bastiones metodológicos a la hora de descubrir patrones estables de comportamiento de la naturaleza, en medio de una maraña de ruido aleatorio, dicho de otra manera han permitido construir buenos procesos para separar la paja del trigo, haciendo conciencia de la coexistencia con el error, pero también

proporcionado siempre una medida de su potencial magnitud y la probabilidad de ocurrencia. Se dio aquí un gran paso para hacer real la bella ecuación que plantea C.R. Rao en su libro “Statistics and True”:



Esta ecuación se ha impuesto en el último siglo y prácticamente todas las comunidades científicas lo aceptan hoy como un paradigma válido para otorgar status científico a los resultados de investigaciones empíricas.

Para la ciencia y los científicos es una verdadera fortuna la consolidación de la estadística como instrumento que da fuerza y poder al método inductivo. En algunas publicaciones científicas que tratan el tema del fraude en la historia de la ciencia, ponen en duda, algunos famosos casos de investigación científica por la “excesiva” coincidencia de los datos experimentales con los modelos teóricos propuestos, a tal punto que hasta ahora, por ejemplo, nadie ha podido lograr experimentalmente unos datos parecidos a los de Galileo, análogamente sucede con los experimentos de Mendel, sugiriendo que estos brillantes científicos “maquillaron” un poco sus datos. Si así fuere, estos pecadillos cometidos se podrían explicar por la ausencia en la comunidad científica de la época de criterios para juzgar la magnitud del error experimental. Hoy, con la maravillosa metodología estadística, estos genios no hubieran necesitado “retocar” sus datos para obtener la credibilidad de la comunidad científica.

Vamos a plantear ahora situaciones de búsqueda del conocimiento más aterrizadas a nuestra realidad, a través de las cuales encontremos aquellos elementos a tomar en cuenta, como nos lo planteaba la epistemología, que aumentan la probabilidad de que nuestros hallazgos pasen los filtros de la comunidad científica y ganen status científico a través de la objetividad intersubjetiva.

# ¿Se puede garantizar que las conclusiones a las que se llegue en una investigación sean válidas? ¿Qué significa validez?

Un criterio que es y ha sido usado para concluir sobre la fiabilidad de un cierto conocimiento, es el criterio de autoridad. Afirmaciones como: "...Esto debe ser cierto porque lo dijo fulano que es una persona muy seria...". "...Yo creo en esto porque fue resultado de una investigación de la universidad de Harvard."

Hay que aceptar que hay credenciales que es razonable asignar a resultados dependiendo de donde provienen, pero no necesariamente por personas o instituciones, sino por la naturaleza del proceso que los genera. Por ejemplo, a la cuestión de si una muestra particular, puede considerarse como una "muestra aleatoria", la propia muestra no tiene la clave de esa respuesta, la clave está en el proceso usado para la obtención de la muestra. De la evaluación crítica de dicho procedimiento, se otorga o no el crédito.

Por otro lado, ¿será posible en una investigación llegar a unas conclusiones y luego en otras realizadas posteriormente, con el mismo cuidado (por lo menos por la misma institución), llegar a conclusiones aparentemente contradictorias?. La Respuesta es afirmativa.

Veamos algunos ejemplos:

## ***Situación N<sup>o</sup> 1. ¿...Y entonces que debo comer?***

En el diario "EL Tiempo", de Abril 5 de 2006, se publicó el artículo: "*...Y entonces que debo comer?*".

Pone en evidencia la "evolución" de la opinión científica, con respecto algunas recomendaciones sobre la conveniencia de la ingesta o no de ciertos alimentos.



Figura 2. ...Y entonces...¿Qué debo comer?. Lo que antes era malo, ahora no lo es?.

**Situación N° 2. Las nueces antes prohibidas por sus contenidos de grasa, ahora mejoran la calidad de vida de los obesos.**



**Figura 3. Ahora comer nueces es recomendable para la salud.**

Las nueces, antes satanizadas por sus contenidos de grasa, ahora resulta que no solo no hacen daño, sino que son benéficas.

**Situación N° 3. Los antioxidantes y algunas vitaminas sintéticas antes tan ponderadas, ahora no solo no sirven sino que hacen daño.**



**Figura 4. Los antioxidantes en lugar de alargar la vida podrían aumentar riesgo de muerte.**

El diario “El Tiempo” de Bogotá, Colombia, en su edición de marzo 14 de 2007, reporta un estudio de la universidad de Copenhague, Dinamarca después de revisar 68 estudios

en los cuales se estudiaron más de 200.000 casos, concluye que no hay evidencia de que los suplementos antioxidantes, tan promocionados en nuestra sociedad, sirvan para algo. Al contrario plantean que hay evidencia para pensar que podrían aumentar el riesgo de muerte.

Esto parece la locura, será que no se puede confiar en nadie? Tanto tiempo privándome de los huevos para librarme del colesterol y protegerme del riesgo de enfermedades del corazón y ahora resulta que no era cierto. Que en realidad no representan riesgo o por lo menos no en la magnitud que se suponía.

Qué bueno era tomar leche, pues con su contenido de calcio, entre otras, fortalecía los huesos contra fracturas. Hoy resulta que otro estudio concluye que mujeres mayores que tomaban leche, no presentaban menos fracturas que las que no tomaban leche.

Otro tanto ha ocurrido con el chocolate y con el café.

Estudios recientes ponen en duda las bondades de la famosa grasa omega 3, que proporciona cierto tipo de pescado.

Después de invertir tanto dinero en antioxidantes, con el propósito de prolongar la vida con vitalidad, ahora un metaestudio danés, reporta que los antioxidantes no solo no han probado ser buenos sino que además podrían ser malos.

Y no es gratis la pregunta : ¿...Y mañana volverán a cambiar estas conclusiones?

Este panorama se aprecia bastante sombrío. ¿Esto es normal? ¿Algo ha estado fallando? Si es así, ¿puede evitarse en un futuro? ¿Qué debería hacerse?

Una primera reflexión, nos lleva a considerar la evolución, no solo de los procesos de búsqueda del conocimiento, sino los avances en la tecnología, las posibilidades de observar cada vez mejor la naturaleza, que hace que hoy podamos medir cosas que antes no podíamos. Ya hablamos del caso del átomo y su proceso de modelación siempre inacabado. No hubo error con el modelo de átomo de antaño, en las investigaciones pasadas, simplemente el modelo propuesto era coherente con las observaciones realizadas en aquella época. Si ahora, con base con posibilidades mayores de medir nuevas características, se

hallan inconsistencias con el modelo, es necesario plantear nuevas hipótesis de comportamiento que incluyan las nuevas observaciones y seguimos con la mente abierta a nuevos hechos potenciales en el futuro que entren en contradicción con el nuevo modelo y así sucesivamente at infinitum. En síntesis, esta dialéctica de permanente y perpetua revisión, es una característica inherente a la búsqueda del conocimiento, es inevitable, por no decir normal. Para la muestra un botón.

#### **Situación 4. Un descubrimiento de impacto promisorio: la Epicatequina en el cacao.**

**'PODRÍA LIBRARNOS DE CUATRO ENFERMEDADES COMUNES EN EL MUNDO OCCIDENTAL'**

## Harvard encontró nuevos beneficios en el cacao

WASHINGTON (EFE) Un compuesto del cacao tiene tantas virtudes curativas que podría compararse con la penicilina y la anestesia en importancia para la salud pública, según un estudio divulgado por la revista *Chemistry & Industry*.

Según Norman Hollenberg, profesor de la Escuela de Medicina de la Universidad de Harvard, se trata de la epicatequina, un compuesto en el brebaje de cacao natural que beben los indígenas de la etnia kuna de Panamá.

Esa sustancia también puede encontrarse en ciertos tipos de cacao, los paros cardíacos, el cáncer y la diabetes se han reducido a solo 10 por ciento entre los miembros de ese grupo étnico.

"Si estas observaciones pronostican el futuro, entonces podemos decir sin temores que son las más importantes en la historia de la medicina", señaló.

"Estamos de acuerdo en que la penicilina y la anestesia son enormemente importantes, pero la epicatequina podría librarnos de cuatro de las cinco enfermedades más comunes en el mundo occidental", añadió.



Edgar Vargas / Archivo EL TIEMPO

Al cacao podría compararse con la penicilina y la anestesia en importancia para la salud pública, según la revista *Chemistry & Industry*.

tiene otras propiedades. Algunos sostienen que estos dos químicos fundamentan la base sobre el tan pregonado efecto estimulante del chocolate.

Está demostrado que el cacao induce a la producción de endorfinas, las cuales producen bienestar y felicidad.

El cacao también contiene antioxidantes naturales, como los polifenoles y el tanino, que ejercen efectos beneficiosos tanto a nivel físico, por su acción de vasodilatación, como a nivel psicológico, proporcionando una sensación de vitalidad y energía.

Sus propiedades hidratantes...

**Figura 5. Harvard descubre la epicatequina en el cacao, la cual podría tener efectos promisorios para la salud.**

Observe por ejemplo, el artículo de la Figura 5, que se publicó en el periódico "El tiempo" de marzo 13 de 2007, anuncia que Harvard ha descubierto ahora la Epicatequina en el cacao, substancia que al parecer tiene efectos espectaculares para la salud, a tal punto que manifiesta que es posible que un buen número de enfermedades letales sea consecuencia de la ausencia de la epicatequina y compara este hallazgo en importancia, con el descubrimiento de la penicilina o de la anestesia.

El hecho de que antes no hubiera podido ser detectada esta sustancia, no es una negación de su existencia.

En algunos casos, se descubren límites para conclusiones que pensábamos mas generales,

ahora nuevos estudios nos dan la posibilidad de incluir nuevas variables, que nos permiten detectar interacciones, que se traducen en que lo que antes creíamos, se sigue cumpliendo, pero siempre y cuando el nivel de otra variable permanezca en cierto rango.

Pensábamos que no había relación entre dos características o dos eventos, pero en el rango estudiado de una variable X, pero resulta que lo que ocurría era que en ese rango era difícil detectarlo, pero ahora que hemos ampliado el rango de observación, pudimos detectar la relación y medir su intensidad.

La muestra estudiada, con base en la cual obtuvimos las conclusiones, era de voluntarios y por lo tanto, se dio cabida a unos sesgos que hacen que no deba generalizarse tanto como quisiéramos. Las conclusiones solo parecen plausibles para un subconjunto de la población que se intenta caracterizar al describir quienes fueron los individuos estudiados.

El tamaño de la muestra se puede comparar con una lente a través de la cual queremos observar un objeto. Cuando la muestra es pequeña, la lente es opaca y logramos detectar rasgos muy gruesos y evidentes, sin embargo no podemos dar cuenta de muchos detalles. A medida que el tamaño de muestra aumenta, la lente es más potente y está más enfocada, estamos más seguros de lo que vemos y además podemos informar sobre mas detalles del objeto estudiado. Por esta razón en algunos estudios realizados con muestras pequeñas, no encontrar ciertos rasgos que permitirían rechazar una hipótesis previamente establecida, no es sinónimo de que dichos rasgos no existan, solo se concluye que no pudieron ser observados. Por eso no es incoherente que en nuevos estudios con más recursos y mejores diseños puedan detectar dichos efectos.

Por otro lado, cuando la muestra es muy grande se pueden detectar diferencias estadísticamente significativas, pero que no tienen ninguna significancia práctica.

La misma naturaleza de los métodos estadísticos que convive con el error y lo mide, tiene como característica inseparable del método la posibilidad de proporcionar conclusiones erradas.

Así por ejemplo, en un proceso investigativo que pretende contrastar ciertas hipótesis, se

realiza un diseño controlando el riesgo de cometer el error tipo I (Rechazar la hipótesis nula, cuando en realidad esta es cierta), de tal manera que hemos aceptado como un parámetro de diseño válido, que la probabilidad de cometer este tipo de error sea del 5% (nivel de significancia  $\alpha = 5\%$ ), esto quiere decir que siendo válida la hipótesis nula, de cada 20 equipos que estén realizando una investigación idéntica a esta en el mundo, en promedio uno de ellos, rechazará por azar, equivocadamente la hipótesis.

Las hipótesis nulas, normalmente son del tipo: “...No existe relación entre...”, “...El agente X, no tiene efecto sobre...”, de tal manera que lo que resulta de interés para el científico y para las revistas de la comunidad científica, es el rechazo de la hipótesis nula.

Cuando no se rechaza la hipótesis, las revistas científicas generalmente no tienen interés en publicar estos resultados, por lo tanto aquel, que por azar, (1/20), llegó a la conclusión equivocada de rechazar la hipótesis nula, tiene mayor probabilidad de que registren su hallazgo en las revistas especializadas sobre el tema y la comunidad científica se quedará sin saber de los 19 de cada 20 investigaciones que llegaron a una conclusión contraria.

Hasta ahora, nos hemos referido a “gajes del oficio”, es decir, circunstancias que están ligadas casi inexorablemente a los procesos de investigación. Muy seguramente con excepción de la última situación que mencionamos, que podría corregirse ahora que la extensión de los artículos no es problema, con las publicaciones en Internet, podrían ser reportadas también las investigaciones que tienen aceptación metodológica por parte de la comunidad académica, pero que no han llegado a resultados de interés, de acuerdo con las expectativas de los investigadores en ese campo.

También es justo tratar las situaciones en la que tienen injerencia los investigadores y que por ignorancia, por descuido o por falta de escrúpulos éticos (la menor parte de las veces), invalidan la metodología de investigación produciendo con culpa o sin ella resultados falaces o amañados, produciendo en no pocas ocasiones impactos graves para la sociedad.

Mencionaré algunas posibilidades, no obstante que más adelante regresaremos con más detalle sobre este tema.

No conocer ampliamente el marco de referencia del problema que se estudia, puede conducir a dejar por fuera de la investigación, características que son cruciales. Descuidar la selección de la muestra, puede llevar a sesgos con consecuencias irreparables. Un instrumento de medición (un aparato físico o un cuestionario o un índice, etc.) no validado, puede hacernos creer que medimos una característica de interés, cuando en realidad estamos midiendo otra cosa, por ejemplo, juzgar la calidad de un hospital, solo preguntando a los pacientes, podría ser un error, pues allí se estaría midiendo “la percepción de la calidad”; muy seguramente la característica “calidad” sea objeto, entre otros, de un conjunto de expertos en el tema de la salud.

Escoger un método de análisis de los datos obtenidos que no es coherente con el diseño del estudio, puede echar al traste, valiosos recursos y llevar a conclusiones erradas.

Ignorar el alcance y las limitaciones de los métodos de análisis y el alcance en la interpretación de las cifras obtenidas, puede ser otra fuente de error.

Un proceso muy bien cuidado, cuyos resultados son mal expresados al momento de la comunicación a la comunidad científica, puede ser objeto de malas interpretaciones con las consecuencias que ello pueda tener.

El asunto no es tan prosaico como para dar la guía perfecta para hacer bien las cosas. Ya se ha dicho que no existen caminos únicos en la ciencia. Sin embargo, aunque cada investigador pueda escoger su propia ruta, todos deben saber que cuando hay semáforo en rojo, hay que detenerse, también que hay que ceder el paso al peatón, etc.

También ahora, cuando hablamos de la búsqueda científica del conocimiento, podemos referirnos a un conjunto básico de principios que todos los investigadores deberían respetar en sus procesos investigativos, aunque este conjunto no sea exhaustivo.

Estamos hablando de un conjunto de principios que contribuyen a definir lo que serían conclusiones fiables, a partir del juicio o valoración crítica de características que tuvo el proceso de búsqueda. Es decir que juzgaremos el resultado a través de juzgar el proceso que lo produjo.

¿Cuáles son estos principios básicos? (No exhaustivos).

## Conceptualizando la validez.

A continuación se presentarán esos principios básicos, cuya consideración va en la dirección de ganar validez en el proceso de investigación. La fisura de alguno de estos elementos, podría ser suficiente para que la comunidad académica, niegue el status de científico al trabajo en cuestión.



Figura 6. Conceptos clave relacionados con la validez de un estudio.

Estos elementos que se describen a continuación están basados y adaptados del desarrollo que hace el profesor William Trochim, de la universidad de Cornell, en su libro “Research Methods Knowledge Base<sup>8</sup>”.

---

<sup>8</sup> Trochim William (2002). “Research Methods Knowledge Base”. Atomic Dog Publishing. Cornell.USA.

En el desarrollo realizado hasta ahora, hemos tocado algunos de los tópicos que presenta la Figura 6, sin embargo vamos a considerarlos con un poco más de detalle.

Estos elementos pueden ser muy útiles al momento de evaluar un trabajo científico y también al momento de la planeación y la ejecución de cualquier proyecto de investigación.

### ***Validez Externa de un estudio.***

Una característica que permite valorar las conclusiones de una investigación es su grado de generalidad. ¿A quienes aplican sus resultados? ¿Hasta dónde pueden generalizarse? La validez externa, está relacionada con la respuesta a estas preguntas. Desde esta perspectiva, la valoración de la validez externa es más de grado que categórica.

Cuando un conocimiento es generado con pocos supuestos y sale airoso de la contrastación en una gama muy amplia de condiciones, se dice que tiene una gran validez externa. Conocimiento de esta clase, son las llamadas “leyes”. Las leyes de Newton, la ley de los gases (Boyle), las leyes de la termodinámica, etc. A continuación ilustraremos con ejemplos.

### ***Situación 5. Arreglo de siembra de Yuca para optimizar su rendimiento.***

Los resultados de una investigación agrícola, se describen en el siguiente artículo científico:

Esta investigación sobre Yuca fue realizada en la Habana, Cuba, y se ensayaron distintos arreglos espaciales (distancia entre surcos y distancia entre plantas), para dos variedades de Yuca: la CMC40 y la “Señorita”, en condición de monocultivo.

Los resultados indican que la variedad CMC40, logró su mayor rendimiento con el arreglo espacial 1,40x0,60 y 0,70x1,0. Este rendimiento fue de 40,9 ton/Ha, Mientras que para la variedad “Señorita”, el mejor rendimiento se dio para 1,80x0,50 y su rendimiento fue de 31,4 ton/Ha.

Las plantaciones se realizaron en suelo ferralítico rojo compactado.



Figura 7. Validez externa. Rendimiento de la Yuca

#### Pregunta:

*Si se replica el mismo experimento para la Yuca “Señorita”, en Palmira (Colombia) o en la Paz (Bolivia), resultará de nuevo que el mejor arreglo espacial es 1,80x0,50? y su rendimiento estará también alrededor de 30 ton/Ha?*

No hay ser demasiado experto en agricultura, para saber que la altura sobre el nivel del mar, es un factor clave para las plantas, a tal punto que existen los llamados “pisos térmicos”, en algunos de los cuales es seguro que es imposible encontrar cierto tipo de plantas. Ya sabemos que el clima es un factor importante para las plantas. Hay plantas más robustas al clima y hay otras muy sensibles. Sabemos que la fisiología de las plantas se mueve alrededor del proceso de fotosíntesis y que por lo tanto la intensidad de luz, es un factor clave. La disponibilidad de agua, en caso de que el cultivo se haga bajo temporal y no bajo riego. También la nubosidad y el brillo solar del sitio donde se haga la siembra son factores importantes.

Hilando más delgado, la composición del terreno, en este caso suelo ferralítico rojo compactado, puede afectar el rendimiento.

En síntesis, la respuesta a la pregunta que hicimos es un triste: “No sabemos”.

Lo que la investigación reporta tiene validez para las condiciones específicas en las cuales se realizó el experimento. Dichas condiciones, acotan la población para la cual pueden generalizarse los resultados y definen el grado de validez externa de esta investigación.

*¿Cómo mejorar la validez externa?*

La validez externa puede ser mejorada replicando el experimento en muchos sitios disímiles del planeta, en variadas condiciones de altitud, de clima, de terrenos, registrando cuidadosamente los valores específicos de cada una de las variables.

Si replicáramos un número muy grande experimentos, podría construirse un modelo que permitiera predecir el rendimiento esperado si se siembra en un sitio con una condiciones especificadas que ingresarían al modelo y harían una estimación local, que podría ser bastante buena, aun para predecir en sitios donde no se han realizado experimentos.

Es claro ahora que la validez externa tiene que ver con el proceso de generación de los datos. En este caso con las condiciones específicas en que ha sido conducido el experimento que da origen a los datos.



*Figura 8. Origen de los datos y Validez Externa*

En el formato de la Figura 9, se evidencia un interés por lograr validez, al definir de manera precisa cuando debe tomarse la muestra. Este caso también está relacionado con la validez

en la medición y en garantizar comparabilidad (validez interna), que trataremos mas adelante.

CRITERIOS DE DESEMPEÑO		CONOCIMIENTOS Y COMPRENSIONES	
<b>Muestras Foliar</b> a) La muestra se toma en el área definida con criterios técnicos establecidos. b) La toma de la muestra se hace en época seca y en horas de la mañana cuando la planta tiene menos actividad fisiológica asegurando validez. c) Las palmas previamente seleccionadas e identificadas presentan un estado de desarrollo representativo con el fin de realizar el muestreo cada año sobre las mismas garantizando la validez de la información.		1. La nomenclatura de los lotes, las áreas de muestreo, sitio y palmas de muestreo. (a),(i) 2. a) y numeración (a),(d),(e),(f) 3. Manejo y remisión al laboratorio de las 4. (),(d),(e),(f),(g),(i),(k),(l),(m), utilizada.(f),(g),(i)	
RANGOS DE APLICACION		EVIDENCIAS REQUERIDAS	
<b>Categoría</b> 1. Edad palma	<b>Clase</b> ✦ Joven, adulta	<b>De Desempeño</b> ✦ Tomar muestras foliares y de suelo con la habilidad y destreza necesaria siendo efectivo para alcanzar los estándares establecidos.	

Figura 9. Cuidando la validez. Cultivo de Palma de aceite.

## La validez externa y los colectivos del muestreo.

Ya hemos visto que la validez externa de un estudio está íntimamente ligada al proceso de generación de datos. Si se generan con base en un diseño experimental, las condiciones del experimento marcan la pauta para saber a quienes pueden generalizarse las conclusiones, si los datos se obtienen de un proceso de muestreo como es clásico en las encuestas, la validez externa queda definida por aquel colectivo cuyas unidades tuvieron la oportunidad de ser seleccionadas al azar, en la muestra, con probabilidad conocida. A este colectivo al cual se puede realizar la generalización estadística le llamamos “Población de estudio o Población Muestreada” y deseamos que coincida con la población de interés, que conocemos como “población objetivo”. Así por ejemplo, hay ocasiones en que un estudio quiere sacar conclusiones para todos los colombianos, sin embargo, hay zonas de muy difícil acceso, bien sea por su geografía, o por su peligrosidad. Estas zonas quedan excluidas de la posibilidad de ser observadas. En algunos casos solo se toma en cuenta la población de las grandes urbes, que constituyan al menos el 70% de la población.

Si queremos conocer la opinión de los estudiantes de la universidad del Valle, acerca de

una nueva política de la administración universitaria, lo que conviene es tomar una muestra aleatoria de la toda la población de estudiantes. Aquí, habrá que definir muy claramente a quien consideramos “estudiantes de la universidad del Valle”, planteando claramente unos criterios de inclusión y seguramente otros de exclusión (quien es y quien no es).

Los estudiantes a distancia son de interés para el estudio? Un estudiante que solo cursa una asignatura será considerado como estudiante? Las personas de la comunidad que toman cursos aislados de extensión son de interés para el estudio? Los de posgrado? Los de las sedes regionales?

Una vez definidos de manera muy precisa, conviene tener un listado de todas las personas que cumplen con la definición convenida para “estudiante de la Universidad del Valle” y a partir de este listado seleccionar una muestra aleatoria. A este instrumento, en este caso un listado, que nos permite materializar el proceso de obtención de nuestra muestra aleatoria, se le conoce en el lenguaje técnico de la estadística como “marco muestral”. Con base en este, haremos la rifa para obtener la “*muestra aleatoria*”.

Note que cuando hablamos de la población de estudio, dijimos que es a esta población a la cual podremos extender o *generalizar estadísticamente* las conclusiones. Porque existe otro tipo de generalizaciones que no son estadísticas, por ejemplo, consideraciones sociológicas, pueden justificar que la parte de la población objetivo que fue excluida, ha quedado bien representada en la población de estudio y con estos argumentos se extienden las conclusiones. Otros investigadores experimentan en conejos o en micos, asumiendo por el conocimiento de su fisiología, que es razonable pensar que un comportamiento similar se tendrá en los seres humanos. Esta inferencia de los conejos o los micos a los humanos, no es una *inferencia estadística*.

No siempre se dispone de un marco muestral perfecto, en el que estén todos los que son y son todos los que están y si no hay una depuración adecuada o se descuida este aspecto, pueden cometerse graves errores como es el histórico caso de la encuesta para valorar la intención de voto, en la elecciones de los Estados Unidos en 1936 y que comentaremos más adelante.

En la Figura 10, se presentan los colectivos que intervienen en el proceso de muestreo aleatorio, en cada uno de los pasos es posible fallar, por ejemplo, es posible definir una población objetivo no apropiada para el interés del investigación, es posible que la población muestreada no represente adecuadamente a la población objetivo; es posible que el marco muestral no sea representativo de la población muestreada y por ultimo también es posible realizar una mala selección de la muestra en el sentido de que no sea estrictamente aleatoria.

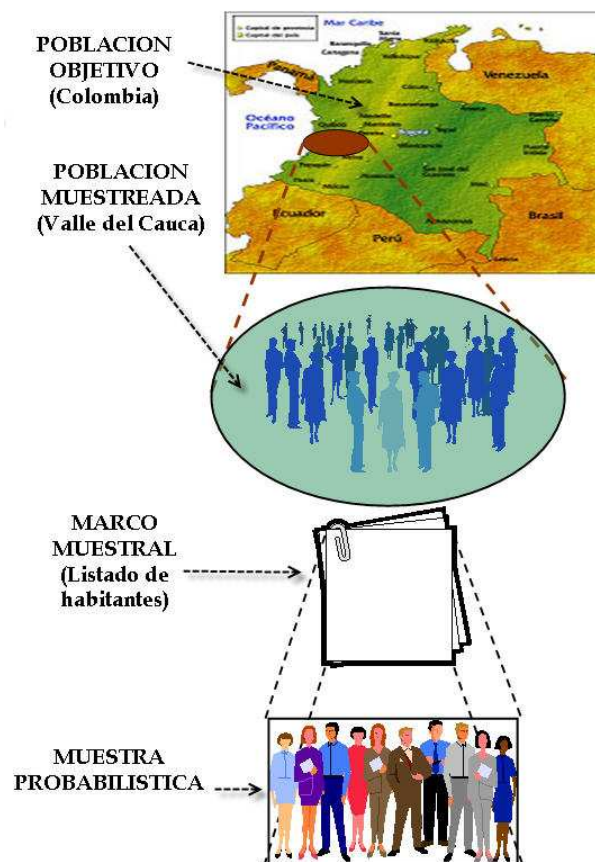


Figura 10. Colectivos a considerar en un proceso de generación de datos por muestreo.

### ***Situación 6. Una famosa Pifia en la validez Externa: Roosevelt Vs Landon en USA.***

Tal vez el caso más famoso de falla en la predicción del candidato ganador basado en sondeos electorales, es el que corresponde a las elecciones presidenciales de los Estados Unidos, en 1936, cuando se enfrentaban los candidatos Roosevelt y Landon. Todavía en 1936, el Literary Digest utilizaba un formulario con respuestas por correspondencia.

Literary Digest venía acertando en sus vaticinios desde 1916, sin embargo en esta vez, sus pronósticos sobre la derrota de Roosevelt obtuvieron un estrepitoso fracaso, no obstante que se basaron en la muestra más grande en toda la historia. Ellos enviaron su formulario por correo a 10 millones de personas “elegidas al azar”. De estas, fueron respondidas cerca de 2.4 millones de respuestas, con base en las cuales Landon ganaba a Roosevelt por 57% contra 43%. Los resultados en realidad fueron 62% para Roosevelt y 38% para Landon. La aplastante victoria de Roosevelt provocó que los republicanos juraran nunca más darle crédito a tales sondeos. Sin embargo, los aún poco conocidos, en aquel tiempo, George Gallup y Elmo Roper ya se basaron en muestras estadísticas y predijeron al triunfador con bastante precisión.

Se tienen pocos antecedentes de una muestra tan grande, y también de un descalabro tan mayúsculo. ¿Qué pudo haber fallado?

La muestra aunque grande, fue resultado de la selección al azar de 10 millones de ciudadanos norteamericanos, obtenida usando como marco muestral el registro de teléfonos y de automóviles de los Estados Unidos, de los cuales 2 millones respondieron por correo postal. ¿Qué ocurrió?

El origen del fiasco en la estimación fue que el marco muestral no representaba la población de votantes, pues en 1936, tener teléfono y/o tener coche era sinónimo de clase social alta, en esa época solo 1 de cada 4 ciudadanos tenía teléfono, es decir que el marco muestral excluía a los numerosos ciudadanos de bajo estrato social, cuya intención de voto favorecía por mucho a Roosevelt sobre Landon.

Otro sesgo importante fue el sesgo por “No respuesta”. Pues cabe preguntarse si la condición de responder o no la encuesta que exigía el uso del correo postal, era independiente de la intención de voto.

La pregunta es: La probabilidad de responder el sondeo por correo era la misma para los simpatizantes de Landon que para los de Roosevelt?, pues de no ser así, las respuestas obtenidas, ni siquiera representarían a los 10 millones seleccionados, lo cual hace mas grave el problema.

Este es pues, un clásico ejemplo, de sesgo en la selección de la muestra debido a un marco muestral defectuoso.

Una gran moraleja, es que el peligroso sesgo de selección no se corrige necesariamente aumentando el tamaño de la muestra.

### **Mitos sobre el tamaño de la muestra.**

*¿El tamaño de la muestra, crece proporcionalmente con el tamaño de la población, cuando se quiere realizar una estimación de la media de una población?*

Sobre el tamaño de una muestra se dicen muchas cosas en la cultura popular, no todas ellas ciertas. Por ejemplo, en ocasiones se rechazan los resultados de un estudio, con el argumento de que “la muestra no es representativa, pues ni siquiera llega a ser el 10% de la población”. Esta cifra del 10% es completamente falaz, sin embargo ha hecho mucha carrera y es conveniente desmitificarla a través de algunas situaciones como las que se comentan a continuación.

### **¿Le hace falta sal a la sopa?**

Para responder esta pregunta el cocinero debe tomar una “muestra representativa” de sopa, probarla (medir la sal con sus papilas) y luego tomar una decisión para toda la sopa. ¿Le parece a usted sensato recomendar al chef que pruebe el 10% de la sopa para que pueda tomar una “buena decisión”? Esta recomendación contradice todas nuestras experiencias, pues la cucharilla que usamos en nuestras casas para probar la sopa que preparamos

habitualmente es la misma que cuando la preparamos en una olla más grande porque tenemos invitados, y esto nunca nos ha sorprendido.

Una acción que no se perdona antes de meter la cucharilla a la sopa es la de hacer un buen meneo con el cucharón, para homogeneizar la sopa, de tal manera que cualquier posible muestra que se tome, brinde la misma información que otra. Es decir, que en esta situación, como en casi todas, se considera que es más importante la variabilidad o la homogeneidad, que el tamaño de la olla. Dicho de otra manera, es mucho más importante mezclar bien antes de tomar la muestra, que aumentar el tamaño de la cuchara



*Figura 11. Ilustración gráfica del impacto del tamaño de la población en el tamaño de la muestra. El caso de la prueba de la Sopa.*

**Doctor, ¿para saber mi grupo sanguíneo me va usted a sacar una muestra con el 10% de mi sangre?**

Es bien sabido que basta con una muestra de una gota de sangre para conocer en forma inequívoca el tipo de sangre de una persona; esto, en virtud de que todas las gotas de sangre de una persona son del mismo tipo, así que vista una, quedan vistas todas. De nuevo vemos aquí que el impacto de la homogeneidad es más importante que el del tamaño de la

población. La misma cantidad de sangre se requiere para un niño recién nacido que para su padre.

**¿Para saber en promedio cuantas patas tienen los cangrejos del Pacífico, necesito capturar el 10%?**

De nuevo aquí se desmitifica la falacia de que una muestra para que sea “representativa” debe contener una cantidad de unidades de acuerdo con el tamaño de la población hacia la que se desea concluir.

En este caso hay un agravante. ¿Quién nos dirá cuanto es el 10% de los cangrejos del Pacífico?

Bueno, ya que hemos recurrido a la intuición ahora vamos a mostrar el verdadero efecto del tamaño “N” de una población en la definición del tamaño “n” que debe tener una muestra para realizar estimaciones con una calidad previamente especificada.

### **Relación del tamaño N de una población con el tamaño n que deberá tener una muestra**

La pregunta ¿cuál debe ser el tamaño de muestra a tomar? se puede formular técnicamente de la siguiente forma: Si queremos estimar la media  $\mu$  de una población que tiene  $N$  unidades, con un margen de error  $\delta$  y con un nivel de confianza  $(1-\alpha)$ , ¿cuál debe ser el tamaño  $n$  de muestra que debemos tomar?

La calidad de la estimación está definida por el margen de error y el nivel de confianza. Supongamos que estamos dispuestos a tolerar un margen de error  $\delta = 1$ , y que el nivel de confianza deseado es del 95%. En una población con desviación estándar  $\sigma=8$ , aplicando la fórmula estadística para encontrar el tamaño de muestra, se obtienen los resultados que se presentan en el Cuadro 1. La estimación en todos los casos tiene la misma calidad. Así por ejemplo apreciamos que si la población estuviera conformada  $N = 500$  unidades, se requeriría un tamaño de muestra de  $n=165$  unidades para lograr la calidad estipulada.

En el Cuadro 1, se repite este cálculo para distintos tamaños  $N$  de la población y se aprecia

muy claramente como a medida que crece la población, su tamaño pierde influencia sobre el tamaño de la muestra. Vemos como cuando la población cambia de tamaño  $N=100$  a  $N=150$ , un incremento de sólo 50 unidades, el cambio de la muestra va de  $n=72$  a  $n=94$ , es decir que se aumenta en 22 unidades, sin embargo, cuando el tamaño de la población cambia de  $N=5.000$  a  $N=10.000$ , es decir, aumenta en 5.000 unidades, el tamaño de la muestra cambia en tan solo 5 unidades.

Cuando la población cambia 100.000 a 50.000.000, el tamaño de muestra, después de hacer el redondeo, se mantiene constante e igual a 246 unidades.

Tamaño $N$ de la población	Tamaño $n$ de la muestra
100	72
150	94
200	111
250	124
500	165
1.000	198
2.000	219
5.000	235
10.000	240
50.000	245
100.000	246
1.000.000	246
50.000.000	246

*Cuadro 1. Impacto del tamaño de la población sobre el tamaño de la muestra para estimar la media con una confianza del 95%*

La Figura 12 muestra gráficamente la relación entre tamaño de la población y tamaño de la muestra.

## Impacto del Tamaño de la Población sobre el tamaño de la muestra

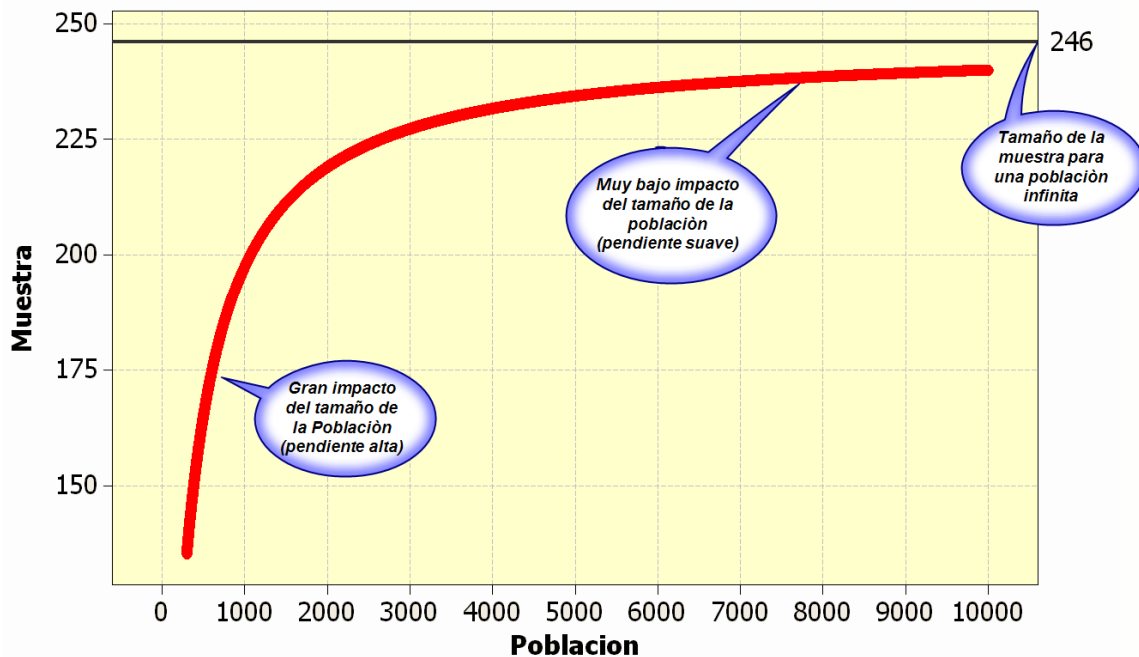


Figura 12. Relación entre el tamaño de la población y el tamaño de la muestra

Nótese la pendiente tan pronunciada en el primer tramo, lo cual indica que cuando el tamaño de la población no es muy grande, este tiene gran impacto sobre el tamaño de la muestra, pero cuando el tamaño de la población es grande, ya no es tan importante, al punto que el tamaño más grande posible para la muestra en la figura es  $n=246$ , para cualquier tamaño de la población.

Estos resultados son verdaderamente sorprendentes y confrontan el sentido común y nuestra propensión natural a usar la regla de tres, que evidentemente no funciona en este caso.

### Muestreo no probabilístico.

Algunas veces, por razones de tipo práctico o por razones éticas, no es posible realizar muestreo probabilístico y se recurre a otras formas de conseguir los datos. En investigaciones en salud, por ejemplo, hay que recurrir a muestras de voluntarios, que por

supuesto deberán cumplir con las restricciones de rigor, que convenga a la investigación, por ejemplo, rango de edad, no haber padecido algunas dolencias, estado de salud, etc.

En estos casos, la pregunta obligada es: ¿A quienes representa la muestra? En esta curiosa situación, habrá que “inventarse” la población, suponerla con base en las características de los individuos estudiados. Por lo tanto debe describirse muy bien las características que comparten las unidades de la muestra, para que el lector de los resultados de la investigación, se haga una idea lo más precisa posible sobre a quienes podría inferir las conclusiones. Estas inferencias en estricto sentido, no serían inferencias estadísticas.

### **Fuentes secundarias para la obtención de datos.**

En muchas ocasiones debemos recurrir a fuentes secundarias para obtener todos los datos o parte de ellos. Las fuentes de estadísticas oficiales de la región o el país, algunas entidades del estado, organismos internacionales, etc.

En otras ocasiones, por ejemplo en ambientes industriales, existe siempre la tentación a recurrir a los archivos de datos históricos de la empresa como sustituto al diseño de algunos experimentos. Este procedimiento puede tener sus riesgos, por la dinámica cambiante de algunos factores, como proveedores de materia prima, procedimientos, máquinas, operarios, etc., que podrían hacer no comparables algunas mediciones. Por otro lado, es posible que el rango en el que se mueven las variables no sea suficiente para detectar algún efecto, lo cual podría mejorarse si se realiza un diseño de experimentos.

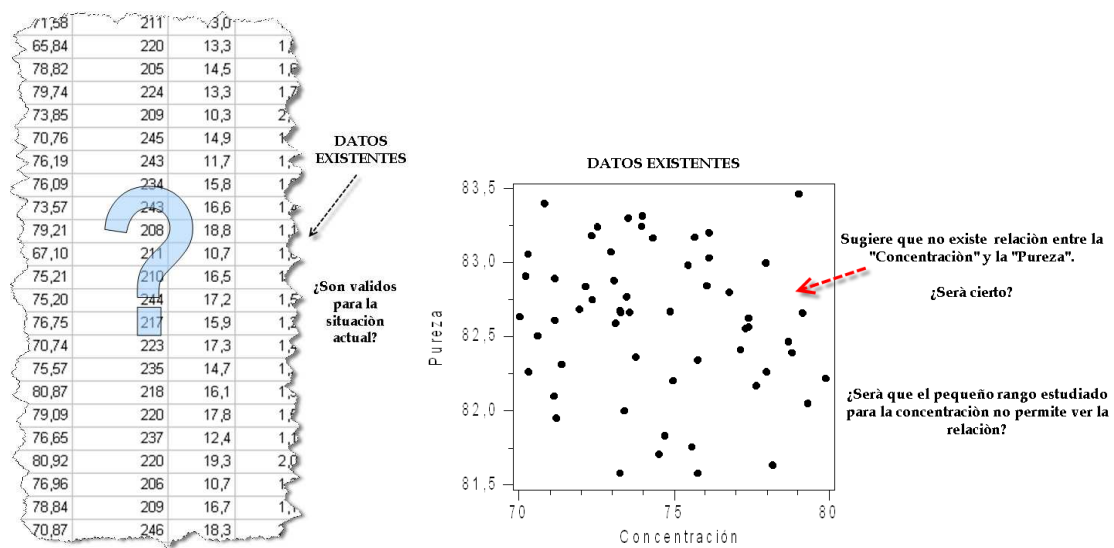


Figura 13. Riesgos de usar datos preexistentes como sustituto a la experimentación.

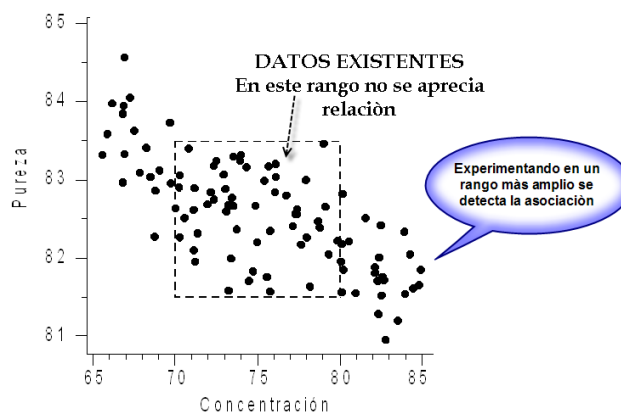


Figura 14. Experimentando puede definirse a voluntad las condiciones apropiadas.

En la Figura 14, se muestra visualiza la relación real existente entre la concentración y la pureza, sin embargo la información la información histórica registrada corresponde a la zona señalada en el recuadro que es la que se ilustra en la Figura 13, en la cual no se aprecia ninguna relación entre las dos variables de interés.

## Muestras aleatorias vs muestras representativas.

***Situación 7. Efecto teflón” de la imagen del presidente de Colombia. ¿Falla en la validez de la metodología para medir la favorabilidad?***

El artículo del periódico “El Tiempo”, marzo 13 de 2007, que se muestra en la Figura 15, se escribe a propósito de la insensibilidad de los índices de favorabilidad del presidente de la república, registrados a través de encuestas, no obstante los eventos negativos que ocurren en el país y que atañen al gobierno, como los famosos escándalos de la llamada “parapolítica”.

La periodista Olga González, intenta explicar el alto índice de permanente favorabilidad del presidente de Álvaro Uribe, con base en los posibles sesgos en las encuestas realizadas.

Los aspectos que toca, desde punto de vista técnico, tienen que ver con el apartado sobre “validez” que estamos tratando en este capítulo, lo cual es una oportunidad para hacer énfasis en la importancia de ciertas consideraciones estadísticas, al momento de realizar encuestas y en particular hacer sondeos de opinión.

Una primera cosa que el artículo pone en duda, es la “representatividad” de las muestras, es decir que si la opinión depende de ciertas categorías socioeconómicas, es necesario tomar en cuenta la participación de la opinión de acuerdo con dichas categorías. Si por ejemplo, la opinión varía de acuerdo con los estratos sociales, es necesario, conocer la participación porcentual de dichos estratos en la población, para el cálculo de los indicadores correspondientes de opinión. Ignorar esta estructura, puede conducir a conclusiones falaces.

Esta consideración, puede tomarse en cuenta, controlando el proceso de selección aleatoria de la muestra, de tal manera que la muestra se refleje la estructura de la población. Esto puede lograrse a través del llamado “Muestreo estratificado”.

De otra manera, si la muestra fue obtenida con algún otro procedimiento aleatorio, conviene hacer ajustes, valorando la opinión en la muestra, de acuerdo a las categorías en cuestión y aplicando los llamados “factores de expansión” que corrigen o ajustan, de acuerdo con la verdadera representatividad de las categorías de interés en la población.

Si este importante hecho no ha sido considerado en los estudios que se critican, la periodista tendría razón en su apreciación.



**Figura 15. Favorabilidad del Presidente de Colombia. Un sondeo que deja demasiadas dudas en su validez.**

Los profesionales de la estadística son muy celosos, en propiciar las condiciones para garantizar una muestra probabilística, pues este es un requisito indispensable para la aplicación de la teoría estadística, condición indispensable para la fiabilidad en las

conclusiones. Sin embargo es necesario aclarar que no siempre una muestra probabilística, o aleatoria, es una muestra representativa de la población. La representatividad hace referencia a la preservación en la muestra de las características más importante de la población, que están relacionadas con las variables que se quieren medir. Por azar, puede suceder que en un muestreo aleatorio, quede sobre representada una categoría en particular, no obstante que cuando la muestra es muy grande, gracias a la ley de los grandes números, puede garantizarse con altísima probabilidad dicha representatividad.

Es en estos casos, en los que la muestra aleatoria, no es representativa de la población, es necesario realizar el ajuste a través de la consideración de ponderadores de dicha estructura en la población.

Por otro lado, la periodista pone en duda la representatividad, basada en el procedimiento usado para la recolección de la información, en este caso la entrevista telefónica, cuestionando que un amplio sector marginado de la población, carece de este servicio, lo cual haría que las conclusiones no podrán extenderse a la opinión de todos los colombianos, pues estos no tuvieron oportunidad de ser seleccionados en la muestra. Además hace alusión a la extensión del cuestionario, 80 preguntas, lo cual limitaría la representatividad de la población, toda vez que no todos los colombianos tendrían el tiempo y la disposición para responder por teléfono un cuestionario tan extenso. La periodista también se refiere a un aspecto muy importante, como es, la forma de realizar las preguntas, pues de no ser elaboradas de manera neutra y muy específica, podrían estar sesgando la opinión. Por último hace una observación relacionada con la ética de quien realiza las encuestas, pues arguye que dicha persona tuvo lazos muy estrechos con el presidente en épocas pasadas.

### ***Situación 8. Rescatando la validez externa para el control de crecimiento y desarrollo infantil.***

El artículo " Talla y peso más reales ", aparecido en el periódico "El Tiempo" del 14 de marzo de 2007, plantea como la Organización Mundial de la Salud después de estudiar los efectos de la lactancia, implementó nuevos parámetros de medición.

CAMBIAN TABLAS DE CRECIMIENTO EN NIÑOS MENORES DE 5 AÑOS

# Talla y peso más reales

Después de estudiar los efectos de la lactancia, la Organización Mundial de la Salud implementó nuevos parámetros de medición.

La Organización Mundial de la Salud (OMS) actualizó recientemente las tablas infantiles de crecimiento y peso y ha creado un nuevo patrón de referencia para los niños entre los 0 y los 5 años de edad. Este parámetro es el resultado de un estudio multiétnico realizado en Brasil, Estados Unidos, Ghana, India, Noruega y Omán, en el que participaron 8.440 pequeños.

Desde hace 30 años, en Co-

**2008**

Es el año en el que se espera implementar, en todo el territorio nacional, el nuevo sistema de medición planteado por la Organización Mundial de la Salud.

tección Social reglamenta su uso. Además, se hace indis-

una forma recomendada y, analizar una muestra de datos cercana a la realidad.

Además, "la referencia NCHS no tenía mediciones en los tres primeros meses de vida, en los cuales el ritmo de crecimiento de los niños es el más acelerado".

En los primeros cuatro años, por ejemplo, el niño alcanza más de la mitad de la estatura que tendrá de adulto.

El nuevo patrón, en cam-

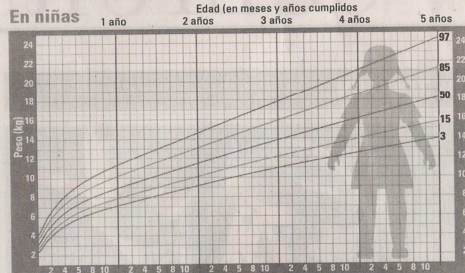
dos exclusivamente durante seis meses y vivían en ambientes favorables para su desarrollo, porque además de no haber fumadores en sus hogares, tenían buenas condiciones de salud.

Según Lutter, "son importantes, sobre todo en los primeros 24 meses de vida, una correcta nutrición con leche materna y alimentación complementaria, un ambiente libre de tóxicos y cuidados es-

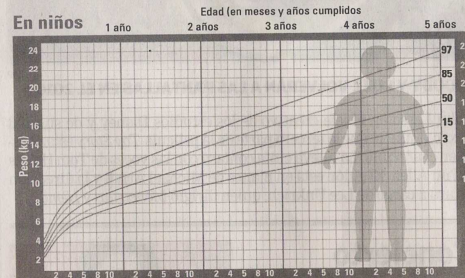
MIÉRCOLES 14 DE MARZO DE 2007 EL TIEMPO

## ABC DEL BEBÉ 2-5

Para ubicar al niño en las tablas de desarrollo es necesario saber su peso, longitud y edad. Luego, se coloca un punto en el lugar donde se cruzan ambos datos para situar al pequeño en un determinado percentil. De esa forma, se hace un seguimiento durante todos los meses y se evalúa la tendencia de crecimiento.



■ Si un niño baja o sube de percentil significativamente quiere decir que, tal vez, exista un trastorno y vale la pena estudiarlo.



■ Siempre hay que tener en cuenta la comparación del percentil con la genética del niño, pues unos padres de baja estatura, por ejemplo, pueden tener niños de percentil bajo, y viceversa. Esto no quiere decir que exista un problema. Gráfico: CEET

**Figura 16. Recuperación de la validez externa para las tablas de Referencia para crecimiento y desarrollo infantil**

continúa el artículo diciendo: “ Desde hace 30 años, en Colombia se han utilizado las gráficas de crecimiento del centro nacional de estadística de salud (NCHS), que se basan en datos obtenidos en niños de EE.UU. y no proporcionan una muestra real con respecto a la

población Colombiana”.

Es realmente increíble, que el crecimiento y el desarrollo de nuestros niños, tome como referente la distribución de pesos, talla y su relación con la edad, contruidos con base en niños norteamericanos. Éste modelo de control, evidentemente falla en su validez externa, pues la población a la cual se aplica, niños colombianos, no coincide con la población de estudio, niños norteamericanos.

“La iniciativa de cambiar el patrón de referencia de crecimiento surgió porque se evidenció que los niños alimentados, según las recomendaciones de la OMS, en cuanto a la lactancia materna exclusiva por seis meses y continuada hasta el segundo año de edad, tenían una trayectoria de crecimiento diferente a las tablas de la NCHS. Es decir, no reflejaban y los patrones de nutrición ni las diferencias étnicas” afirma Chessa Lutter, asesora regional de la unidad nutrición de la Organización Panamericana de la Salud.

Es plausible pensar que las tablas de referencia para el control y desarrollo de los niños, dependa de la raza, pues el fenotipo de los caucásicos, los chinos, los negros, los indígenas, etc. son distintos; además las condiciones ambientales y de desarrollo de los entornos de los distintos países también marca diferencias.

Se recibe con beneplácito esta noticia, pues estas tablas de referencia permiten detectar problemas agudos o crónicos que afectan la talla y el peso. Además pueden evidenciar trastornos alimentarios como desnutrición u obesidad. ¡ Bienvenido el rescate de la validez externa!.

### ***Validez del constructo.***

El término “constructo” hace referencia al concepto que se desea medir. En general un constructo es una construcción mental sobre una característica de un objeto de interés, que en algún caso podría ser no observable de manera directa, por ejemplo la capacidad de gestión del gobierno en una gran ciudad, o podría ser muy específico y medible objetivamente, así por ejemplo en el área de los materiales de construcción un importante constructo es la “Resistencia del concreto”.

## Situación 9. La medición en resistencia de materiales

Se ha definido de manera precisa como preparar el cilindro con el concreto al cual se desea medir la “resistencia a la compresión” (probeta). Una máquina especialmente diseñada para proporcionar carga (fuerza) sobre la superficie superior del cilindro de concreto, registra el valor correspondiente a la fuerza al momento de reventar el cilindro, como se muestra en la Figura 17. Está también establecida la relación entre el resultado de la resistencia en el laboratorio y la resistencia “in situ”. La resistencia a la compresión es en realidad una medida de presión, razón por la cual el valor de la fuerza que informa la máquina, se debe dividir entre el área del cilindro que soportó la fuerza.



Figura 17. Medición del constructo “Resistencia a la compresión” del concreto.

La fiabilidad del sistema de medición se podrá conocer al hacer mediciones repetidas a varios cilindros de prueba confeccionados con la misma mezcla homogénea de concreto.

En general, en ingeniería, las cosas están razonablemente claras y casi todos los

constructos de interés pueden ser planteados en función de magnitudes objetivas y claramente definidas, desde las ciencias básicas.

Un caso distinto ocurre en las ciencias humanas o ciencias sociales. Todos tenemos alguna idea de lo que significa “autoestima”, sin embargo, si nos pidieran ordenar un conjunto de personas de mayor a menor de acuerdo con su autoestima, no estaría tan claro el procedimiento.

El constructo “autoestima” a diferencia de la “Resistencia a la compresión”, no puede medirse en forma directa, corresponde entonces “aterrizarlo” descomponiendo el constructo en características más tangibles a partir de las cuales se pueda construir un instrumento, por ejemplo, un test, una entrevista estructurada, o un indicador o sistema de indicadores, etc.

Lo mismo ocurre con constructos como “inteligencia”, y otros menos intangibles pero igualmente complejos como la “pobreza”.

La manera de lograr construir un indicador que sirva para medir el constructo, no es tarea fácil y tiene su propio desarrollo en la Psicología o en las ciencias sociales y económicas, en los ejemplos mencionados.

Una posible amenaza a la validez del constructo es una inadecuada definición pre operacional u operativa, según sea el caso.

### ***Situación 10. Validez del constructo “Pobreza”***

Para los gobernantes y para la población de un país o una región, características como “la pobreza”, el “desempleo”, el “nivel de ingreso”, “analfabetismo”, “injusticia social”, “inequidad”, “costo de vida”, “inflación” etc., son muy importantes, a tal punto que casi con seguridad estarán en boca de todos los políticos en campaña.

Es de mucho interés para quien adquiere el compromiso de liderar los destinos de un pueblo, lograr mitigar estas características.



Figura 18. ¿Qué significa "pobreza"? ¿Cómo medirla?.

¿Cómo saber que durante el ejercicio de un gobierno cuanto se ha mejorado en realidad en cuanto a la *pobreza* de la población?

La respuesta es: "...Se debe medir...".

El constructo "pobreza" es medido adecuadamente? Los indicadores propuestos miden lo que se quiere medir?. Valoremos algunas de las medidas propuestas:

Definición 1: "*Pobre es quien vive con menos de dos (2) dólares al día*".

Es posible que la intención de medir la pobreza usando dólares en la definición haya tenido la sana intención de hacer más estable y comparable la medida, sin embargo esta manera de medir el constructo "pobreza", tiene serios reparos. Baste con decir que un ciudadano puede acostarse a dormir siendo pobre y amanecer no siéndolo, con tan solo que se

produzca una fluctuación del dólar. Así por ejemplo, esto ocurriría para una persona que vive con \$4500 diarios. Si el dólar ayer valía \$2500, este individuo, ganaba menos de 2 dólares y por lo tanto era pobre. Si hoy el dólar vale \$2200, el mismo individuo ha dejado de ser pobre, sin que su nivel de vida haya cambiado en nada. Sin embargo, si se hace el conteo de los pobres en ese instante, podremos concluir que la pobreza ha disminuido. Falla esta medida en lo que llamamos validez del constructo. De hecho en Colombia desde hace algún tiempo, el peso colombiano se está revaluando a una velocidad impresionante, lo cual significa, con la definición propuesta, que sin haber mejorado en nada las condiciones de vida, el porcentaje de pobres puede descender vertiginosamente.

Características como la pobreza, la autoestima, la inteligencia, no son atributos que puedan registrarse en los individuos de manera directa. Ellos no son observables como la estatura, el color de los ojos, etc. Para realizar su medición es necesario construirlos a partir de otros rasgos observables y ese proceso exige especial atención, pues se corre el riesgo de simplificar demasiado o de excluir rasgos importantes, haciendo que la final no se esté midiendo lo que se quiere medir.

En el artículo del periódico “El tiempo”, de marzo 1 de 2006, referido en la Figura 18, se afirma que existen 9 metodologías para medir la pobreza y que aun no se ha decidido cuál de ellas es “la mejor”. Se afirma que existen dudas no solo técnicas sino políticas, pues algunos piensan que es posible que la elección de “la mejor” manera de medir la pobreza, se adopte con base en criterios que favorezcan la imagen del gobierno de turno.

Estas decisiones son muy importantes, pues de hacerse con criterios técnicos y objetivos, el país podrá conocer si las políticas y los programas diseñados para combatir la pobreza están teniendo el impacto pronosticado o si es necesario cambiar el rumbo. Por eso no es un problema de poca valía.

### ***Situación 11 . Los “Sniffers” o “Supernarices”, olorosa profesión.***

Ya hemos hablado la medición en física o en ingeniería, hemos hablado de las mediciones en las ciencias sociales, cuando nos referimos a la pobreza, ahora nos referiremos a una situación, que aunque es muy concreta, su registro cuantitativo está impregnado de una

buena dosis de subjetividad.



**Figura 19. Medición de los olores del cuerpo humano.**

Curioso oficio el de los llamados “Sniffers”, que son contratados por las empresas productoras de desodorantes, para medir el impacto que sobre los malos olores del sudor, tiene sus productos.

Cómo validar las mediciones realizadas por estos individuos?

La empresa que fabrica desodorantes, experimenta con las proporciones de sus insumos, para lograr la fórmula óptima en cuanto a eficiencia (neutralizar malos olores producto del

sudor) y en cuanto a economía, de la manera más barata.

La Figura 19, describe el procedimiento. Unas personas se someten a experimentación, para ello se les hace medición previa de la cantidad de sudoración en un sauna con temperatura controlada, con el propósito de medir posteriormente el efecto antitranspirante del producto. Pasado un tiempo se someten de nuevo al sauna, aplicando el producto en una axila y teniendo la otra axila como control, se recogen los humores en esponjas atrapadas en las axilas por un tiempo establecido. Se guardan de manera hermética estas esponjas, que luego serán sometidos a las narices de los “Sniffers”, quienes asignarán un número del 1 al 5, que representa una escala que va desde “casi inodoro” (1) hasta “terriblemente fétido” (5).

¿Cómo saber a cuales narices contratar? ¿Cómo medir el poder discernidor de olores de una nariz?

Casi como lo podemos imaginar, deben existir unos “olores estándar” cuya escala está previamente establecida. En forma secreta, se coloca un código a 30 esponjas, las cuales se impregnan con olores en escala de 1 al 5, seis repeticiones de cada una. Luego son entregadas secuencialmente al azar a cada uno de los “sniffers” para su correspondiente clasificación.

El “sniffer” será fiable si clasifica las esponjas de un mismo grupo con el mismo número, así este número sea equivocado, es decir se valora su consistencia. En la Figura 20 la situación (a) y (d) corresponden a “sniffers” fiables, sin embargo en la situación (a), se presenta un sesgo, por ejemplo las esponjas que estaban secretamente marcadas con el número 1, las clasificó todas en el 2, mientras que la situación (d) no sólo fue consistente en su clasificación, sino que las clasificó correctamente.

Estas características tienen distintos nombres en la literatura, a la consistencia se le suele llamar fiabilidad o precisión y a la correspondencia con la medición se le suele llamar validez o exactitud. En la Figura 20 se presentan las distintas posibilidades.

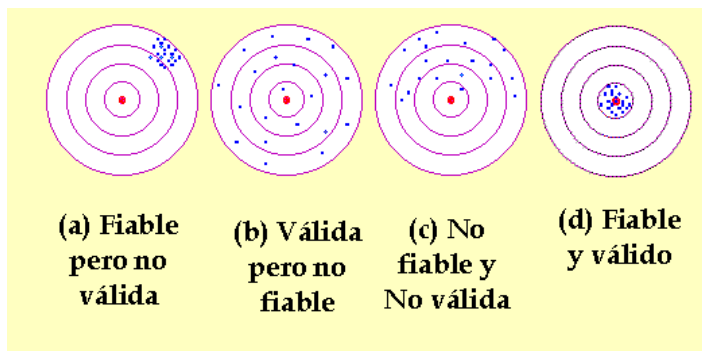


Figura 20. Valoración de la Fiabilidad y la validez (exactitud) de una medición.

Existen técnicas con base teórica, para valorar sistemas de medición, tanto en constructos abstractos como los que aparecen en psicología y las ciencias sociales, hasta otros más objetivos definidos en las ciencias fácticas y en la ingeniería.

### Naturaleza de las características que se desea medir.

Entender la naturaleza de las variables y definir una clasificación apropiada de acuerdo con sus escalas de medición es muy importante en estadística, pues muchos métodos y procedimientos están definidos dependiendo de la escala en que se mida la variable en cuestión.

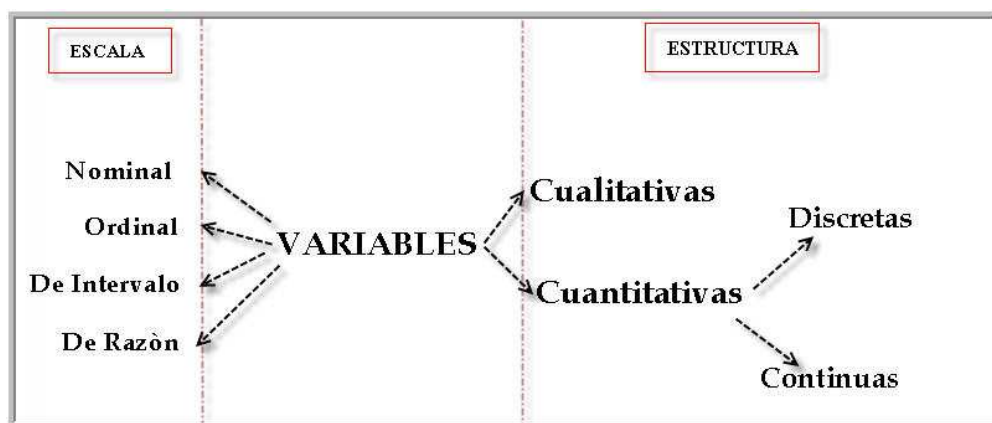


Figura 21. Clasificación de las variables según su escala de medición y su estructura.

Una primera clasificación de las variables es en: Cualitativas y cuantitativas.

**Cualitativa:** se refiere a aquellas variables para la cual valores posibles que puede

asumir no corresponden a números sino a categorías o atributos. Por ejemplo, los atributos de la variable "estado civil" son: soltero, casado, viudo, divorciado y separado. Aunque para propósitos de análisis le podemos asignar números a estas categorías, estos en sí mismos no tienen ningún significado cuando se describe el status civil de la persona, como explicaremos en la escala nominal.

**Cuantitativas o numéricas:** se refiere a variables cuyos valores observables son números. Por ejemplo: el salario de una persona o su edad se expresan naturalmente como números y estos números puestos en una escala se convierten en información.

En las escalas de intervalo y de razón algunas veces es necesario establecer la diferenciación de las variables por su naturaleza, entonces se habla de variables discretas y variables continuas.

**Variable discreta,** es aquella cuya naturaleza hace que el conjunto de valores que puede tomar la variable sea finito o infinito numerable.

Por ejemplo, la variable: número de personas por hogar, el conjunto de valores que puede asumir ésta son:

$$\{1, 2, 3, 4, \dots, M\} \quad \text{donde } M \text{ es finito}$$

Otros ejemplos son los siguientes: número de consultas al médico durante un año, número de clientes que llegan a un banco durante una hora, número de ensayos realizados hasta obtener el primer éxito.

**Variable continua,** es aquella, cuya naturaleza hace que exista un intervalo de puntos, los cuales son valores que puede tomar la variable. Por ejemplo, la estatura de una persona, esta variable puede tomar cualquier valor en el intervalo (1.50 m, 2.30m). El tiempo entre dos llegadas consecutivas al servicio de urgencias de un hospital. El área cultivada de trigo en las fincas del valle del río Cauca .

Esta clasificación no tiene en cuenta la población en la cual va a ser observada la variable, es decir, no interesa en la clasificación, si la población es finita o infinita, puesto que de

acuerdo con la definición una variable es discreta o continua por sí misma. Tampoco juega papel alguno el instrumento de medición que se use.

La materia prima de la Estadística son los datos, los cuales son el resultado de la "observación" de alguna(s) característica(s) de los elementos de interés en cierto estudio. La naturaleza de la característica y el instrumento que dispone para registrar la misma, definirá el tipo de **escala de medición** que se ajuste a la situación dada.

### ***Escalas de medición.***

Cuando se hace referencia a las escalas se trata de asociar números a las características con el propósito de transformarlas y obtener nuevo conocimiento sobre las características del estudio.

La caracterización de la escala de una variable se hace en términos de cinco (5) factores: **Exclusividad Mutua, Exhaustividad, Orden, Unidades Estándar de Medición** y presencia del "**Cero absoluto**".

La presencia (o ausencia) de estos factores define las diferentes escalas. Estas escalas tienen una jerarquía y podríamos ordenarlas de la escala más débil a la escala más fuerte, de tal manera que cuando una variable posea uno de los atributos, también posee los anteriores, por ejemplo, si la variable posee un cero absoluto también posee los cuatro atributos anteriores.

**Exclusividad mutua y Exhaustividad:** estos dos atributos se tratan en conjunto y son las características principales de las variables de escala nominal. La exclusividad se refiere a que ningún objeto de medida puede ser clasificado simultáneamente en dos o más categorías. La exhaustividad se refiere a que se han provisto suficientes categorías para todas las posibles mediciones. Por ejemplo, si preguntas a qué partido perteneces y no incluyes la categoría "Polo Democrático", tu escala no es exhaustiva aunque puede ser mutuamente exclusiva, esto es, no puedes ser "Conservador" y al mismo tiempo "Liberal" a la vez. Aquí alguien puede estar pensando en algunos casos de tráfugas, sobre todo entre políticos que salen elegidos por un partido político y luego trabajan para otro. Esto obliga a

tener definiciones operativas muy claras y precisas, de tal manera que aún en estos difíciles casos se pueda definir en una y solo una categoría.

**Escala Nominal.** Es la escala más débil pues en esta se usan los números para dar nombre a los atributos o categorías en los que pueden clasificarse los objetos de medición de acuerdo con alguna propiedad cualitativa. El número asignado a una clase sólo actúa como un rótulo o código para diferenciar los elementos de esa clase con los de otra. Por ejemplo si se clasifica un conjunto de objetos por su color, las categorías pueden ser: azul, amarillo, rojo, verde, a las cuales podemos asociar respectivamente los números 1,2,3,4 y se hablará de la categoría 1 para hacer referencia al grupo de objetos de color azul o 4 para el verde, pero los números aquí, sólo son códigos para nombrar los elementos de una clase. No se pueden sumar o restar o hacer operaciones con estos números.

**Orden:** Se refiere a que las categorías de las variables reflejan algún orden significativo, por ejemplo, el "amor que tú sientes por tu novio/a" es: ninguno, poco o mucho. Las variables que tienen esta característica se llaman ordinales.

**La escala ordinal,** hace uso de los números para clasificar los elementos de un conjunto en categorías en los cuales los números no sólo sirven para nombrar sino que son base para comparaciones de la forma: "mas grande", "igual", "menor", es decir, que el valor numérico de la medida se usa para indicar el orden que ocupa un elemento al comparar el tamaño relativo de sus medidas, del más grande al más pequeño, de allí el nombre de la escala. Por ejemplo, cuando a una persona se le pide ordenar de la más importante a la menos importante, asignando números de 1 a 4, a las siguientes necesidades: empleo, salud, vivienda, servicios públicos. Aquí el número se usa para representar la prioridad de las necesidades; de esta manera si un individuo asigna el número 1 a la vivienda y el 4 al empleo, indicará que para él es "más importante" la vivienda que el empleo.

**Unidad estándar:** Se refiere a que el espacio entre las unidades de medición es el mismo. En las ciencias sociales no hay muchas variables que posean esta característica de forma exclusiva (esto es, sin que a la vez sean también variables "razones"). Las puntuaciones estándar de pruebas psicométricas es un ejemplo de este tipo de variable que se conoce como de intervalos. Por ejemplo, el espacio que existe entre un cociente de Inteligencia

IQ de 75 y uno de 85 son 10 unidades iguales. Ese mismo espacio existe entre un IQ de 115 y 125.

Observe que el valor "cero" en estas escalas no significa que la propiedad no existe. Por ejemplo un IQ de "cero" no significa que no tiene inteligencia, sino que tiene ciertamente muy poca. De igual forma cero grados centígrados (0°C) no significa que no hace frío o que no hace calor.

**La escala de intervalo**, considera pertinente información no sólo sobre el orden relativo de las necesidades, como en la escala ordinal, sino también del tamaño del intervalo entre mediciones, esto es, el tamaño de la diferencia (resta) entre dos medidas. La escala de intervalo involucra el concepto de una unidad de distancia. Por ejemplo la escala con la cual casualmente representamos la temperatura; un incremento en una unidad (grado) de la temperatura está definido por un cambio particular en el volumen de mercurio en el interior del termómetro, de esta manera, la diferencia entre dos temperaturas puede ser medida en unidades (grados). El valor numérico de una temperatura es meramente una comparación con un punto arbitrario llamado "**cero grados**". La escala de intervalo requiere un punto cero, como también, una unidad de distancia, pero no importa cual punto se define como cero ni cual unidad es la unidad de distancia. La temperatura ha sido medida adecuadamente por mucho tiempo en las escalas fahrenheit y centígrada, las cuales tienen diferente temperatura cero y diferentes definiciones de 1 grado o unidad. El principio de la medida de intervalo no es violado por cambios en la escala o en la localización.

**Cero absoluto:** Se refiere a escalas en las que el cero es significativo. Estas variables llevan el nombre de "razones". Un buen número de variables numéricas poseen esta propiedad, especialmente los conteos discretos, aunque muchas variables continuas también la poseen. Por ejemplo el número de actos de violencia doméstica es una variable numérica discreta para la cual "cero" actos de violencia doméstica tiene un único significado. De igual forma, salario, una variable numérica continua puede ser medida de forma significativa como \$0.00 si no ha recibido ningún salario.

**La escala de razón**, es usada cuando no solamente el orden y el tamaño del intervalo entre medidas son importantes, sino también la razón (o cociente) entre dos medidas. Si es

razonable hablar de que una cantidad es "dos veces" otra cantidad, entonces la escala de razón es apropiada para la medición, como cuando medimos distancias, pesos, alturas, etc. Realmente la única diferencia entre la escala de razón y la escala de intervalo, es que la escala de razón tiene un punto cero natural, mientras que en la escala de intervalo éste es arbitrario. En ambas escalas la unidad de distancia es arbitrariamente definida.

Es muy importante tener presente la escala de medición cuando se realiza un estudio, puesto que las pruebas estadísticas varían dependiendo de la escala de medición de las características en referencia.

En general puede decirse que la escala de razón es la que tiene a su disposición una mayor cantidad de herramientas estadísticas para su tratamiento.

A continuación presentamos una síntesis de las características de las diferentes escalas de medición.

Escala	Mutuosmente Exclusivos Exhaustivos	Categorías están Ordenadas	Unidad de Medición Estándar	El valor "cero" es significativo
Nominal	Si	No	No	No
Ordinal	Si	Si	No	No
Intervalos	Si	Si	Si	No
Razones	Si	Si	Si	Si

Figura 22. Cuadro Resumen de las características de las escalas de medición.



A continuación aparece una lista de variables. Utilizando marcas de “visto bueno” (✓) indica lo siguiente: si es (C)ualitativa o (N)umérica; si es numérica indica si es (D)iscreta o (Cont)ínua; por último identifica cuál es la escala o nivel de medición de la variable, esto es

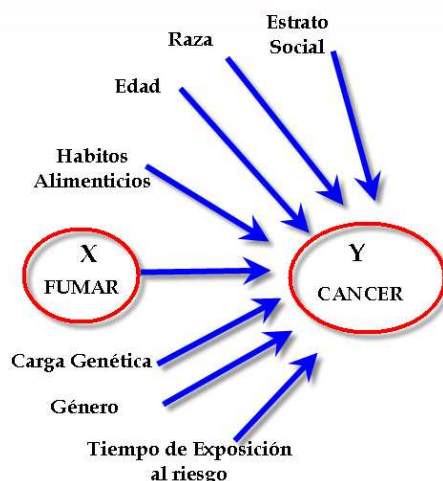
si es (N)ominal, (O)rdinal, (I)ntervalo o (R)azón:

Variable	Cualitativa	Cuantitativa		Escalas de Medición			
		D	Cont	N	O	I	R
Tiempo de Servicio							
Prestigio							
Número hijos							
Marca de tu PC							
Precio de un libro							
El grado de afecto por una persona							
Código postal							
Dureza de un material							
Índice de felicidad							
Índice de pobreza de un pueblo.							
Producto interno bruto PIB							
Velocidad							

## **Validez interna y Control de factores de confusión.**

La dimensión de la validez que se aborda con esta denominación, intenta garantizar que se compare solo lo que es comparable. Así por ejemplo, si un equipo científico está interesado en valorar el impacto de fumar sobre la incidencia de cáncer, su ideal sería comparar dos grupos de personas idénticas en todo, excepto que un grupo es de fumadores y el otro grupo es de no fumadores, para tener la confianza que el efecto observado (la diferencia de la tasa de incidencia) se debe al hábito de fumar y no a otra cosa. Esto sintetiza la idea de validez interna. Al principio de este escrito, abordamos la discusión acerca de la lógica ortodoxa, determinística, del tipo P implica Q, cuyo significado es que siempre que ocurra P, necesariamente ocurrirá Q. Decidimos que este paradigma no es suficiente para modelar los complejos fenómenos de la realidad, pues lo que ocurre en la práctica, es que no siempre un efecto es el resultado de una única causa. Distintas posibilidades se plantearon en la Figura 1. Algunas *Formas posibles de relación causal entre X y Y*).

Para el ejemplo que nos ocupa, podremos visualizar la situación de la siguiente manera:



**Figura 23. Fumar vs. Cáncer. Multicausalidad**

En el diagrama de la Figura 23, se muestran otras variables distintas al hábito de fumar, que podrían influir sobre la tasa de incidencia de cáncer. No todas podrían calificarse en sentido estricto como “causas”, sin embargo, el nivel que ellas asuman, podría facilitar o dificultar la presencia de la enfermedad. Suena razonable, por ejemplo, que un grupo de personas que es fumador habitual por más de 15 años, su tasa de incidencia de cáncer sea mayor que la de un grupo de fumadores que tienen el hábito hace tan sólo un año.

Los esfuerzos que se realizan para intentar aislar el efecto de esas otras variables, distintas al tratamiento de interés, va en dirección de lograr la validez interna. Podemos imaginar el escenario en el cual estamos presentando las conclusiones de la investigación a un grupo de científicos y nos sentimos preparados para cualquier cuestionamiento que se relacione con la comparabilidad. Nos referimos a cuestionamientos del tipo: “¿... *no sería que la diferencia observada en la variable respuesta se debe más bien a...*?”

La validez interna va en el sentido de neutralizar explicaciones alternativas a la que se quiere contrastar.

A estos otros factores, que de no ser controlados de alguna manera, podrían dar pie a interpretaciones alternativas de los cambios en la variable respuesta, menguando el impacto de nuestra explicación, se les llama **factores de confusión**.

De esta manera, ganar la validez interna, es controlar los factores de confusión.

En forma sintética, podríamos decir que:

*Un factor de confusión es una característica que aunque no es de estricto interés en el estudio, afecta de manera diferencial a los individuos de los grupos que se pretenden comparar.*

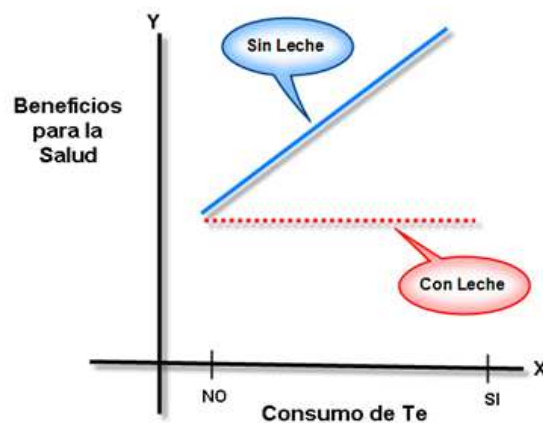
Algunas de las aparentes contradicciones, que se encuentran en los estudios que se reportan en la literatura científica, y que mencionamos anteriormente podrían explicarse en algunos casos, por la falta de control de algún factor de confusión, que fue ignorado en alguno de los estudios.

## ¿Cómo controlar los factores de confusión?

### **Situación 10. Impacto del consumo de Te sobre la salud.**

Veamos cómo podría explicarse la contradicción de algunos estudios, con respecto a los beneficios que produce el Té.

En enero 10 de 2007, el periódico “El tiempo”, da la noticia sobre un informe de una investigación publicada en el European Heart Journal , que el Té tiene buen impacto sobre la salud, solamente si se toma con leche.



#### **INTERACCION.**

*El te con leche es como no tomarlo.*

**El te sin leche tiene buen impacto sobre la salud**

**¿Es Bueno el Te?**

**Depende....**

Figura 24. El significado de la interacción. El caso del Te

Imagine por un momento, que se hacen dos estudios de manera independiente, ambos con el objetivo de evaluar el impacto del té sobre un parámetro de la salud. Uno de los estudios toma una gran muestra de consumidores de té, casi todos con leche; mientras que el otro realiza su estudio con consumidores de té, la mayoría sin leche. Aquí tendríamos muy posiblemente un par de estudios con resultados contradictorios. El primero diría que no tiene ningún efecto, mientras que el otro diría que si lo tiene.

En la Figura 24, se ilustra el impacto de no considerar en un estudio un factor de confusión.

Si se observa el diagrama de la Figura 24, la recta horizontal representa el comportamiento del efecto de tomar té con leche. Esta recta no sube ni baja. Significa que tomar té con leche no produce ningún impacto en el parámetro de interés. Mientras que la otra recta, la que corresponde al té sin leche, tiene pendiente positiva, revelando un incremento en la respuesta, cuando se pasa de “No toma Té” a “Sí toma Té”.

Este diagrama que valora la interacción entre el factor X: Tomar Té y el factor W: Agregar Leche, es una manera de impedir que un factor de confusión afecte la validez. En este caso se ha involucrado este factor de confusión al modelo, midiendo su impacto y deslindarlo del impacto del té.

Responda la siguiente pregunta, observando el diagrama de la Figura 24:

*¿Cuál es el impacto sobre la salud de ingerir té?.*

*La respuesta es: ... Depende.*

*Pues si el té se consume con leche, el efecto es nulo, pero si el té se consume sin leche hay un efecto importante.*

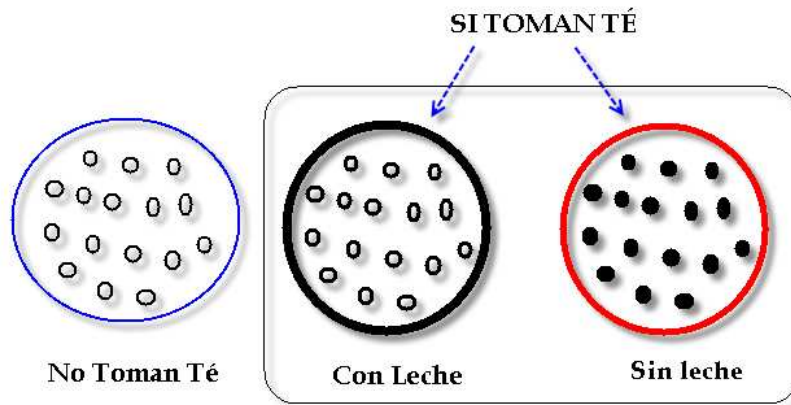
Cuando nos hagamos una pregunta de este tipo y la respuesta es " depende", estamos ante la presencia de una **interacción**.

Cuando hagamos un gráfico como el de la Figura 24 y los segmentos de recta no resulten

paralelos, estamos ante la presencia de *interacción*.

El lenguaje simple, la diferencia entre tomar y no tomar té, depende de si es con leche o no.

**Diseño del estudio para controlar un factor de confusión que interactúa con el factor de interés.**



**Figura 25. Diseño del estudio para controlar un factor que interactúa con el factor de estudio.**

Si se ignorara, como parece haber ocurrido en el pasado, el factor “Leche”, es como si la comparación se hubiera hecho comparando los que no toman té, con los que si toman extrayendo una muestra de las personas que consumen té. Si en la población de estudio la razón de los que consumen te con leche con respecto a los sin leche, fuera de 9 a 1, es decir, que entre los que toman té, el 90% lo hace con leche, entonces en una muestra aleatoria sería fácil que los de Té con leche sean alta mayoría, como se ilustra en la Figura 26.

De esta manera la muestra de los que toman Té , se parecería más a los que toman Te con leche y por lo tanto el efecto de tomar Té se diluiría, llegando posiblemente a la conclusión de que el Te no tiene impacto positivo sobre la salud.

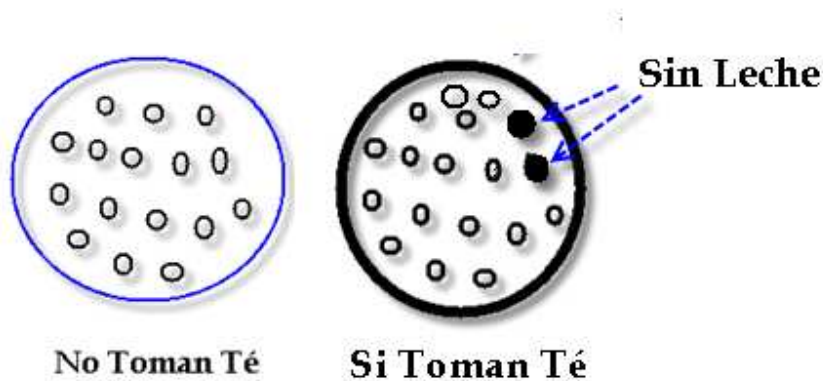


Figura 26. Diseño que ignora el efecto de agregar leche al Té.

### Control de Factores de confusión usando el concepto de Bloques.

En algunas ocasiones existe uno o varios factores distintos de los factores de interés en el estudio, que afecta la respuesta aunque no interactúan con estos. En éste caso, una manera de controlar estos posibles factores de confusión, es a través del uso de “Bloques”. A continuación expondremos un ejemplo, en el área de la agricultura que es donde nacieron estas ideas.

#### **Situación 11. Comparación de dos Dosis de Abono. El diseño de Bloques.**

Supongamos que se desea valorar el impacto de la aplicación de un cierto tipo de abono, digamos A. Para ello se dispone de un gran lote de terreno cerca de la orilla de un río. Allí, serán dispuestas 16 parcelas que serán rifadas estrictamente al azar, para decidir en cuales 8 de ellas se pone abono (A) y en cuales 8 de ellas no (N). Por puro azar, la ubicación de los tratamientos podría quedar como se muestra la Figura 27. Este arreglo, aunque fue producto del azar, y es balanceado, puesto que hay el mismo número de parcelas con abono (A) que sin abono (N), podría presentar inconvenientes por el gradiente de humedad que existe. Las parcelas más cerca de la orilla del río, tendrán más humedad y en la hilera más próxima al río, que se supone tiene más humedad, la gran mayoría corresponden a parcelas con el tratamiento de abono. Lo contrario por supuesto, ocurre en la hilera más lejana al río.

Al final quedarán dudas, sobre si el efecto encontrado se debe a que las parcelas tenían

abono, o a que tenían mayor humedad. Aquí el gradiente de humedad actúa como un factor de confusión.

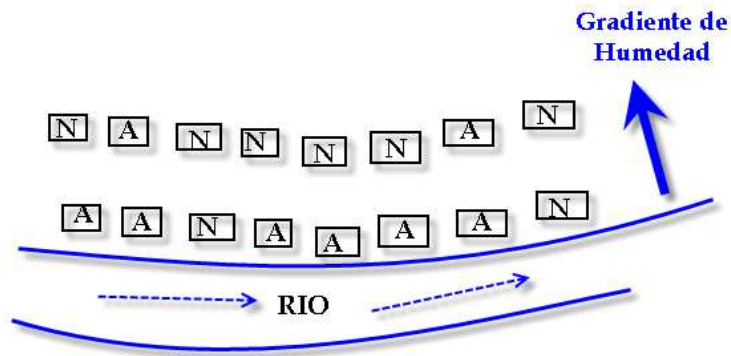


Figura 27. Diseño completamente al azar para la estimación del efecto de aplicación de abono (A).

Una manera de controlar el factor de confusión “gradiente de humedad” es restringiendo la aleatoriedad usando el criterio de Bloques, como se muestra en la Figura 28.

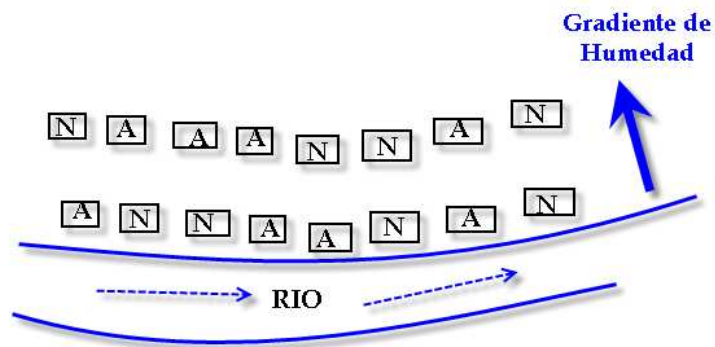


Figura 28. Control de factores de confusión usando Bloques

Ahora se ha garantizado que en cada nivel de humedad estén justamente representados los dos tratamientos. Esto se ha logrado rifando en cada hilera las parcelas de tal manera que queden cuatro con abono y cuatro sin abono. Análogamente se hizo con la segunda hilera.

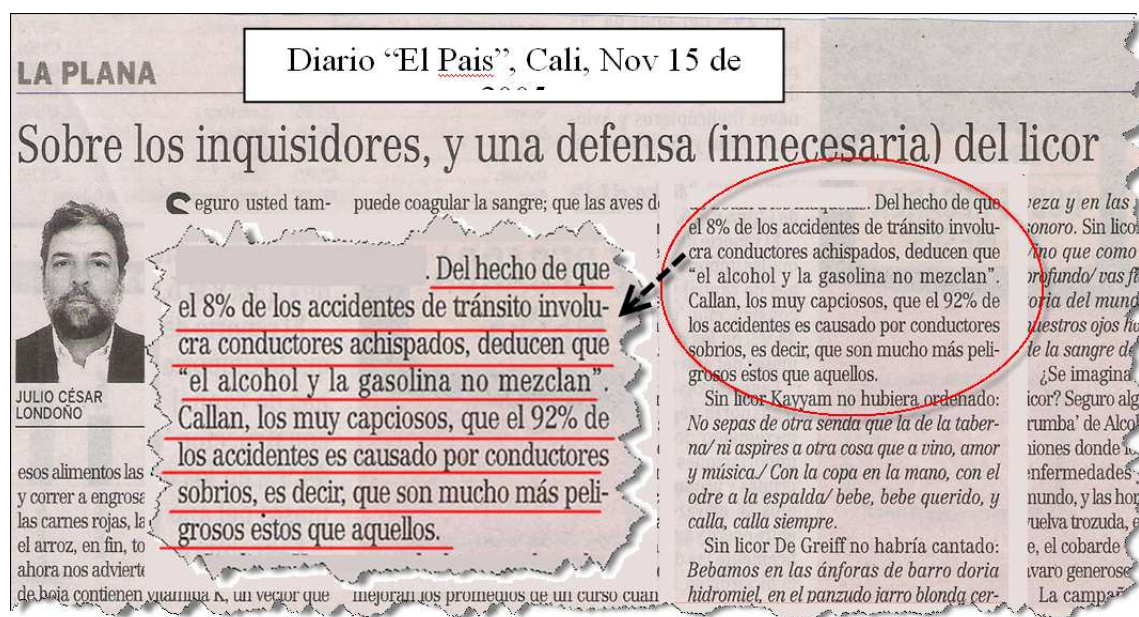
Ahora el problema del factor de confusión está resuelto.

### **Validez en las conclusiones.**

En un estudio, es posible que hayamos hecho una muy buena selección de la muestra, un buen diseño del estudio y hayamos definido medidas apropiadas y válidas para nuestros constructores, sin embargo al pasar de los datos o de los resultados numéricos a las conclusiones podemos ser desafortunados, en el sentido de usar herramientas de análisis inapropiadas para el diseño planteado; un caso en el que esto sucedería por ejemplo, será cuando unos datos son recogidos usando muestreo estratificado aleatorio y se pretenden analizar usando las expresiones que corresponden al muestreo aleatorio simple. Aquí hay una falla en el análisis diremos que falla la validez en las conclusiones.

Otra situación en la que falla la validez en las conclusiones, se da cuando hemos hecho muy bien los cálculos y a partir de los resultados sacamos conclusiones equivocadas, hacemos afirmaciones que no pueden deducirse de los resultados numéricos. Veamos un ejemplo.

**Situación 12. ¿Son más peligrosos los conductores sobrios que los ebrios?**



**Figura 29. Falla la validez en las conclusiones. ¿Son más peligrosos los conductores sobrios?**

El escritor Julio César Londoño, escribe en el diario “El país” de Cali, Colombia, el artículo al que se refiere la Figura 29. Aunque muy seguramente, este escritor cuyas obras están cargadas de buen humor, haya hecho las afirmaciones a manera de broma, nos da una

buena oportunidad para ilustrar el significado de validez en las conclusiones, usando un ejemplo en el que se viola dicha validez.

En síntesis se afirma que dado que de los accidentados “solo” el 8% estaba alicorado, mientras que el restante 92% estaba sobrio, se deduce incorrectamente que son más peligrosos los sobrios que los ebrios.

En primer lugar, no pueden obtenerse conclusiones sobre riesgo, observando únicamente a los que han padecido el siniestro. Se requiere siempre conocer la población expuesta al riesgo, cuantificar de dicha población cuántos padecen el riesgo en un período de tiempo determinado, y de este cociente puede hacerse una idea más realista de la probabilidad de padecer el riesgo de cierto evento. Para ilustrarlo de una manera simple y contundente hemos construido en la Figura 30, una situación en la que se cumplan las cifras del artículo que se critica, y sin embargo la conclusión en la opuesta a la que se deduce.

Observe detalladamente la Figura 30. Las cifras que están encerradas en el cuadrado grande, son las únicas que se usan en el artículo que estamos criticando. Se ve claramente como de 12 accidentados, uno estaba alicorado, lo cual representa aproximadamente el 8%, los restantes 11 accidentados, estaban sobrios y representan el 92% restante. Sin embargo en el ejemplo que se ilustra inventando una situación en la cual de los 10.000 conductores observados, 10 estaban alicorados y los restantes 9990 estaban sobrios.

De esta manera podemos decir que el único accidentados alicorado, provino de 10 expuestos al riesgo, lo cual representa una probabilidad del 10%, mientras que los restantes 11 accidentados sobrios, provienen de una población de 9990 conductores, lo cual representa un 0,11%, que es un riesgo 100 veces menor que el de los alicorados, mostrándose de esta forma la falacia de la conclusión.

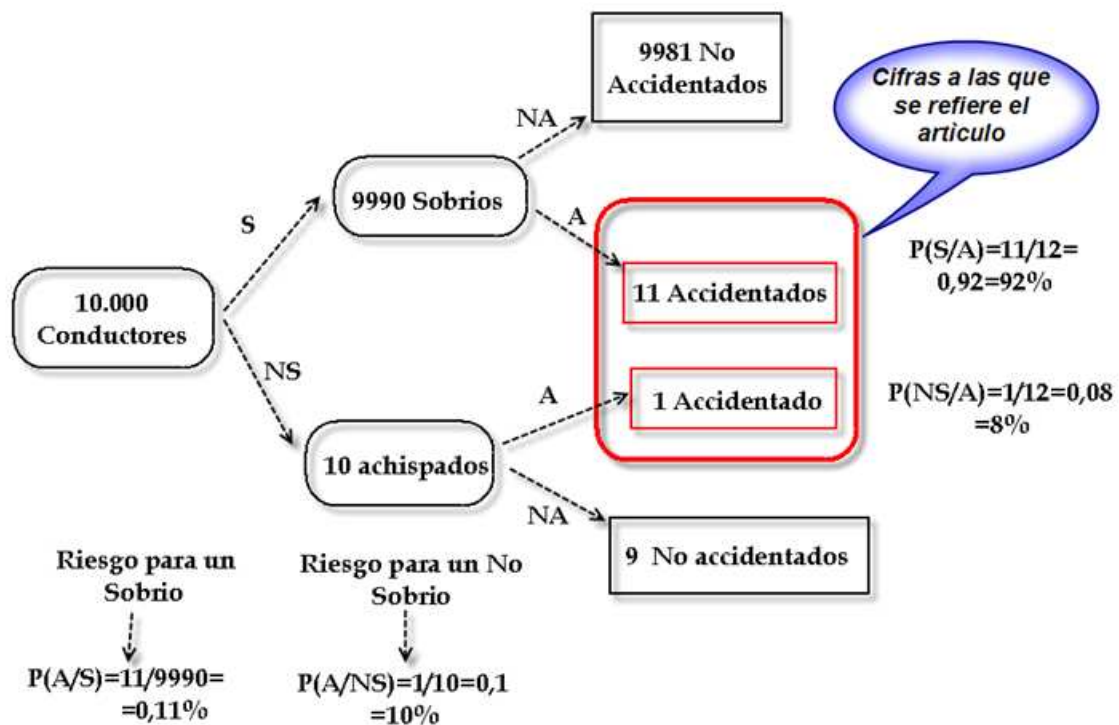


Figura 30. Falla en la validez en la conclusión. Un ejemplo ilustrativo

### Validez en la comunicación.

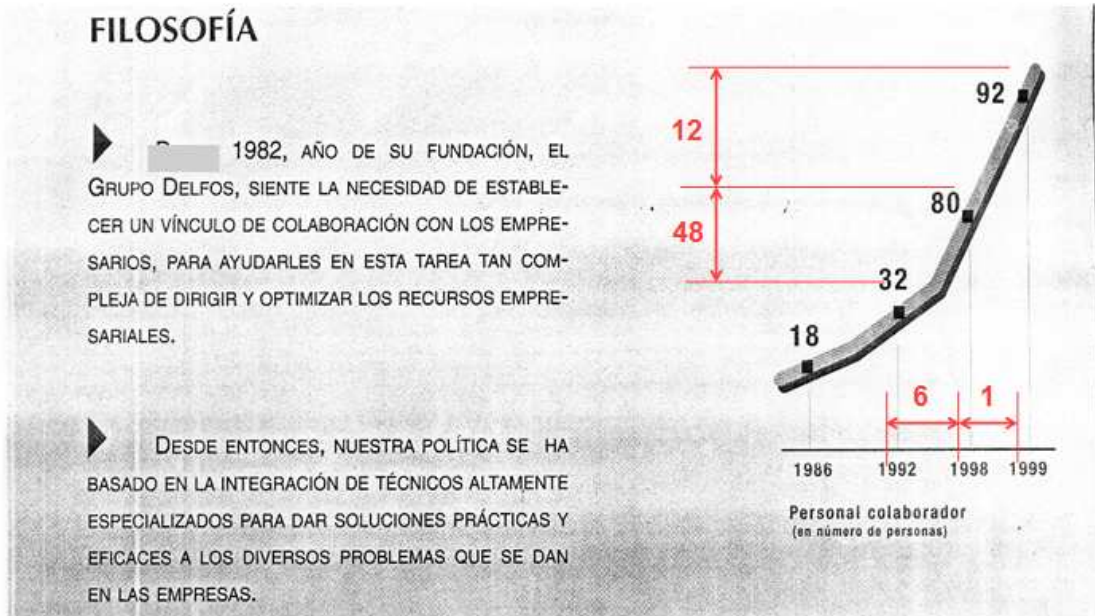
En esta parte, nos ocupamos del adecuado uso de los recursos para expresar a la comunidad académica los hallazgos de nuestras investigaciones. En algunas ocasiones se ha hecho un gran esfuerzo para garantizar la validez externa, la validez del constructo, la validez interna, la validez en las conclusiones y se falla al momento de la comunicación, al hacer uso por ejemplo de un gráfico desafortunado, que hace que el lector se forme una idea inapropiada.

Ilustremos esto con una situación.

#### Situación 13. Ilustración del crecimiento de una empresa. Falla de la validez en la comunicación.

En la gráfica de la figura, se pretende ilustrar la evolución del crecimiento de una empresa en términos del número de colaboradores en el tiempo. Esta fue tomada del folleto de una

empresa consultora.



Tomado del folleto de presentación de una empresa consultora

**Figura 31. Ilustración del crecimiento de una empresa. Falla en la validez de la comunicación.**

Como puede apreciarse hay una distorsión en la pendiente del último tramo, generado por la falta de proporcionalidad en la escala del tiempo registrada en el eje horizontal. Se ve por ejemplo, que las longitudes del segmento que representa seis años (1992-1998) es comparable con el segmento que representa un año (1998-1999), cuando el primero debería ser seis veces el segundo. Esto ocasiona desde el punto vista visual una pendiente exageradamente elevada, lo cual es grave si tomamos en cuenta que cuando se hace un gráfico se espera un impacto visual. Evidentemente el efecto que se logra presente sobre representar el crecimiento. Falla la validez en la comunicación.

# Pensamiento Estadístico

## *Introducción*

En los últimos tiempos en la literatura estadística se hace referencia a “*pensamiento estadístico*”. Así por ejemplo en la *International Statistical Review* de la ISI, en su número de Agosto de 1999, aparece un artículo de Dransfield S.B., Fisher N.I. and Vogel N.J., titulado “*Using Statistics and Statistical thinking to improve Organizational Performance*”. En éste, se cita a Mallows (1998)<sup>9</sup>, quien revisó un número de definiciones sobre *pensamiento estadístico*, una de ellas, es la que define la *American Society Quality* (ASQ 1996) - una filosofía de aprendizaje y acción basada en tres principios fundamentales:

- Todo trabajo ocurre en un sistema de procesos interconectados.
- La variación existe en todos los procesos.
- Entender y reducir la variación es la clave del éxito.

Mallows critica esta definición, porque “...no enfatiza la necesidad del pensamiento estadístico cuando los datos puedan ser irrelevantes y no admite que tal pensamiento sea necesario cuando la variabilidad es tan pequeña (o los datos son muy abundantes) que la variabilidad, no sea un tema central”.

Mallows propone la siguiente definición:

“El pensamiento estadístico, tiene que ver con la relación de datos cuantitativos a un problema del mundo real, a menudo en presencia de variabilidad e incertidumbre. Intenta hacer preciso y explícito lo que los datos dicen sobre el problema de interés.”

---

<sup>9</sup> Mallows C. (1998). “The Zeroth Problem”. *American Statistician* 52, 1-9.

Más recientemente la misma revista en su número de diciembre de 1999, presenta un artículo de Wild C.J. y Pfannkuch M.<sup>10</sup> “*Statistical Thinking in Empirical Enquiry*”, en el cual intentan descubrir los procesos de pensamiento que tienen lugar en la solución de un problema estadístico en un amplio sentido, desde la formulación del problema hasta las conclusiones. El estudio se basa en la literatura sobre el tema y en la realización de entrevistas en profundidad de estudiantes y estadísticos practicantes. El desarrollo de las ideas sobre pensamiento estadístico que viene a continuación corresponde a una traducción y adaptación libres de apartes de dicho artículo.

Los autores dicen “...el pensamiento estadístico es la encarnación del sentido común..” “Nosotros lo reconocemos en cuanto lo vemos” o mejor “..su ausencia es a menudo claramente obvia” y “..para la mayoría de nosotros, ella es producto de la experiencia”.

Se refieren al trabajo de Moore (1997)<sup>11</sup> en el cual presentó la siguiente lista de elementos del pensamiento estadístico, como aprobados por el *Board of the American Statistical Association* (ASA) en respuesta a recomendaciones del *Joint Curriculum Committee* de el ASA y de la *Mathematical Association of America*:

- Necesidad de los datos.
- La importancia de la producción de datos.
- La omnipresencia de la variabilidad.
- La medición y el modelamiento de la variabilidad.

Sin embargo Wild y Pfannkuch (1999) consideran que estos elementos son solo parte de lo que se ha dado en llamar “pensamiento estadístico”. En el área de la calidad, más propiamente en mejoramiento organizacional y de procesos, mucho se ha escrito pero para

---

<sup>10</sup> Wild C.J. and M. Pfannkuch. (1999). “ Statistical Thinking in Empirical Enquiry”. *International Statistical Review*, 67,3, 223-265. ISI.

<sup>11</sup> Moore D. (1997). “New Pedagogy and New Content: El Case of Statistics.”. *Interational Statistical review*, 65, 123-165.

una audiencia específica. Snee (1990)<sup>12</sup> define pensamiento estadístico como “...El proceso de pensamiento que reconoce la variación que existe alrededor de nosotros y presente en todo lo que hacemos, considerar que todo trabajo es una serie de procesos interconectados, identificar, caracterizar, cuantificar, controlar y reducir la variación para proveer oportunidades para el mejoramiento”.

Enfatizan que un extenso marco de trabajo ha sido desarrollado en lo que a modelación estadística se refiere, con aspectos técnicos del diseño y el análisis que son aplicables una vez que el problema ha sido formulado, las variables han sido determinadas y el diseño básico del estudio ha sido definido. Sin embargo, una enorme cantidad de pensamiento estadístico ha sido realizado antes de llegar a esta etapa, descubriendo nexos entre la información en los datos y el conocimiento del contexto a través del proceso estadístico global.

Wild y Pfannkuch (1999), a partir de su experiencia y de la observación empírica, intentan explorar la complejidad de los procesos de pensamiento involucrados en la solución de problemas reales que usan la estadística para mejorar tales soluciones. Intentan desarrollar un marco para los patrones de pensamiento considerados en la solución de problemas, estrategias y la integración de los elementos estadísticos con los de la solución de problemas.

Wald y Pfannkuch (1999) identifica en su estudio empírico cuatro dimensiones del pensamiento estadístico: (a) La dimensión del ciclo de investigación, (b) Tipos de pensamiento, (c) Ciclo de cuestionamiento, (d) Disposición o actitud. Estas se desarrollan a continuación matizadas por el autor de este escrito.

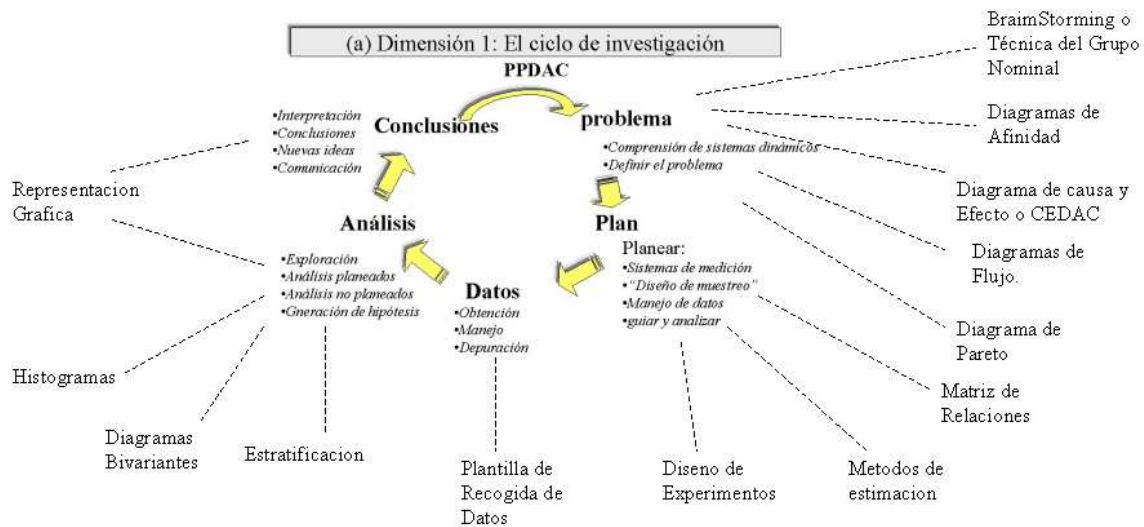
---

<sup>12</sup> Snee. (1990). “Statistical Thinking and its contributions to Quality”. The American Statistician 44, 116-121.

## Un marco para el pensamiento estadístico

### Dimensión 1: El ciclo investigativo

La primera dimensión se relaciona con la manera como uno actúa y lo que piensa durante el desarrollo de una investigación estadística. Las siglas PPDAC, corresponden a Problema, Plan, Datos, Análisis, Conclusiones. (Figura 32)



**Figura 32. Dimensión 1 del pensamiento estadístico: El ciclo de investigación**

El ciclo PPDAC, se refiere a la abstracción y solución de un problema estadístico inherente a una problemática real. La mayoría de los problemas están enmarcados en el deseo de cambiar un sistema para mejorarlo. La curiosidad que guía la investigación está justificada generalmente, en el convencimiento de que dichos cambios conducirán a un beneficio práctico en el largo plazo.

Una solución basada en el conocimiento del problema real requiere un mejor entendimiento de cómo funciona el sistema y quizás también de cómo reaccionará a posibles cambios en las entradas (*inputs*), en los propios componentes y en el ambiente. Ciertos objetivos de aprendizaje deben haber sido logrados para llegar al nivel deseado de entendimiento. Un ciclo de investigación PPDAC, deberá ser realizado para lograr cada una de las metas de aprendizaje. El conocimiento ganado y las necesidades identificadas, se convierten en

puntos de partida para nuevos ciclos PPDAC.

Las personas que trabajan con estadística, se sentirán muy familiarizados con este ciclo y no hallarán en él muchos nuevos elementos.

## Dimensión 2: Tipos de pensamiento

Los tipos de pensamiento están categorizados según los planteamientos de Wild y Pfannkuch (1999) y sobre estos descansan los fundamentos del pensamiento estadístico. (Ver Figura 33)

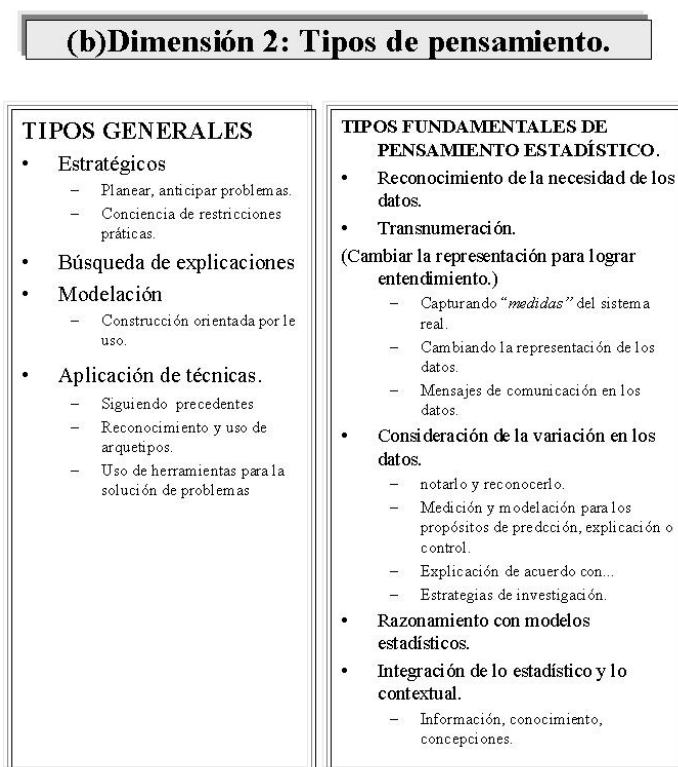


Figura 33. Dimensión 2: Tipos de pensamiento estadístico

### Reconocimiento de la necesidad de los datos

El reconocimiento de lo inapropiado de las experiencias personales y de la evidencia anecdotal, que conducen al deseo de basar las decisiones sobre datos colectados

deliberadamente, es un impulso estadístico. La necesidad de pasar de las anécdotas a las afirmaciones soportadas por datos válidos.

## **Transnumeración**

La idea más fundamental en el enfoque estadístico para el aprendizaje, es la de formar y cambiar las representaciones de datos sobre aspectos de un sistema para llegar a su mejor entendimiento. La *transnumeración* ocurre cuando encontramos la manera de obtener los datos (a través de medición o clasificación) y capturamos los elementos significativos del sistema real. Esto permea todos los análisis estadísticos. Ocurre cada vez que cambiamos nuestra manera de mirar los datos, con la esperanza de encontrar nuevos significados. Podemos mirar a través de muchas representaciones gráficas, hasta encontrar algunas realmente informativas. Podemos expresar los datos vía transformaciones y reclasificaciones para obtener nuevas ópticas. Podemos tratar con una variedad de modelos estadísticos. Y al final del proceso, la transnumeración ocurre de nuevo, cuando intentamos hallar representaciones adecuadas para comunicar nuestros hallazgos a otros. Transnumeración es un proceso dinámico de cambio en las representaciones para lograr entendimiento.

## **Variación**

El pensamiento estadístico, en el sentido moderno, se relaciona con la toma de decisiones bajo incertidumbre. La causa de esta incertidumbre es la variación, cuyo estudio se constituye en elemento esencial, pero sin olvidar la razón del mismo.

Se aborda el estudio de la variabilidad, con el propósito de explicar, predecir o controlar un proceso. Definir el propósito y tenerlo presente, es indispensable, pues no se mide y se modela en el vacío.

Se mide y se modela para investigar estrategias tales como la aleatorización y el bloqueo.

## **Un conjunto específico de modelos.**

Toda forma de pensamiento hace uso de modelos. La principal contribución de la

estadística al pensamiento ha sido su propio conjunto de modelos o marcos, para pensar en ciertos aspectos de la investigación de una manera genérica. En particular han sido desarrollados métodos para el diseño de estudio y análisis. Modelos matemáticos que involucran componentes aleatorias. Grandes partes del proceso investigativo, tales como análisis de problemas y medición, han sido abandonados por los estadísticos, y los educadores de la estadística para el reino de lo particular, quizás para ser desarrollado separadamente dentro de otras disciplinas. Hay necesidad de expandir la riqueza de los modelos estadísticos.

### **Conocimiento contextual, conocimiento estadístico y síntesis.**

Las materias primas con las cuales trabaja el pensamiento estadístico son el conocimiento estadístico, el conocimiento del contexto y la información contenida en los datos. El pensamiento en sí mismo, es una síntesis de estos elementos para producir implicaciones, indicios y conjeturas. No se puede ejercer el pensamiento estadístico, sin disponer del conocimiento del contexto. Uno tiene que traer a colación todo el conocimiento relevante, sin importar la fuente, sobre la situación de interés, para luego hacer conexiones entre el conocimiento existente sobre el contexto y los resultados del análisis para obtener hechos con sentido o significantes. Idealmente, todo este conocimiento debería ser residente en el mismo cerebro, pero frecuentemente esto no es posible, por tal razón, las investigaciones deben ser trabajo de equipo, en los cuales concurren personas de diferentes experticias. La Figura 34 pone de manifiesto la síntesis entre ideas e información del área de contexto y la estadística.

## Interacción entre contexto y estadística

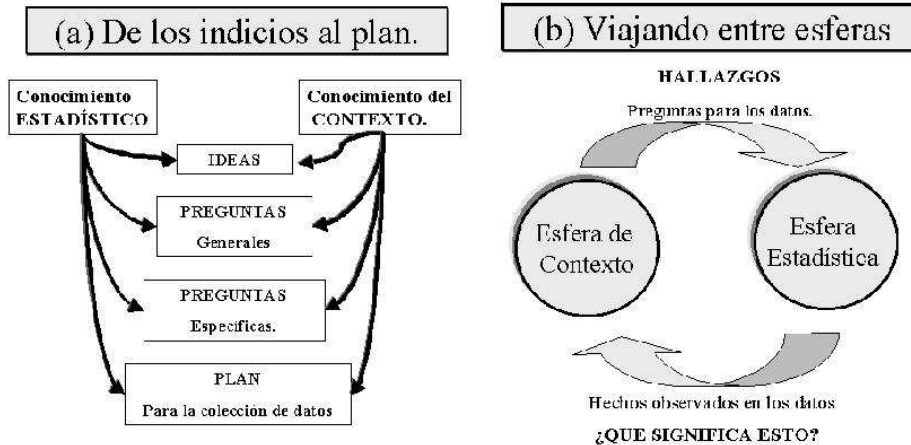


Figura 34. Interacción entre contexto y estadística

En la parte (a) de la Figura 34 se muestra la evolución de una idea, desde el más temprano indicio hasta la formulación de preguntas estadísticas, suficientemente precisas, para ser respondidas por algún conjunto de datos, cuya obtención debe ser cuidadosamente planeada. Las primeras etapas son guiadas casi completamente, por el conocimiento del contexto. El pensamiento estadístico contribuye más como un cristizador del pensamiento.

La parte (b) de la Figura 34 ilustra el continuo movimiento entre el pensamiento en la esfera del contexto y la esfera estadística. Esto se desarrolla todo el tiempo a través del PPDAC. Por ejemplo el análisis en la etapa de preguntas que son sugeridas por el conocimiento del contexto, requerirán consultar los datos, los cuales temporalmente empujarán a la esfera estadística, donde se visualizarán hechos que nos impulsarán a regresar a la esfera del contexto para responder las preguntas: “¿Por qué ocurre esto?” y “¿Qué significa?”.

### Tipos generales de pensamiento aplicados en el contexto estadístico

#### *Pensamiento estratégico*

Se entiende por pensamiento estratégico, aquel que está orientado a decidir qué hacer, ahora o en el futuro, y como se hará. Esto incluye cosas como: Planear cómo enfrentar una tarea; fraccionar tareas en subtareas; establecer guías para la realización de las subtareas; división del trabajo; anticiparse a los problemas y planear cómo evitar los posibles obstáculos.

Una parte muy importante del pensamiento estratégico es tener conciencia de las restricciones y tomarlas en cuenta en la planeación.

La estadística real tiene que ver poco con la búsqueda de la respuesta “correcta” y tiene mucho que ver con encontrar la mejor (o una útil) solución dentro de ciertas restricciones.

Muchos factores limitan la calidad y la efectividad del pensamiento estadístico. Algunos de estos factores son internos al propio pensador. La *falta de conocimiento*, constituye obviamente una importante restricción. Desafortunadamente lo que “sabemos”, no solo es nuestro más caro tesoro, sino que puede ser nuestra mas grande maldición, pues el fundamento de lo que nosotros “sabemos” a menudo no está sólidamente fundamentado.

Nuestras preconcepciones pueden llevarnos por mal camino de muchas maneras, por ejemplo cegándonos a considerar algunas posibilidades, debido a que lo que nosotros “sabemos” determina donde miramos y nos puede hacer poco sensibles a información relevante.

Retar lo que “sabemos” y damos por garantizado, puede remover un obstáculo y llevarnos a un nuevo punto de vista. Esto ocurre generalmente cuando gente con diferentes disciplinas discuten sobre el mismo problema.

Tenemos la tendencia a resolver problemas usando las experiencias precedentes. En investigación aplicada, esto ocurre todo el tiempo y a menudo los métodos estadísticos de sus precedentes son inadecuados.

Como se discutirá cuando se considere la dimensión 3, sobre las *actitudes*, algunas personas que no son curiosas, ni imaginativas o escépticas, serán menos efectivas que otras

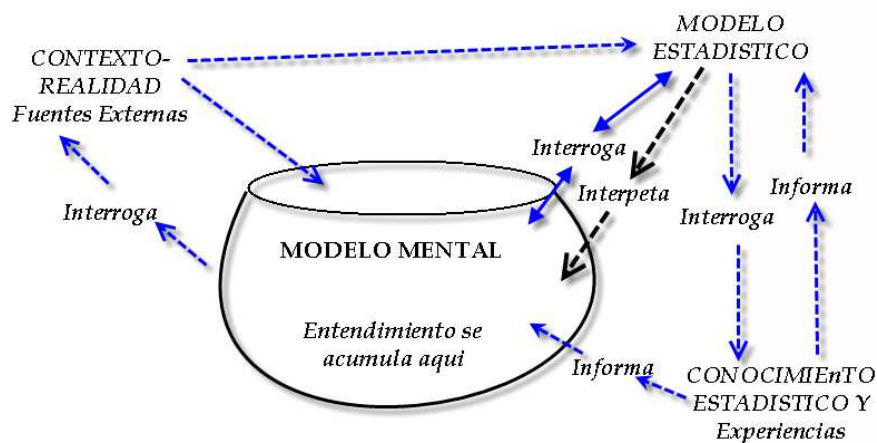
que si lo son:

Tampoco puede desconocerse un factor de *habilidad*, que puede diferenciar a dos profesionales. Dos personas de cara al mismo conjunto de información, puede darse que una de ellas establezca más y mejores conexiones útiles sobre hechos esenciales que la otra persona. Por otro lado, inadecuadas habilidades de comunicación limitarán las posibilidades de interactuar con los miembros del equipo de trabajo, perdiendo oportunidades de extraer de los otros, información vital para la solución de la problemática que enfrentan.

### **Modelación**

Construir modelos y usarlos para entender o predecir el comportamiento de características que nos interesan del mundo parece ser una manera completamente general del pensamiento.

Todos los modelos son simplificaciones de la realidad, en las cuales necesariamente descartamos información. Tenemos la esperanza de haber capturado la información esencial para la problemática que se pretende abordar y que la pérdida de información, no invalide las conclusiones obtenidas.



**Figura 35. Aprendiendo vía estadística**

La Figura 35 ilustra la manera en la cual nosotros aprendemos del contexto de la realidad, como procede en una investigación estadística.

“Entender” es construir modelos mentales del contexto de la realidad. Estos modelos son alimentados con información sobre el mundo real, por ejemplo incorporando el “conocimiento experto”.

En un mundo ideal se debería verificar continuamente la pertinencia de la correspondencia entre el modelo y la realidad, “interrogando” el contexto de la realidad. Parte de la información que nosotros buscamos y conseguimos de la realidad, está en los datos estadísticos. Nosotros construimos modelos estadísticos para sacar provecho al visualizar esta información (“interpretar”), lo cual retroalimenta el modelo mental. “Modelos estadísticos” aquí, tiene un carácter más general que, por ejemplo, regresión o diseño de experimentos, se refiere a toda nuestra concepción estadística del problema, que afecta la manera como recolectamos los datos y como los analizamos. En la Figura 35 también se incorpora el rol del conocimiento estadístico y de la experiencia.

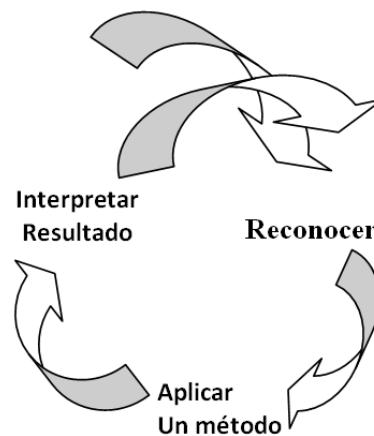
Adicionalmente, dependiendo del problema, de la educación y experiencia del pensador, los elementos estadísticos, pueden también ser parte de una manera de pensar acerca de nuestro mundo y será parte integral de nuestros modelos mentales del contexto de la realidad.

### ***Aplicación de técnicas***

Una técnica básica en la solución de problemas en la ciencias matemáticas, es encontrar una manera de plantear un nuevo problema, en los términos de uno que ya ha sido resuelto, para aplicar el mismo esquema de solución o adaptarlo. La disciplina de la estadística es en sí misma una manifestación de ésta estrategia. La teoría estadística, hace eficiente el proceso de mapeo, al crear problemas arquetipos y ligar a ellos el método de solución. Para usar la estadística, nosotros primero reconocemos elementos de nuestro contexto que pueden ser útilmente trasladados a nuestro modelo (Un proceso de abstracción de lo

particular a lo genérico), opera dentro de ese modelo y luego trasladamos de nuevo los resultados al contexto. (De lo genérico a lo particular). (Adicionalmente, los estadísticos, están siempre tomando prestadas ideas sobre solución de problemas de experiencias anteriores, para ser aplicadas a otros problemas y otros conjuntos de datos).

Implementación de las estrategias de problemas arquetipos y la aplicación práctica de cualquier técnica, algoritmo o concepto, involucra los tres pasos que muestra la Figura 36



**Figura 36. Usando una técnica.**

Es característica de la estadística, que apliquemos modelos estadísticos relativamente sofisticados para analizar diseños experimentales. Sin embargo, parece ser que solo aquellos que se dedican al mejoramiento de la calidad usan herramientas (como diagramas de causa y efecto y técnicas de diagrama de proceso) para analizar la naturaleza del problema en sí. Para los demás, parece ser un proceso de construcción imaginativa de un modelo mental del sistema, sin organización discernible.

### **Dimensión 3: El ciclo interrogativo**

El ciclo interrogativo ilustrado en la Figura 37 es un proceso de pensamiento genérico en constante uso en la solución de problemas estadísticos. Parece ser que el investigador permanece siempre en un estado de interrogación mientras resuelve el problema. El ciclo se aplica a niveles macro, pero también a muy detallados niveles de pensamiento porque el

ciclo es recursivo. Subciclos están inicialmente dentro de un ciclo mayor, por ejemplo, el paso de “chequear” de cualquier ciclo, puede dar origen a un subciclo de interrogación completo.

El esquema ordenado sobre una rueda, es una idealización de lo que quizás debería ocurrir. A continuación se exploran las componentes del ciclo de interrogación con más detalle.

## Generar

Por esto se entiende, la capacidad de imaginar y de proponer posibilidades, bien sea de manera individual o en grupo. Puede aplicarse a la búsqueda de causas, explicaciones y mecanismos, a las partes de un sistema que pueden interrelacionarse, y a otros bloques de construcciones mentales y modelos estadísticos. Puede también aplicarse a tipos de información que es necesario buscar para completar los datos que se requieren para verificar una idea o para planear un enfoque para resolver un problema. La generación de posibilidades puede ser del contexto, datos o conocimiento estadístico, aplicados al presente problema o puede referirse a futuras investigaciones. (Generación de hipótesis).

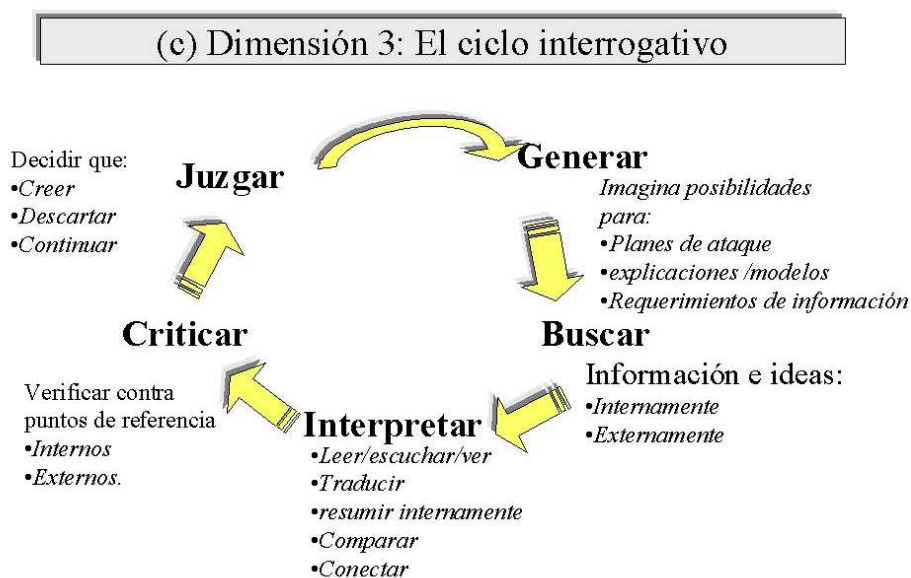


Figura 37. Dimensión 3: Ciclo interrogativo

## **Buscar**

La generación tiende a ser seguida por la búsqueda de información, interna o externa. Para la búsqueda interna, podemos observar a la gente pensando “..yo se algo de esto..” y hurgando en sus memorias conocimiento relevante. La búsqueda externa consiste en obtener información ajena al propio pensador o al equipo de investigación. Los estadísticos que están trabajando en el problema, lo comentan con otras personas, como clientes, colegas, expertos en el contexto, personas que “trabajan en el sistema”. La búsqueda incluye la lectura de literatura relevante. A nivel macro, incluye la colección de datos estadísticos, mientras que para un nivel de más detalle incluye el cuestionamiento de los datos que dispone.

## **Interpretar**

El significado de esta fase es tomar y procesar los resultados de la búsqueda.

Leer/ver/escuchar → Traducir → Resumir internamente → Comparar → Conectar

Este proceso se aplica a todas las formas de información incluidos gráficos, resúmenes y otros productos del análisis estadístico. “Conectar”, punto final de “interpretar”, se refiere a la interconexión de las nuevas ideas con nuestros modelos mentales existentes y la prolongación de nuestros modelos mentales para acoplar o encajar estas interrelaciones.

## **Criticar**

La fase de crítica puede ser aplicada al ingreso de información y a las ideas involucradas en el proceso de verificación de la consistencia interna y compatibilidad con puntos de referencia. Se pregunta: “¿Esto es correcto?” “¿Esto tiene sentido?” “¿Esto está de acuerdo con lo que sabemos?”. Verificamos contra nuestros puntos de referencia internos, arguyendo contra nosotros mismos, sopesando nuestro propio conocimiento del contexto, contra nuestros conocimientos estadísticos, contra las restricciones que el trabajo nos impone y anticipamos potenciales problemas que serían consecuencia de una elección particular. Verificamos contra puntos de referencia externos, tales como otras personas

(clientes, colegas, expertos, etc.); contra la literatura disponible, contra otras fuentes de datos (históricas por ejemplo).

Similarmente podemos intentar tomar distancia, para monitorear nuestro propio pensamiento. Puntos de referencia para hacer chequeos en este sentido, podrían incluir aspectos relativos al (1) *propósito del pensamiento*: “¿Estamos andando en la dirección de la pregunta que mi cliente quiere responder?” o puede referirse a una especie de acuerdos objetivos: (2) *Sistema de creencias* “¿Estoy siendo guiado en forma permisiva o indulgente por preconcepciones no garantizadas, en favor de mi cliente o de mi comunidad? (3) *Respuestas emocionales*.

## Juzgar

Esta es la decisión, que constituye el punto final de la crítica. Qué conservamos, qué descartamos o ignoramos y qué mantenemos tentativamente, qué creemos ahora. Aplicamos el juicio para cosas tales como: La fiabilidad de la información, la utilidad de las ideas, la factibilidad de los planes, lo apropiado de un “encapsulamiento”.

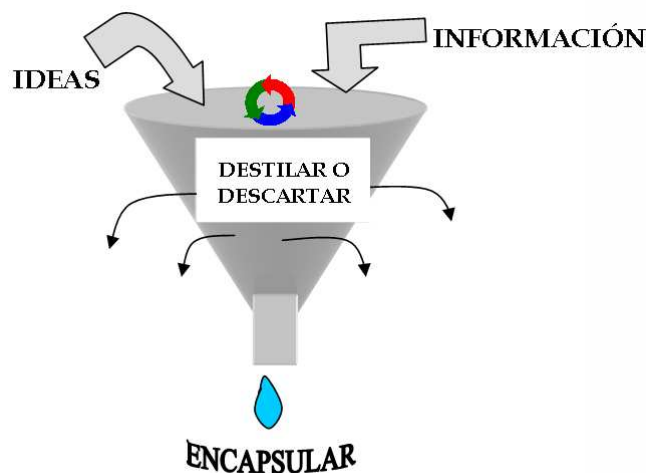


Figura 38. Destilación y encapsulamiento

Se relaciona con el contexto y el entendimiento estadístico, la relativa plausibilidad y competencia de las explicaciones, el conjunto más probable de escenarios, la necesidad de

más investigación, y muchas otras decisiones relacionadas con la construcción y razonamiento de modelos

El resultado del engranaje en el proceso interrogativo es una *destilación y encapsulamiento*, de ideas e información. Los ciclos interrogativos internos, nos ayudan a extraer la esencia de los *inputs*, descartando distracciones y detalles a lo largo del camino. (Ver Figura 38)

#### **Dimensión 4: Disposiciones (Actitudes)**

En esta sección se discutirán las cualidades personales, categorizadas en la Figura 39. Aunque estas características son genéricas, aquí se discutirán en relación con el pensamiento estadístico.

#### **Dimensión 4: Disposiciones (Actitudes y aptitudes).**

- Escepticismo.
- Imaginación.
- Curiosidad y conciencia
  - Observar, notar.
- Mentalidad abierta.
  - a ideas que reten sus preconcepciones.
- Propensión a la búsqueda de significados profundos.
- Ser lógico.
- Compromiso
- Perseverancia

**Figura 39. Disposiciones. (Actitudes y aptitudes)**

#### **Curiosidad y Conciencia.**

Los descubrimientos surgen porque alguien nota algo y reacciona con preguntas internas como “¿Por qué?” “¿Cómo pudo pasar?”, “¿Esto ocurre mas generalmente?”, “¿Cómo puedo explotar este hecho?”. Ser observador (consciente) y curioso, es un perfil apropiado para el proceso de generación de preguntas que el aprendizaje innovativo requiere. Wild (1994), planteó el lema “las preguntas son más importantes que las respuestas”, para

enfatar este punto. Este es un elemento muy básico del pensamiento estadístico y la raíz de casi la mayor parte de la investigación científica. “Notar y preguntar por qué...” es también un aspecto muy crítico de la exploración de datos y el análisis.

### **Compromiso (involucramiento)**

Cuando los investigadores, están intensamente interesados en un problema o área, una elevada sensibilidad y conciencia se desarrolla en la periferia de nuestra experiencia, que puede ser relacionada con el problema.

Las personas son más observadoras en aquellas áreas por las que tienen mayor interés. Un elevado grado de compromiso intensifica la curiosidad y la sensibilidad, la imaginación y la perseverancia.

¿Cómo puede lograrse el compromiso? El interés espontáneo es innato. El conocimiento previo ayuda; es muy difícil estar interesado en algo de lo cual uno no conoce nada. Que el trabajo sea remunerado también ayuda.

### **Imaginación**

Es muy clara la importancia de la imaginación para el pensamiento estadístico. La formación de modelos mentales que atrapen la dinámica esencial de un problema es un proceso profundamente imaginativo, como lo es, ver una situación desde diferentes perspectivas y generar posibles explicaciones para el fenómeno y los hechos que sugieren los datos.

### **Escepticismo**

Por escepticismo, se entiende la tendencia a estar constantemente atento por la lógica y la validez de las acciones que se realizan cada vez que se reciben nuevas ideas e información. Algunos autores, se refieren a esta cualidad, como “adoptar una actitud crítica”. Gal et Al.

(1995)<sup>13</sup> y Pfannkuch (1996)<sup>14</sup> discuten sobre el pensamiento crítico, basados en la interpretación de reportes basados en estadística. “¿Son justificadas las conclusiones obtenidas?”. Se puso especial atención a la motivación, predisposición y objetividad del escritor, lo cual afecta la confianza en algo que ha sido planteado. Los estadísticos experimentados, muy probablemente, evocan en forma automática, “preguntas técnicas preocupantes”, que se relacionan con lo pertinente de las mediciones realizadas o del diseño del estudio, la calidad de los datos, la disponibilidad de un método de análisis y si las conclusiones son realmente soportadas por los datos. Las explicaciones propuestas, crean preocupaciones acerca de si esta es la única opción o existen otras explicaciones posibles.

Otro aspecto incluye el *sentido del número* y escepticismo. Un paso previo en este sentido es “¿Es *justificada* esta información/conclusión?”. “¿Es creíble esta información/conclusión?”. Vale la pena notar, que uno está menos inclinado a ser escéptico, cuando las conclusiones compatibilizan con nuestras preconcepciones. Se hace necesario un esfuerzo consciente, para controlar esta tendencia.

## **Ser lógico**

La capacidad para detectar cuando una idea se sigue de otras y cuando no, y la capacidad de construir un argumento lógico, es una característica clave en todo pensamiento. La síntesis de nueva información con el conocimiento existente es un sólido camino para la búsqueda de *implicaciones*. Razonamientos lógicos aseguran la manera de llegar a conclusiones válidas. Para que el escepticismo sea útil es necesario que esté soportado por el

---

<sup>13</sup> Gal I., Ahlgren, C. Burril, G., Landwher, J., Rich., W&Begg, A. (1995). Working Group: Assessment of Interpretive Skills”. Writing group Draft Summaries Conference on Assessment Issues in Statistics Education. Philadelphia University of Pennsylvania.

<sup>14</sup> Pfannkuch, M. (1996). “Statistical Interpretation of Media Reports”. In New Zeland Statistical Association Research in the Learning of Statistics Conference Proceedings, Eds. J. Neyland and M. Clark, pp 67-76. Wllintong: Victoria University.

razonamiento; asumir cosas o información para que las implicaciones puedan ser contrastadas con los datos.

## **Tendencia a buscar significados profundos**

Significa, no solo tomar lo de valor que se encuentre en la superficie, sino estar preparado para cavar un poco más profundo. Por otro lado, otras deseables propensiones son: tener *mente abierta*, pues esto nos ayuda a considerar nuevas ideas e información, aunque esta pueda estar en conflicto con nuestras propias suposiciones y la *perseverancia*, que habla por si misma.

¿Las “disposiciones” (Aptitudes y actitudes) pueden ser enseñadas?

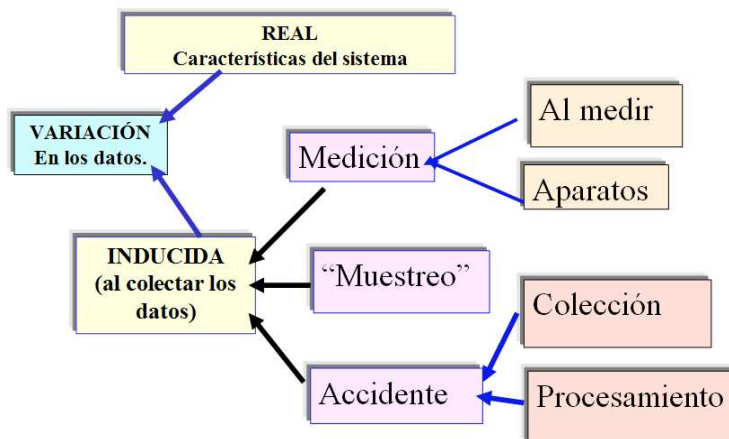
No obstante que existe la provocación de considerar estas características como innatas al pensador, puede considerarse que estas varían dependiendo del nivel de *compromiso* o *involucramiento*. Las disposiciones de una persona, son dependientes de la naturaleza del problema que se enfrenta, ellas cambian en la medida en que la persona está comprometida con una problemática y la conoce. Así por ejemplo la credulidad o el escepticismo, en un área particular, es el resultado de la ignorancia. En la medida en que se gana experiencia, aparecen más elementos y formas que podrían fundamentar la desconfianza sobre la validez de una información o conclusión, siendo de esta manera, más escéptico.

## ***Variación, Aleatoriedad y Modelos Estadísticos***

### **Variación como punto de partida**

La pieza central de la calidad y de las definiciones de la ASA, sobre pensamiento estadístico es “Variación” o “variabilidad”. Cualquier discusión seria sobre pensamiento estadístico debe examinar el papel de la “variación”. Los primeros tres mensajes de variación son: (1) La variación es omnipresente. (2) La variación puede tener serias consecuencias prácticas. (3) la estadística es el medio para entender y actuar sobre la variación.

### **Omnipresencia de la variabilidad.**



**Figura 40. Fuentes de variación en los datos**

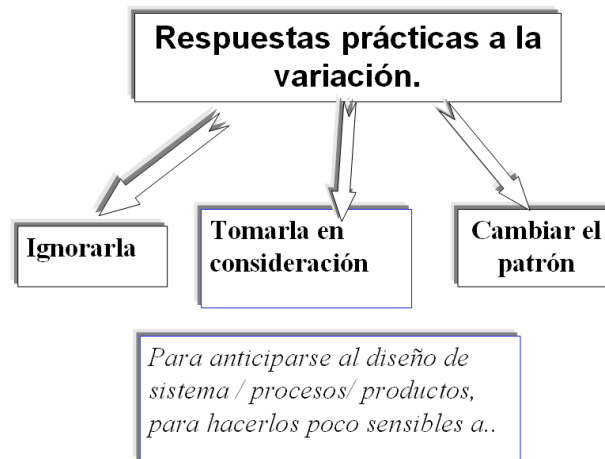
La variación es una realidad observable. Está en toda parte y en todas las cosas. La variabilidad afecta todos los aspectos de la vida.

Ningún par de artículos manufacturados son idénticos, ningún par de organismos son idénticos, ni reaccionan de idéntica manera. En realidad los organismos son sistemas en constante cambio. Todos los aspectos mencionados se refieren a variaciones inherentes al sistema. La Figura 40 muestra cómo, cuando coleccionamos los datos del sistema, su variación real es complementada con la variación adicionada de varias maneras por el proceso de recolección de datos.

### **Impacto práctico**

Habiendo establecido que la variación está en toda parte, debe entonces demostrarse la importancia práctica del impacto de esta variación, sobre la vida de las personas y sobre la manera como ellas trabajan. Es la variación la que hace impredecible los resultados de sus acciones. La variación hace difícil responder preguntas sobre causa y efecto, hace difícil descubrir los mecanismos. La variación es la razón por la cual la gente ha desarrollado métodos estadísticos sofisticados, para poder filtrar los mensajes que los datos contienen y separarlos del ruido, con que se mezclan.

### **Predecir, explicar y controlar.**



**Figura 41. Respuestas prácticas a la variación.**

La Figura 41 categoriza las respuestas racionales a la variación en el mundo de la acción. Esta es una idealización. La manera como la gente reacciona a la variación puede ser una historia muy diferente. (Ver Joiner<sup>15</sup>, 1994, Capítulo 10).

---

<sup>15</sup> Joiner, B. (1994). "Fourth generation management". New York: McGraw-Hill Inc.



**Figura 42. Impacto de Ignorar la variabilidad . (Talla única)**

Una primera respuesta es pretender que la variación no existe, es decir comportarse como si todos los objetos u organismos fueran idénticos o difirieran entre si de una manera determinística conocida. En algunas circunstancias esto funciona admirablemente bien. Si no fuera así debería rescribirse casi todos los resultados de las llamadas matemáticas aplicadas.

Una segunda manera de responder, es investigar el patrón de variabilidad existente, con el propósito de definir la manera de obrar sobre nuestro sistema, de la misma manera que se hace un vestido y zapatos a medida. Así, la variación es considerada en la etapa de diseño en el enfoque de ingeniería de calidad, para diseñar productos robustos a la variabilidad en los usos y condiciones en que serán tratados. Una tercera manera de responder, puede ser cambiar el patrón de variación por uno más deseable, por ejemplo aumentar el rendimiento medio de un cultivo o disminuir la tasa de fallos (o de mortalidad), aislando las causas manipulables o aplicando tratamientos externos.

El primer enfoque es usado en el área de mejoramiento de la calidad o en salud pública y el

segundo es usado frecuentemente en agricultura o en investigación médica orientada al tratamiento de pacientes individuales. Los estadísticos modelan la variación para propósitos de predicción, explicación o control.

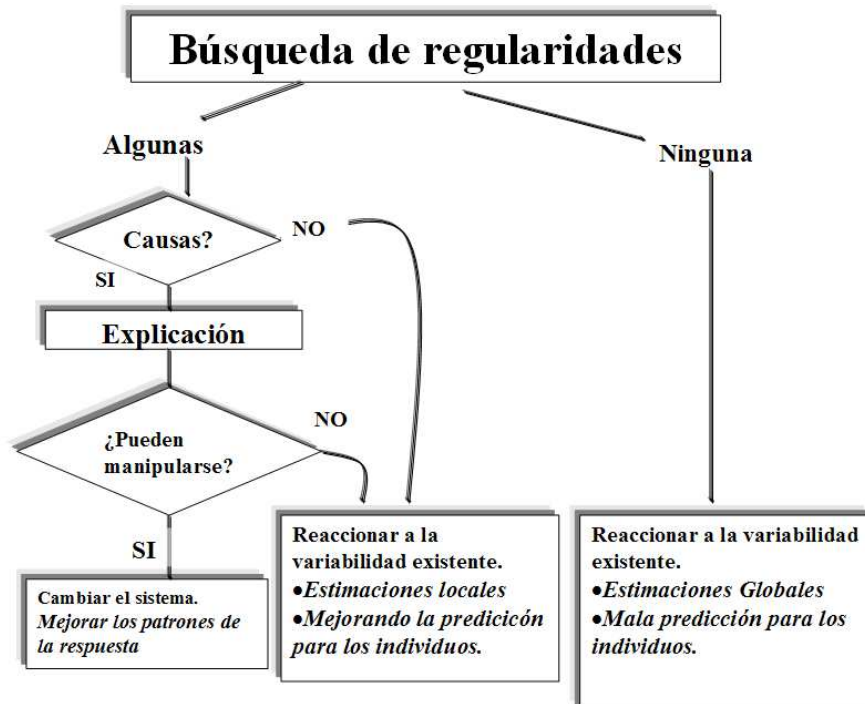


Figura 43. La búsqueda de regularidades y su ganancia

*Control* es cambiar el patrón de variación por otro más deseable. *Predicción* es considerar la variación para obrar en consonancia. *Explicación* para ganar en el entendimiento del por qué diferentes unidades responden de manera diferente, mejorando nuestra capacidad de hacer buenas predicciones y si fuese necesario para el control. Explicación causal y mecanicista, es la meta de la ciencia básica en contraste con la aplicada. En el momento en que preguntamos “¿Por qué?”, estamos buscando causas. La variación no controlada en un sistema nos incapacita para descubrir causas. Nosotros podemos buscar patrones en la variación como se muestra en la Figura 43.

Los estadísticos buscan fuentes de variabilidad para encontrar patrones y relaciones entre variables (“regularidades”). Si no encuentran ninguna, lo mejor que pueden hacer es medir el grado de variabilidad y obrar con la conciencia de su magnitud. Las regularidades no

necesariamente corresponden a causas. En términos de la solución de problemas reales, las causas que no pueden ser manipuladas, son equivalentes a cualquier otra regularidad observada, aunque ellas nos darán más confianza en nuestras predicciones. La presencia de regularidades nos capacita para hacer predicciones y medir la variabilidad de una manera localmente más relevante, por ejemplo más pertinente a un paciente individual. Las causas manipulables abren la opción al control.

### **La búsqueda de las causas**

Los problemas reales, suponen un problema estadístico que muy frecuentemente involucra la búsqueda y aislamiento de causas de alguna respuesta. Siempre y cuando los estudiantes tengan un conocimiento contextual de la situación y sus experiencias les hayan proporcionado una noción de la naturaleza del problema real, ellos plantearán tarde o temprano, un abanico de posibles explicaciones causales. Este parece ser un bien desarrollado impulso que no ha sido explícitamente enseñado. Es un primer paso directo hacia una solución primaria del problema.

El propósito real de muchas de las ideas nuevas que enseñamos es simplemente moderar la búsqueda de las causas para prevenir saltos prematuros hacia las conclusiones. Este rol es secundario. Por lo tanto, probablemente no es sorprendente que después de alguna instrucción estadística, estén mucho más débiles en aquellos estudiantes que tienen el impulso a postular causas. El pensamiento probabilístico, no es tanto, una alternativa al pensamiento determinístico como algunos educadores y estadísticos han sugerido, sino que es algo para encajar en la parte superior de los modos de pensamiento natural, que se enfoca directamente hacia el problema primario. Como un apartado interesante, si una explicación o causa ha sido sugerida a los estudiantes para un conjunto de datos o si los datos han sido estratificados de una manera particular, esto puede constituirse en una gran incitación para los estudiantes para que vayan más allá de la explicación dada, a pensar sobre otras explicaciones y empezar a aportar nuevas ideas.

¿Qué tiene que decir la educación estadística sobre el proceso de búsqueda de causas?. El mensaje mas chillón que podemos dar es “Correlación no es causación”. Esta es la

estadística de Cassandra, el trágico anuncio dice: “Por aquí vamos hacia el desastre”. La verdad es que nosotros usualmente, no tenemos duda que la aleatorización de nuestro experimento es la manera más convincente de establecer que una relación es causal. Pero como lo plantean algunos estadísticos de otros campos distintos a la calidad y la epidemiología, está muy mal vendida la verdadera importancia de la búsqueda de causas. La solución a la mayoría de los problemas prácticos incluye hallar y calibrar agentes de cambio. La educación estadística debería realmente decir a los estudiantes, algo que todo científico sabe: “La búsqueda de las causas es el juego más importante en el pueblo”. Debería decirse: “Así es como la estadística ayuda en esa búsqueda. Estas son algunas de las estrategias y algunas posibles fallas que deben ser evitadas durante el camino...”. Esto podría prevenir a la gente del salto hacia falsas conclusiones, pero también sería la guía hacia conclusiones válidas, conclusiones usables, reemplazando a Cassandra por la literatura de detectives favorita.

### **Pensando acerca de las causas.**

Es irónico que la variabilidad no controlada en un sistema, nos proporcione las mejores oportunidades para hacer el trabajo del detective que descubre las causas. Para establecer relaciones entre las variables aguas arriba y las respuestas aguas abajo, identificamos posibles factores causales. La observación precede a la experimentación. Todas las *ideas* sobre posibles relaciones causales, se originan en la observación, bien sea en forma anecdótica o a través de estudios formales. Los experimentos aleatorizados proporcionan la manera más convincente de confirmar o rechazar la naturaleza causal de una relación observada.

Conducir cualquier clase de estudio para detectar causas y estimar sus efectos, consiste en proveer ideas acerca de los sitios más provechosos para mirar, ideas que son sacadas casi exclusivamente del conocimiento de la temática del contexto y de la intuición. Ideas sobre posibles causas y otros factores que pueden ser importantes predictores del comportamiento de la respuesta, son trasladados a un conjunto de variables a medir (transnumeración) y los datos son tomados para facilitar la investigación de relaciones entre las variables medidas y las respuestas de interés. Las herramientas primarias de análisis en la búsqueda de causas

son los modelos del tipo regresión, es decir, modelos para explorar como Y se comporta ante cambios en el comportamiento de X. (El modesto diagrama de dispersión cae en esta clase).

Cox (1992)<sup>16</sup> distingue entre variables de respuesta (aquellas a cuyo comportamiento queremos encontrar explicación); variables de respuesta intermedias (las cuales miden el efecto intermedio que ocurre a lo largo del camino entre el estado inicial y el estado de la respuesta) y las variables explicatorias (aquellas que queremos usar para explicar o predecir el comportamiento de la respuesta). Las variables explicatorias son además categorizadas en posibles variables causales, propiedades intrínsecas de las entidades bajo estudio y características no específicas. (Por ejemplo, pertenecer a países diferentes). Variables intrínsecas son aquellas cuyos valores no pueden ser manipulados. Las variables intrínsecas son incluidas generalmente, para mejorar nuestra capacidad de detectar relaciones, mejorar la precisión de la estimación de los efectos y explorar cómo las causas pueden actuar de manera diferente en distintos tipos de entidades (interacciones). Detección de fuertes relaciones entre variables no específicas y variables de respuesta, nos llevan a encontrar nuevas variables explicatorias, variables asociadas con las variables no específicas, las cuales podrían explicar la respuesta. Por ejemplo, cuando existen tasas de enfermedad diferentes entre países, empezamos a buscar entre los factores que diferencian los países, como posibles causas. Debe notarse que la distinción hecha entre variables es mucho más que distinción para el análisis. Como conjunto, ellas constituyen una herramienta general de pensamiento, la cual proporciona claridad a la manera como es concebido el contexto del problema. Ellas son un ejemplo de cómo el conocimiento estadístico puede alimentar un modelo mental del contexto de la realidad que es entendible por estadísticos y no estadísticos, es la flecha del “informa” ligada al conocimiento estadístico en el modelo mental de la Figura 35.

---

<sup>16</sup> Cox, D.R. (1992). “Causalidad: Algunos Aspectos estadísticos”. J.R. Statist. Soc. A; 155, part 2, 291-301.

## **Algunas consecuencias de la complejidad**

La mayoría de los sistemas reales son enormemente complejos con variación en un gran número de sus componentes, cada una de las cuales contribuye a la respuesta de interés. Nosotros somos incapaces de manejar tal complejidad y requerimos de estrategias para “filtrar” la gran cantidad de posibilidades que surgen y lograr un número promisorio más pequeño. La distinción que se hace en control de calidad entre variación debido a una causa especial o a una *causa común*, puede ser vista en este sentido. Esto da la idea de distinguir situaciones (variación debido a una causa especial) en las que al parecer el instinto humano reacciona a mirar alrededor de alguna causa inusual que haya podido ocurrir en el sistema antes de que se presentara el evento problema. Esta es probablemente una provechosa estrategia para localizar la causa y separarla de otras situaciones (causas comunes de variación).

La principal estrategia estadística para producir el mencionado “filtrado”, es restringir la atención a variables las cuales tienen fuerte correlación con la respuesta de interés. Nosotros no tenemos esperanza de identificar una causa y caracterizar su efecto, si ella actúa complejamente, de diferente manera, en distintos individuos a través del tiempo. Las únicas causas que esperamos hallar son aquellas que actúan de una manera uniformemente regular. Más aún, solo detectaremos la existencia de una causa, si pensamos alguna manera de mirar la situación que permita revelar dicha regularidad (transnumeración). Debe haber suficiente “variabilidad natural” en la variable\_ causa, en el sistema, para que el efecto de dicha variabilidad sobre la respuesta, pueda ser visto. Las variables causales que se omiten con el procedimiento explicado antes, tienen pocas posibilidades de ser buenos agentes que produzcan cambios sustanciales, a menos que la mentada variable, asuma valores lejos del rango de variabilidad visto en los datos.

## **De asociación a causación**

Aquí es donde Cassandra hace su entrada. El mundo realmente necesita sus advertencias. Es claro, que la gente tiende a saltar demasiado rápidamente a las conclusiones. Pero “Correlación no es causación”, es simplemente una advertencia: “Hey, no tan de prisa”,

pero es necesario suministrar caminos alternativos para moverse de allí. El proceso de búsqueda no nos provee de un conjunto de causas, solo nos proporciona un conjunto de candidatos promisorios para ganar el status de causal. Nuestra principal preocupación en este punto se desprende del hecho que no hemos considerado el universo de variables relevantes, sino solamente el subconjunto de dicho universo que vino a nuestra mente. Estamos preocupados de que otros factores no considerados, aquellas siniestras tan famosas en los libros de texto, puedan estar produciendo las relaciones que estamos viendo. Entonces, retamos la suposición causal. Sacudimos nuestro cerebro en busca de otras explicaciones posibles y de estrategias para contrastar dichas explicaciones. Este comportamiento debe ser aprendido, no obstante que es una forma natural de pensamiento, para algunos, muy pocos, estudiantes. La meta del científico, es avanzar hasta la posición en la cual, no existen otras explicaciones plausibles con el actual nivel de entendimiento de la problemática. Para esto, se requieren estrategias que usen la observación, la experimentación y el análisis para descartar todas las otras posibles explicaciones.

### **Niveles de “prueba causal”**

Las decisiones sobre la definición de una acción, tienden a estar hechas sobre la base de la “mejor conjetura” a la luz de la información disponible. Rara vez se espera una evidencia incontrovertible de causalidad. Los resultados pueden ser espectacularmente buenos. Consideremos el caso de muertes en la cuna en Nueva Zelanda. La investigación mostró una muy fuerte relación entre la tasa de muertes en la cuna y ciertos comportamientos, por ejemplo, la forma como se ponía a dormir al bebé boca abajo. No había pruebas incontrovertibles de que dichos comportamientos fueran la causa, pero la idea fue suficientemente plausible para montar una campaña publicitaria y de consejo de los doctores a las nuevas madres. La tasa de muertes en la cuna, se redujo a la mitad. Hay un nivel de seguridad que permite a quien debe tomar las decisiones, estar preparado para asumir un pequeño riesgo y actuar. Hay muchos factores que afectan este nivel de seguridad. El grado de prueba causal depende de muchos factores, entre los cuales está la dificultad de hacer (echar atrás) cambios en el sistema, las consecuencias de equivocarse y el número de personas que deben estar convencidas de la decisión antes de tomar la acción. Consideremos el debate fumar\_cancer, como un ejemplo primario de las

dificultades de establecer causalidad. En el debate había (y hay) poderosos intereses creados con elevados beneficios políticos. No es sorprendente que el nivel de prueba requerido en estas circunstancias sea extremadamente alto.

## **Modelación de la variación**

Un buen número de estadísticos ha dicho que la más grande contribución de la estadística es el aislamiento y modelación de la “señal” en presencia de “ruido”. El problema básico con los datos estadísticos, es dar algún tipo de sentido de información. El principal enfoque para resolver este problema comienza con tratar de encontrar patrones en los datos. El conocimiento del contexto, nos dará algunas ideas sobre dónde mirar y qué esperar. La metodología estadística nos da herramientas para ser usadas en esta búsqueda. Experiencias comunes nos dicen que estudios conducidos bajo muy similares condiciones siempre dan resultados que se diferencian en detalles y no en la amplia evidencia. Los patrones vistos en los datos de un estudio, nunca son repetidos idénticamente en otro. La base del problema es entonces poder acopiar estrategias que permitan separar fenómenos que parecen persistir mas generalmente de aquellos que son puramente locales, para filtrar lo permanente de lo efímero. Los patrones de lo persistente proveerán la base para los pronósticos y para control. Los estadísticos han desarrollado conjuntos de estrategias particulares para la solución de este problema, estrategias basadas principalmente en la modelación probabilística. A menudo decimos que una importante función de los modelos probabilísticos y de la inferencia estadística, es contrarrestar la tendencia humana de ver patrones donde no existen. No es completamente cierto que en los fenómenos puramente aleatorios no aparezcan patrones. Estos patrones son reales para el cerebro, en el sentido en que nosotros podemos reconocer algunos hechos que podrían ayudarnos a reproducirlos. Sin embargo, tales patrones son (i) efímeros y (ii) no nos dicen nada útil acerca del problema bajo estudio, en otras palabras, son irrelevantes. Parte de nuestro razonamiento de los modelos aleatorios, es no considerar un comportamiento de los datos como permanente, si se parece mucho a algo que aparece frecuentemente a partir de un modelo puramente aleatorio.

La distinción entre variación “explicada” y “no explicada” es importante aquí.

Generalmente tratamos de encontrar significado a la variación explicada, los patrones que no son considerados como efímeros, la “señal”. La variación no explicada o “ruido” es la que queda después de haber “removido” todos los patrones. Es entonces, por definición, variación en la que no es posible encontrar patrones. Modelamos la variación no explicada como si fuese generada por un (inestructurado) proceso aleatorio. En realidad no tenemos ni idea si dicha variación será realmente aleatoria, pero eso no es cosa que nos preocupe. Si se ha realizado realmente un muestreo aleatorio, entonces hay un elemento de aleatoriedad en el ruido. Sin embargo, la medida del error y las componentes de la variación en el proceso, contribuyen típicamente a la variación no explicada y no hay manera de saber si su comportamiento es aleatorio o no. De hecho, la aleatoriedad es un conjunto de ideas, un modelo abstracto, una invención humana que nosotros usamos para modelar la variación en la cual no podemos ver patrones. Muchos de los modelos físicos que usamos para ilustrar la aleatoriedad son realmente determinísticos. Todo es parte de nuestro intento por tratar la complejidad, que de otra manera nos desbordaría. El nivel al cual imponemos la aleatoriedad al modelo, es el nivel de la capacidad de plantearnos cierto tipo de preguntas relacionadas con significación y causación.

Lenguaje como “real o aleatorio” o referente a la posibilidad de que “la diferencia observada es debida al azar” obscurece activamente la distinción entre el verdadero problema abordado y un enfoque para su solución. Hablando sobre un proyecto sobre un manglar que un estudiante realizó, dijo: *“Mi profesor me explicó (t-test) que los resultados que yo obtuve eran debidos al azar. Yo creo que esa afirmación no tiene ningún sentido. Yo entiendo que el azar está en el lanzamiento de un dado. Yo realmente no creo que el azar esté presente cuando haces relación a los datos biológicos. Cada cosa que usted posiblemente mida, es debida a algún impacto ambiental”*.

Algunos escritores han optado por decir “Toda variación es causada”. Estas afirmaciones parecen predominantemente motivadas por lo que concierne a la naturaleza humana. Mañana, con nueva información, con una nueva forma de ver las cosas, con nueva tecnología, podremos estar en capacidad de encontrar patrones en lo que hoy en día miramos como aleatorio, para rastrear las causas de aquellos patrones y mejorar el sistema.

## Referencias Bibliográficas.

Ceberio, Marcelo y Paul Watzlawick. 1998. La Construcción del Universo. Herder. Barcelona.

Cinta de Moebio No. 18. Diciembre 2003. Facultad de Ciencias Sociales. Universidad de Chile <http://www.moebio.uchile.cl/18/jaramillo1.htm>

Cox, D.R. (1992). "Causalidad: Algunos Aspectos estadísticos". J.R. Statist. Soc. A; 155, part 2, 291-301.

Gal I., Ahlgren, C. Burril, G., Landwher, J., Rich., W&Begg, A. (1995). Working Group: Assessment of Interpretive Skills". Writing group Draft Summaries Conference on Assessment Issues in Statistics Education. Philadelphia University of Pennsylvania.

Maffesoli, Michel. 1999. El Nomadismo fundador. En: Revista Nómadas. No 10. Santafé de Bogotá. Departamento de Investigaciones Universidad Central. Abril. Pág. 126 -142.

Mallows C. (1998). "The Zeroth Problem". American Statistician 52, 1-9.

Mardones, José M. 1991. Filosofía de las Ciencias Humanas y Sociales. Materiales para una fundamentación científica. Barcelona: Anthropos.

Moliner María (1998). "Diccionario de Uso del Español". Segunda edición. Madrid, España.

Moore D. (1997). "New Pedagogy and New Content: El Case of Statistics.". Interational Statistical review, 65, 123-165.

Osorio, Francisco. 1998. Qué es Epistemología. Conferencia dictada para el Magíster en Antropología Mención Desarrollo. Departamento de Antropología. Universidad de Chile. Abril 7 de 1998.

Pfannkuch, M. (1996). "Statistical Interpretation of Media Reports". In New Zealand Statistical Association Research in the Learning of Statistics Conference Proceedings, Eds. J. Neyland and M. Clark, pp 67-76. Wllintong: Victoria University.

Rao C.R: (1989 ). "Statistics And True"..CSIR. Nueva Delhi, India.

Ricci, Ricardo Teodoro. 1999. Acerca de una Epistemología Integradora. En: Revista Cinta de Moebio No.5. Abril. Facultad de Ciencias Sociales. Universidad de Chile.

Snee. (1990). "Statistical Thinking and its contributions to Quality". The American Statistician 44, 116-121.

Trochim William (2002). "Research Methods Knowlodge Base". Atomic Dog Publishing. Cornell.USA.

Wild C.J. and M. Pfannkuch. (1999). " Statistical Thinking in Empirical Enquiry". International Statistical Review, 67,3, 223-265. ISI.

# Apéndice

## ***Algunas Referencias sobre la medición y validez del constructo en Investigación social.<sup>17</sup>***

ALVIRA, F.: "Criterios para evaluar diferentes diseños de investigación". En GARCÍA FERRANDO, M.; IBAÑEZ, J. Y ALVIRA, F.: El análisis de la realidad social. Madrid, Alianza, 1989.

ALWIN, D.F. Y JACKSON, D.J.: "Measurement Models for Response Errors in Survey: Issues and Applications". En SCHUESSLER, K.F. (ED.): Sociological Methodology: 1980. San Francisco, Jossey BAs, 1980.

ARNAU, J.: "Validez y diseño de experimentación". En Métodos de investigación en las ciencias humanas. Barcelona, Omega, 1978, págs. 179-207.

ASHER, H.A.: "Some Consequences of Measurement Error in Survey Data". En American Journal of Political Science, nº 18, 1974, págs. 469-485.

BAILEY, K.D.: "Measuring. The Validity and Reliability". En Methods of Social Research. Nueva York, The Free Press, 1986, págs. 66-78.

BECHTOLDT, H.: "Construct Validity: A Critique". En American Psychologist. Vol. XIV, 1959, págs. 619-629.

BERKOVITZ, L. Y DONNERSTEIN, E.: "External Validity is More than Skin Deep: Some Answers to Criticism of Laboratory Experiments". En American Psychologist, nº 37, 1982, págs. 257-275.

---

<sup>17</sup> Basado en la pagina Wen "Métodos Y Técnicas de Investigación Social", MITYS [http://selene.uab.es/mytis/btis\\_i.htm](http://selene.uab.es/mytis/btis_i.htm) .

BLOOM, M. Y OTROS: Taxonomy of Educacional Objectives. Handbook I: Cognitive Domain. Nueva York, David McKay, 1956.

BOHRNSTEDT, G.W.: "Evaluación de la confiabilidad y validez en la medición de actitudes". En SUMMERS, G.F.: Medición de actitudes. México, Trillas, 1984 (e.o. 1976), págs. 103-127.

BOHRNSTED, G.W.: "Reliability and Validity Assesments in Attitude Measurement". En SUMMERS, G.F.: Attitude Measurement. Skokie, Ill., Rand McNally, 1970, págs. 80-89.

BRINBERG, D. Y McGRATH, J.E.: Validity and the Research Process. Newbury Park, Cal, Gage, 1985.

BRINBERG, D. Y McGRATH, J.E.: "A Network of Validity Concepts within Reseach Process". En BRINBERG, D. Y KIDDER, L. (EDS.): New Direction for Methodology of Social and Behavioral Sciences: Forms of Validity in Research. San Francisco, Jossey-Bass. 1982.

BRIONES, G.: "Confiabilidad y validez". En Métodos y técnicas de investigación para las ciencias sociales. México, Trillas, 1982, págs. 128-132.

CALDER, B.J. Y OTROS: "Beyond External Validity". En Journal of Consumer Research, vol. 10, nº 1, 1983, págs. 112-114.

CAMPBELL, D.T. Y FISKE, D.W.: "Las validaciones convergente y discriminante mediante la matriz multirrasgo-multimétodo". En SUMMERS, G.F.: Medición de actitudes. México, Trillas, 1984,(e.o. 1976), págs. 128-154.

CARMINES, E.G. Y ZELLER, R.A.: Reliability and Validity Assessment. Beverly Hills, Cal., Sage, 1979.

CEBERIO M. y WATZLAWICK, P. La Construcción del Universo. Herder. Barcelona. 1998

CROMBACH, L.: "Test Validation". EN THORNDIKE, R. (ED.): Educational

Measurement. Washington,, American Council on Education, 1971, págs. 453-507.

CROMBACH, L. Y MEEHL, P.: "Construct Validity of Psychological Tests". En Psychological Bulletin, LII (1955), págs. 281-302.

CROMBACH, L.; GLENER, G.C.; NANDA, H. Y RAJARATNAM, N.: The Dependability of Behavioral Measurements. Nueva York, Wiley, 1972.

EBEL, R.: "Estimation of the Reliability of Ratings". En Psychometrika, vol. XVI, 1951, págs. 407-424.

EBEL, R.: "Obtaining and Reporting Evidence on Content Validity". En Educational and Psychological Measurement, Vol. XVI, 1956, págs. 269-282.

FELDT, L.S. Y BRENNAN, R.L.: "Reliability". En LINN, R.L. (ED.): Educational Measurement. Nueva York, Macmillan, 1989.

FRANK, P.: The Validation of Scientific Theories. Boston, The Beacon Press, 1956.

GANTUNG, J.: "El problema de la validez". En Teoría y método de la investigación social. Buenos Aires, Eudeba, 1973.

GOODE, W.J. Y HATT, P.K.: "Confiabilidad y validez". En Métodos de investigación social. México, Trillas, 1966.

GRAWITZ, M.: "Nociones de fidelidad y validez". En Métodos y Técnicas de ciencias sociales. Barcelona, Hispano Europea, 1975, págs. 311-319.

GRAZIANO, A.M. Y RAULIN, M.L.: "Validity and the Threats of Validity". En Research Methods. A Process of Inquiry. Nueva York, Harper and Row, 1989.

GUTTMAN, L.: "A Basis for Analyzing Test-Retest Reliability". Psychometrika, vol. 10, 1945, págs. 255-282.

HARDRYCK, C. Y PETRINOVICH, L.F.: "Confiabilidad y validez". En Investigación en

ciencias sociales. México, Interamericana, 1977.

HEISE, A.R.: "Separating reliability and Stability in test-retest Correlation". En BLALOCK, H.M. (ED.): Causal Models in the Social Sciences. Chicago, Aldine-Altherton, 1971.

HOYT, C.: "Test Reliability Estimated by Analysis of Variance". En Psychometrika, vol. VI, 1941, págs. 153-160.

JARAMILLO, J. "¿Qué es la Epistemología?". Cinta de Moebio No. 18. Diciembre 2003. <http://www.moebio.uchile.cl/18/jaramillo1.htm> Facultad de Ciencias Sociales. Universidad de Chile

JUDD, C.M. Y KENNY, D.A.: "Research Design and Research Validity". En BRINBERG, D. Y KIDDER, L. (EDS.): New Directions for Methodology of Social and Behavioral Sciences: Forms of Validity in Research. San Francisco, Jossey-Bass, 1982.

KAPLAN, A.: "Precision and Reliability". En The Conduct of Inquiry. Methodology for Behavioral Sciences. Nueva York, Chandler, 1964.

KERLINGER, F.N.: "Confiabilidad y validez". En Investigación del comportamiento. Técnicas y Metodología. México, Interamericana, 1987, págs. 311-335.

KERLINGER, F. Y KAYA, E.: "Elaboración y validación analítico factorial de escalas para medir actitudes hacia la educación". En SUMMERS, G.E.: Medición de actitudes. México, Trillas, 1984, Pa'gs. 311-323.

KERLINGER, F.: "A Social Attitude Scale: Evidence on Reliability and Validity". En Psychological Reports, Vol. XXVI, 1970, págs. 379-383.

KIRK, J. Y MILLER, M.L.: Reliability and Validity in Qualitative Research. Beverly Hills, Cal, Sage, 1986.

KRATHWOHL, D. Y OTROS: Taxonomy of Educational Objectives. Handbook II:

Affective Domains. Nueva York, David McKay, 1964.

KUDER, G.F. Y RICHARDSON, M.W.: "The Theory of the Estimation of Test Reliability". Psychometrika, vol. 2, 1937, págs. 135-138.

LENNON, R.: "Assumption Underlying the Use of Content Validity". En Educational and Psychological Measurement, Vol. XVI, 1956, págs. 294-304.

LUNDBERG, G.A.: "Reliability and Validity of Questionnaires". En Social Research. A Study in Methods and Gathering Data. Nueva York, Greenwood Press, 1962, págs. 198-202.

LYNCH, J.G.: "On the External Validity of Experiments in Consumer Research". En Journal of Consumer Research, Vol. IX, nº 3, 1982, págs. 225-235.

LYNCH, J.G.: "The Role of External Validity in Theoretical Research". En Journal of Consumer Research, Vol X, nº 1, 1983, págs. 109-111.

MAFFESOLI, M. "El Nomadismo fundador". Revista Nómadas. No 10. Santafé de Bogotá. Departamento de Investigaciones Universidad Central. Abril. Pág. 126 -142. 1999

MANHEIM, H.L.: "Fiabilidad y validez". En Investigación sociológica. Filosofía y métodos. Barcelona, Ceac, 1982.

MANHEIM, J.B. Y RICH, R.C.: "Validez y fiabilidad". En Análisis político empírico. Madrid, Alianza, 1988, págs. 85-95.

MARDONES, J. M. "Filosofía de las Ciencias Humanas y Sociales. Materiales para una fundamentación científica". Barcelona: Anthropos. 1991.

MAYNTZ, R.; HOLM, K. Y HÜBNER, P.: "Problemas de validez y fiabilidad". En Introducción a los métodos de la sociología empírica. Madrid, Alianza, 1969, págs. 31-32.

McGRATH, J.E. Y BRINBERG, D.: "External Validity and The Research Process: A Third Party Comment on the Calder/Lynch dialogue". En Journal of Consumer Research, Vol. X, nº

1, 1983, págs. 111-124.

MESSICK, S.: "Validity". En LINN, L. (ED.): Educational Measurement. Nueva York, Macmillan, 1989.

MORALES DOMINGUEZ, F.: "Validez interna y externa de los diseños experimentales" y "Validez ecológica de los diseños experimentales". En Metodología de la Psicología. Madrid, UNED, 1987.

NOVICK, M.R. Y LEWIS, C.: "Coefficient Alpha and the Reliability of Composite Measurement". En Psychometrika, vol. 32, 1967, págs. 1-13.

OSORIO, F. "¿Qué es Epistemología?". Conferencia dictada para el Magíster en Antropología Mención Desarrollo. Departamento de Antropología. Universidad de Chile. Abril 7 de 1998.

RAO C.R. "Statistics And True". CSIR. New Delhi, India, 1989.

RICCI, R T... Acerca de una Epistemología Integradora. En: Revista Cinta de Moebio No.5. Abril. Facultad de Ciencias Sociales. Universidad de Chile, 1999

ROSENTHAL, R.: "Validity Interpretation of Quantitative Research". En BRINBERG, D. Y KIDDER, L. (EDS.): New Directions for Methodology of Social and Behavioral Sciences: Forms of Validity in Research. San Francisco, Jossey-Bass, 1982.

RUBIN, A. Y BABBIE, E. : "Reliability and Validity" En Research Methods for Social Work. Belmont, Cal., Wadsworth, 1989.

SCHMID, C.F.: "Measures of Reliability and Significance". En YOUNG, P. Y SCHMID, C.F. (EDS.): Scientific Social Surveys and Research. Englewood, N.J., Prentice-Hall, 1966, págs. 334-338.

SELTIZ, C.: WRIGHTSMAN, L.S. Y COOK, S.W.: "La validez de las medidas". En Métodos de investigación en las relaciones sociales. Madrid, Rialp, 1980, 241-276.

SIERRA BRAVO, R.: "Validez de una escala". En Técnicas de Investigación

Social. Teoría y ejercicios. Madrid, Paraninfo, 1979, págs. 332-334.

SJOBERG, G. Y NETT, R.: "Confiabilidad y validez de la evidencia". En Metodología de la investigación social. México, Trillas, 1980, págs. 350-361.

SUTCLIFFE, J.P.: "A Probability Model for Errors of Classification. I. General Considerations". En Psychometrika, vol. 30, 1965, págs. 73-96.

THORNDIKE, R. Y HAGE, E.: Measurement and Evaluation in Psychology and Educación. Nueva York, Wiley, 1969.

TRAXEL, W.: "Examen del modo de obtención de los datos: Fidelidad y validez". En La psicología y sus métodos. Barcelona, Herder, 1970.

TROCHIM, W. "Research Methods Knowledge Base". Atomic Dog Publishing. Cornell.USA. 2002

TRYON, R.: "Reliability and Behavior Domain Validity: Reformulation and Historical Critique". Psychological Bulletin, vol. LIV, 1957, págs. 229-249.

VISAUTA, B.: "Fiabilidad y validez de las escalas". En Técnicas de investigación social: Barcelona, Promociones y Publicaciones de la Universidad, 1989, págs. 211-217.

WERTS, C.E., JÖRESKOG, K.G. Y LINN, R.L.: "Comment on the Estimation of Measurement Error in Panel Data". En American Sociological Review, vol. 36, 1971, págs. 110-112.

WILEY, D.E. Y WILEY, J.A.: "The Estimation of Measurement Error in Panel Data". En American Sociological Review, vol. 35, 1970, págs. 112-117.

WINMSATT, W.C.: "Robutness, Reliability and Overdetermination". En BREWER, M.B. Y COLLINS, B.E. (EDS.): Scientific Inquiry and the Social Sciences. San Francisco, Jossey Bass, 1981.