
MALOS USOS EN ESTADISTICA (WHAT IS NOT WHAT IN STATISTICS)

Louis Guttman

Nota del traductor

La estadística ofrece actualmente un repertorio espléndido de instrumentos de análisis para el sociólogo. Si los utiliza adecuadamente, el sociólogo puede estudiar un número creciente de problemas de investigación. Ahora bien, la estadística, para que pueda cumplir su papel adecuadamente, debe utilizarse en conjunción con la teoría y con los métodos de investigación. La «tecnificación» exagerada de la investigación sociológica que se observa en la actualidad ha conducido, sin embargo, a una mitificación de la estadística, que la ha sacado de su contexto, produciendo situaciones que necesitan con urgencia ser transformadas.

El artículo de Guttman que presentamos al lector en lengua castellana es, en nuestra opinión, el más desmitificador que conocemos sobre el uso y abuso de la estadística en las ciencias sociales. Escrito por Guttman en el cénit de su carrera investigadora, no contiene una sola cita. Recoge, ni más ni menos, algunos de los problemas que la estadística matemática no ha podido aún resolver en el campo de las ciencias sociales, y algunos de los abusos más corrientes que se deslizan en la práctica sociológica de la investigación que utiliza la estadística inferencial, y que Guttman ha ido recogiendo a lo largo de su dilatada tarea como investigador.

Conviene recordar que la mayoría de las técnicas estadísticas utilizadas actualmente en el campo de la sociología, han sido desarrolladas para problemas no sociológicos por no sociólogos. El análisis de varianza se desarrolló originariamente para la investigación agraria, el análisis de regresión se diseñó para los estudios genéticos y, en general, buena parte de los análisis multivariantes surgieron como respuestas a problemas que planteaban las investigaciones en el campo de la biología y otras ciencias naturales. Regresión, causación, correlación, análisis de camino, varianza, etc., son términos que inducen a error con mucha frecuencia al sociólogo cuando al utilizar las técnicas correspondientes en problemas de investigación sociológica, no tiene en cuenta el origen de tales técnicas. Guttman, con un tono claramente irónico, señala, por ejemplo, que la sociología es actualmente la ciencia que ofrece un mayor número de descubrimientos «causales», debido al uso y abuso del análisis de correlaciones como si realmente analizase «causas», y no meramente correlaciones.

El sociólogo debe ser más cuidadoso con el lenguaje estadístico que utiliza. Así, por ejemplo, cuando habla de variables «independientes», conviene que tenga en cuenta que tales variables no son en realidad «independientes», ya que si así lo fueran no tendría sentido el análisis de dependencia. Y es que la estadística matemática y la estadística real usan términos semejantes con significados distintos.

Quizá uno de los términos que pueden conducir a mayor equívoco sea el de probabilidad. En el lenguaje matemático, «probabilidad» es una fracción matemática. No hay nada de «probable» en ella. Está libre de error. Es un término que expresa la certeza de que un ítem se produzca dentro de un modelo matemático. Sin embargo en la estadística que utiliza el sociólogo, la «probabilidad» tiene un significado bien diferente*. Guttman no utiliza el ejemplo del término «probabilidad» en el presente artículo, pero en cambio ofrece una lista de otros conceptos estadísticos que suelen utilizarse equivocadamente por los sociólogos. Esta lista, una vez leída, nos obligará a ser más cautos en nuestras futuras investigaciones.

Por otra parte, quiero señalar el carácter plenamente positivo de la crítica de Guttman. Los problemas se presentan para que detengamos nuestra atención en ellos, para buscar soluciones y obtener mayor rigor en nuestro trabajo científico. Se evita la fácil crítica, negativa, de manifestar que el carácter de la sociología impide su tratamiento cuantitativo, y de igual manera que ocurre en el campo de la medicina, en el que el hecho de que no se haya vencido al simple y común resfriado no conduce al desarrollo de actitudes de escepticismo sobre el progreso médico, así ocurre también con la investigación sociológica, en donde la existencia de problemas hasta ahora irreso-

* Para una detallada exposición de este argumento, se puede consultar el artículo de Thomas F. MOSIMANN "Mathematical Statistics and Real Statistics", *IAISI, Estadística*, junio 1957, págs. 390-394.

lubles, no impide a Guttman que desarrolle una gran labor de clarificación metodológica que permitirá algún día, gracias a sus esfuerzos y al de otros sociólogos, ir dando respuesta a tales problemas.

Finalmente, una nota de advertencia sobre el acto mismo de la traducción. Aunque me he esforzado en todo momento por lograr una versión castellana lo más limpia posible de anglicismos, me encuentro bastante lejos de estar satisfecho con la tarea realizada. Y es que al tratar de verter al idioma castellano, literariamente tan fértil pero científicamente tan pobre, los términos científicos que una lengua tan innovadora, desde el punto de vista científico y tecnológico, como la inglesa produce continuamente, nos encontramos con expresiones para las que existen pocos referentes en nuestro idioma. Piénsese en expresiones tales como «stepwise regression procedures», «path analysis», «belongingness coefficients», «partial out», o en el propio título del artículo, «What is not what», y se comprobará que su «castellanización» no es nada fácil. Al hilo de esta advertencia final, y ya para terminar, un ruego. Si alguno de los lectores conoce términos o expresiones en castellano que traduzcan con mayor fidelidad las utilizadas en la presente traducción, que las haga públicas. De este modo, ayudará a enriquecer nuestro vocabulario científico, tan necesitado como está de nuevas sugerencias.

Manuel GARCIA FERRANDO
Departamento de Sociología
Universidad Autónoma de Madrid.¹

INTRODUCCION

Hace unos cuarenta años, Harold Hotelling señaló que los libros de texto de estadística de aquel período estaban escritos, en su mayoría, por no matemáticos. Aquellos libros estaban repletos de conceptos erróneos, e ignoraban en su mayoría los nuevos e impresionantes desarrollos de la disciplina matemática de la estadística inferencial. La situación mejoró paulatinamente a medida que más estadísticos matemáticos comenzaron a publicar libros de texto, de tal modo que hoy en día el péndulo parece haber ido demasiado lejos. En algunos lugares, los símbolos más que la sustancia parecen ser los dueños. Esto es especialmente verdad en las ciencias sociales —con las que estoy más familiarizado— y a las cuales va dirigido (aunque no exclusivamente) este trabajo. Por ejemplo, consejeros y editores de algunas revistas insisten en decorar las tablas de diversos tipos de datos con estrellas y dobles estrellas, y en presentar listados de «errores estándar», a pesar de que las probabilidades implícitas de significación o confianza son claramente erróneas desde el punto de vista de la estadística inferencial (ver los problemas 3 y 1, más adelante).

Junto al mal uso de los nuevos desarrollos, persisten todavía muchos de los antiguos falsos conceptos en los libros de texto y revistas actuales debido a la extremadamente pobre terminología que ha sido conservada, por razones históricas, por los estadísticos matemáticos. Los matemáticos están habituados a tratar correctamente simbolizaciones arbitrarias e incluso confusas, dado que están preparados para centrarse directamente en los conceptos que señalan, los cuales, por otra parte, están bien definidos. No ocurre así con los no matemáticos, quienes más bien están inclinados a reaccionar ante las etiquetas verbales como si tuvieran unas implicaciones y un significado diferentes y más allá de los conceptos técnicamente designados. Por ejemplo, el término «regresión» emergió originalmente en el contexto de las investigaciones genéticas de Francis Galton, anteriores a la difusión de la teoría de los genes, y ha sido conservado desde entonces por los matemáticos para algo que no tiene que ver nada con los procesos genéticos, ni con ningún tipo de proceso. Es cierto que es deseable tener una palabra única para «una serie de medias aritméticas condicionales», pero al retener la palabra «regresión» para tal serie se da a los no matemáticos una idea de procesos dinámicos y leyes de la naturaleza en contextos para los que tales ideas son totalmente erróneas. (Irónicamente, incluso hoy en día algunos genetistas confunden el concepto estadístico de regresión con una teoría genética de herencia biológica, y en consecuencia alcanzan conclusiones falsas.)

Uno de los objetivos del presente trabajo es poner de manifiesto algunos de los problemas básicos aún no resueltos en la estadística inferencial. La discusión de estos problemas sirve como introducción para un segundo objetivo, presentar una lista de hechos sobre la terminología estadística y las ideas equivocadas que merecen ser sometidas a la atención, tanto de los matemáticos

como de los no matemáticos. Los hechos se formulan negativamente: *que no es que*. Tal lista puede servir como una forma rápida de alertar a los no matemáticos de algunas de las trampas en las que han caído algunos colegas. Espero que la lista y los problemas estimulen a los estadísticos matemáticos para que se centren en la resolución de temas que son de gran importancia para la práctica estadística.

De igual manera que el resfriado común se ha resistido a ser conquistado por la ciencia médica, así ocurre con algunos de los problemas más comunes de la investigación social que han eludido la solución de los estadísticos matemáticos. Esta puede ser una de las razones de la permanencia de algunos antiguos falsos conceptos y de la creación de algunos nuevos; los profesionales tratan de arreglárselas con las herramientas inadecuadas, dado que necesitan tener respuestas. Se destacarán seis clases de problemas comunes aún no resueltos. A continuación se presenta la lista de lo *que no es que*. Cada ítem en la lista viene acompañado de una breve explicación. Las explicaciones son bastante autosuficientes, pero el profesor de estadística interesado en el tema puede ampliarlas fácilmente. Naturalmente, la lista se puede extender fácilmente, y serán bienvenidos los comentarios críticos.

No se incluyen referencias bibliográficas en este artículo, dado que la discusión se refiere a lo que no existe. La prueba empírica de la no existencia es un principio difícil, mientras que la prueba de la existencia es comparativamente más simple, requiere que se exhiba un solo ejemplo. Por eso, cuando realizamos más adelante una afirmación de la forma: «Ningún libro de texto prueba que...», si uno quisiera documentar tal declaración, tendría que referirse a todos los libros de texto existentes. Por otra parte, si un lector creyese que podía demostrar que la afirmación era falsa, todo lo que necesitaría hacer es suministrar una sola referencia correcta. Realmente me encantaría recibir cualquier referencia correctora de cualquier lector interesado, para cualquier «no es» de los que se incluyen más adelante.

Una reacción inicial de algunos lectores puede ser la de considerar que el presente artículo pretende ser polémico. Pero esto no es así en absoluto. Señalar que el emperador no viste ningún tipo de ropa puede resultar un tanto turbador. Ahora bien, la experiencia ha mostrado que la polémica puede venir más bien de la dirección opuesta, esto es, de los firmes creyentes en prácticas no fundamentadas. A tales devotos se les suele utilizar como referencias y jueces científicos, y no escatiman la acumulación de críticas irrelevantes y decisiones negativas sobre los nuevos desarrollos que no contienen sus favoritas concepciones equivocadas. El presente artículo puede servir, como uno de sus objetivos, para evitar que en el futuro se produzcan tales situaciones kálfianas.

Además, este artículo no pretende ser tan sólo un ejercicio sobre terminología. Se han omitido algunos de los ítems «que no es» en la discusión que sigue, con el fin de dedicar más espacio a los temas que estimo son más per-

judiciales. Algunos de los ítems incluidos han obstaculizado notablemente el progreso de las ciencias sociales, permitiendo a veces que se hayan malgastado decenas de miles de dólares en investigaciones no provechosas, por no hablar de la enorme cantidad de tiempo y esfuerzos intelectuales desperdiciados.

ALGUNOS PROBLEMAS TODAVIA NO RESUELTOS DE LA ESTADISTICA INFERENCIAL

Problema 1. *Regiones de Confianza Multivariante.*—Muchos de los problemas inferenciales en las ciencias sociales son inicialmente multivariantes, y, sin embargo, no suelen tratarse como tales. Considérese cualquier serie de datos reunidos a partir de un cuestionario demográfico o de actitudes, o por medio de un test mental o de logro compuesto de diversos ítems. ¿Cómo se puede establecer una región de confianza para una tabla contingente de población a partir de una tabulación cruzada (cross-tabulation) de tales datos? Esto requiere la especificación simultánea de una serie de intervalos de confianza para muchos parámetros de una distribución multimodal, pero con un solo nivel de confianza para la serie entera. Las tablas de contingencia se encuentran entre las formas más corrientes de datos observados, y, sin embargo, no existe todavía solución para este problema: los libros de texto ni siquiera lo mencionan. En la práctica, los «errores estándar» se calculan a menudo para valores estadísticos separados en una tabla de ese tipo, sin que nadie haya mostrado qué relevancia tienen para el problema tales valores. El mismo abuso se mantiene en el caso de la región de confianza para una serie de medias aritméticas. Los proyectos de investigación sociológica y psicológica pueden implicar muchas variables numéricas simultáneamente, y es de interés el establecimiento de límites para cada una de las medias aritméticas de la población. Incluso para el caso de distribuciones normales multivariantes, el uso del «error estándar» con cada una de las medias de la muestra no ha podido dar lugar a una región de confianza para todas las medias de la población simultáneamente. ¿Cuál es la forma correcta de establecer tal región? No es de menor interés la región de confianza para una matriz de coeficientes de correlación entre diversas variables. Resulta estimulante saber que algunos estadísticos matemáticos están comenzando a comprobar estas cuestiones. Estos problemas en los que intervienen proporciones, medias aritméticas y coeficientes de correlación son todos ellos casos especiales de un problema general: si $\theta_1, \theta_2, \dots, \theta_n$ son n parámetros de una población con una distribución multivariante de variables mutuamente dependientes, definir los valores estadísticos $a_1, b_1, a_2, b_2, \dots, a_n, b_n$ a partir de una sola muestra tal que, para un nivel de confianza dado α

$$\text{Prob} \{ a_1 \leq \theta_1 \leq b_1, a_2 \leq \theta_2 \leq b_2, \dots, a_n \leq \theta_n \leq b_n \} = 1 - \alpha$$

y con alguna condición de optimización para la elección de a_1 y b_1 . Ningún

libro de texto se interesa por estos obvios y clásicos problemas de estadística inferencial, como tampoco se interesan por los anteriores en la forma en que realmente ocurren en la práctica, si es que llegan siquiera a mencionar los problemas. La solución de tales cuestiones dejará todavía abierto el no menos básico problema de la replicación tal como se bosqueja a continuación.

Problema 2. *Replicación.*—Tanto la estimación como la comprobación de hipótesis se han restringido usualmente como si se tratase de experimentos que se realizan una sola vez, tanto en la teoría como en la práctica. Pero la esencia de la ciencia es la replicación: un científico debe estar continuamente interesado con lo que ocurra cuando él u otro científico repita su experimento. Por ejemplo, supongamos que se establece un intervalo de confianza para la media de una población en base a un único experimento: ¿cuál es la probabilidad de que la media de una muestra de un próximo experimento cabrá dentro de dicho intervalo? El nivel de confianza del primer experimento no dice nada al respecto. O de nuevo, supongamos que se calcula una ecuación de regresión a partir de una muestra aleatoria incondicional: ¿cuál es la varianza de las predicciones realizadas a partir de una nueva muestra aleatoria incondicional de la misma población en base a la ecuación anterior? La respuesta de esta última pregunta es desconocida: muchos psicólogos, sabedores de ello, no dependen de una sola muestra y realizan validación cruzada empírica. El mismo tipo de problema, con un giro diferente, se presenta en la prueba de hipótesis. Supongamos que un científico rechaza una hipótesis nula en favor de una alternativa dada: ¿cuál es la probabilidad de que en el próximo experimento científico hará lo mismo? El próximo experimento del científico no será, por lo general, independiente del primero, dado que la repetición no se hubiera emprendido si en el primer experimento se hubiera retenido la hipótesis nula. Este es uno de los problemas más realistas de la inferencia, y espera una respuesta. El asunto no es puramente matemático, ya que el comportamiento real de los científicos debe tenerse en cuenta. Enfrentados con tales problemas reales de la replicación, pueden surgir dudas sobre la llamada perspectiva Bayesiana en la estadística inferencial.

Problema 3. *Niveles simultáneos de Significación e Hipótesis Simultáneas.*—Una dificultad intrínseca del problema precedente, y de muchos otros problemas reales de inferencia, es la complicación de los hábitos de los investigadores. Estos normalmente no fijan con anterioridad a sus experimentos un error del tipo I. La fijación preliminar de tal valor es requerida por la lógica de la teoría de Neyman-Pearson, pero el *cómo* hacerlo no forma parte de la teoría. Dado que los investigadores gustan de encontrar instrucciones precisas, insisten en que se les diga *cómo* seleccionar un nivel de significación, a pesar de que no es de la incumbencia del matemático el hacerlo. Forzados a formular una respuesta, el estadístico matemático puede carraspear, y finalmente decir:

«Pruebe algo tal como .05 ó .01 o incluso .001.» En tiempos anteriores podría haber sugerido: «Tome algo como más o menos dos o tres errores estándar.» El estadístico matemático podría olvidarse de recordar al investigador que tomara *uno y sólo uno* de tales números —y *con anterioridad*— para el problema. En cualquier situación, dadas varias opciones, el investigador las acepta *todas* y las utiliza simultáneamente, y habitualmente *después* del hecho. Esta práctica por sí misma puede hacer completamente insoluble el problema 2 anterior. La situación se torna más confusa cuando los niveles del *omnibus* se aplican todos ellos simultáneamente *tal cual* a una *serie* de hipótesis simultáneas. ¿Qué solución puede existir para los procedimientos utilizados en la práctica? ¿Cómo pueden convencerse los autores y editores de revistas científicas de que cuando rellenan sus tablas de datos con una galaxia de estrellas, dobles estrellas, e incluso triples estrellas, no están probando hipótesis, sino rechazando la propia estadística inferencial?

Problema 4. *Elección de las Hipótesis Nula o Alternativa.*—La teoría de Neyman-Pearson para la prueba de hipótesis requiere una formulación previa de —y distinción entre— las hipótesis nula y alternativa. No es competencia de la teoría decir *cómo* realizar esta distinción preliminar, dejando de nuevo al investigador en la duda. La retención, por razones históricas, del infortunado adjetivo «nula» es a estos efectos contraproducente. Una terminología más aclaratoria podría ser: hipótesis «obligatoria» («incumbent») *versus* hipótesis «desafiante» o «cuestionante» («challenging»). Una hipótesis nula es la obligatoria, que no se modificará al menos que exista una evidencia abrumadora en contra (de ahí que la probabilidad pueda ser 99 contra 1 para el error del tipo I, en favor de la hipótesis obligatoria). En muchas áreas de las ciencias sociales, las hipótesis simplistas —tal como *no* diferencia o *no* correlación— son habitualmente desafiantes en campos bien documentados de investigación. Tomemos el caso de los tests de inteligencia: nadie hasta ahora ha sido capaz de realizar un diseño *a priori* con dos tests mentales diferentes y seguros que correlacionan cero entre sí: ésta es una tarea desafiante (casi todas las correlaciones entre tests mentales observados durante los últimos sesenta años son positivas). O de nuevo, la hipótesis de Charles Spearman del único factor común, fue una innovación desafiante (últimamente rechazada por él mismo) para un fenómeno tan complejo como es la inteligencia. El tener un pequeño número de factores comunes continúa siendo una hipótesis desafiante en contra de la usual hipótesis obligatoria de un número más amplio de factores comunes. Tales casos pueden contrastarse con campos más problemáticos e interesantes, tales como la parapsicología y la grafología, por ejemplo, para los cuales la nulidad persiste como una apropiada hipótesis nula. Una ilustración de este punto en otra área es la ley de Newton sobre el movimiento de un cuerpo en línea recta con velocidad constante: ¡desde luego que ésta sí fue una hipótesis desafiante! ¿Cuál fue la hipó-

tesis nula desafiada por Newton? ¿Y cuándo en la historia se convirtió la hipótesis de Newton en obligatoria, para hacer frente a un nuevo desafiador? El cambio en el tiempo de los papeles de las hipótesis de alternativa a nula es un proceso importante a elucidar para la estadística inferencial. No hay necesidad de hacerse bayesiano para ello; realmente, la tarea cae fuera del dominio de los estadísticos matemáticos. (Como ya se indicó anteriormente, el problema 2 sobre replicación plantea cuestiones sobre el realismo de la perspectiva bayesiana, cuestiones similares a aquellas que hubiera podido conducir al propio reverendo Thomas Bayes a no recomendar lo que actualmente es «Bayesiano».) El problema práctico que persiste es que muchos estadísticos matemáticos han hecho creer a los investigadores que, por ejemplo, la linealidad de la regresión es una hipótesis obligatoria, a pesar de su rareza y de su carácter desafiante en muchas áreas de la ciencia empírica. Aquí puede existir confusión entre el concepto de «primera aproximación» y el de «hipótesis nula» —los dos son esencialmente contradictorios. Lo mismo ocurre con la falta de interacción en el análisis de varianza y con la falta de correlación en las distribuciones bivariantes —tales nulidades serían fenómenos bastante sorprendentes en las normalmente complejas interacciones de la vida social—. ¿Cómo se podría enseñar a los investigadores empíricos que, sin un conocimiento sustantivo de sus campos respectivos, no existen fundamentos para asignar papeles a las hipótesis «nula» o «alternativa»? ¿Y que una primera aproximación no es la hipótesis nula de que hablan los libros de texto?

Problema 5. *Ortogonalidad*.—La búsqueda de «contribuciones independientes» a partir de diversos componentes correlacionados es una empresa perenne entre los no matemáticos. La creencia en la realidad de tal milagro estadístico se ha visto reforzada por la noción de ortogonalidad en el diseño de experimentos. El diseñador puede forzar la ortogonalidad, y hacerlo así si es que puede, debido a la teoría distribucional simplificada que resulta. Muchos no matemáticos creen que un diseño *debe* generar ortogonalidad, ¡pues piensan que si no es así van en contra de la teoría estadística! Los matemáticos saben que tal ortogonalidad no es otra cosa que un artificio creado por el diseñador de experimentos, y puede no tener nada que ver con las interrelaciones de los fenómenos naturales. De manera similar, el estadístico crea ortogonalidad cuando utiliza mínimos cuadrados para predecir una variable numérica: la predicción y el error de predicción son ortogonales entre sí. Hay que decir que muchos contextos en los que la ortogonalidad ocurre en estadística están creados por el análisis estadístico, y que la ortogonalidad no tiene necesariamente una implicación o interpretación «natural». Una cuestión interesante sería: ¿existe algún tipo de ortogonalidad en datos que no hayan sido creados por el estadístico? Una respuesta posible es un coeficiente de correlación observado de valor cero (la elección conocida de una hipótesis «nula» discutida en el anterior problema 4). En la correlación múltiple, uno quisiera

a menudo tener los predictores no correlacionados entre sí: si lo estuvieran, podrían considerarse como si hicieran contribuciones «independientes» a la regresión múltiple. Pero, en general, los predictores se correlacionan entre sí, y no existe una operación inmediata para definir contribuciones «independientes» en este caso. Incluso para el caso de predictores no correlacionados, no existen garantías de que no se pueda encontrar otro predictor que *correlacionará* con el anterior predictor, restaurando la imposibilidad de conceder confianza independiente a cada uno de los predictores por separado.

Problema 6. *Diseño de datos, análisis de datos y primeras aproximaciones.*—R. A. Fisher puso de manifiesto cómo la inferencia estadística debe basarse en el diseño experimental. ¿Cómo puede extenderse esta forma de razonamiento al análisis de datos más generales para los que los estadísticos matemáticos no tienen todavía respuestas inferenciales (y puede que no las tengan durante mucho tiempo)? ¿Por qué en los sondeos sociológicos y en los tests mentales los ítems que contienen se construyen sin el mismo cuidado y formalización que se emplean en el diseño de la muestra de la población a la que se les va a administrar? ¿Y por qué el análisis de los datos debe conducirse de acuerdo con tal diseño del contenido? Hacer esto requiere el desarrollo de una teoría de muestras (estratificada) para construir variables para un universo de contenidos, de igual modo que la teoría ordinaria de muestras discute la selección de sujetos individuales de una población. El muestreo estratificado (incluso dentro de estratos) no puede mantenerse para la construcción de ítems de tests de inteligencia o de actitud. Lentamente se ha ido desarrollando una teoría que pueda dar una respuesta parcial a este problema, especialmente en el contexto de las teorías de la estructura de las intercorrelaciones y de lo que Lee Cronbach llama «generalizabilidad». No puede suponerse que las técnicas asociadas para el análisis de datos se ajusten a tests «exactos» de significación, sean o no paramétricos. Más bien sugieren que la atención se dirija de nuevo a la propia inferencia: ¿por qué debemos interesarnos en un *nivel «exacto» de significación o confianza*? El análisis de datos no inferencial se contenta con ser descriptivo, y a menudo solamente como una «primera aproximación» con alguna indicación de cuán *aproximadamente es exacto*. (No se puede asegurar el opuesto, esto es, exactamente cuánta aproximación, sin conocer la respuesta exacta, en cuyo caso la aproximación sería superflua.) De una manera más general, ¿por qué no estar satisfechos con un *nivel aproximado de aproximación*? ¿Por qué debe estar el investigador enfrentado perennemente con la contradictoria y paradójica cuestión: exactamente cuán aproximado es su trabajo? Esencialmente, la replicación es el test de la ciencia, y las repeticiones repetidas —por más que aproximadas— valen más la pena que el tratar de fijar la «exactitud» de un nivel de aproximación de una o dos pruebas. La manera en que se puedan obtener inferencias estadísticas correctas sobre los parámetros cuando solamente se utilizan

primeras aproximaciones, parece que es en buena medida un territorio inexplorado por parte de los estadísticos matemáticos. W. Edwards Deming y otros han hecho trabajo de caballería al señalar docenas de fuentes de error no muestral, que deberían sensibilizar en particular a los investigadores en el problema de la aproximación. A pesar de ello, la confusión aparece ampliamente extendida entre los investigadores en lo que se refiere a errores de muestreo, *versus* errores de aproximación.

UNA LISTA DE LO QUE NO ES QUE

La siguiente lista de hechos (negativos) amplía y se añade a las seis clases precedentes de temas no resueltos. Como ha mostrado la discusión anterior, los problemas inferenciales pueden ser oscurecidos por rasgos no inferenciales. Así, una de las fuentes de falsas interpretaciones para el profesional es la dificultad de especificar dónde comienza formalmente la inferencia (el análisis de varianza es un buen ejemplo de ello). He tratado de resistir a la tentación de tratar de clasificar las variedades de falsas interpretaciones y confusiones en la lista. Unas tropiezan con otras de manera sutil y no sutil. Puede que sea consustancial a tales falsas interpretaciones el que todo intento de su clasificación sea confuso, y pueda conducir a mayores equivocaciones. Por ello he formulado cada ítem sucintamente como un hecho por sí mismo, y sólo intento ligeramente hacer alguna referencia cruzada, tanto dentro de la lista como con las seis áreas de problemas precedentes.

1. Las medias no miden tendencias centrales.

(En el concepto de media no está implícito un proceso dinámico, como sugiere erróneamente la palabra no matemática «tendencia». Considérese la distribución en forma de U. Una media puede definirse como un valor que minimiza una función residual, y cualquier valor en el área de alcance de una variable es una media según alguna función residual. Ver también el punto 3 más adelante.)

2. La dispersión o desviación de una distribución no se define necesariamente como si estuviese alrededor de una media.

(Considérese el valor esperado de $|x_p - x_q|$, en donde x_p y x_q son valores de los miembros p y q de una población en una variable numérica x . Analistas de varianza tomen nota. Al contrario, el valor esperado de $(x_p - x_q)^2$ ocurre que es proporcional a la varianza alrededor de la media aritmética. Ver también los ítems 4 y 20, más adelante.)

3. No existe regresión a la media.

(De igual modo que no existe un proceso dinámico para la media [ver el ítem 1, anterior]. El verbo «regresar» no tiene definición matemática, aun-

que el nombre «regresión» desgraciadamente va unido a él. Una regresión es sencillamente una serie de medias condicionales, usualmente de medias aritméticas.)

4. El concepto de correlación no depende necesariamente del concepto de regresión.

(Considérense los coeficientes libres de regresión de monotonicidad entre dos variables numéricas x e y para una población p , tal como μ_2 :

$$\mu_2 = \frac{\sum_{p \in P} \sum_{q \in P} (x_p - x_q) (y_p - y_q)}{\sum_{p \in P} \sum_{q \in P} |x_p - x_q| |y_p - y_q|}$$

Tal coeficiente de correlación varía entre -1 y $+1$, alcanzándose estos valores extremos cuando se obtiene perfecta monotonicidad, sin especificación de la forma exacta de la función monotónica, y sin referencia a las medidas condicionales de una variable a partir de la otra. Esto amplía el concepto de media-libre de dispersión del anterior punto 2. Los conceptos de correlación libres de regresión están mostrando ser útiles en los análisis de datos sobre series temporales, así como en los análisis corrientes de distribuciones bivariantes y multivariantes. Tales coeficientes son las bases de los programas de ordenador para el análisis de datos no paramétricos, tales como el análisis del espacio mínimo.)

5. Una primera aproximación no es una hipótesis nula.

(Puede tratarse de una hipótesis *aproximada*, nula o alternativa, si se prueba una hipótesis. La teoría actual de Neyman-Pearson no se ocupa de hipótesis aproximadas, por lo que no resulta apropiada para primeras aproximaciones. Por ejemplo, si se considera que la linealidad es tan sólo una primera aproximación a la estructura de una curva de regresión de una población, esto es, en efecto, un rechazo automático de la hipótesis de linealidad, y es necesario que el investigador decida exactamente qué es lo que está comprobando y contra qué alternativa, si quiere utilizar la teoría actual de Neyman-Pearson para hablar de la estructura de la regresión de la población. La utilización de ideas de aproximación puede entrar en colisión con las ideas de hipótesis nulas *versus* hipótesis alternativas.)

6. No existe generalmente salida de la linealidad de una regresión.

(El tomar la linealidad como punto de partida equivale atribuir a un fenómeno tan poco común el papel generalmente incorrecto de una hipótesis nula obligatoria. En las ciencias sociales, al menos, la linealidad debiera ser contemplada como una salida de la no linealidad, y no viceversa. Ver el proble-

ma 4, anterior: también los puntos 4 y 20. Esto provoca una cuestión interesante sobre cómo desarrollar una prueba realista de significación, o si la inferencia estadística debiera siquiera mezclarse con problemas de aproximación. Ver el problema 6, anterior.)

7. Una diferencia que se afirma que es «significativa al nivel 0.1» no es significativa al nivel .01.

(Esto es un hecho para cualquier nivel —el nivel .01 se ha tomado como ejemplo —y para cualquier valor estadístico, no sólo para una diferencia—. Tal afirmación implica que el «nivel» se determinó después de calcular el valor estadístico. Al ensayar una hipótesis, el nivel [y sólo un nivel] debe fijarse con anterioridad a la investigación; la hipótesis nula se rechaza o no subsecuentemente, de acuerdo con el valor estadístico observado y la región de rechazo. Proclamar un «nivel de significación» después de calcular un valor estadístico implica un valor incorrecto para la probabilidad del error tipo I, y de hecho hace que la probabilidad quede indeterminada. Ver el problema 3, anterior. Si un investigador desea destacar su pertenencia a la minoría de los que realmente ensayan hipótesis, debiera decir algo parecido a lo que sigue: «El nivel .01 de significatividad se eligió previamente a la investigación, y el estadístico obtenido es significativo a dicho nivel.» Además, no debería poner ninguna estrella.)

8. Un ensayo de significación estadística no es un ensayo de importancia científica.

(Este es un hecho que se necesita enseñar a cada generación de estudiantes. Puede que haya pasado desapercibido a una parte de la generación anterior que se ha convertido en editora y miembros del consejo de redacción de las revistas científicas en campos tales como psicología experimental y sociología experimental dedicadas *de facto* a asuntos tales como «diferencias pequeñas pero significativas» y «efectos significativos». Hasta ahora nadie ha publicado una ley científica en las ciencias sociales que haya sido desarrollada, modelada o efectivamente sustentada en base a ensayos de significación. Las leyes básicas de la física, desde luego, no se desarrollaron de esta forma. La estimación y la aproximación pueden ser más provechosas que la significación en el desarrollo científico, sin olvidar la replicación. Téngase en cuenta la ley radex para la inteligencia y la ley cilindrex para las actitudes, o las leyes de regresión politónica para los componentes principales de las actitudes.)

9. Un intervalo de confianza para la media de una población no vale para predecir la media de una nueva muestra.

(Y lo que es peor, una ecuación de regresión lineal múltiple calculada a partir de una muestra aleatoria no condicional puede con frecuencia hacer más daño en la predicción de una nueva muestra aleatoria no condicional

que la que pueda hacer una simple ponderación. Ver el problema 2 anterior sobre replicación.)

10. La distribución normal no es un fenómeno empírico normal.

(Raras veces, si alguna, se observa en la naturaleza. En buena medida se genera por los estadísticos cuando desarrollan las matemáticas de la teoría de las muestras. Este hecho se ha enseñado correctamente durante mucho tiempo, pero necesita que se repita constantemente a los estudiantes después de haber pasado por cursos de inferencia estadística.)

11. La correlación parcial no parcializa [*partial out*] nada.

(De igual manera que la probabilidad condicional no parcializa nada. Todas las correlaciones bivariantes son correlaciones parciales: cada una de ellas es condicional a la población de la que se calcula. Proponer nuevas condiciones implica la estratificación en subpoblaciones, y las correlaciones condicionales resultantes pueden variar ampliamente entre tales subpoblaciones. Una mejor y menos confusa terminología sería decir siempre «correlación condicional» en vez de «correlación parcial», de igual modo que los matemáticos dicen «probabilidad condicional» y no «probabilidad parcial».)

12. Las variables «independientes» de una regresión son habitualmente no independientes estadísticamente de la variable que va a predecirse a través de la regresión.

(Si lo fueran, la regresión sería inútil para la predicción. Estas variables «independientes» son generalmente mutuamente dependientes. Una terminología más adecuada consistiría en llamarlas variables «condicionales» de la regresión, o los «predictores». Hotelling sugirió llamar a la variable incondicional, es decir, la que se va a predecir, la «predictante» («predictant») de la regresión: los psicólogos a menudo la llaman el «criterio».)

13. Las contribuciones «independientes» a una regresión múltiple son habitualmente dependientes.

(Incluso en el caso de que los predictores sean independientes estadísticamente entre sí, no existen garantías de que se pueda encontrar un predictor ulterior que introduzca dependencia estadística. Cuando hay dependencia estadística existen muchas formas de resolverla en componentes independientes estadísticamente; la elección entre estas diversas formas no es un problema estadístico. Ver la discusión del milagro de la ortogonalidad en el problema 5 anterior: ver también el punto 21 más adelante sobre regresión «stepwise».)

14. Cuando se calcula a partir de un solo ensayo en una muestra una estimación de un coeficiente de fiabilidad para la población, es generalmente inconsistente (usualmente una subestimación).

(Se necesitan al menos dos ensayos sobre la misma muestra para conseguir una estimación consistente —y por supuesto no sesgada— de un coeficiente de fiabilidad de población, incluso para la fiabilidad de una suma o mitades partidas. Los intentos comunes de salir del paso con un solo ensayo introducen supuestos que son falsos y que no anulan los sesgos respectivos. Los sesgos de los supuestos habituales son acumulativos, y a menudo conducen a serias subestimaciones. Los autores algunas veces se percatan de esto cuando «corrigen para atenuar», y obtienen un coeficiente de correlación mayor que la unidad; el sesgo es bastante universal y puede ser drástico incluso si una «corrección» en un caso particular no conduce a un absurdo inmediato. Muchos libros de texto en psicología educativa y en áreas similares tratan erróneamente las fórmulas convencionales del coeficiente de fiabilidad como si fueran consistentes, mientras que estas fórmulas no son usualmente otra cosa que estimaciones de límites bajos del coeficiente de fiabilidad en cuestión.)

15. En un ensayo de items de elección múltiple, las diversas respuestas falsas no tienen generalmente la misma probabilidad de ser elegidas.

(Si las respuestas falsas son igualmente probables para cada miembro de una población amplia, y si existe independencia experimental entre los miembros, entonces la proporción observada de población que elija una respuesta falsa particular será igual a la proporción observada que elija cada una de las otras respuestas falsas. Habiendo revisado a lo largo de los años cientos de items de elección múltiple, aún no he observado empíricamente tales proporciones iguales, y no conozco a nadie que haya manifestado haber observado tal fenómeno empírico. En la práctica, los distraedores distraen diferencialmente. La tan extendida hipótesis de que existen «intuiciones» en los ensayos prácticos que conducen a probabilidades iguales es un ejemplo de supuestos «matemáticos» falsos e innecesarios anclados en algunos lugares de las ciencias sociales. Los supuestos «matemáticos» no son un sustituto para el estudio real de la conducta humana.)

16. El estadístico chi-cuadrado para los ensayos de independencia estadística entre dos variables no es una medida de dependencia.

(Por ejemplo, este valor estadístico no indica cuándo existe perfecta dependencia monotónica. Existe un solo tipo de independencia estadística, pero muchas variedades de dependencia perfecta, y cada una de éstas requiere su propia función residual [*loss function*]. Es por esto también que la prueba del chi-cuadrado, tal como se usa habitualmente, es bastante débil: no tiene una hipótesis alternativa particular. Generalmente se puede hacer una prueba mejor cuando se especifica el tipo de dependencia.)

17. El concepto de «variable aleatoria» no se define en términos de muestreo aleatorio.

(El recíproco es cierto. Una «variable aleatoria» es realmente una función, esto es, una función que tiene como dominio una población con una medida de probabilidad. El que la teoría estadística se ocupe desde el principio del concepto de «función» puede ser una de las fuentes de dificultad en la enseñanza de estadística elemental: al menos se deben considerar simultáneamente, y desde el principio, dos azares.)

18. Nada ocurre por azar.

(«Azar» no es un término estadístico. Algunos autores usan «azar» para referirse a sucesos con igual probabilidad, otros pueden estar pensando en la independencia estadística entre variables, mientras que algunos lo utilizan para indicar simplemente que no se conoce todavía una ley definitiva. Se puede evitar mejor la palabra en discusiones técnicas. Igualmente, nada «ocurre al azar» aunque el muestreo aleatorio sea posible —a saber, al generar una distribución multivariantes de variables estadísticamente independientes e idénticamente distribuidas.)

19. Un valor esperado no es esperado generalmente.

(Se trata simplemente de la media aritmética de una población.)

20. Una frase teóricamente diseñada («mapping sentence») no es una teoría (*).

(El concepto de frase teóricamente diseñada generaliza simplemente el diseño de experimentos de R. A. Fischer al diseño de cualquier observación, con el rasgo adicional de una conexión verbal informal que es necesaria para el trabajo empírico real. Tal diseño ampliado define el contenido de las observaciones y de este modo puede servir como base para el establecimiento y comprobación de una hipótesis o una teoría. Tal como señaló Fischer, el diseño de los datos debiera ser una parte explícita de una teoría. Se puede definir una teoría, en este sentido [y de una manera bastante general], como una hipótesis, con su lógica, de la unión correspondiente entre un sistema —o diseño— definicional para un universo de observaciones. Ver el problema 6: también los puntos 47 y 50, más adelante.)

21. Generalmente no se puede muestrear aleatoriamente un universo de variables para una población dada.

(Generalmente no existe una distribución de probabilidad para un diseño de contenido de facetas. La replicación de una muestra se consigue al construir nuevas variables según el mismo diseño de contenido de facetas. Un

* N. T.: La "mapping sentence" es una técnica básica de la teoría de las facetas ("facet theory"). Representa uno de los intentos más recientes que se están llevando a cabo en el campo de las ciencias sociales para formalizar conjuntamente la teoría y la investigación empírica, tarea en la que participa activamente Guttman.

caso especial de tal construcción y replicación es la traducción en diversas lenguas y la comparación entre diversas culturas.)

22. Por lo general, una hipótesis nula no debería hipotetizar el carácter de ser idénticamente nula (*nullity*).

(Lo idénticamente nulo debiera ser generalmente una hipótesis alternativa: ver el problema 4 anterior.)

23. El análisis de varianza no analiza varianza.

(Analiza la forma de la regresión de una variable numérica sobre otra variable [predicador] condicional numérica o categórica. Las varianzas y los grados de libertad aparecen en escena para ayudar en el estudio del error de muestreo con fines inferenciales: no son esenciales para el reparto básico del predictor numérico en la regresión [«entre»] y desviación de la regresión [«dentro de»]. El diseño factorial representa el caso general de las condiciones categóricas: pero, tradicionalmente, la forma más general y posible de regresión para este diseño no se estudia. Por ejemplo, dado un diseño tridimensional para realizar observaciones en una variable numérica t , los analistas de varianza tradicionalmente consideran solamente una tautología de la forma

$$t_{ijk} = t_{ij.} + t_{i.k} + t_{.jk} + t_{i..} + t_{.j.} + t_{..k} + \text{residuo}$$

y buscan la ortogonalidad [problema 5], mientras que esta forma es tan sólo un caso especial de una tautología más general

$$t_{ijk} = \mu_{ij} v_{jk} w_{ik} + \text{residuo}$$

en donde μ_{ij} , v_{jk} , y w_{ik} pueden a su vez ser descompuestos de formas diversas. La regresión de la población, naturalmente, es la serie de valores esperados condicionales de t_{ijk} , y esta serie no necesita estimarse consistentemente con el uso de una tautología restringida. A pesar de ello, los autores habitualmente tratan de ensayar hipótesis en las que se mantiene una forma más simplificada que cualquiera de las anteriores —usualmente varias hipótesis simultáneamente—. Suelen calcular una serie de valores estadísticos [«ratios de varianza»] F_1, F_2, \dots, F_m —siendo cada F_i una especie de ayuda para el ensayo de la hipótesis nula H_{oi} en algún aspectos de la regresión— y realizan afirmaciones referentes a «niveles de significación» a_i , en donde probablemente

$$a_i = \text{Prob} \{ F_i \geq \lambda_i \mid H_{oi} \} (i = 1, 2, \dots, m)$$

el coeficiente λ_i se refiere al límite entre los intervalos de rechazo y aceptación de la hipótesis H_{oi} . Tales afirmaciones son típicamente erróneas, tal

como se ha discutido en el problema 3 y en el punto 7. Sería más apropiado especificar una región de rechazo R , un valor estadístico multivariante r , y un nivel de significación α para todas las hipótesis simultáneamente, tal que

$$\alpha = \text{Prob} \{ r \in R \mid H_{01}, H_{02}, \dots, H_{0m} \}$$

y donde r y R minimizan el tipo II de error para dar una serie de hipótesis alternativas. Algunos estadísticos matemáticos han prestado atención a casos especiales de este problema, destacando nulidades como hipótesis nulas, y normalmente sin ofrecer alternativas especializadas. Esto, naturalmente, nos conduce de nuevo al problema 4. A pesar de ello, los autores a menudo muestran que en lo que están interesados realmente es en la estimación de la forma de la regresión, y que utilizan la prueba de la hipótesis como una técnica para la estimación. Tales autores tratan las hipótesis secuencialmente, pero sin utilizar inferencia secuencial. Esto no es diferente a la regresión «stepwise» con variables condicionales numéricas [ver el punto siguiente]. Se combinan diversas cosas y se recalculan «probabilidades», en un desenfado olvido de que la inferencia estadística es denegada con ello. Incluso algo más básico se niega cuando los autores y editores de revistas se quedan tan enamorados con el aparato técnico de las sumas de cuadrados y grados de libertad que en efecto publican, pero deciden ahorrar espacio —o simplemente olvidan— y no publican la estimación de la regresión final que era el objetivo del trabajo: se fijan en el baño, pero no en el bebé. Incluso cuando se ahorra espacio de publicación, sería generalmente más útil publicar al menos el ratio de correlación asociada con la regresión, con el fin de ayudar al lector a comprender inmediatamente el poder predictivo relativo de la regresión tal como se estima a partir de los datos. Ver también el punto 2 anterior para algo parecido.)

24. La regresión secuencial óptima («stepwise regression»), tal como se practica corrientemente, no es ni inferencia ni teoría inteligentes.

(Hacer ulteriores cálculos condicionales con los ensayos de «significación» de cálculos anteriores no produce las probabilidades implícitas con fines inferenciales. Todavía no se conoce una prueba secuencial correcta. Alternativamente, el fijarse en todas las regresiones posibles simultáneamente crea otro problema de inferencia que tampoco se ha resuelto aún: ver el problema 3. Más importante todavía: la búsqueda de una regresión más simplificada se hace probablemente para usos prácticos en una nueva muestra. Nadie ha mostrado que ninguna técnica para reducir regresiones —incluyendo el análisis de varianza tal como se ha discutido anteriormente— tenga ninguna cualidad óptima para tratar el problema de la nueva muestra; ver el problema 2. Enfrentados con este estado de ignorancia inferencial, nada puede ser más práctico para lograr regresiones simplificadas que una teoría sustantiva para la estructura de la matriz entera de covarianza —predictante y predictor

juntos— que puede probarse aproximadamente por los datos de la muestra. La evidencia cros-validada empírica y matemáticamente, indica que es interesante la búsqueda de un número mínimo de predictores para la predicción práctica. Demasiados predictores pueden dar predicciones sin valor en la muestra siguiente; estropean una regresión al añadir más error muestral que otra cosa. Además, unos simples pesos constantes para los predictores pueden ser mejores para la predicción en una nueva muestra que lo puedan ser los coeficientes de regresión de la muestra anterior, debido a la inestabilidad de los coeficientes de regresión. Algunos autores consideran la regresión secuencial no como un problema práctico, sino como una herramienta teórica para comprobar las contribuciones de incrementos «independientes» a una regresión. Al hacer esto caen en la trampa de la ortogonalidad discutida en el problema 5, punto 12 y punto 24. El uso de la regresión secuencial es de hecho una confesión de ignorancia teórica en lo referente a la matriz de correlación. Si se conoce la estructura, las formas apropiadas de la regresión se pueden predecir con anterioridad; simples ilustraciones de esto son las inversas de las matrices de covarianza «simplex» y «circumplex». Para el desarrollo de teoría sustantiva puede ser más conveniente considerar la estructura de la matriz de covarianza como un todo, a la luz del diseño definicional de todas las variables implicadas. Ver el problema 6 y el punto 20.)

25. La correlación generalmente no indica causación.

(Este hecho se ha enseñado adecuadamente durante bastante tiempo. Pero la esperanza brota eterna en algunos cuarteles sociológicos: ver los puntos 26, 27, 39 y 40 sobre «determinación», «explicación», «análisis causal» y «análisis de camino».)

26. Un coeficiente de determinación no indica determinación.

(El cuadrado de un coeficiente de correlación de Pearson o ratio de correlación se llama a menudo un coeficiente de «determinación», y con frecuencia se dice erróneamente que expresa la «proporción» de una variable que es «determinada» por otras. Es obvio que cualquier variable puede tener correlaciones diferentes de cero en muchos contextos, por lo que la suma de todas las «proporciones de determinación» posibles para cualquier variable dada es generalmente infinita. Se enseña corrientemente que «la correlación no significa necesariamente causación»; por alguna razón, al cambiar la palabra «causación» por «determinación» se anula esta enseñanza.)

27. La proporción (o porcentaje) de varianza nunca se explica.

(La palabra «explicación» juega aquí el mismo papel que «determinación» o «causación» en el punto anterior. Ninguna de estas palabras tiene un significado técnico matemático; su uso representa un «deseo» [«wishful thinking»] acerca de la relativa predictibilidad de una variable en un contexto dado,

que generalmente conduce a porcentajes de «explicación» que suman varias veces 100 por 100 para la variable en cuestión.)

28. La correlación no determina el contenido.

(Al igual que la correlación no implica causación. De otra manera existiría siempre una respuesta obvia a una pregunta como: «supóngase que, para una población dada, una variable X correlaciona .60 con la altura de la gente. ¿Cuál es el contenido de la variable X?».)

29. «Item analysis» no analiza items.

(Intenta simplemente «ensayar» la hipótesis —¡desafiante!— de que todas las correlaciones inter-ítem son cero, y habitualmente a través de una técnica de correlación de puntuación total de los items correcta. Igualmente implica el «buen deseo» [«wishful thinking»] de que las correlaciones debieran determinar el contenido.)

30. La escalabilidad no debe ser deseada o construida.

(Decir que uno «quiere construir» una escala de actitudes hacia alguna cosa, o de logro en algún área es casi análogo a decir que no «quiere» que el mundo sea plano. Los items son las cosas a construir —no la escalabilidad; la escalabilidad es una hipótesis empírica para un universo de items en una población dada [normalmente una hipótesis alternativa a la hipótesis nula de multidimensionalidad; ver el punto 33]—. El rechazo de items que no «se ajustan» a la unidimensionalidad es como rechazar la evidencia de que el mundo es redondo.)

31. Si todas las correlaciones entre los items son positivas, ello no presupone la presencia de un solo factor común (incluso cuando los coeficientes son muy elevados).

(Al contrario, la observación de que todas las correlaciones eran positivas condujo a Charles Spearman a desarrollar —y a desaprobar— la hipótesis de un solo factor común para la inteligencia. De esta forma se desarrolló el análisis del factor común *múltiple*. Incluso cuando todas las intercorrelaciones son muy altas —del orden de 0.95— no se puede inferir nada sobre dimensionalidad a partir de esta sola información. Se puede hipotetizar que todas las correlaciones sean positivas cuando las variables tienen un rango común —considérese, por ejemplo, las primeras leyes de inteligencia y actitudes—. Un rango común no debe confundirse con un factor común.)

32. Que el número de factores comunes sea pequeño no es generalmente una hipótesis nula.

(Esto continúa siendo una hipótesis desafiante en el campo de la inteligencia y de otras áreas de la conducta social. Véase también el epígrafe siguiente.)

33. La escalabilidad no es generalmente una hipótesis nula.

(Es por ello por lo que se ha desarrollado el análisis del escalograma multidimensional. Las distribuciones multivariantes de ítems en las ciencias sociales, tanto estén basadas en diseños de observaciones intuitivas o formales, han mostrado ser generalmente multidimensionales en vez de unidimensionales. Hasta ahora nadie ha sugerido una explicación básica para un universo de ítems, y para la población a observar a partir de él, según la cual la unidimensionalidad sea la regla en vez de la excepción. Los editores de libros de texto y revistas científicas parecen desconocer el hecho de que el análisis de escalograma multidimensional puede ser más apropiado que los tan extendidos intentos por «forzar» la escalabilidad [ver el epígrafe 30]. El análisis de escalograma multidimensional no debe confundirse con el llamado análisis multidimensional de escalas, que será discutido en el epígrafe 46.)

34. El espacio euclidiano puede definirse sin un sistema coordenado.

(En efecto, así es como lo hizo Euclides. Descartes vino unos siglos más tarde. Hoy se puede obtener una perspectiva conveniente libre de coordenadas a través de ideas sobre vectores o distancias. Es curioso cómo algunos editores de artículos que incluyen técnicas de análisis de datos como el Análisis del Espacio Mínimo [*Smallest Space Analysis*] continúan preguntando por una presentación y/o interpretación de ejes coordenados, a pesar del hecho de que tales ejes son completamente irrelevantes al problema. Ver también los puntos siguientes sobre análisis factorial, y los puntos 46 y 47.)

35. El espacio euclidiano bidimensional posee un número infinito de dimensiones.

(Esta es una de las razones por las que deben ensayarse hipótesis regionales, relacionadas con diseños de facetas, en vez de tratar de encontrar tan sólo un par de ejes coordenados «significativos». Lo mismo se puede decir en el caso de espacios n-dimensionales, euclidianos o no, cuando $n \geq 2$.)

36. Los autores que utilizan análisis factorial no analizan, en la práctica, factores.

(No hallan una serie de puntuaciones de factores que, al mantenerse constantes, produzcan correlaciones condicionales de valor cero [o independencia estadística local] entre las variables observadas. Todo lo más, analizan parcialmente la matriz de correlación observada —no las puntuaciones observadas— al calcular los supuestos «factores o coeficientes de saturación» [«factor loadings»] de los coeficientes para la reproducción aproximada de aquella matriz. Nadie ha demostrado la utilidad de los actuales cálculos indirectos de los coeficientes de saturación de la matriz de correlación observada, cuando el problema real es el de las puntuaciones observadas. Tampoco ha de-

mostrado nadie que las matemáticas subyacentes en las rutinas actuales del ordenador sean consistentes con las matemáticas de la teoría de la puntuación del factor [«factor score theory»]. Las matemáticas del análisis factorial demuestran que aunque se fijan los coeficientes de saturación de una manera consistente, se continuará dejando abierta la cuestión de las puntuaciones de factor que deben acompañar a los coeficientes de saturación: existen generalmente soluciones de puntuaciones alternativas muy diferentes que son consistentes precisamente con los mismos coeficientes de saturación. Muchos libros de texto no mencionan estos problemas de indeterminación e inconsistencia de las puntuaciones de factor, que se encuentran en la base de la teoría analítica factorial, y los programas de ordenador existentes ignoran estos problemas.)

37. El análisis factorial no es un instrumento poderoso ni exploratorio.

(Llamar «exploratorio» al análisis factorial equivale a afirmar que los autores que utilizan análisis factorial no practican análisis factorial [ver el punto anterior], sino algo para lo que no fue diseñada la teoría analítica factorial. Las ideas no métricas encajan mejor con fines exploratorios en algo que no sea tan rígido como el marco de una teoría factorial. Todo lo más, los analistas de factores exploran parcialmente la matriz de correlaciones —aunque esta matriz sea meramente incidental a la teoría factorial— al buscar un sistema de coordenadas para las variables sin tener que factorializar puntuaciones para los individuos. Esta exploración es bastante limitada; por ejemplo: todos los programas de ordenador existentes etiquetados como «análisis factorial» no dan información elemental sobre una matriz de correlación: ¿son todos sus items de un signo o no? En esta cuestión del signo es donde comenzó históricamente el análisis factorial: ver el punto 31 anterior. Se ha olvidado la hipótesis de L. L. Thurstone de un «positivo múltiple» para todos los signos positivos. Análogamente, los programas no dan información sistemática sobre los tamaños relativos de los coeficientes de correlación observados, y están engranados siempre para pasar por alto una estructura simple y otras configuraciones simples que se sabe que existen en diversas matrices de correlación empírica. Los programas tampoco se benefician de ningún diseño tipo «faceta» para las variables observadas [incluyendo otros diseños factoriales]. Todos estos analistas adoptan la posición estrecha arbitraria de que un espacio euclidiano [para las variables] debe «comprenderse» en términos de un sistema coordinado [ver el punto 34 anterior], cerrando sus ojos a otras posibilidades regionales y libres de coordenadas. Están cegados por su insistencia en las coordenadas cartesianas, ignorando los sistemas cilíndricos y otros sistemas coordinados que han mostrado su utilidad en otras formas de análisis de datos, si es que las coordenadas son útiles para algo.)

38. La teoría de las estructuras latentes no es una teoría estructural.

(Es una teoría de desviación de una estructura. El argumento principal de esta perspectiva es que una población puede estratificarse en subpoblaciones, dentro de cada una de las cuales se adquiere independencia estadística para el universo de los ítems. No forma parte de la teoría el cómo estratificar —o la especificación estructural—, y debe decidirse de nuevo para cada problema a través de consideraciones externas. Es por ello por lo que no pueden haber programas de ordenador estándar para el análisis de las estructuras latentes. En este y otros aspectos, el análisis factorial y el análisis de las estructuras latentes pertenecen a la misma familia; en particular, ambos comparten el problema básico de la indeterminación de los valores estructurales o puntuaciones para los individuos, incluso después de especificar la estructura sobre los ítems. Comparar con el punto 36.)

39. El análisis causal no analiza causas.

(Incluso no ofrece una definición del término «causa». Tampoco ofrece una necesaria o suficiente condición empírica para la prueba de la «causalidad» de las relaciones. Si se propusiese cualquiera de las dos condiciones, conduciría indudablemente a muchas cosas que serían «causadas» repetidas veces [compárese con los puntos 27 y 40 sobre la «explicación de la varianza y del análisis de camino»]. A pesar de ello, se ha producido un «florecimiento» de descubrimiento «causales» en sociología a un ritmo desconocido en el campo de las ciencias naturales. Virtualmente cada mes, las revistas actuales publican nuevos «análisis causales» y «modelos causales» que sin duda colocan a la sociología a la cabeza de todas las ciencias en términos de la frecuencia de descubrimientos de relaciones fundamentales. De hecho, las ciencias no sociológicas se las han arreglado para seguir adelante sin «causación». Según Sir Isaac Newton, la «causación» pudiera ser que no denotase siquiera un concepto científico.)

40. El análisis de camino («path analysis») no analiza caminos no genéticos.

(Sewall Wright sugirió inicialmente el análisis de camino como un algoritmo para calcular varianzas genéticas bajo ciertas condiciones cuando se conoce el camino de herencia de genes de una generación a otra. El término «path analysis» ha sido tomado prestado por algunos investigadores para usos no genéticos, principalmente para referirse a algunos cálculos lineales algebraicos para los que los «caminos» no existen aparte de la propia álgebra, y sin ninguna definición de lo que [análogamente a los genes] se supone que se transmite en el tiempo a lo largo del «camino». Incluso en genética, si se introdujera el entorno en el análisis de «caminos», no existiría una lógica clara para el camino en el tiempo; al ampliar las ecuaciones genéticas de esta manera puede suponerse que los genes se modifican o se generan por el

entorno. Unidades de tiempo y/o secuencias de generaciones están generalmente ausentes de los análisis de «camino» de datos sociológicos y no genéticos, a pesar del hecho de que el problema básico en el estudio del movimiento en el tiempo sobre los caminos —presuponiéndose que existen caminos conocidos para ser estudiados—. La genética sólo tiene un modesto marco para los caminos. Sin embargo, y de acuerdo con algunas revistas actuales, los sociólogos continúan descubriendo nuevos marcos de caminos fundamentales cada mes; y a los estudiantes graduados de sociología se les obliga rutinariamente, como ejercicios de clase individuales, a que se manejen en descubrimientos que iguallen a los de Gregor Mendel. Ver también los puntos 27 y 39, sobre «explicación» de la varianza y sobre análisis «causal».)

41. Las regiones no son generalmente conglomerados («clusters»).

(Dos puntos pertenecientes a regiones diferentes de un espacio, pueden estar más cerca entre sí que dos puntos de una misma región. Las regiones para el análisis de datos deben definirse usualmente a través de consideraciones de contenido, no por análisis tentativos de «conglomerados» de distancias entre puntos. Las regiones están indicadas por —y generalmente comparten— puntos límites, y no están generalmente separadas por espacios vacíos como sugiere el término «conglomerados».)

42. La «conglomeración» («clustering») no define el contenido.

(De igual forma que la correlación no define el contenido. Un test verbal y un test aritmético pueden estar más juntos entre sí que lo puedan estar dos tests aritméticos o dos tests verbales.)

43. No existe una definición ampliamente aceptada del concepto «conglomerado» para el análisis de datos.

(Difícilmente puede haber una, especialmente para las ciencias sociales, dado que las teorías acerca de los espacios físicos [incluyendo teorías no geográficas y no ecológicas] generalmente requieren continuidad, sin «vacíos» o separaciones claras entre las regiones del espacio social o psicológico. Las diversas técnicas de análisis de datos subsumidas bajo el nombre de «análisis de conglomerados» generalmente no justifican de una forma lógica por qué deben esperarse «conglomerados sistemáticos», por lo que no existe una lógica para su definición. Se usa el término «conglomerado» cuando es más apropiado el término «región», y además requiere un criterio externo para la delimitación de los límites. Ver el punto 41.)

44. Las escalas nominales, de intervalo, y de ratios no son escalas.

(Una «escala nominal» está desordenada por definición, por lo que no es una escala por definición, puesto que el orden es una parte esencial de la noción de una «escala». En psicofísica, la «escala de intervalos» y la «escala

de ratios» son nombres para hipótesis acerca de algún aspecto de ciertas curvas de regresión experimentales. Algunos no-psicofísicos han tomado prestada esta terminología desafortunada para contextos menos apropiados —y por supuesto indefinidos—, y puede que desconozcan el problema de la regresión psicofísica experimental original. Existe un folklore ampliamente difundido acerca de «reglas» estadísticas míticas, que prohíben o permiten cálculos que incluyen «escalas», siendo tales reglas independientes del contexto. Ver el punto siguiente. Quizá los psicofísicos podrían sugerir una palabra mejor que «escala» para sus hipótesis de regresión bivalente.)

45. No se requieren permisos en el análisis de datos.

(Lo que se requiere es una función residual [«loss function»] para ser minimizada. Los autores suelen demandar «reglas» *a priori* sobre lo que está «permitido» hacer con sus observaciones numéricas, ordenadas o desordenadas, sin referencia a una función residual comprensiva para su problema. En vez de ello, deberían decir al matemático: aquí está mi función residual: ¿cómo hago para minimizarla? La minimización puede requerir el tratamiento de datos desordenados de una forma numérica y los datos numéricos de una forma desordenada. Si el matemático concede o retiene el «permiso» sin referencia a una función residual, puede ser accesoria la ayuda al investigador para escapar de la realidad de definir el problema de la investigación.)

46. El análisis de escalas multidimensionales no métricas, no escala dimensiones.

(Todo lo más, escala distancias. Transforma monotónicamente información inter-puntos del tipo métricamente-ordenada, en el lenguaje de Clyde Coombs, en una función de distancia [Euclidea o no Euclidea] que relaciona puntos. Warren Torgerson empleó originalmente el término «escalabilidad multidimensional» en un análisis totalmente métrico de distancias observadas entre puntos, con la intención de «escalar» realmente dimensiones, esto es, de hallar una serie de coordenadas, que fueran cada una de ellas «significativas» al análisis factorial métrico, y con la menor dimensionalidad, para reproducir los coeficientes de distancia observados. Los intentos no métricos de tratar las desemejanzas se centran solamente en el aspecto de hallar un espacio de dimensionalidad mínima, y en este sentido están libres de coordenadas. En efecto, el crecimiento acumulativo de hallazgos de estructuras que responden a leyes en los tests de datos actitudinales y mentales —entre otros— ha sido posible por el uso de conceptos *regionales* para el espacio mínimo, y no por la búsqueda de dimensiones significativas. La «escala» se usa técnicamente tan sólo para una variable unidimensional [la distancia es siempre unidimensional, incluso dentro de un espacio multidimensional], por lo que la «escalabilidad multidimensional» puede ser una terminología contradictoria en contextos no métricos y en otros libres de coordenadas. Puede resultar apropiado

para el análisis factorial de multifacetadas [«multi-modal»], y para otras perspectivas que insisten en el hallazgo de dimensiones significativas. El término es innecesariamente erróneo en contextos en los que sólo se intenta un análisis del espacio mínimo, confundiendo a los investigadores —y editores de revistas— de nuevo acerca del punto 34, anterior.)

47. El número de facetas («facets») no determina la dimensionalidad.

(Considérese el ejemplo del diseño factorial de tres facetas del punto 23. Si ninguno de los términos en la tautología tradicional tiene una varianza cero, y si se mantiene la ortogonalidad, entonces la regresión tiene seis dimensiones ortogonales para las tres facetas. La hipótesis de que todas las interacciones desaparecen, equivale a la hipótesis de que la dimensionalidad de la regresión no será mayor que el número de facetas. Igualmente, en el análisis del espacio mínimo de una matriz de correlaciones, la dimensionalidad mínima obtenida no tiene necesariamente una conexión con el número de facetas de contenido en la frase teóricamente diseñada [«mapping sentence»] para las observaciones: la dimensionalidad puede ser más grande, igual, o más pequeña que el número de facetas. En efecto, uno de los problemas principales en la construcción de teoría sustantiva es el de racionalizar hipótesis viables acerca de las relaciones de las facetas de contenido con la dimensionalidad, y otros aspectos de los datos. Ver los puntos 20 y 50.)

48. El análisis de datos no métricos es generalmente métrico.

(El *input* puede ser completamente no numérico, o también un aspecto no numérico de datos numéricos; pero el *output* es generalmente un espacio métrico, a menudo un espacio euclideo. En el caso especial de que tanto el *input* como el *output* sean métricos, pero sólo se conserva la monotonicidad —como en el análisis de espacios mínimos y otras técnicas análogas—, el diagrama de Shepard realmente representa la naturaleza métrica de la función monotónica implícita. En último término, una función explícitamente monotónica podría especificarse como un resultado del análisis: T. W. Anderson hizo ya tal cosa en 1958 para el «radex».)

49. En principio, no existe contradicción entre el análisis de datos métricos y el análisis de datos no métricos.

(Cada análisis métrico consistente debe retener rasgos no métricos del *input* de datos, y simplemente añadir restricciones posteriores. Esta es la razón por la que un análisis dedicado tan sólo a aspectos no métricos, da lugar a un espacio más pequeño que el análisis métrico más restrictivo de los mismos datos. Paradójicamente, cuando un análisis métrico aproximado puede calcularse más rápidamente que un análisis no métrico, los cálculos métricos son a menudo una primera aproximación útil en repeticiones hacia una solución no métrica. Las diferencias en principio ocurren *dentro* de procedimientos

métricos y, en consecuencia, dentro de los correspondientes procedimientos no métricos: diferencias sobre aspectos del *input* de datos debieran representarse en el *output* como puntos, como vectores, como distancias, como ángulos, como regiones, etc.)

50. Las funciones residuales (*loss functions*) utilizadas normalmente en el análisis de datos son incompletas.

(Los coeficientes de ajuste, tales como el de reproductibilidad, contigüedad, alienación, y similares —basados bien en los mínimos cuadrados, en el principio del valor absoluto, en el principio del rango de imagen, o cualquier otro— se utilizan indiscriminadamente en consideraciones de contenido. No incorporan residuos asociados con alejamientos de una teoría sustantiva sobre la estructura de los datos, y consiguientemente necesitan una modificación. Ver el problema 6; también los puntos 20 y 47. En particular, estas deficiencias se mantienen en mi propio trabajo hasta ahora: pero espero gradualmente remediar el asunto a la luz de los nuevos desarrollos en la teoría de las facetas.)

NOTAS